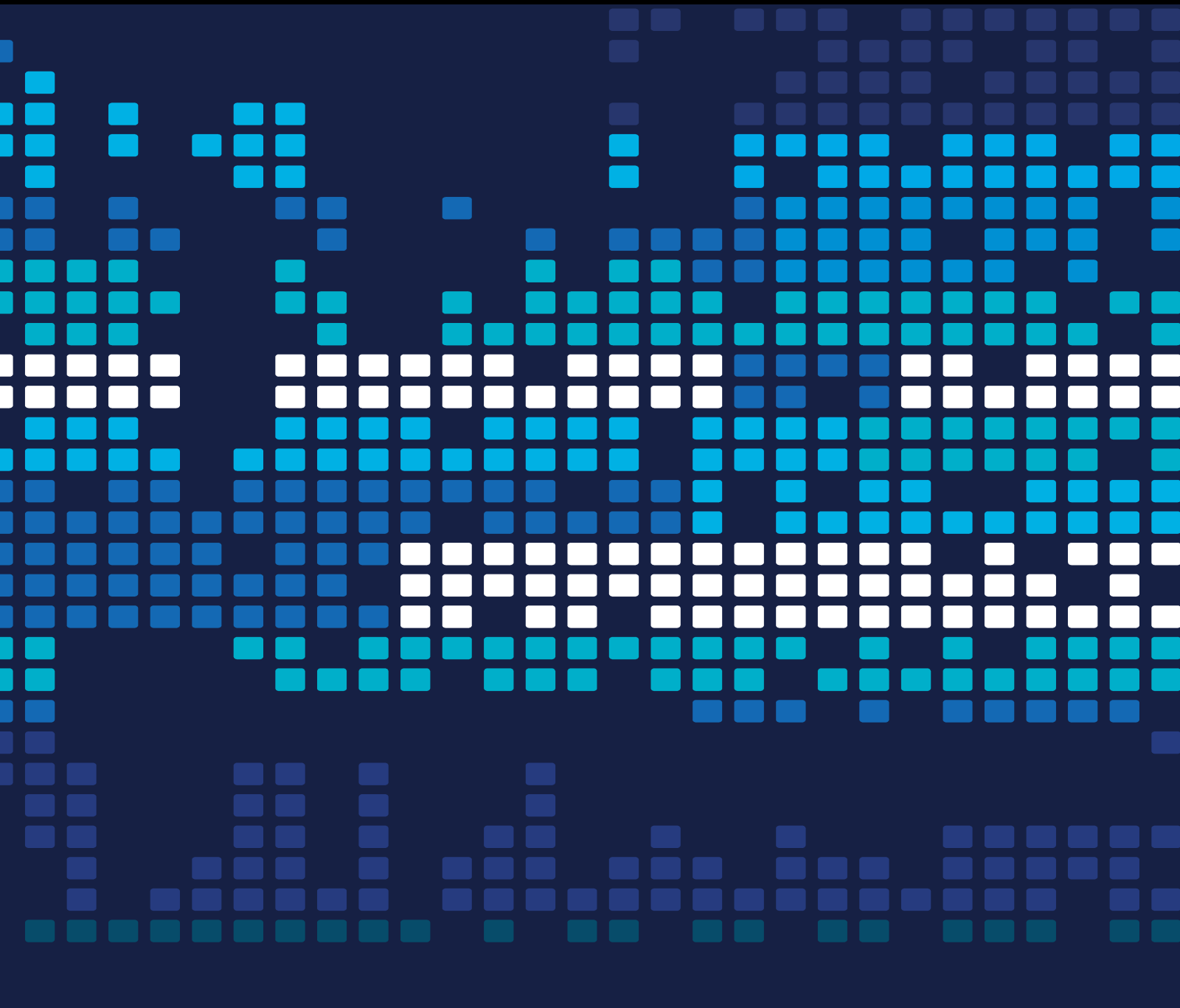# Machine Learning in Image and Video Processing

Lead Guest Editor: Bai Yuan Ding
Guest Editors: Yang Gu, Wei Jiang, and Chongyi Li

Machine Learning in Image and Video Processing

Machine Learning in Image and Video Processing

Lead Guest Editor: Bai Yuan Ding
Guest Editors: Yang Gu, Wei Jiang, and Chongyi Li

Contents

Contents

Contents

Contents

Contents

Contents

Retraction

Retracted: Research on Network Layer Recursive Reduction Model Compression for Image Recognition

Scientific Programming

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] H. Ling, W. Zhang, Y. Tao, and M. Zhou, "Research on Network Layer Recursive Reduction Model Compression for Image Recognition," *Scientific Programming*, vol. 2021, Article ID 4054435, 11 pages, 2021.

Retraction

Retracted: Application Analysis of 3D Printing Technology in Design Field: Taking Shoe Design as an Example

Scientific Programming

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope

(2) Discrepancies in the description of the research reported

(3) Discrepancies between the availability of data and the research described

(4) Inappropriate citations

(5) Incoherent, meaningless and/or irrelevant content included in the article

(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] T. Gong and L. Kang, "Application Analysis of 3D Printing Technology in Design Field: Taking Shoe Design as an Example," *Scientific Programming*, vol. 2021, Article ID 5662460, 8 pages, 2021.

Retraction

Retracted: Research on the Application of Environmental Art Design Based on the Combination of VR and Panoramic Video Technology

Scientific Programming

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope

(2) Discrepancies in the description of the research reported

(3) Discrepancies between the availability of data and the research described

(4) Inappropriate citations

(5) Incoherent, meaningless and/or irrelevant content included in the article

(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] Y. Kong, "Research on the Application of Environmental Art Design Based on the Combination of VR and Panoramic Video Technology," *Scientific Programming*, vol. 2021, Article ID 9820550, 11 pages, 2021.

Retraction

Retracted: Application of Computer Simulation Analysis in Green City Garden Plant Landscape Design

Scientific Programming

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope

(2) Discrepancies in the description of the research reported

(3) Discrepancies between the availability of data and the research described

(4) Inappropriate citations

(5) Incoherent, meaningless and/or irrelevant content included in the article

(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] J. Zhang, "Application of Computer Simulation Analysis in Green City Garden Plant Landscape Design," *Scientific Programming*, vol. 2021, Article ID 9422417, 9 pages, 2021.

Retraction

Retracted: Career Recommendation for College Students Based on Deep Learning and Machine Learning

Scientific Programming

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope

(2) Discrepancies in the description of the research reported

(3) Discrepancies between the availability of data and the research described

(4) Inappropriate citations

(5) Incoherent, meaningless and/or irrelevant content included in the article

(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] Q. Wan and L. Ye, "Career Recommendation for College Students Based on Deep Learning and Machine Learning," *Scientific Programming*, vol. 2022, Article ID 3437139, 10 pages, 2022.

Retraction

Retracted: Research on the Generation and Design Method of Residential Community Scheme Based on Improving Design

Scientific Programming

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope

(2) Discrepancies in the description of the research reported

(3) Discrepancies between the availability of data and the research described

(4) Inappropriate citations

(5) Incoherent, meaningless and/or irrelevant content included in the article

(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] C. Li and L. Wang, "Research on the Generation and Design Method of Residential Community Scheme Based on Improving Design," *Scientific Programming*, vol. 2021, Article ID 5053494, 9 pages, 2021.

Retraction

Retracted: Construction of the Open Oral Evaluation Model Based on the Neural Network

Scientific Programming

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope

(2) Discrepancies in the description of the research reported

(3) Discrepancies between the availability of data and the research described

(4) Inappropriate citations

(5) Incoherent, meaningless and/or irrelevant content included in the article

(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] Z. Chen, X. Zhang, Z. Li, and A. Li, "Construction of the Open Oral Evaluation Model Based on the Neural Network," *Scientific Programming*, vol. 2021, Article ID 3928246, 11 pages, 2021.

Retraction

Retracted: Willingness and Evaluation Model of College Students' Online Learning Behavior Based on Distributed Cognition

Scientific Programming

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope

(2) Discrepancies in the description of the research reported

(3) Discrepancies between the availability of data and the research described

(4) Inappropriate citations

(5) Incoherent, meaningless and/or irrelevant content included in the article

(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] L. Chen and S. Huang, "Willingness and Evaluation Model of College Students' Online Learning Behavior Based on Distributed Cognition," *Scientific Programming*, vol. 2021, Article ID 6386455, 12 pages, 2021.

Retraction

Retracted: Apparel Design and Development Based on 3D Scanning Technology

Scientific Programming

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope

(2) Discrepancies in the description of the research reported

(3) Discrepancies between the availability of data and the research described

(4) Inappropriate citations

(5) Incoherent, meaningless and/or irrelevant content included in the article

(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] G. Zhang, "Apparel Design and Development Based on 3D Scanning Technology," *Scientific Programming*, vol. 2021, Article ID 7933206, 7 pages, 2021.

Retraction

Retracted: Analysis of the 3D Application Evaluation System of Landscape Based on Hybrid Cooperation of VR and AI

Scientific Programming

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope

(2) Discrepancies in the description of the research reported

(3) Discrepancies between the availability of data and the research described

(4) Inappropriate citations

(5) Incoherent, meaningless and/or irrelevant content included in the article

(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] L. Niu and F. Han, "Analysis of the 3D Application Evaluation System of Landscape Based on Hybrid Cooperation of VR and AI," *Scientific Programming*, vol. 2021, Article ID 4595808, 11 pages, 2021.

Retraction

Retracted: Mind Map Construction for English Grammar Teaching Based on Knowledge Map

Scientific Programming

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] Q. Fu and S. Kuang, "Mind Map Construction for English Grammar Teaching Based on Knowledge Map," *Scientific Programming*, vol. 2021, Article ID 4921825, 10 pages, 2021.

Retraction

Retracted: Research on Intelligent English Translation Method Based on the Improved Attention Mechanism Model

Scientific Programming

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope

(2) Discrepancies in the description of the research reported

(3) Discrepancies between the availability of data and the research described

(4) Inappropriate citations

(5) Incoherent, meaningless and/or irrelevant content included in the article

(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] R. Wang, "Research on Intelligent English Translation Method Based on the Improved Attention Mechanism Model," *Scientific Programming*, vol. 2021, Article ID 9667255, 8 pages, 2021.

Retraction

Retracted: Application Research on Roller Skater Detection, Tracking, and Trajectory Prediction Based on Video Stream

Scientific Programming

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] S. Duan, L. Meng, D. Ma, and L. Mi, "Application Research on Roller Skater Detection, Tracking, and Trajectory Prediction Based on Video Stream," *Scientific Programming*, vol. 2021, Article ID 2702272, 9 pages, 2021.

Retraction

Retracted: Application of Thematic Context-Based Deep Learning in Foreign Language Teaching

Scientific Programming

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] Q. Zhao, "Application of Thematic Context-Based Deep Learning in Foreign Language Teaching," *Scientific Programming*, vol. 2021, Article ID 8664219, 11 pages, 2021.

Hindawi

Retraction

Retracted: Construction of College English Mobile Learning Model Based on Distributed Terminal

Scientific Programming

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] J. Wang and L. Lei, "Construction of College English Mobile Learning Model Based on Distributed Terminal," *Scientific Programming*, vol. 2021, Article ID 2220096, 7 pages, 2021.

Retraction

Retracted: An Intelligent Evaluation Method of Information Course Teaching Effect Based on Image Analysis

Scientific Programming

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope

(2) Discrepancies in the description of the research reported

(3) Discrepancies between the availability of data and the research described

(4) Inappropriate citations

(5) Incoherent, meaningless and/or irrelevant content included in the article

(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] H. Chen and J. Zhang, "An Intelligent Evaluation Method of Information Course Teaching Effect Based on Image Analysis," *Scientific Programming*, vol. 2021, Article ID 3200865, 9 pages, 2021.

Hindawi

Retraction

Retracted: Design of Online Music Teaching System Based on B/S Architecture

Scientific Programming

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope

(2) Discrepancies in the description of the research reported

(3) Discrepancies between the availability of data and the research described

(4) Inappropriate citations

(5) Incoherent, meaningless and/or irrelevant content included in the article

(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] D. D. Dai, "Design of Online Music Teaching System Based on B/S Architecture," *Scientific Programming*, vol. 2021, Article ID 1297019, 6 pages, 2021.

Hindawi

Retraction

Retracted: Construction of Machine Learning Model Based on Text Mining and Ranking of Meituan Merchants

Scientific Programming

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] Y. Tang, D. Liao, S. Huang, Q. Fan, and L. Liu, "Construction of Machine Learning Model Based on Text Mining and Ranking of Meituan Merchants," *Scientific Programming*, vol. 2021, Article ID 5165115, 9 pages, 2021.

Research Article

A Target Recognition Method Based on Multiview Infrared Images

Junyi Zhang ⬤ and Yuan Rao

School of Software Engineering, Xi'an Jiaotong University, Xi'an 710049, China

Correspondence should be addressed to Junyi Zhang; zhangjunyi0806@stu.xjtu.edu.cn

Infrared image target recognition provides an important means of night traffic management and battlefield environment monitoring. With the improvement of the performance of infrared sensors and the popularization of applications, it becomes possible to obtain multiview infrared images of the same target in the same scene. A target recognition method combining multiview infrared images is proposed. At first, the internal correlation analysis of multiview infrared images is performed based on the nonlinear correlation information entropy (NCIE). The view subset from all the multiview images with the largest NCIE is selected as candidate samples for the subsequent target recognition. The joint sparse representation (JSR) is used to classify all infrared images in the candidate view subset. JSR can effectively investigate the internal correlation of multiple related sparse representation problems and improve the reconstruction accuracy and classification capabilities. In the experiments, the tests are performed on the collected infrared images of multiple types of traffic vehicles, under the conditions of original, noisy, and occluded samples. The effectiveness and robustness of the proposed method can be verified by comparative analysis.

1. Introduction

Compared with visible light observation, infrared imaging can work in night scenario, providing a powerful tool for all-day monitoring, which is widely used in military and civilian fields [1–4]. In the military field, the use of infrared imaging and processing can assist in monitoring the battlefield environment at night to achieve target recognition and precision strikes. In the civil field, infrared imaging can be used for night traffic control. It can accurately analyze and identify the thermal effects of different types of vehicles and provide auxiliary decision making for drivers at night. Therefore, the classification and identification of typical objects has important meaning in both military and civilian fields. At present, the research on infrared image vehicle recognition is mainly based on the classic pattern recognition ideas, generally using a two-phase procedure: feature extraction and classifier. In the phase of feature extraction, researchers employed or developed various algorithms [5–9], including the intensity- or geometric-based ones such as histogram of oriented gradients (HOG), region moment, target boundary, and so on. In general, all these methods involve manually extracted features. It usually requires professional knowledge to design these features in order to maintain the effectiveness. However, the design process has some uncertainties, so the discrimination is often limited. In terms of the classifiers, the infrared image recognition, like other pattern recognition problems, mainly employs classical and robust classifiers [10–12], such as support vector machines (SVMs), neural networks, and sparse representation-based classification (SRC). With the development of deep learning theory [13–16], different types of deep learning models have also been applied in the field of infrared image target recognition, and their effectiveness has also been verified [17–21].

The development and maturity of infrared sensing technology provide rich samples for target observation and identification. In view of the target recognition problem, it is possible to obtain infrared images of the same target from different aspects in the same scene. In this sense, it has become an effective technical approach to improve the recognition accuracy and robustness by combining multiview infrared images for comprehensive analysis. In this paper, a multiview infrared target recognition method is

proposed. First, the multiview infrared image of the same target is analyzed based on the nonlinear correlation information entropy (NCIE) [22–25]. The NCIE reflects the inner relevance of the selected views. Therefore, the view subset with the strongest correlation can be obtained according to the principle of the maximum NCIE. To exploit such correlations, this paper uses joint sparse representation (JSR) [26–30] as the classification algorithm in the classification stage to determine the target label according to the overall reconstruction errors of the multiview infrared images. In the experiment, the proposed method is tested and comparatively analyzed based on the infrared image set of several types of traffic vehicle targets. The results confirm the effectiveness and superiority of the proposed method.

2. Selection of Candidate Views Using NCIE

In order to effectively analyze and screen multiview infrared images, this paper chooses NCIE as the basic evaluation criterion to measure the internal correlation of different views [22–25]. First, the traditional image correlation coefficient is adopted as the similarity measure between two infrared images. Afterwards, the correlation matrix between different views is constructed as follows:

$$R = E + \widetilde{R}$$

$$= \begin{bmatrix} 1 & r_{12} & \cdots & r_{1N} \\ r_{21} & 1 & \cdots & r_{2N} \\ \vdots & \vdots & \vdots \\ r_{N1} & r_{N2} & \cdots & 1 \end{bmatrix}, \tag{1}$$

where E is an identity matrix and $\widetilde{R}$ represents the cross-correlation matrix between the infrared images from N different views. According to the eigenvalues λ_t of R, the NCIE denoted by H_R is defined as follows:

$$H_R = 1 + \sum_{t=1}^{T} \frac{\lambda_t}{T} \log_T \frac{\lambda_t}{T}. \tag{2}$$

According to equation (2), when all the images from different views share completely different distributions, the correlation coefficient matrix is a unit with all the eigenvalues as 1. At this time, the NCIE is the minimum of 0. When the similarity between different is larger than 0, the eigenvalues of the correlation coefficient matrix are no longer equal. When different views share exactly the same distribution, the NCIE is the maximum of 1. Therefore, according to the resulting NCIE, the inherent correlations between a certain number of views can be obtained.

In this paper, the NCIE defined in equation (2) is used to select the optimal view subset. For M infrared images from different views, they are randomly combined to obtain P subsets. Afterwards, the NCIEs of different subsets are calculated. Finally, based on the principle of the maximum entropy, the optimal subset can be found. The multiview images in the selected subset share strong internal correlation and can be effectively used in subsequent target recognition.

3. Sparse Representation for Target Recognition

3.1. SRC. Sparse representation is developed based on the linear representation theory, and the representation accuracy is improved by introducing sparse constraints. Specifically, in the field of target recognition, SRC uses training samples to build a global dictionary $A = [A_1, A_2, \ldots, A_C] \in R^{d \times N}$, where A_i represents the N_i atoms corresponding to the training sample in the ith class [31–33]. The test sample y is represented based on the global dictionary as follows:

$$\widehat{x} = \arg \min_x \|x\|_0 \text{ s.t.} \|y - Ax\|_2^2 \leq \varepsilon, \tag{3}$$

where x is the sparse coefficient vector to be solved and ε is the set error threshold.

By solving the problem in equation (3), the sparse representation coefficient vector can be obtained. On this basis, the reconstruction error calculation is performed for each training class, as shown in the following equation:

$$r(i) = \|y - 0 A_i x_i\|_2^2 \, (i = 1, 2, \ldots, C), \tag{4}$$

where x_i represents the part of coefficient vector corresponding to the ith class. Accordingly, the decision of the target label can be made by comparing the errors from different training classes.

3.2. JSR. For multiple related sparse representation problems, when they are solved independently according to the traditional sparse representations, their correlation information cannot be properly considered. As a remedy, researchers proposed the JSR model to solve multiple sparse representation problems simultaneously under a unified framework. Taking three related inputs as an example, denoted as $y^{(1)}, y^{(2)}, y^{(3)}$, the problem of JSR can be preliminarily expressed as follows:

$$\min_\beta \left\{ g(\beta) = \sum_{k=1}^{3} \left\| y^{(k)} - A^{(k)} \alpha^{(k)} \right\| \right\}, \tag{5}$$

where $A^{(k)}$ is the global dictionary corresponding to the kth ($k = 1, 2, 3$) input; $\alpha^{(k)}$ is the corresponding coefficient vector; and $\beta = \begin{bmatrix} \alpha^{(1)} & \alpha^{(2)} & \alpha^{(3)} \end{bmatrix}$ is the coefficient matrix.

It can be seen from equation (5) that although the representation process of different problems is examined uniformly, the results have no differences with the independent solutions. Also, there is no consideration of the correlation between different inputs. For this reason, the JSR model restricts the structure and distribution of the coefficient matrix and updates the objective function as follows:

$$\min_\beta g(\beta) + \lambda \|\beta\|_{2,1}, \tag{6}$$

where λ is a positive parameter. The correlation between different inputs can be reflected through the constraint of ℓ_1/ℓ_2 norm on the matrix β.

Algorithms such as orthogonal matching pursuit or multitask compressive sensing can be used to solve the

optimization problem in equation (6) [34, 35]. Afterwards, the reconstruction errors of different classes for the three inputs can be calculated, respectively. The final decision is made according to the principle of the minimum error as follows:

$$\text{identity}(y) = \min_i \sum_{k=1}^{3} \left\| y^{(k)} - A_i^{(k)} \alpha_i^{(k)} \right\|, \qquad (7)$$

where $A_i^{(k)}$ is the dictionary corresponding to the kth input and the ith class and $\alpha_i^{(k)}$ is the corresponding coefficient vector.

According to the specific steps of the proposed method, the recognition process shown in Figure 1 is designed. First, the multiview infrared images of the same target are analyzed based on the NCIE to obtain a subset of views for subsequent classification. Afterwards, for the chosen views, the JSR is employed to obtain the corresponding reconstruction errors of different training classes. Finally, the target label is determined based on the comparison of different reconstruction errors.

4. Experiments and Analysis

4.1. Preparation.
The proposed method is tested based on the infrared dataset of several traffic vehicles. These images are acquired by night infrared sensors and have a certain degree of randomness. After preprocessing, 2000 bus images, 3000 car images, 1200 truck images, and 1600 pickup truck images are obtained. Images of the four types of targets are all collected in real-world conditions by different sensors from multiple aspects. In the experiments, half of the samples of various targets are randomly selected for training, and the remaining ones are used as test samples.

During the experiment, 8 views of infrared images are selected as typical multiview conditions to test the performance of the proposed method. The view subset is selected based on the NCIE. At the same time, four types of comparison methods are selected from the existing literature, including the HOG-based method, target boundary-based method, SRC-based method, and CNN-based method. The average recognition rate is used as the measurement criterion of the recognition accuracy, which is defined as the proportion of the number of correctly recognized samples in the total test samples.

4.2. Original Samples.
First, the original images of the four types of targets are recognized. The original infrared images of the targets have good visibility and quality after preprocessing, and the distinction between various types of targets is strong, so the recognition difficulty is relatively low. Table 1 shows the specific recognition results of the proposed method for the four types of targets. The recognition rates of buses, cars, trucks, and pickup trucks are 96.25%, 96.80%, 96.67%, and 96.63%, respectively. After calculation, the average recognition rate of the proposed method under current condition is 96.60%. Table 2 lists the average

recognition rates of various methods. The comparison shows that the proposed method achieves the highest average recognition rate. On the one hand, this paper uses multiple complementary information from different views, which is more discriminatory than the traditional single view. On the other hand, this paper also explores the internal correlations of multiview images, so the recognition performance can be further improved. Among the four types of comparison methods, the deep learning method has certain advantages, which shows the effectiveness of the deep networks and the deep features for infrared target recognition.

4.3. Noisy Samples.
Like other types of images, the infrared image acquisition process is also susceptible to noise, which leads to a decrease in the overall signal-to-noise ratio (SNR) and brings obstacles to correct recognition. In this experiment, we obtained the test set at different SNRs by adding noises. Specifically, the overall energy of the original image to be processed is calculated, and the Gaussian white noises are generated according to a preset SNR, which are added to the original image to obtain a corresponding noisy sample. For the test sets with different noise levels, the proposed method and the four types of reference methods classify them, and the statistical results are shown in Figure 2. It can be seen that the noise interference has a significant impact on the performance of various methods. In contrast, the proposed method maintains the highest average recognition rates under various noise conditions, showing its robustness. From the results of the four types of reference methods, the performance of the SRC-based method is much better, reflecting the noise robustness of sparse representation. The proposed method combines the complementarity of multiple views and the robustness of sparse representation of noise to further improve the overall recognition performance.

4.4. Occluded Samples.
Despite the influence of noises, in reality, due to the influence of occlusions, the target may not be completely reflected in the acquired infrared image. For this reason, the effectiveness of the recognition method under occlusion conditions is very important. In the simulation process, this paper takes the complete target in the test set as a reference and occludes some of its areas, respectively. The occlusion level is defined according to proportion of the target being occluded. We specifically construct four occlusion levels, i.e., 10%, 30%, 50%, and 70%. Afterwards, the average recognition rates of different methods are obtained as shown in Figure 3. Similar to the results of noise interference, the proposed method has the best performance under the current test condition. From the results of the reference methods, it can also be seen that the sparse representation is more robust to partial occlusions. The multiple views obtained in this paper have good complementarity and can effectively improve the robustness to the occlusion conditions. Furthermore, the sparse representation enhances the overall robustness to occluded samples of the proposed method.
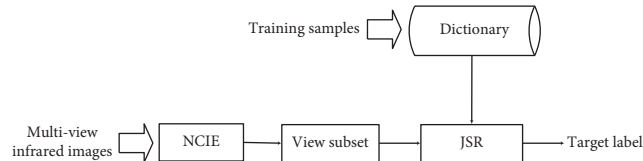
Figure 1: Process of infrared image target recognition based on multiple views.

Table 1: Recognition results of the proposed method on the original samples.

Class	Bus	Car	Truck	Pickup trucks	Recognition rate (%)
Bus	1925	24	37	14	96.25
Car	13	2904	27	56	96.80
Truck	27	7	1160	6	96.67
Pickup trucks	13	32	9	1546	96.63
Average recognition rate (%)					96.60

Table 2: Performance comparison of different methods on original samples.

Method type	Average recognition rate (%)
Proposed	96.60
HOG	94.98
Target boundary	95.12
SRC	95.23
CNN	96.02
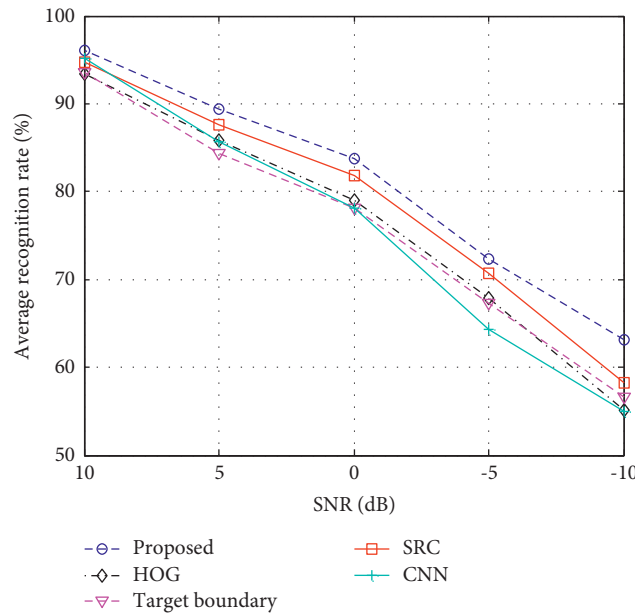
Figure 2: Performance comparison of different methods on noisy samples.
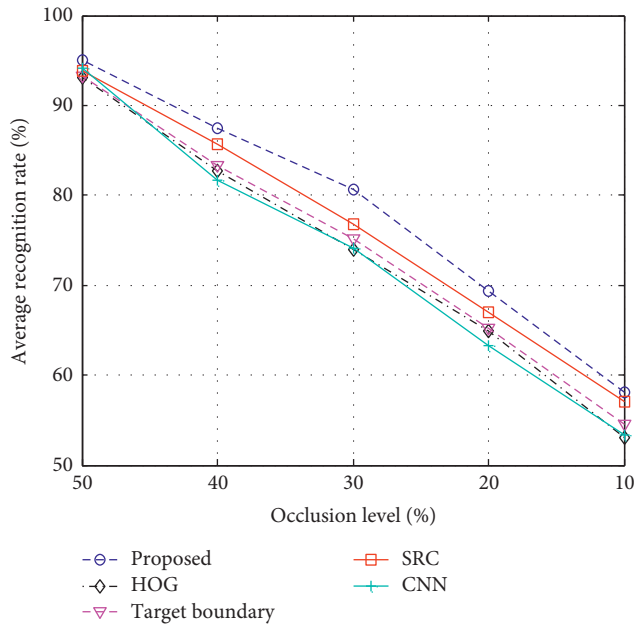
Figure 3: Performance comparison of different methods on occluded samples.

5. Conclusion

This paper proposes a multiview infrared image target recognition method. The different views of the same target can reflect the characteristics of the target from different aspects. The proposed method first uses classical image correlation and NCIE as a criterion to obtain a view subset with images of high correlations. The JSR model is used to analyze the images in the chosen subset, and the overall representation accuracy is improved by investigating the inner correlation. Finally, based on reconstructions from JSR, a reliable recognition result can be reached. The experiment is carried out with multiview infrared images of four types of traffic vehicles as the training and test sets. The original, noisy, and occluded samples are tested, respectively. According to the experimental results, the proposed method is more effective than some reference methods.

Data Availability

The dataset used can be accessed upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

References

[1] X. Dai, Y. Duan, J. Hu et al., "Near infrared nighttime road pedestrians recognition based on convolutional neural network," *Infrared Physics & Technology*, vol. 97, pp. 25–32, 2019.

[2] C. Gao, Y. Du, J. Liu, J. Lv, L. Yang, and D. Meng, "InfAR dataset: infrared action recognition at different times," *Neurocomputing*, vol. 212, pp. 36–47, 2016.

[3] H. Deng, X. Sun, M. Liu, C. Ye, and X. Zhou, "Small infrared target detection based on weighted local difference measure," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 7, pp. 4204–4214, 2016.

[4] S. Kim, W.-J. Song, and S.-H. Kim, "Infrared variation optimized deep convolutional neural network for robust automatic ground target recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 195–202, IEEE, Honolulu, HI, USA, July 2017.

[5] H. Wang, X. Yang, and H. Ding, "Method of features extraction for infrared image recognition based on image moment," in *Proceedings of the International Conference on Computer Application and System Modeling*, pp. 443–446, IEEE, Taiyuan, China, October 2010.

[6] S.-G. Sun, "Automatic target recognition using boundary partitioning and invariant features in forward-looking infrared images," *Optical Engineering*, vol. 42, no. 2, pp. 524–534, 2003.

[7] M. N. A. Khan, G. Fan, D. R. Heisterkamp, and L. Yu, "Automatic target recognition in infrared imagery using dense hog features and relevance grouping of vocabulary," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 293–298, IEEE, Columbus, OH, USA, June 2014.

[8] Y. Cho, S. Shin, S. Yim, K. Kong, H. W. Cho, and W. J. Song, "Multistage fusion with dissimilarity regularization for SAR/IR target recognition," *IEEE Access*, vol. 7, p. 728, 2019.

[9] G. Lin, G. Fan, L. Yu, X. Kang, and E. Zhang, "Heterogeneous structure fusion for target recognition in infrared imagery," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 118–125, IEEE, Boston, MA, USA, June 2015.

[10] A. Apatean, A. Rogozan, and A. Bensrhair, "SVM-based obstacle classification in visible and infrared images," in *Proceedings of the IEEE European Signal Processing Conference*, pp. 293–297, IEEE, Baden-Baden, Germany, August 2009.

[11] C. Mu, J. Wang, Z. Yuan, X. Zhang, and C. Han, "The research of the ATR system based on infrared images and L-M BP neural network," in *Proceedings of the 7th International Conference on Image and Graphics*, pp. 801–805, IEEE, Qingdao, China, July 2013.

[12] S. Zhang, J. Gong, D. Chen, L. Xu, and L. Yan, "Sparsity-motivated multi-scale histograms of oriented gradients feature for SRC," in *Proceedings of the IEEE International Conference on Unmanned Systems (ICUS)*, pp. 389–393, IEEE, Beijing, China, October 2017.

[13] X. Zhu, D. Tuia, L. Mou et al., "Deep learning in remote sensing: a comprehensive review and list of resources," *IEEE Geoscience and Remote Sensing Magazine*, vol. 5, no. 4, pp. 8–36, 2017.

[14] S. Chen, H. Wang, F. Xu, and Y. Jin, "Target classification using the deep convolutional networks for SAR images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 6, pp. 1685–1697, 2016.

[15] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the 2016 Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, IEEE, Las Vegas, NV, USA, June 2016.

[16] L. Xu and Q. Chen, "Remote-sensing image usability assessment based on ResNet by combining edge and texture maps," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 12, no. 6, pp. 1825–1834, 2016.

[17] A. Akula, A. K. Shah, and R. Ghosh, "Deep learning approach for human action recognition in infrared images," *Cognitive Systems Research*, vol. 50, pp. 146–154, 2018.

[18] Z. Ding, N. M. Nasrabadi, and Y. Fu, "Deep transfer learning for automatic target classification: MWIR to LWIR," *Proceedings of SPIE International Society for Optics and Photonics*, vol. 9844, Article ID 984408, 2016.

[19] A. D'Acremont, R. Fablet, A. Baussard, and G. Quin, "CNN-based target recognition and identification for infrared imaging in defense systems," *Sensors*, vol. 19, p. 2040, 2019.

[20] F. He, X. Hu, B. Liu, and Z. Decai, "Infrared image recognition technology based on visual processing and deep learning," in *Proceedings of the Chinese Automation Congress (CAC)*, pp. 641–645, IEEE, Shanghai, China, November 2020.

[21] A. Akula and H. K. Sardana, "Deep CNN-based feature extractor for target recognition in thermal images," in *Proceedings of the 2019 IEEE Region 10 Conference (TENCON)*, pp. 2370–2375, IEEE, Kochi, India, October 2019.

[22] H. Wang and X. Yao, "Objective reduction based on nonlinear correlation information entropy," *Methodologies and Application*, vol. 20, pp. 2393–2407, 2016.

[23] Z. Shen, Y. Shen, and Q. Wang, "Medical ultrasound signal denoise based on ensemble empirical mode decomposition and nonlinear correlation information entropy," in *Proceedings of the 2009 IEEE Youth Conference on Information, Computing and Telecommunication*, pp. 19–22, IEEE, Beijing, China, September 2009.

[24] Q. Wang, Y. Shen, and J. Q. Zhang, "A nonlinear correlation measure for multivariable data set," *Physica D: Nonlinear Phenomena*, vol. 200, pp. 287–295, 2005.

[25] E. Pereda, R. Q. Quiroga, and J. Bhattacharya, "Nonlinear multivariate analysis of neurophysiological signals," *Prog Neurobiol*, vol. 77, no. 1, pp. 1–37, 2005.

[26] H. Zhang, N. Nasrabadi, Y. Zhang, and T. Huang, "Multi-view automatic target recognition using joint sparse representation," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 48, no. 3, pp. 2481–2497, 2012.

[27] G. Dong, G. Kuang, N. Wang, L. Zhao, and J. Lu, "SAR target recognition via joint sparse representation of monogenic signal," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 8, no. 7, pp. 3316–3328, 2015.

[28] S. Liu and J. Yang, "Target recognition in synthetic aperture radar images via joint multifeature decision fusion," *Journal of Applied Remote Sensing*, vol. 12, no. 1, Article ID 016012, 2018.

[29] G. Dong and G. Kuang, "Classification on the monogenic scale space: application to target recognition in SAR image," *IEEE Transactions on Image Processing*, vol. 24, no. 8, pp. 2527–2538, 2015.

[30] B. Ding and G. Wen, "Exploiting multi-view SAR images for robust target recognition," *Remote Sensing*, vol. 9, p. 1150, 2017.

[31] J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 2, pp. 210–227, 2009.

[32] J. J. Thiagaraianm, K. N. Ramamurthy, P. Knee, A. Spanias, and V. Berisha, "Sparse representations for automatic target classification in SAR images," in *Proceedings of the 4th International Symposium on Communications, Control and Signal Processing (ISCCSP)*, pp. 1–4, IEEE, Limassol, Cyprus, March 2010.

[33] H. Song, K. Ji, Y. Zhang, X. Xing, and H. Zou, "Sparse representation-based SAR image target classification on the 10-class MSTAR data set," *Applied Sciences*, vol. 6, no. 1, p. 26, 2016.

[34] J. A. Tropp, A. C. Gilbert, and M. J. Strauss, "Algorithms for simultaneous sparse approximation," *EURASIP Journal on Applied Signal Processing*, vol. 86, no. 3, pp. 589–602, 2006.

[35] S. Ji, D. Dunson, and L. Carin, "Multitask compressive sensing," *IEEE Transactions on Signal Processing*, vol. 57, no. 1, pp. 92–106, 2009.

Research Article

The Stereo Communication Method of Artistic Visual Image Based on CAD Platform

Kai Gao [ID] **and Yong Wang**

Department of Art Design, School of Education and Art, Shandong Institute of Petroleum and Chemical Technology, Dongying City, Shandong Province 257061, China

Correspondence should be addressed to Kai Gao; 000420@slcupc.edu.cn

Due to the large amount of noise in the image background, the traditional methods have the problems of poor image communication effect, slow image communication efficiency, and poor noise processing effect. A three-dimensional communication method of art visual image based on CAD platform is proposed. Wavelet transform is used to extract artistic visual image features, and the fuzzy clustering method is used to standardize the obtained image feature data. Based on the data processing results, the CAD platform is constructed, the software program is designed, and the image chain coding is realized at the same time. The particle swarm optimization algorithm is used to convey the image of the obtained CAD general vectorization file, so as to complete the three-dimensional transmission of artistic visual image. The experimental results show that the image communication effect of this method is better, the image communication efficiency is higher, the noise suppression effect is better, the visual communication effect of the image is improved, and the performance of the designed method is superior.

1. Introduction

In recent years, the worldwide informatization process has been accelerating, and the development of all kinds of data has shown a soaring trend. As a key type of data in all kinds of information, image data can more intuitively record the state of human life. The subsequent technical problems of data analysis have gradually attracted people's attention, which has promoted the development of image visual communication research [1–3]. In the research of traditional image visual communication methods, a more scientific image extraction method is selected to obtain the image data, continuously track the data information, ensure the safety of the data in the acquisition process, and use the multifunctional algorithm to analyze the collected information, so as to enhance the reliability of the data information [4–6]. The data protection operation according to the corresponding data management rules improves the effect of image visual communication, but the processing effect of image features in the operation process is poor, and the image data information of the remaining part cannot be

mastered. Therefore, it is necessary to study effective image visual communication methods to analyze and solve the above problems [7, 8].

In order to improve the effect of image visual communication, after a long time of research and development, some image communication methods have been formed. Reference [9] proposes an optimization method of plane visual communication effect based on wavelet change. The image is decomposed by wavelet and reconstructed by wavelet. In the process of reconstruction, the modulus diagram and phase angle diagram are calculated, and the edge images of each scale are extracted. Through the edge image, the corresponding edge points of the semireconstructed image are enhanced. On the basis of the above, the SLIP model is used to calculate the graphic beautification vector. In order to simplify the operation, the above calculation method is converted into a simple mathematical operation, and the image visual transmission effect is optimized through the reflected light graphic mode. The experimental results show that this method effectively optimizes the cost of image plane visual transmission, but some image details

are not clear and the visual effect is not good. Reference [10] proposes a plane visual communication design method based on user experience effect. Computer vision imaging technology is used to sample the image information of plane visual communication, combined with the edge contour extraction method to detect the image contour of plane visual communication design, and extract the multiscale local structure feature information of the plane visual image. According to the needs of user experience effect, the boundary feature detection and multilevel structure decomposition in the process of plane visual communication design are carried out, the low-level visual structure of plane visual communication design is reconstructed by the method of adjacent pixel information fusion, and the user experience effect evaluation model of plane visual design image is established. According to the user experience effect, realize the optimization of plane visual communication design. The simulation test results show that this method has a good user experience effect and improves the design effect of plane visual communication, but the noise in the image background will affect the visual communication results. Reference [11] designed a visual communication system of animated character graphics and images in a virtual reality environment. In the hardware design, various adaptation parameters of the renderer motherboard were designed to optimize the experience of visual communication. In the software design, by introducing a Sobel edge operator, the gray function is established to solve the gradient amplitude, and the threshold is selected to compare it to complete the recognition and thinning of the edge data of animated character graphics and images. Design the motion capture module, establish the behavior control model, generate and manage the motion capture files, and complete the overall design of the system. The effectiveness of the system is verified by simulation experiments, but the efficiency of image visual communication is not high due to the large number of images.

Considering the large number of images and the noise in the image background, in order to improve the image communication effect, improve the image communication efficiency, and reduce the influence of noise factors on images, a three-dimensional communication method of artistic visual image based on the CAD platform is proposed.

2. Art Visual Image Preprocessing

There is a certain complementary relationship between the three-dimensional communication effect of an artistic visual image and color, among which the most obvious characteristics are reflected in the matching of color and the presentation of light feeling. The beauty of visual art is mainly reflected in its certain color openness, which is also the embodiment of a free concept in aesthetics and the theoretical basis of the three-dimensional communication method of artistic visual image.

2.1. Feature Extraction of Art Visual Image Based on Wavelet Transform. Visual art is a kind of humanities juxtaposed with architecture, music, literature, film, and other forms. It is an art form that can be perceived and understood by vision. It is a language means based on vision and guided by artistic visual images, such as paintings, photographic pictures, and film and television pictures. All visual artworks and images have highly important language as the intervention point. Artistic visual image is a visual cultural symbol that visually presents the law of art, artistic concept, artistic style, and characteristics of the times. Artistic visual images are full of deep cultural connotations. Image refers to the picture with visual effect. It is a way of information transmission different from words created by human beings through imitation or imagination of the natural world from the level of visual cognition and feeling. It is one of the most commonly used information carriers in human social activities. Artistic image is a kind of beautiful creation. It visually records the process of human civilization and is the main part of visual art. Its production and consumption are always the most basic communication activities in human social life and an invisible force to condense the society. Image chain coding is a binary image representation method based on the region boundary of binary image. This method can transform two-dimensional image into one-dimensional digital sequence. For large-scale artistic visual images, image chain coding can greatly save storage space and improve processing speed. The chain coding diagram in four directions is shown in Figure 1.

In Figure 1, the coding of using edges to calibrate the image area is image chain coding. Four chain codes 1, 2, 3, and 4 are used to represent the up, left, down, and right directions. Starting from any pixel on the image boundary, walk counterclockwise along the boundary pixel of the image and return to the starting point. The walking direction on the recording edge forms an ordered chain composed of direction chain code. With the starting point coordinates, the image boundary can be uniquely determined.

Wavelet transform is a local transform of time and frequency. It has the characteristics of multiresolution analysis and has the ability to characterize the local characteristics of signals in time domain and frequency domain. Wavelet transform can concentrate the energy of the original image on a small part of wavelet coefficients, and the decomposed wavelet coefficients have high local correlation in the detail components in three directions, which provides a powerful condition for feature extraction. Texture feature extraction using wavelet transform has been widely used in texture analysis, image compression, and surface defect detection of industrial products. Wavelet transform is often used for signal multiresolution decomposition [12], assuming the signal is $g(x)$; the discrete wavelet decomposition formula is

$$g(x) == \sum_{i=1}^{n} e(x_i) \times \sum_{j=1}^{n} e(x_j), \tag{1}$$

where $e(x_i)$ represents the scaling function, $e(x_j)$ represents the wavelet function, i represents the scaling coefficient, and j represents the wavelet coefficient.

As a two-dimensional function, the image is constructed by one two-dimensional scaling function $\mu(a, b)$ and three
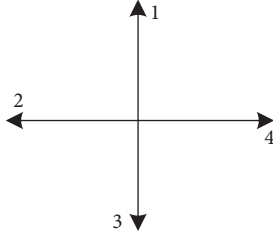
FIGURE 1: Chain coding diagram in four directions.

two-dimensional wavelet functions $\vartheta^S(a, b)$, $\vartheta^K(a, b)$, and $\vartheta^L(a, b)$. The wavelet function measures the change of image gray in different directions: $\vartheta^S(a, b)$ measures the change of gray along the column direction, $\vartheta^K(a, b)$ measures the change of gray along the row direction, and $\vartheta^L(a, b)$ measures the change of gray along the diagonal direction [13]. Therefore, the discrete wavelet transform of the image $F(a, b)$ of size $m \times n$ is

$$F(a, b) = \sum_{k=1}^{n} e(k) + \mu[e(k) + e(k-1)], \quad (2)$$

where $e(k)$ represents the high-frequency coefficient and $e(k-1)$ represents the low-frequency coefficient.

According to formula (2), it can be seen that the wavelet transform decomposes the original image into a low-frequency part and a high-frequency part, and different levels of wavelet decomposition can obtain low-frequency images of different scales. After the k-level wavelet decomposition, reset the high-frequency part to 0, and only keep the low-frequency part, reconstruct the original image according to the modified high-frequency information and the unmodified low-frequency information, and obtain an approximate image $F'(a, b)$ of the original image $F(a, b)$. This image contains the low-frequency information of the original image [14]. In the logarithmic domain, the difference between the original image and the approximate image is used to obtain the reflection component map $R(a, b)$ of the original image. Assuming that the original image undergoes K-level wavelet decomposition, K_{ab} approximate images at different scales will be generated, and then, R_{ab} reflection component maps will be obtained, combined with R_{ab} reflection component maps to construct a visual image multiscale reflection model [15], that is,

$$R_{ab} = \sqrt{\frac{\sum_{i=1}^{m} \sum_{j=1}^{n} \left[r_{ij}(x) - g(x)\right]^2}{dx}}, \quad (3)$$

where $r_{ij}(x)$ represents a nonlinear function, and this article chooses the arctangent function, because the arctangent function can reduce high-frequency noise to a certain extent.

The new media era is an image-centered information age. The mainstream consumption form of visual culture has been transformed into the production and consumption of artistic images based on digital and Internet technology. It can be said that digital and Internet technology is the technical basis of the new media era. It is leading to a revolution of visual culture, affecting people's production and consumption of artistic images at an unimaginable speed, thus imperceptibly changing people's inherent values and social ideology. While art image consumption has gradually become a popular way in this era, it is also pregnant with new art image forms and art language, and is tending to create art image culture that meets the needs of the new media era. In the new media era, the consumer of visual culture has gradually transformed into an image-centered consumer behavior. Due to the convenience of digital and Internet technology for visual information transmission, it greatly improves its superiority for carrying image text, so as to transform the traditional consumption mode of visual culture into the consumption mode dominated by image.

In the high-frequency information obtained by wavelet decomposition of the original image, most of the information corresponds to the reflection components of the objects in the image scene, and the high-frequency information corresponding to the shadow part is relatively small. Therefore, the high-frequency information representing the shadow part can be regarded as the noise in the high-frequency information, and the noise can be detected by using the wavelet-based search method [16–18].

Let $U_{(p,v)}$ denote the wavelet coefficients obtained after the p-level wavelet decomposition of the artistic visual image, where $v = 0, 1, 2, 3$ corresponds to different subbands. For each pixel (a, b) in the high-frequency subband, the threshold T is used to determine whether it corresponds to the frequency of the shaded part.

$$\sigma_{(p,v)}(a, b) = \begin{cases} 1\left(U_{(p,v)}(a, b)\right)^2, & < T, \\ 0, & \text{others,} \end{cases} \quad (4)$$

where $\sigma_{(p,v)}$ represents the detection result of whether the pixels in the subband v represent shadows after the p-level decomposition of the artistic vision image. Among them, the calculation formula of the threshold T is

$$T = \frac{y_i^u - x_j^u}{r_a}, \quad (5)$$

where y_i^u represents the calculation result of the difference in the brightness value of the pixel; x_j^u represents the calculation result of the threshold value of the brightness of the pixel; and r_a represents the shadow intensity index.

According to the updated wavelet coefficients, reconstruct the approximate image $F'(a, b)$ of the original image $F(a, b)$, that is,

$$F'(a, b) = S^{-1} x^v x^u, \quad (6)$$

where S^{-1} represents the inverse wavelet transform operator.

The final artistic visual image features can be extracted by the multiscale reflection model:

$$S_h = \sum_{j=1}^{n} \partial\left[F(a, b) - F'_j(a, b)\right]. \quad (7)$$

The artistic visual image features are obtained by wavelet transform. These features are expressed as global features and local features. According to the artistic visual image

content, the global features can be divided into color features, texture features, shape features, and spatial features. Local features are stable enough and have good discrimination in artistic visual images; Local features and local features are important aspects of art visual image feature acquisition, which provide parameter basis for further image processing.

2.2. Data Processing of Artistic Vision Image Processing Results. Before the artistic visual image is communicated, the acquired image feature data are standardized. Because the actual result data obtained are not only different in dimension meaning, but also in data variables. Therefore, in order to improve the validity and authenticity of the data processing results, the problem of excessive data difference is solved from a mathematical point of view.

Use fuzzy clustering method to transform the data; then, the data processing process is expressed by the following formula:

$$G_k = \rho \varepsilon^2 - (R_\varepsilon + S_\varepsilon), \tag{8}$$

where ε^2 represents the sample data set matrix and ρ represents the initial variable. Through the analysis of this formula, the clustering data processing formula can be obtained, as follows:

$$V_j = \frac{\left[\sum_{i=1}^n \left(x_{ij} - \overline{x_{ij}} \right)^2 \right]}{N^2}, \tag{9}$$

where n represents the number of sample data processing times and V_j represents the sample data collection after clustering X_{ij}. Expand this formula to obtain the data cluster center; the specific formula is as follows:

$$X_{ij} = \frac{x_{ij} - \overline{x_{ij}}}{V_j}. \tag{10}$$

Using the above formula, the standard sample data in X_{ij} can be obtained, and the processed sample data matrix can be expressed as

$$X = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \dots & \dots & \dots & \dots \\ x_{m1} & x_{m2} & \dots & x_{mn} \end{bmatrix}. \tag{11}$$

Through the above process, the basic data processing is completed. In the process of artistic visual image processing and analysis, the data are processed hierarchically through data statistics and dimensionality reduction. Due to the variety and amount of data, the original sample data are processed dimensionless. The general data processing methods include standardized processing and average processing. In the design of this method, MATLAB 6.5 statistical software package is used as the carrier of data processing to complete the high-precision data processing process.

3. Stereo Communication Method of Artistic Visual Image Based on CAD Platform

Based on the processed artistic visual image data, carry out the visual communication operation of artistic visual image, and use the computer and its graphics equipment to help the designers carry out the design work. Computer-aided design (CAD) is an effective means for the quality control of industrial parts design, manufacturing, and data processing. It emphasizes human-computer interaction design, operation setting, graphic display, and so on [19–21]. The CAD platform usually displays terminal graphics input version, drawing, scanning, printer, tape drive, and various software platforms with interactive calculation with graphics function. The overall structure of the CAD platform adopts a typical client/server network structure. The server and all clients are connected to the LAN, and then the collaborative scheme of distributed and centralized is adopted; that is, most functions are distributed on each site, such as interface programs, CAD software, and common information and program components. Only the site loading management module and shared information module of the initiator or organizer are designed, which are expressed as certain server functions. The client and server first establish communication with their respective CAD platforms through the CAD platform interface function and then establish a network connection based on TCP protocol, which is convenient for the collaborative design program of the client and server to form the connection between the three-dimensional CAD platforms at both ends, and establish an artistic visual image to convey the design information. Since CAD technology can provide a feature source for visual recognition, the CAD platform can be used to comprehensively process images. Therefore, the CAD platform method is used to obtain the three-dimensional communication strategy of artistic visual images to complete different visual tasks.

3.1. CAD Platform Construction. In the traditional sense, related platforms can only process G code files during the design process. If there is a CAD/CAM interface beforehand, CAD standard dwg/dxf files can be preprocessed and converted into G code input files, and then handed over to the system deal with. But when the system encounters hand-drawn sketches, scanned images, engineering blueprints or digital photos, traditional systems lack effective recognition methods [22]. Therefore, this article focuses on the preprocessing technology and method of the CAD platform, including the database storage and extraction technology and the import and export graphics and image technology. The following is a specific discussion and analysis.

3.1.1. CAD Platform Software Program Design. In view of the defects and shortcomings of the previous CAD platform, a new CAD platform was further constructed and realized. It has some characteristics such as integration, interaction, popularization, networking, and parallelization. These are also the requirements of a new generation of software [23]. The CAD platform uses object-oriented technology, with

Visual C ++6.0 as the development tool, SQL 2000 as the data library support, AutoCAD 2000 for the production of the library, and C/S mode development. Figure 2 is a schematic diagram of the platform architecture.

According to Figure 2, the following is a detailed analysis of the database library storage and extraction technology and the import and export graphics and image technology of the designed CAD platform:

(1) Database storage and extraction technology: the CAD platform needs to call a large number of functional component symbols, browse, and call, using the client/server C/S (Client/Server) mode, and the graphic data module is stored on the server for remote computer access. Common database interfaces on Windows systems include ODBC (Open Database Interconnection), MFC (Microsoft Foundation Class), ODBC class, DAO (Data Access Object), RDO (Remote Data Object), OLEDB (Object Link Embedded Database), and ADO (ActiveX Data Object) [24]. ADO provides a database programming object model, simplifies OLEDB, belongs to the high-level database interface, and provides an automation interface. Therefore, use ADO mode to initialize OLE/COM library environment A fxO leInit (); the definition of ADO class is stored in ADO DLL (m sado15. d ll) as a resource, and then the ADO library file is imported with direct import symbol #import in stdafx.h file to enable the compiler to compile correctly.

(2) Import and export graphics and image technology: each type of graphics software has its own storage format and file type. In order to share and interact with information between different systems, different file formats are required to be exchanged or converted. At present, the most commonly used data exchange method is star exchange; that is, each system uses standard data exchange format for indirect exchange [25]. With the help of the AutoCAD's DXF file structure, this article refers to the DXF file format: title section, table section, block section, element section, and file end, and establishes a CDXF class that can read and write the dxf file format. In addition, it also realizes image calling and processing.

3.1.2. Image Chain Coding Processing.

After the art vision image is refined by the CAD platform, it is transformed into a single-pixel image, but it is still a binary image. To become a vectorized image, it needs to be coded and simplified by chain coding. The encoded image information can be conveniently used to extract the geometric features of the image to pave the way for the subsequent stereoscopic transmission of the image [26]. In addition, the chain-coded images can greatly save storage space and can store massive CAD data. The Freeman chain code method is a universal method widely used internationally to turn a single-pixel graphic into a chain code. The principle of Freeman chain
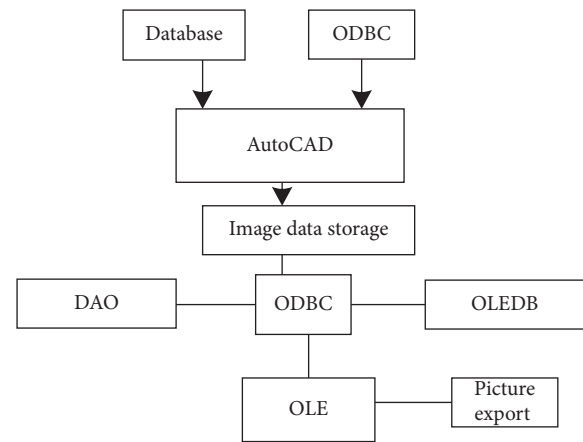
FIGURE 2: Schematic diagram of CAD platform structure.

code is to treat a binary image as a raster image composed of grid points. The method of chain coding is to express a continuous curve with the direction between adjacent pixels in the grid point theory. It can be divided into 4-domain chain codes and 8-domain chain codes. The chain coding process in this article adopts the 8-domain Freeman chain code method. Figure 3 is a schematic diagram of 8-domain chain codes.

Using the method from bottom to top, from left to right, scan the artistic vision image in turn, and the curves in the final image are represented as a two-dimensional array. The rows of the array indicate that there are several continuous curves in the image, and the columns of the array indicate the chain code trend of each curve. In the process of chain code scanning, the starting position of the curve must first be judged [27, 28]. The pixel degree is calculated to determine, which pixel is the starting point of the chain code and which pixel is the ending point of the chain code. And mark the start point and end point, respectively, and record their coordinates. When scanning to the starting point, it will follow the direction of the starting point adjacent to the target pixel to the target end, and at the same time record the direction number of the current pixel relative to the previous pixel, and finally complete the chain coding of a curve [29, 30].

At this point, the CAD preprocessing process is set up, and the generated dxf file is the CAD universal vectorized file format, which can be imported into CAPP software for subsequent image stereo transmission processing.

3.2. Realization of Stereo Communication of Artistic Vision Images Based on Particle Swarm Optimization Algorithm.

On the basis of using the CAD platform to obtain the CAD general vector file format, the image transmission method based on the particle swarm optimization algorithm is used to complete the three-dimensional transmission of artistic visual images [31]. To transform the art visual image stereo communication problem into the classification problem of image target foreground and image target background, the art visual image needs to be stereo transmitted under visual communication. The ratio function between the image
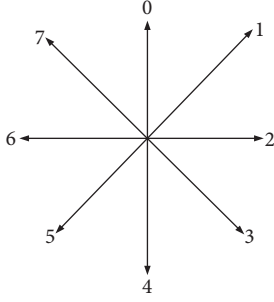
Figure 3: Original image.

feature distribution variance in the target foreground range is the separation coefficient to judge the foreground range of the image target, and the ratio function between the variance of image feature distribution in the target background range is the separation coefficient to judge the target background range of the image. In order to complete the three-dimensional communication of art visual image, particle swarm optimization algorithm is used to weight the feature components of art visual image and carry out real-time image target visual communication [32]. The process is as follows.

In order to understand the distinguishability of each range of artistic visual image, the corresponding log-likelihood ratio function variance formula is obtained on the premise of variance calculation:

$$y(t) = e^{\omega} A \frac{z}{\sqrt{z^2 + c^2}}, \quad (12)$$

where the log-likelihood ratio function is e^{ω}, and the foreground range scale and the background range scale of the artistic visual image are z^2 and c^2, respectively. Use a vector to describe the weight ς:

$$\varsigma = [\varsigma_1, \varsigma_2, \varsigma_3]. \quad (13)$$

In the above formula, the feature vector of artistic visual image is $\varsigma_1, \varsigma_2, \varsigma_3$. Assuming that ς' is the optimal solution of ς, the particle swarm optimization algorithm is used to extract the optimal solution ς' of different feature component weights of artistic vision images.

The steps of particle swarm optimization algorithm to extract ς' are as follows.

Assuming that there are η particles $\{\eta_1, \eta_2, \eta_n\}$ and ς in the same dimensional space, each particle η_i has a solution for ς, and the orientation of each particle in the group is determined by the fitness function value corresponding to the particle [33, 34].

Suppose the dimension of ς is μ, the size of the particle swarm is σ^2, each particle in the swarm is 1, and the orientation of each particle in the t step iteration is

$$\varsigma_{\mu}(\sigma^2) = (\varsigma_{\mu 1}, \varsigma_{\mu 2}, \ldots, \varsigma_{\mu i}, \ldots, \varsigma_{\mu N})^t. \quad (14)$$

The local optimal solution corresponding to the orientation of the best fitness of each particle is

$$Q_{\mu}(\sigma^2) = (q_{\mu 1}, q_{\mu 2}, \ldots, q_{\mu i}, \ldots, q_{\mu N}). \quad (15)$$

The solution that reflects the best fitness of the particle in the local optimal solution is the current particle swarm optimal solution Q_k. The population adjusts the current collection speed and orientation of each particle through multiple iterations [35], and the iteration expression is

$$Q_k = E_{eq} k_s \sqrt{K_D} \times \frac{f^{\beta} y^{\beta}}{V_c} + E_{eq} \sqrt{K_D} K_0 \times \frac{f^{\gamma} y^{\gamma}}{V_l}, \quad (16)$$

$$Q_{\mu}(t + 1) = Q_{\mu}(t) + \varsigma_{\mu}(t + 1). \quad (17)$$

In formulas (16) and (17), the learning factors are f^{β} and f^{γ}, the uniform random numbers in the [0, 1] area are y^{β} and y^{γ}, the collection speed of each particle in the t step is K_D, the each particle collection speed in the $t + 1$ is K_0, and the current azimuths of the population particles in steps t and $t + 1$ are V_c and V_l, respectively.

The optimal solution Q_k sought in this article is obtained after the iterative exploration of all particles in the population is completed, and there is no change in the orientation of each particle. It successfully realizes the tracking of the target feature of the artistic vision image and completes the stereoscopic vision transmission of the image target feature. Then, use the transmission value to sort the image data, which is easy to operate, reduces the complexity of image processing and completes the three-dimensional transmission of artistic visual images.

4. Simulation Experiment Analysis

In order to test the research effect of art visual image three-dimensional communication method based on CAD platform, the simulation experiment needs to be carried out on the CAD platform, and the experimental results are compared in the form of pictures. In order to build a CAD platform that met the experimental conditions, the development script language of 3D core engine is java language. The editor tool can realize the high integration of various functions of the CAD platform, fuse and match the virtual scene in the artistic visual image with the task graphic image, and use the editor tool to reduce the operation complexity of the experiment. The 3D player loads the system content in the artistic visual image in a component-based way. Through the 3D player, you can experience the virtual reality effect, which provides an effective way for the display of artistic visual image effect. After the construction of the experimental platform, the simulation experiment is carried out on the platform. The research effect is compared with the traditional plane visual communication effect optimization method based on wavelet change and the plane visual communication design method based on user experience effect, and the experimental results are analyzed.

4.1. Experimental Data Set. The images used in this experiment are from CIFAR open-source database [36]. CIFAR is collected and sorted by Alex Krizhevsky, Vinod Nair, and Geoffrey Hinton; 60000 of the 800000 images in visual dictionary are selected and divided into CIFAR-10 and CIFAR-100. CIFAR-10 dataset contains 60000 $32 * 32$ color

Original image

(a)   (a)   (a)
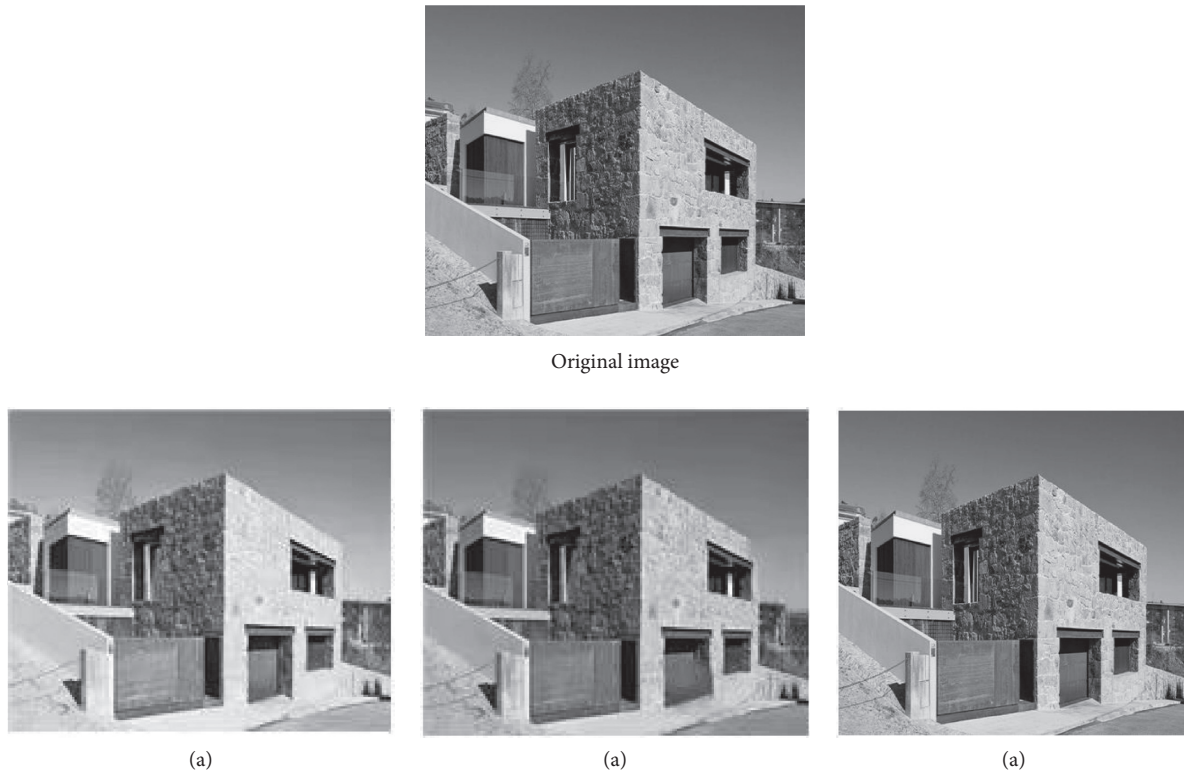
FIGURE 4: Comparison of image visual communication effects.

images, with 10 categories in total. There are 50000 training images and 10000 test images. The dataset is divided into five training blocks and one test block, and each block has 10000 images. The test block contains 1000 images randomly selected from each class, the training block contains these images in random order, and the training block contains 5000 images per class. The CIFAR-100 dataset contains 100 subclasses, each of which contains 600 images, including 500 training images and 100 test images. 100 categories are divided into 20 categories. Different types of artistic visual images are randomly selected from the database. In order to ensure the consistency of experimental conditions, 680 pixels are ensured ×480 pixels is the size of each film and television image.

4.2. Analysis of Experimental Results

4.2.1. Image Visual Transmission Effect. Two images are arbitrarily selected from the CIFAR database, and three methods are used to visually communicate them. The original image is shown in Figure 3, and the image processing result is shown in Figure 4.

By analyzing Figures 3 and 4, it can be seen that the visual effect obtained by using the method in this article is good, which not only retains the image detail information, but also does not have the problem of image blur. However, when using traditional methods to process the image, the image has some fuzziness, edge fuzziness, and poor visual communication effect. Therefore, the image processing effect of this method is better.

4.2.2. Image Visual Communication Effect under the Influence of Noise. An arbitrary image is selected as the input in the experimental data set, and the peak signal-to-noise ratio PSNR is used as the objective criterion of evaluation. Set the noise probability to 0.3, 0.6, and 0.9, respectively, and compare the three methods under different conditions. The experimental results are shown in Figure 5.

It can be seen from the curve in Figure 5 that the three methods are not suitable for the case of strong noise. The higher the noise probability, the lower the peak signal-to-noise ratio. The comparison shows that under different noise probabilities, the peak signal-to-noise ratio of this method is higher than that of the traditional method, and it is obviously better than the traditional method in the whole noise probability range. It shows that in the presence of noise, the traditional methods cannot remove the bright and dark points caused by noise from the image, but this method has a good effect on noise suppression and solves this problem, and the subjective visual effect of the image is greatly improved. In the experiment, it is found that the image visual effect is the best when the noise probability is 0.3. Therefore, this method has good antinoise performance and good visual communication effect.

4.2.3. The Efficiency of Image Visual Transmission. In addition to the visual communication effect of the image itself, the image processing efficiency is also an important index to measure the method. Due to the rapid improvement of the number of images, the rapid visual communication processing of a large number of images is a key problem for

(a)

(b)

(c)
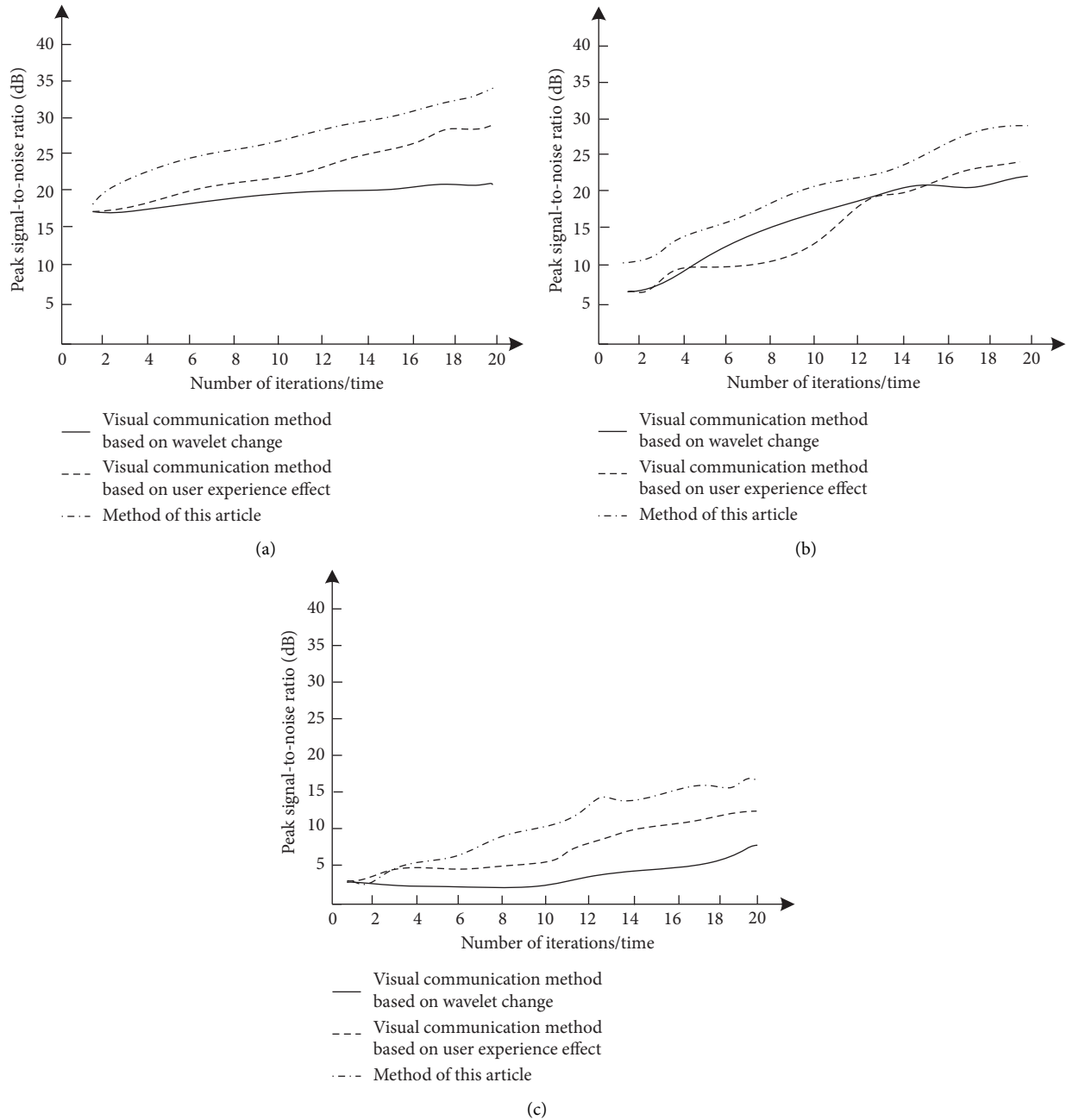
Figure 5: Comparison of antinoise performance. (a) Noise probability 0.3. (b) Noise probability 0.6. (c) Noise probability 0.9.

studies at present. Therefore, taking the image visual communication efficiency as the experimental index, the different methods are compared, and the results are shown in Figure 6.

According to the analysis of Figure 6, with the increase in the number of images, the visual communication efficiency of different methods shows a downward trend, indicating that the visual communication efficiency of images is affected by the number of images. Comparing the visual communication efficiency of different methods, it can be seen that the visual communication efficiency of this method is always higher than the two traditional methods, the highest visual communication efficiency reaches 84%, and

the downward trend is slower than the traditional methods. The visual communication efficiency of the plane visual communication effect optimization method based on wavelet change and the plane visual communication design method based on user experience effect is not only significantly lower than that of this method, but also decreases significantly with the increase of the number of images, indicating that the two traditional methods are not suitable for a large number of image processing and have certain limitations.

The artistic visual image stereo communication method is used to perceive the technical action posture data at the gymnastics scene, and the visual reconstruction analysis
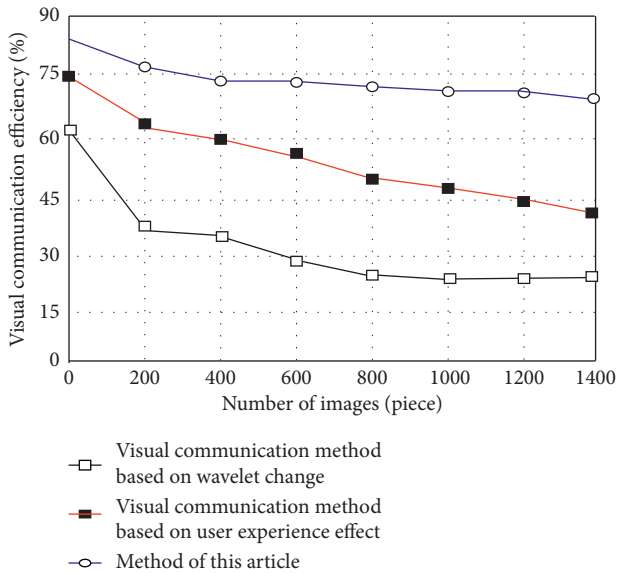
FIGURE 6: Comparison of image visual communication efficiency.
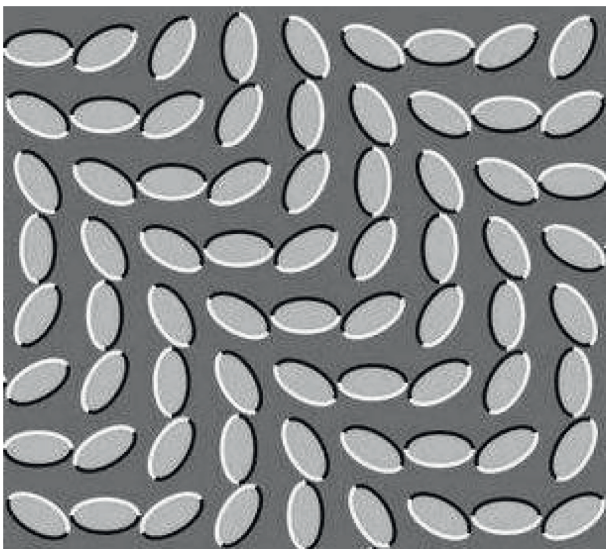
FIGURE 7: Visual reconstruction of art image feature data.

image data and improve the visual communication efficiency, and it has better research value to enhance the performance of image information protection and improve the clarity of image transmission.

5. Conclusion

In order to study an effective image processing method, in order to improve the image communication effect, improve the image communication efficiency, and reduce the influence of noise factors on the image, this article proposes a three-dimensional communication method of artistic visual image based on CAD platform. The simulation results show that compared with the traditional methods, the image processing effect of this method is good, which can improve the image transmission effect to a certain extent and improve the clarity and integrity of the image. The performance of the designed method is superior. In the experiment, it is found that when the noise probability is 0.3, the image visual effect is the best, satisfactory effect can be obtained, and the antinoise performance is good. The visual communication efficiency is always higher than the two traditional methods, and its highest visual communication efficiency reaches 84%, and the downward trend is more slow.

Data Availability

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Conflicts of Interest

The authors declare that they have no conflicts of interest regarding this work.

Acknowledgments

results of artistic image feature data are obtained, as shown in Figure 7.

It can be seen from the results in Figure 7 that through the method designed in this article, the visual reconstruction effect of art image feature data is good, the visual effect of art image can be displayed accurately, and the performance is stable. It provides visual support for the analysis of visual reconstruction of art image feature data, and the performance of the designed method is superior.

Through the above comparative analysis, it can be seen that the image definition, visual communication efficiency, and denoising effect of this method are higher than the traditional plane visual communication effect optimization method based on wavelet change and the plane visual communication design method based on user experience effect. This method can better grasp the basic information of

References

[1] C. Yu, "Retracted article: climate environment of coastline and urban visual communication art design from the perspective of GIS," *Arabian Journal of Geosciences*, vol. 14, no. 4, p. 310, 2021.

[2] T. Li, Y. Wang, R. Hong, M. Wang, and X. Wu, "pDisVPL: probabilistic discriminative visual Part Learning for image classification," *IEEE MultiMedia*, vol. 25, no. 4, pp. 34–45, 2019.

[3] Z. Tang, H. Zhang, C. M. Pun, M. Yu, C. Yu, and X. Zhang, "Robust image hashing with visual attention model and invariant moments," *IET Image Processing*, vol. 14, no. 5, pp. 901–908, 2020.

[4] Z. Hao, X. Wang, and S. Zheng, "Recognition of basketball players' action detection based on visual image and Harris corner extraction algorithm," *Journal of Intelligent and Fuzzy Systems*, vol. 40, no. 1, pp. 1–11, 2020.

[5] K. Kyamakya, A. Haj Mosa, F. A. Machot, and J. C. Chedjou, "Document-image related visual sensors and machine learning techniques," *Sensors*, vol. 21, no. 17, p. 5849, 2021.

[6] D. D. Rodriguez, C. Goulart, C. Valadão, D. Funayama, and T. Bastos, "Visual and thermal image processing for facial specific landmark detection to infer emotions in a child-robot interaction," *Sensors*, vol. 19, no. 13, p. 2844, 2019.

[7] J. Yin and J. H. Yang, "Virtual reconstruction method of regional 3D image based on visual transmission effect," *Complexity*, vol. 2021, Article ID 5616826, 12 pages, 2021.

[8] A. Kosovicheva, K. Sridhar, and P. J. Bex, "Image predictors of visual localization in natural scenes," *Journal of Vision*, vol. 20, no. 11, p. 183, 2020.

[9] M. M. Liu, "Optimization of plane visual communication effect simulation based on graphic beautification technology," *Computer Simulation*, vol. 36, no. 9, pp. 426–429, 2019.

[10] C. Li, "Research on graphic visual communication design based on user experience effect," *Modern Electronics Technique*, vol. 43, no. 11, pp. 111–114, 2020.

[11] X. Q. Zhu, "Innovative design of graphic image visual communication system for animated character in virtual reality environment," *Modern Electronics Technique*, vol. 561, no. 10, pp. 178–180, 2020.

[12] J.-J. Liaw, C.-P. Lu, Y.-F. Huang, Y.-H. Liao, and S.-C. Huang, "Improving census transform by high-pass with haar wavelet transform and edge detection," *Sensors*, vol. 20, no. 9, p. 2537, 2020.

[13] X. Dong, G. Li, Y. Jia, and K. Xu, "Multiscale feature extraction from the perspective of graph for hob fault diagnosis using spectral graph wavelet transform combined with improved random forest," *Measurement*, vol. 176, no. 2, Article ID 109178, 2021.

[14] F. Muoz-Muoz and A. Rodrigo-Mor, "Partial discharges and noise discrimination using magnetic antennas, the cross wavelet transform and support vector machines," *Sensors*, vol. 20, no. 11, p. 3180, 2020.

[15] A. Rezapour, A. Ortega, and M. Sahimi, "Upscaling of geological models of oil reservoirs with unstructured grids using lifting-based graph wavelet transforms," *Transport in Porous Media*, vol. 127, no. 3, pp. 661–684, 2019.

[16] J. Pushparaj and M. Malarvel, "Panchromatic image denoising by a log-normal-distribution-based anisotropic diffusion model," *Journal of Applied Remote Sensing*, vol. 13, no. 1, p. 1, 2019.

[17] S. Seo, "Image denoising and refinement based on an iteratively reweighted least squares filter," *KSCE journal of civil engineering*, vol. 24, no. 3, pp. 943–953, 2020.

[18] Y. Wang, D. Chang, and Y. Zhao, "A new blind image denoising method based on asymmetric generative adversarial network," *IET Image Processing*, vol. 15, no. 6, pp. 1260–1272, 2021.

[19] M. H. Rahman, C. Schimpf, C. Xie, and Z. Sha, "A CAD-based research platform for data-driven design thinking studies," *Journal of Mechanical Design*, vol. 141, no. 12, p. 1, 2019.

[20] C. Dasgupta, A. J. Magana, and C. Vieira, "Investigating the affordances of a CAD enabled learning environment for promoting integrated STEM learning," *Computers & Education*, vol. 129, no. 2, pp. 122–142, 2019.

[21] A. Jhld, "Fabricating a crown under an existing removable partial denture with impression scanning and CAD-CAM technology," *The Journal of Prosthetic Dentistry*, vol. 124, no. 2, pp. 148–152, 2020.

[22] Z. Chen, Y. Wang, and Z. Song, "Classification of motor imagery electroencephalography signals based on image processing method," *Sensors*, vol. 21, no. 14, p. 4646, 2021.

[23] H. Zhang and J. Yao, "Automatic focusing method of microscopes based on image processing," *Mathematical Problems in Engineering*, vol. 2021, Article ID 8243072, 9 pages, 2021.

[24] W. Zhang, X. Li, Q. Song, and W. Lu, "A face detection method based on image processing and improved adaptive boosting algorithm," *Traitement du Signal*, vol. 37, no. 3, pp. 395–403, 2020.

[25] F. Han, J. Yao, H. Zhu, and C. Wang, "Underwater image processing and object detection based on deep CNN method," *Journal of Sensors*, vol. 2020, Article ID 6707328, 20 pages, 2020.

[26] N. H. Hai, H. L. Minh, D. N. H. Thanh, N. V. Son, and S. Prasath, "An adaptive image inpainting method based on the weighted mean," *Informatica*, vol. 43, no. 4, pp. 507–513, 2019.

[27] B. Nga, B. Sta, B. Mkea, M. N. Mab, and B. Nla, "Image postprocessing adoption trends in clinical medical imaging," *Journal of the American College of Radiology*, vol. 16, no. 7, pp. 945–951, 2019.

[28] B. Hadhami, R. Slama, R. Gaha, and A. Benamara, "Proposal of new eco-manufacturing feature interaction-based methodology in CAD phase," *International Journal of Advanced Manufacturing Technology*, vol. 106, no. 3, pp. 1057–1068, 2020.

[29] J. Qin, J. Chen, X. Xiang, Y. Tan, W. Ma, and J. Wang, "A privacy-preserving image retrieval method based on deep learning and adaptive weighted fusion," *Journal of Real-Time Image Processing*, vol. 17, no. 1, pp. 161–173, 2020.

[30] S. Vaez and Z. Minaei, "Pulse extraction of pulselike ground motions based on particle swarm optimization algorithm," *Scientia Iranica*, vol. 27, no. 1, pp. 134–158, 2020.

[31] J. Zhang, Z. Sun, N. Qiu, Y. Zhang, and F. Li, "3-D effetive elastic thickness inversion of subduction zone based on particle swarm optimization algorithm," *Chinese Journal of Geophysics- Chinese Edition*, vol. 62, no. 12, pp. 4738–4749, 2019.

[32] X. N. Bui, P. Jaroonpattanapong, H. Nguyen, Q. H. Tran, and N. Q. Long, "A novel hybrid model for predicting blast-induced ground vibration based on k-nearest neighbors and particle swarm optimization," *Scientific Reports*, vol. 9, no. 1, pp. 1–14, 2019.

[33] W. F. Fihri, H. E. Ghazi, and B. Majd, "A multi-objective particle swarm optimization based algorithm for primary user emulation attack detection," *Wireless Personal Communications*, vol. 117, no. 8, pp. 1–20, 2021.

[34] K. Malik and F. Kim, "Optimal travel route recommendation mechanism based on neural networks and particle swarm optimization for efficient tourism using tourist vehicular data," *Sustainability*, vol. 11, no. 12, p. 3357, 2019.

[35] V. Santucci, M. Baioletti, and A. Milani, "Tackling permutation-based optimization problems with an algebraic particle swarm optimization algorithm," *Fundamenta Informaticae*, vol. 167, no. 1-2, pp. 133–158, 2019.

[36] B. Xiao and S.-C. Kang, "Development of an image data set of construction machines for deep learning object detection," *Journal of Computing in Civil Engineering*, vol. 35, no. 2, Article ID 05020005, 2021.

Retraction

Retracted: Career Recommendation for College Students Based on Deep Learning and Machine Learning

Scientific Programming

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope

(2) Discrepancies in the description of the research reported

(3) Discrepancies between the availability of data and the research described

(4) Inappropriate citations

(5) Incoherent, meaningless and/or irrelevant content included in the article

(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] Q. Wan and L. Ye, "Career Recommendation for College Students Based on Deep Learning and Machine Learning," *Scientific Programming*, vol. 2022, Article ID 3437139, 10 pages, 2022.

Research Article

Career Recommendation for College Students Based on Deep Learning and Machine Learning

Qing Wan [ID] and Lin Ye

Shaanxi Institute of International Trade&Commerce, Shaanxi 712046, China

Correspondence should be addressed to Qing Wan; 368009050@qq.com

With the popularization of higher education, China's higher education has moved from the stage of elite education to the stage of universal education, and the number of graduates is also increasing. At present, college students are facing tremendous pressure on employment. One is the huge number of jobs, the other is the difference in professional needs, and the third is the spread of job information, which makes it difficult for college students to find a job that suits them. In order to solve this dilemma, this paper analyzes the related technologies and in-depth basic theories of data mining. After introducing several traditional recommendation algorithms, the traditional convolutional network method is improved from three aspects: activation function, pooling strategy, and loss function. Finally, using the hybrid convolutional neural network, a career recommendation model for college students based on deep learning and machine learning is proposed, and simulation experiments are carried out on it. The main research work is as follows: (1) a hybrid convolutional neural network is proposed, which uses convolution operation to learn high-level features to achieve personalized employment recommendation; (2) the training optimization strategy of the hybrid convolutional neural network is studied, aiming at the activation function, pooling processing, and loss function, and the feasibility of the optimization method is verified through simulation experiments; (3) finally, according to the evaluation index of the recommendation algorithm (recall rate and $F1$-Score), the recall rate of the algorithm in this paper is nearly 15% higher than that of the DNN model. The experiment is compared with the traditional commonly used recommendation algorithm, and the comparative analysis of the experimental results proves the effectiveness of the algorithm for the employment recommendation of college students.

1. Introduction

As the enrollment scale of colleges and universities continues to expand, as of 2020, the gross enrollment rate of higher education in China has exceeded 50%, and higher education has entered the stage of universal education from the stage of elite education [1]. The employment demand of graduates is increasing year by year [2], and enterprises have higher and higher requirements for the professional skills and comprehensive quality of graduates. The traditional employment guidance model in the past can no longer adapt to the new employment supply and demand relationship. How to provide graduates with personalized and precise employment services has become the focus of the work of colleges and universities. Data show that in 2020, the

number of college graduates nationwide will reach 8.74 million [3]. However, with the impact of the new crown epidemic and the intensified world trade protection, China's economy is under great downward pressure, which has gradually led to the emergence of college students'' employment problems mainly realizing the following aspects:

(1) The number of employed people is huge, the employment peak lasts for a long time, and the situation is severe. The employment pressure of fresh graduates will not weaken in a relatively long period of time in the future.

(2) In the context of the rapid development of China economy, the demand for talents is constantly expanding, but some majors have not matched the

needs of China industry, resulting in structural unemployment for college students.

(3) Under the massive data situation of 'information overload' faced by college students when choosing employment, a wide range of job information makes it easy for college students to fall into information scrutiny fatigue and unable to clarify their job search needs, which leads to the need to spend more energy to find what is really suitable for them. The post information caused employment difficulties in the talent market [4]. The lack of a unified platform support and the inability to deeply match their own characteristics with the needs of enterprises result in a lack of scientific advice on big data employment guidance for college graduates.

In recent years, scholars from various universities and research institutions in China have also begun to explore related issues such as job matching and job matching. Kumar and Gupta and others put forward the job matching system theory [5]. Sajjadi et al. proposed the method basis of job matching in 2002 [6], and Figueiredo et al. took a different approach [7] and proposed the use of dynamic research methods to study the matching relationship between candidates and organizations. In 2006, under the leadership of Hiriyannaiah et al. and others, the dynamic person-post matching model became more and more mature, and dynamic fitting was proposed [8]. Koehn and Da'u, respectively, used fuzzy mathematics theory and grey system theory in their research and proposed a new matching model [9, 10]. With the advent of the data age, the simple theory of job matching with applicants can no longer provide accurate advice for college students' employment choices. A large amount of data has made the inherent theoretical avatars useless. Many scholars have begun to introduce recommendation algorithms from other fields and analyze big data. Applied to job recommendation, traditional recommendation algorithms mainly include three types: collaborative filtering recommendation algorithm, content-based recommendation algorithm, and hybrid recommendation algorithm. The collaborative filtering recommendation algorithm can use existing historical activity interaction information and a series of user comparison information with the same behavior to make collaborative similarity matching recommendation without relying on user or item content

information [11]. The content-based recommendation algorithm is an algorithm scheme that uses the user's profile or project description information to recommend an algorithm scheme [12], which combines the attributes and characteristics of the project by means of data mining or information retrieval to build the user's profile model. The hybrid recommendation algorithm aims to obtain better prediction or recommendation performance by combining the content-based recommendation algorithm and collaborative filtering recommendation algorithm, taking the advantages of the two. Traditional recommendation algorithms need to perform a lot of experiments on the selection of combination methods and the sequence of combinations to find a better combination method, and the weight distribution of the recommendation results obtained by different methods also needs to be experimentally tested and analyzed in order to make up for this. For this kind of defect, scholars introduced the concept of deep learning and started the research of the recommendation algorithm based on deep learning. Covington et al. proposed a deep neural network for YouTube video recommendation [13]. Cheng et al. proposed a wide and deep model to provide application recommendations for Google's application mall [14]. Okura et al. proposed using the RNN model to recommend news information for Yahoo News [15].

This paper draws on the research results of Zhu Wei's flight delay prediction algorithm combined with convolution processing [14], hoping to use the high-level feature learning ability of convolution processing to improve the quality of employment recommendation of college students and solve the "data trap" problem faced by college students in graduation selection.

2. Related Work

2.1. Hybrid Convolutional Neural Network. Convolutional neural network can be understood as a hierarchical model. The original data are used as the input of the model. Through a series of operation layers such as convolution operation, pooling operation, and nonlinear activation function mapping, the high-level abstract information is removed from the original extracted from the data input layer. Among them, the level of abstract operation is the "feedforward operation" in the convolutional neural network. The specific mathematical expression is as follows:

$$x^1 \longrightarrow \omega^1 \longrightarrow x^2 \longrightarrow \cdots \longrightarrow x^{L-1} \longrightarrow \omega^{L-1} \longrightarrow x^L \longrightarrow \omega^L \longrightarrow z,$$
$$z = f(x^L, y),$$

(1)

where x^L represents the data input of the Lth layer, ω^L represents the weights related to the Lth layer, z represents the loss function of this calculation process, y represents the final true classification mark, and the function f uses ω^L as the calculation parameter.

2.1.1. Receptive Domain. In a convolutional neural network, the receptive field is expressed as the size of the region where the output features of each layer in the network are mapped to the input space. The receptive field represented by a certain feature can be described by the center position of the

receptive field and its size. Figure 1 shows the receptive field performance of a convolution operation with a size of 3×3 and a step size of 1. The convolution kernel can obtain the same scale of receptive field as the large convolution kernel through multilayer superposition. The use of the small convolution kernel can deepen the network depth to enhance the capacity and complexity of the network and also reduce the number of training parameters.

2.1.2. Distributed Representation. In deep learning, convolutional neural networks will exhibit important characteristics of distributed representation. The characteristics of distributed representation can indicate that the convolutional neural network can form different representation modes by different convolution kernels, and different responses can be generated from different modes, and different feature representations can be abstracted according to different convolution kernels. Therefore, this feature can be used to implement different abstract feature representations using different convolution kernels to obtain more property feedback related to the input data.

2.1.3. Local Connection and Weight Sharing. The local connection method and the weight sharing method are both based on the local receptive domain, and the feature input satisfies the sparse but locally dense nature and is used to model the input data [17]. The local connection completes the feedback of various properties of the input data characteristics through the local information, and it is also the theoretical support for the distributed representation of the convolutional neural network. Each neuron in the convolutional neural network model only needs to be connected to a part of the forward layer. The local connection can not only obtain the different properties of the input data, but also reduce the risk of overfitting. The operation diagram of the local connection is shown in Figure 2.

The proposed weight sharing is used for the reduction of model training parameters, and the number of training parameters required for model training is reduced by sharing the convolution kernel [18]. Weight sharing uses convolution kernels with the same internal weights, and by using a common mode for feature detection, it can be understood that weight sharing is to use a template to abstract a certain feature of the data. Although the weight sharing mode can effectively reduce the scale of parameters required for model training, weight sharing can only extract a common feature property feedback, which is not as rich as the locally connected multiproperty feedback, and cannot mine more latent association representations at one time. The operation diagram of weight sharing is shown in Figure 3.

2.2. Traditional Recommendation Algorithm. The use of recommendation algorithms is currently the main effective measure to deal with the problem of information overload. It can provide users with appropriate personalized options in the case of massive data, avoid users from getting into information review fatigue, and save a lot of time and energy.

2.2.1. User-Based Collaborative Filtering Recommendation Algorithm. The user-based collaborative filtering recommendation algorithm mainly implements similar recommendations for users to find other users with similar behaviors or constructs a user's own preference model and performs a predictive score based on their own preference model to complete personalized recommendation. The core of the user-based collaborative filtering recommendation algorithm is similarity calculation. The main function of similarity calculation is to measure the similarity $S_{u,v}$ between user u and user v. The mathematical expression of Pearson similarity calculation is shown as follows:

$$S_{u,v} = \frac{\sum (z_{u,i} - \overline{Z_u})(z_{v,i} - \overline{Z_v})}{\sqrt{\sum (z_{u,i} - \overline{Z_u})^2}\sqrt{\sum (z_{v,i} - \overline{Z_v})^2}}. \tag{2}$$

$z_{u,i}$ and $z_{v,i}$ represent the actual scores of user u and user v on item i and $\overline{Z_u}$ and $\overline{Z_v}$ represent the average of all item scores of user u and user v, respectively.

2.2.2. Content-Based Recommendation Algorithm. The content-based recommendation algorithm is to use the content that users are interested in to calculate the similarity to achieve relevant recommendations. By means of data mining or information retrieval, a data file model belonging to the user is constructed by combining the attributes and characteristics of the project. The recommendation algorithm adopted by this algorithm is to use the content that the user is interested in to calculate the similarity to realize the recommendation. The algorithm uses x to represent the similarity evaluation of item s recommended to user c, where x is based on s, indicating that the recommendation of item c is obtained by quantitative calculation of other items with similar meta-information to item s.

2.2.3. Model-Based Collaborative Filtering Recommendation Algorithm. PMF is a model-based collaborative filtering recommendation algorithm. It reduces the dimension of the high-dimensional evaluation matrix to obtain the user hidden feature matrix and the item hidden feature matrix and further predicts the missing value of the evaluation matrix by calculating the inner product of the two. Its mathematical expression is as follows.

Given an $m \times n$ evaluation matrix R, two low-dimensional matrices U and V with rank d are used for fitting, and the fitting formula is as follows:

$$R \approx UV^T. \tag{3}$$

$U \in m \times d$; $V \in d \times n$, $d \leq \min\{m, n\}$; U represents the tendency of each user to d features; V represents the existence of d features in each item.
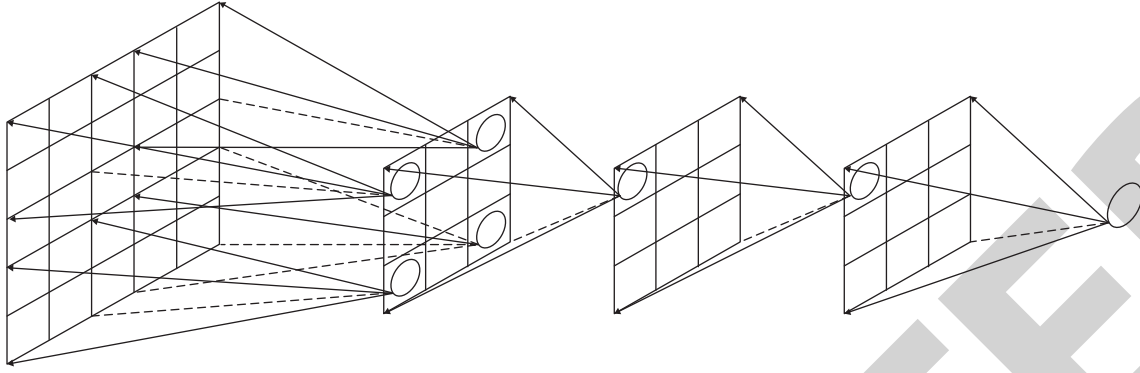
FIGURE 1: The receptive field with a size of 3×3 and a step length of 1 [16].
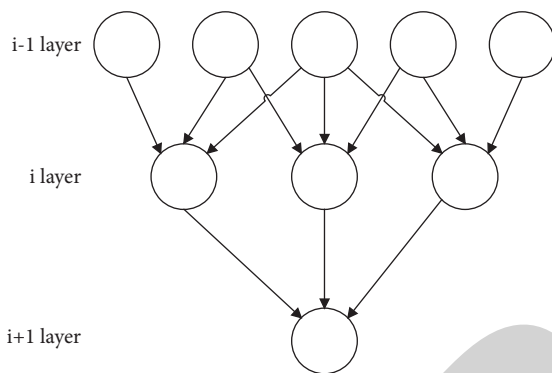
FIGURE 2: Schematic diagram of partial connection operation.
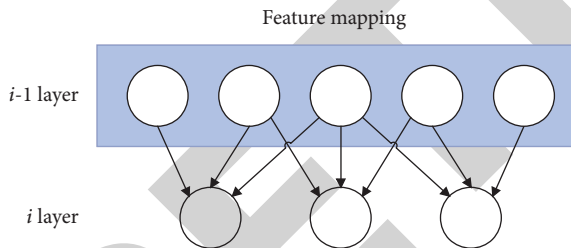
FIGURE 3: Schematic diagram of convolution weight sharing operation.

2.3. Data Feature Extraction. In this paper, gradient boosting tree is used for data feature processing, and the input features are filtered and processed in an integrated manner to eliminate the influence of useless features. The specific feature conversion flowchart is shown in Figure 4.

Before using the gradient boosting tree for feature conversion processing, it is necessary to perform feature dimensionality reduction processing on the type attributes of the one-hot encoding format, use the function of the Embedding operation to map low-level features to different feature spaces, and use high-level features as the mapping representation. The dimensionality reduction processing usually uses a neural network for training, and its standard algorithm expression is shown in the following formula:

$$y_z = \mathscr{F}\left(W_z x_z + b_z\right). \tag{4}$$

Among them, z is used to represent the index value of the feature x_z that needs dimensionality reduction in the input features, W_z represents the matrix of $m_z \times n_z$, b_z represents the deviation term of the vector, y_z is the vector obtained by dimensionality reduction, and $\mathscr{F}(\cdot)$ represents nonlinearity activation function.

3. Employment Model for College Students Based on Deep Learning and Machine Learning

This paper draws on the research results of Zhu Wei's flight delay prediction algorithm combined with convolution processing [18], hoping to improve the recommendation quality with the help of the high-level feature learning ability of convolution processing. The hybrid convolutional neural network model is selected, and the transformed features are used as input to train the hybrid convolutional neural network model through model stacking integration to generate recommended results.

3.1. Model Structure Design. The composition of the hybrid convolutional neural network model is mainly divided into a multichannel hybrid convolution submodel and a convolution and local connection hybrid submodel. The overall model structure diagram is shown in Figure 5.

Among them, Σ is represented as a data combination and splicing operation, which is used to combine the processing and generation results of multichannel convolution and is used for subsequent convolution and local connection mixing operations. The loss function used in the training process of the hybrid convolutional neural network is the cross-entropy loss function, which is used to determine how close the current actual output is to the expected output. Taking the binary classification task as an example, the mathematical expression of the cross-entropy loss function is shown in formula (5). Among them, y represents the true category label of the sample and p represents the predicted probability of the sample.
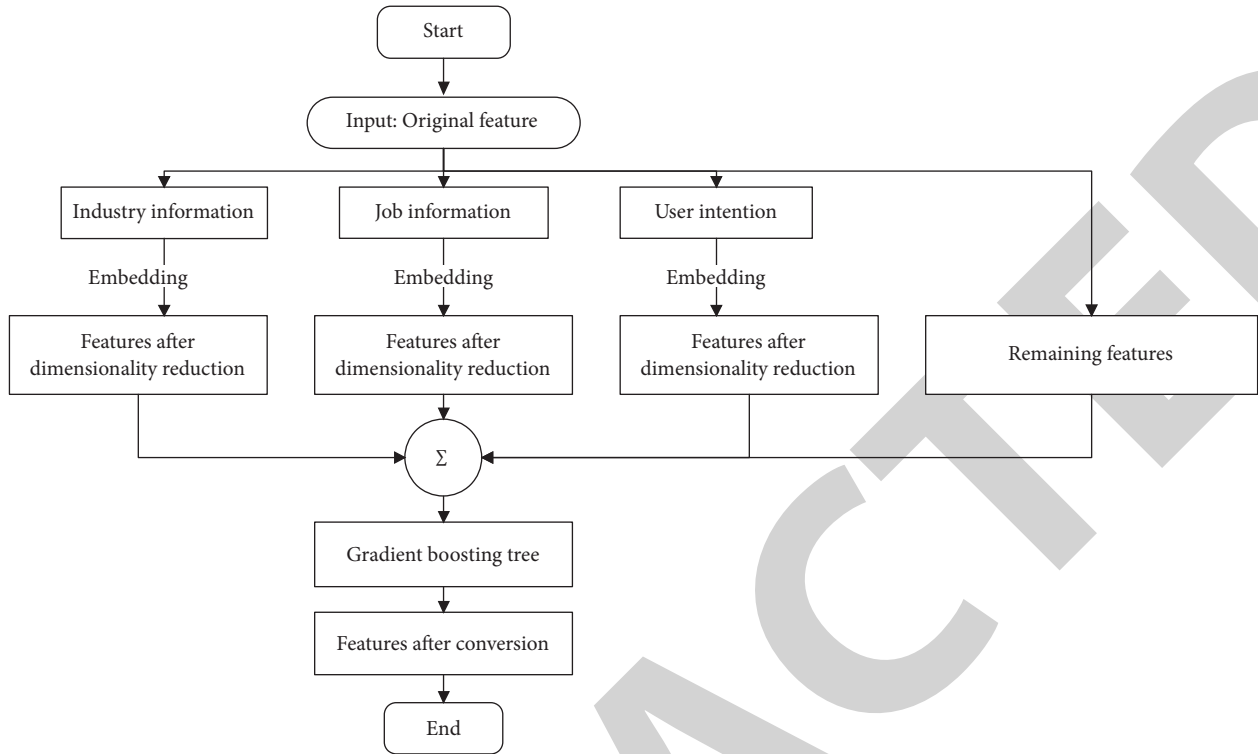
Figure 4: Feature extraction model.
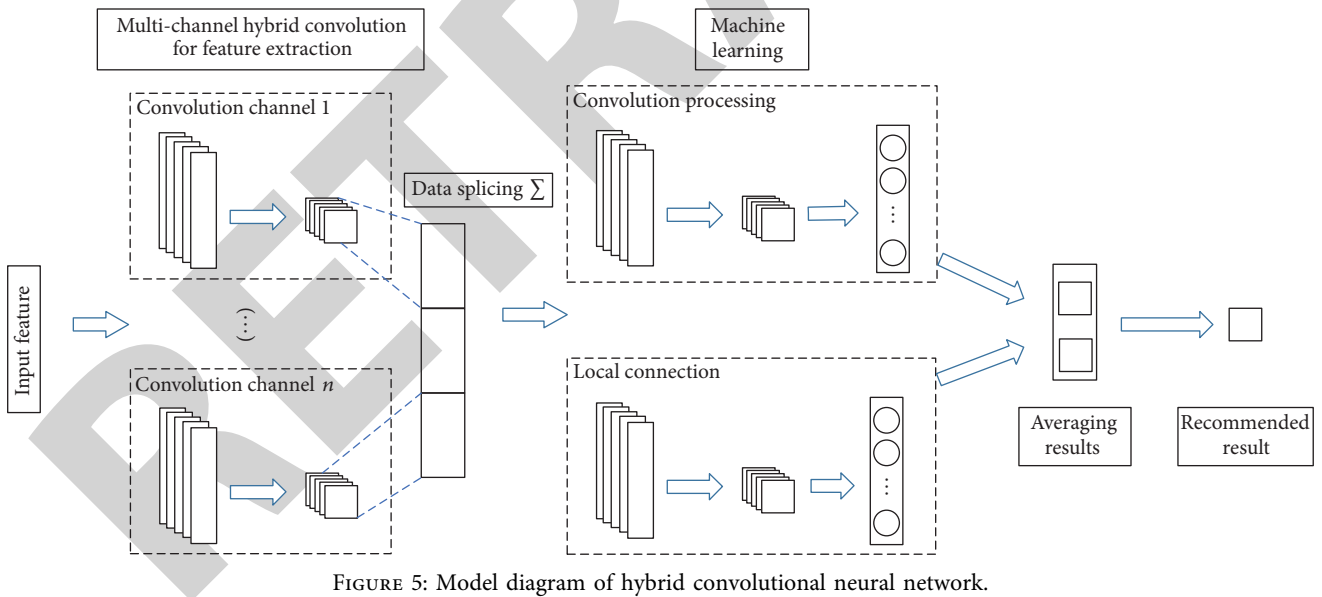
Figure 5: Model diagram of hybrid convolutional neural network.

$$\mathscr{L} = -\sum_{n} (y \log p + (1 - y)\log(1 - p)). \tag{5}$$

The description of the training algorithm of the hybrid convolutional neural network model is shown in Algorithm 1:

3.2. Model Training Optimization. The abovementioned hybrid convolutional neural network model is the benchmark model structure of the recommendation algorithm in this article, which can be used to recommend employment for college students, but it is found that there is still room for improvement in the process of model training.

3.2.1. Activation Function Optimization. Currently, commonly used nonlinear activation functions mainly include three functions: Sigmoid, Tanh, and ReLU. In the convolutional neural network model, the activation function of

---

Input: input feature D
Output: hybrid convolutional neural network model
(1) Initialization parameters: the number of iterations t, the number of convolution channels n, the learning rate ℓ of the Adam
    algorithm, and the hyperparameters β_1, β_2, and ε
(2) for $i = 1$ to t do
(3)   for $j = 1$ to n do//n represents the number of convolution channels
(4)   Calculate the extracted features of the j-th CNN channel model, and the output is F^ω
(5)   end for
(6) Integrate the extracted features of multichannel convolution $\{F^{(1)}, F^{(2)}, \ldots, F^{(n)}\}$, get F
(7) Input the F input convolution and local connection hybrid processing model to obtain the recommended prediction probabilities
    p_{CNN} and p_{LC}, respectively
(8) Averaging p_{CNN} and p_{LC} to get the current prediction result
(9) Calculate the error ℓ^i between the predicted result and the actual recommended result according to formula (4)
(10) Calculate the reciprocal of model parameters W and b according to back propagation and chain rule

---

ALGORITHM 1: Description of hybrid convolutional neural network training algorithm.

ReLU is optimized, and the main function is to alleviate the defect of network convergence caused by the death of neurons with input signal $x \leq 0$. The activation optimization used in this paper is the deformation optimization of ReLU to realize ELU. When dealing with negative input, the exponential term calculation method is used to alleviate the death problem caused by neuron death [19]. The specific algorithm expression of ELU is as follows:

$$ELU(x) = \begin{cases} x, & x > 0, \\ \alpha(\exp(x) - 1), & x \leq 0. \end{cases} \quad (6)$$

Using the activation function method of ELU for optimization processing can better obtain the representation information of the features, and compared with other ReLU improvements, ELU is more robust to input changes or noise.

3.2.2. Pooling Strategy Optimization. In the hybrid convolutional neural network benchmark model in this article, the pooling strategy adopted by the convolutional hierarchical model is the maximum pooling strategy of uniform size, which is a single pooling strategy. In this paper, hybrid pooling is used to make up for the shortcomings of feature loss caused by single pooling and to alleviate the model loss caused by feature loss. Integrating the advantages of the maximum pooling strategy and the intermediate value pooling strategy, on the basis of retaining the most salient feature representation in the pooling area, the pooling process of the hybrid convolutional neural network model is optimized through the intermediate value pooling strategy. The specific hybrid pooling integration implementation

structure diagram is as shown in Figure 6. $\overline{y}$ represents the averaging process and $\sum$ represents the combined process, used to splice the maximum pooling result and the intermediate value pooling result, as the output of the mixed pooling operation.

3.2.3. Loss Function Improvement. The model training loss function used in this paper is the cross-entropy loss function, and the mathematical expression is shown in (4), which is used to determine the closeness of the current actual output to the expected output. After analyzing the prediction result data, He et al. put forward the drawback of "misleading loss" for the situation that the number of samples that are easy to determine is too large [19]. If the number of samples that are easy to determine is too large, it may conceal the effects of other types of samples and make a major contribution to the decrease of the loss value, which may dominate the overall gradient update direction of the model, leading to poor model convergence results. In response to this problem, He et al. put forward the idea that samples are difficult and easy to distinguish, by improving the loss function to reduce the loss contribution of easy-to-classify samples and alleviate the problem of loss misleading. This paper draws on the idea of difficult-to-differentiation of the algorithm and uses the threshold to distinguish the classification difficulty of samples, so that the model pays more attention to the learning of difficult-to-differentiate samples [20]. In order to better classify sample categories, penalty weights are introduced to deal with the problem of imbalanced sample categories. The improved loss function expression is shown in the following equation:

$$\mathscr{L} = -\sum_n \left(\beta_+ \theta(p - m) y \log p + \beta_- \theta(1 - p - m)(1 - y)\log(1 - p) \right). \quad (7)$$

Y is the true classification category of the sample, p is the category prediction rate of the sample, β_+ and β_-

represent the penalty weights of positive and negative samples, respectively, m represents the set training
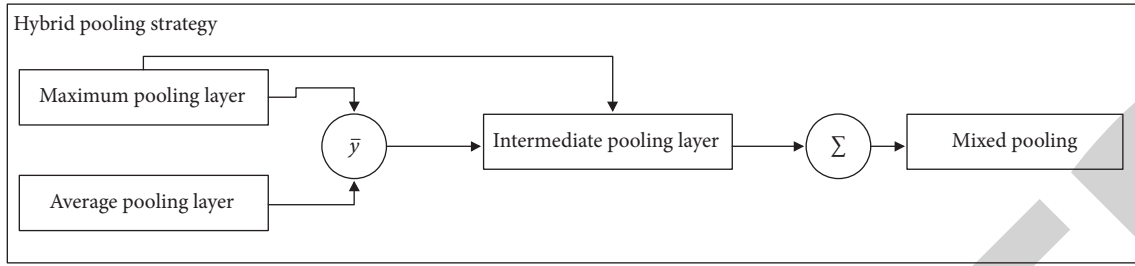
Figure 6: Implementation diagram of hybrid pooling strategy.

threshold, and $\theta(\cdot)$ represents the processing function for comparing the predicted probability with the threshold. The expression is shown in as follows:

$$\theta(x) = \begin{cases} 0, & x > 0, \\ 1, & x \leq 0. \end{cases} \tag{8}$$

4. Simulation Experiment

4.1. Recommended Quality Evaluation Standards. In the recommendation field, the commonly used recommendation quality evaluation standards include Precision, Recall, $F1$-Score, Mean Absolute Error (MAE), and Root Mean Square Error (RMSE). Among them, MAE and RMSE are generally applied to the recommendation method of the scoring type, and the quality of the recommendation is judged by calculating the error between the predicted recommendation score and the actual score. In this paper, recall rate and $F1$-Score are selected to evaluate the quality of algorithm recommendation. The calculation formula of recall rate is shown in formula (9), and the calculation formula of $F1$-Score is shown in formula (10).

$$\text{Recall} = \frac{TR}{TR + FNR}, \tag{9}$$

$$F1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}. \tag{10}$$

TR indicates the number of positions that are correctly recommended to college students who are really interested, FNR indicates the number of positions that are truly interested in incorrectly predicted as nonrecommended categories, and FR indicates that positions that are not of interest to college students incorrectly predicted as recommended categories.

4.2. Experimental Data. The experimental data set used in this article is collected from the employment guidance platform for college students. The data set used in the experiment mainly contains information about college students themselves (gender, grades, interests, and so on), job requirements information, and user-post interaction behavior information. Since this article converts the recommendation question into a "recommendation/ nonrecommendation" two-category question, the collected data are divided into two types such as positive samples and

negative samples based on behavioral information. The sample is marked as a negative sample, and the sample that has undergone job browsing, job collection, and job application is marked as a positive sample. The total number of job seekers is 7547, the total number of jobs is 15023, the number of recommended positive samples is 9870, and the number of recommended negative samples is 5153.

4.3. Experimental Results and Analysis. This paper divides the collected employment data set of college students in proportion, divides 90% of the data set into the algorithm training data set, and uses the remaining 10% as the algorithm test data set. The experiment will verify the model optimization strategy and the performance comparison between the algorithm in this paper and the existing recommended algorithm. The experimental results will be analyzed, and the experimental conclusions will be drawn.

4.3.1. Activation Function Optimization Verification. The benchmark model uses ReLU as the activation function. Aiming at the dead zone problem caused by the ReLU activation function on negative inputs, this paper proposes an ELU optimization method based on ReLU and uses exponential terms to alleviate the neuron suppression phenomenon of negative inputs. This article will compare ReLU (benchmark model) and ELU on the basis of the recommended length of the post of 70 (regardless of the changes in the learning rate in the model, the loss function starts to stabilize when the number of iterations is 70). Among them, the recommended index for the benchmark model using ReLU is marked as X, and the recommended index for other activation methods is marked as Y. All models adopt the maximum single pooling strategy. The results of the comparative experiment are shown in Table 1.

From the comparison of the experimental results in Table 1, the recommended recall rate and $F1$-Score obtained by all models with optimized activation methods are better than ReLU, which proves the necessity of alleviating the neuron suppression problem that exists during model training. Optimizing the defect of the zero-value output of the negative input can effectively help the prediction of the model recommendation.

4.3.2. Pooling Strategy Optimization Verification. This section mainly conducts experimental comparisons for different pooling strategies. All pooling comparison

Table 1: Recommendation quality evaluation under different activation function modes.

Activation function type	Recommended quality evaluation index		Recall	F1-Score
ReLU	X	0.7305	—	
ELU (alpha = 0.01)	Y	0.7356	+1.88%	

experiments are based on the used ELU activation function model. Based on the maximum pooling, average pooling, and intermediate value pooling, this article mainly divides the pooling strategy into two methods: single pooling strategy and mixed pooling strategy. The specific pooling strategy is divided into the following seven kinds:

(1) Maximum single pooling strategy

(2) Average single pooling strategy

(3) A single pooling strategy for the median value

(4) Maximum-average hybrid pooling strategy

(5) Average-median mixed pooling strategy

(6) Maximum-medium mixed pooling strategy

(7) Max-median-mean hybrid pooling strategy

The verification experiment will compare the Recall rate and *F*1-Score of the above seven pooling strategies based on the recommended length of the post of 70. The specific experimental results are compared as shown in Table 2.

From the comparison of the experimental results in Table 2, it can be concluded that on the whole, the recall rate and *F*1-Score of the hybrid pooling strategy are higher than those of the single pooling strategy. The recall rate obtained by the max pooling strategy is higher than that of average pooling and median pooling, and the recall rate of median pooling is also higher than that of average pooling, which proves that maximum pooling and median pooling are relative to average pooling. It can have a better extraction effect on feature information. By comparing the *F*1-Score of each pooling strategy, it can be obtained that the *F*1-Score of the maximum-median mixed pooling is higher than other pooling strategies. Therefore, this paper adopts the maximum-median hybrid pooling strategy that performs better in terms of performance and recommendation quality as the model pooling strategy.

4.3.3. Loss Function Improvement Verification.
In this paper, the cross-entropy loss function is improved, the classification difficulty of samples is distinguished by the way of threshold setting, and the learning of difficult to distinguish samples is paid more attention to improve the quality of recommendation. This section compares the recall rate and *F*1-Score of the algorithm before and after the loss function improvement to verify the feasibility of the loss function improvement. The comparison experiment is based on the ELU activation method and the maximum-median hybrid pooling strategy. The threshold is set to 0.65. The specific experimental results are shown in Figures 7and 8.

Table 2: Comparison of experimental results of different pooling strategies.

Pooling strategy	Recommended quality evaluation index	
	Recall	F1-Score
Maximum single pooling strategy	0.7964	0.7356
Average single pooling strategy	0.7854	0.7334
A single pooling strategy for the median value	0.7915	0.7341
Maximum-average hybrid pooling strategy	0.7972	0.7363
Average-median mixed pooling strategy	0.8054	0.7376
Maximum-medium mixed pooling strategy	0.8104	0.7414
Max-median-mean hybrid pooling strategy	0.8031	0.7368
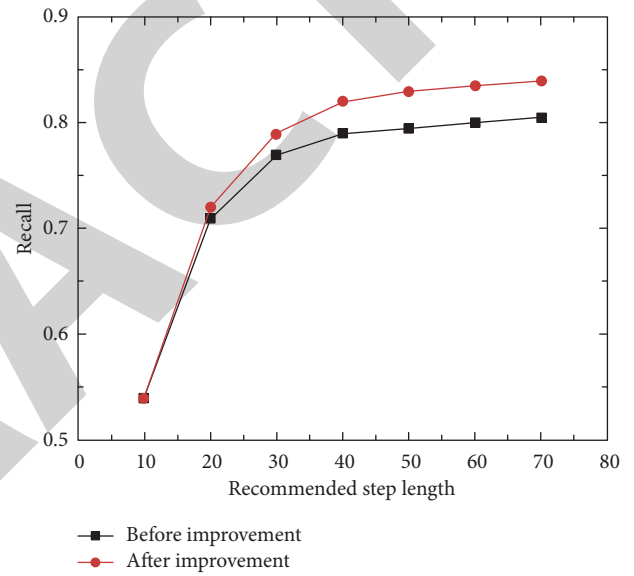
Figure 7: Recall experiment comparison chart before and after improvement.

The optimal values of recall rate and *F*1-Score before loss function improvement are 0.8104 and 0.7414. After using the improved loss function, the optimal values of recall rate and *F*1-Score are 0.8198 and 0.7432, which are increased by 1.159% and 0.243%, respectively. And the corresponding algorithm training time is reduced by about 9.6%. The comparison results prove that the model using the improved loss function can pay more attention to the learning of indistinguishable samples, which is beneficial to improve the quality of recommendations, and verifies the effectiveness of the improved loss function.

4.3.4. Algorithm Comparison Results.
The selected comparative recommendation algorithms include user-based collaborative filtering recommendation algorithm (UserCF), content-based recommendation algorithm (CBF), probability-based matrix factorization (PMF), and four-layer deep neural network (DNN). The specific experimental comparison is shown in Figures 9 and 10.

Figure 8: F1-Score experiment comparison chart before and after improvement.

Figure 9: Experimental comparison of recall rates of various algorithms.

Figure 10: F1-Score experimental comparison of each algorithm.

the data matrix for similarity calculation depending on the number of college students and jobs in the school. The larger the corresponding data scale, the more space it will take up. Therefore, the algorithm in this paper has a certain improvement in the time and space consumption of prediction and can optimize the recommended prediction.

From the experimental results in Figures 9 and 10, it can be seen that the algorithm in this paper has achieved better results in recall rate and F1-Score compared with the commonly used recommendation algorithms, which proves that the gradient boosting tree model and the feasibility of combining convolutional neural network models. With the help of the gradient boosting tree model for feature conversion and rich overall feature abstraction through multichannel convolution, combined with convolution processing and local connection processing, it can better carry out a deeper correlation between the information of college students and job information mining, and learning high-level abstract feature information can improve the recommendation quality of the model on human resources, which proves the effectiveness of the algorithm in this paper to improve the quality of human resources recommendation.

5. Conclusion

This paper establishes a college student employment data set and uses distributed data processing to extract data features through the analysis of college students' personal analysis and job information mining, so as to avoid the problem of "information flooding." At the same time, a hybrid convolutional neural network model is proposed for employment recommendation of college students. By improving the activation function, pooling strategy, and loss function in the algorithm, the quality of model prediction is greatly improved. Compared with the traditional recommendation

In a stand-alone environment, the recommendation time based on the test set, the algorithm in this paper, takes 36.33 seconds, and the DNN takes 16.67 seconds due to less calculation parameters, but the recommendation effect is not as good as the algorithm in this paper, and the recall rate of the algorithm in this paper increases by nearly 15%. For the calculation of similarity between a large number of college students and their positions, the recommended prediction time consumed is hour-level; for space occupation, the space consumption of the algorithm and DNN in this paper is mainly based on the network-level hyperparameters set by the model, while other algorithms require complementing

Research Article

Research on the Recognition Algorithm of Basketball Technical Action Based on BP Neural System

Xiangfeng Hou [iD][1] **and Qing Ji**[2]

[1]*School of Physical Education, Shanxi University, Taiyuan, Shanxi 030006, China*
[2]*School of Physical Education, Yanshan University, Qinhuangdao, Hebei 066000, China*

Correspondence should be addressed to Xiangfeng Hou; hxf50006@sxu.edu.cn

Vision-based intelligent human action recognition is the most challenging direction in the field of computer vision in recent years. It detects human actions in video sequences, extracts action features and learns action features, and then recognizes human actions in videos. This paper is based on BP neural network's basketball technique action recognition and experimental verification. First, design a basketball technique action recognition method based on BP neural network, analyze basketball actions, collect relevant test data, and divide the methods of basketball action recognition. Finally, analyze the action characteristics and waveform conditions of the upper- and lower-limb movements of the basketball action and analyze the key basketball action recognition data. The designed classification method realizes the effective recognition of basketball actions; then, the basketball recognition method used in this article is experimentally verified, and the feasibility and effectiveness of the recognition method selected in this article are verified by recognizing basketball technical actions, and the experimental results are carried out. Compared with other related studies, this method proposes a division of unit actions to complete the cycle division of basketball actions. The division results do not include the overlap of other actions, avoiding repeated calculations of actions and greatly reducing the amount of calculation of the system. In addition, the method for the recognition of basketball movement includes the separate recognition of upper- and lower-limb movements, comprehensive consideration of arm and leg movements, and a more comprehensive and accurate analysis of basketball movements.

1. Introduction

The development of modern basketball has been popularized all over the world. It not only is a competition between athletes, but also has gradually become a comprehensive competition for technological development in various countries. The breakthrough of various records in basketball matches is not only a manifestation of humans' breakthroughs in physiological limits, but also a manifestation of sports technology innovation. Therefore, the disciplines that intersect with sports science have received more and more attention from scholars. How to improve the competitive level of basketball has become the research focus of biomechanics, psychophysiology, sports medicine, and computer science. Among them, from the perspective of computer science, by extracting various parameters and

action recognition in the training process of athletes and basketball players, it can improve the scientificity of daily basketball skill training. In the study of basketball movement recognition technology based on acceleration sensors, how the sensor collects the smooth and effective acceleration data of basketball gestures is the premise, and which basketball movement recognition method is used is the key to whether the basketball movement can be effectively recognized. As a commonly used deep learning model in the field of action recognition, BP neural network has been widely used in various scenarios, such as signal processing or pattern recognition, the construction of expert systems, and the production of robots. It has good learning performance and recognition accuracy. Based on the advantages of the BP network algorithm, the recognition of basketball technical actions can be better realized. Therefore, this article mainly

studies basketball actions based on BP neural network and validates the recognition method through basketball action experiments.

2. Related Work

With the rapid development of machine learning technology, people's demand for automatic identification and monitoring of various sports image objects is increasing, so as to realize the statistics of sports data and bring more scientific training methods for the improvement of athletes' later skills. Cunado et al. [1] developed a 3D convolutional neural network architecture for human behavior recognition. When viewing motion images, the computer captures the human body motion trajectory and motion trend in real time to determine the position and shape of human body parts, then through the computer analyze the technical characteristics of the action, and report the analysis results to the coach or athlete. Trung et al. [2] proposed a progressive motion detection algorithm based on deep residual network and instance search, focusing on achieving high-precision and fine positioning of athletes' movements. Preece et al. [3] used the distance information between all joints contained in the previous image and each joint point in the current image. The distance information of the joint points corresponding to the reference action is used to describe the athlete's offset characteristics, posture characteristics, and movement characteristics. Shutler et al. [4] used the deep learning algorithm model to study the changes in the load of the hind lower-limb joints and the foot posture of running during long-distance running. Scientific running training method is used for joint damage. It can be seen from this that, with the rapid development of machine learning technology, people's demand for automatic recognition and monitoring of various sports image objects is increasing, so as to realize the statistics of sports data and bring more scientific improvement to the later skills of athletes. The research content of BP neural network is quite extensive, reflecting the characteristics of multidisciplinary and interdisciplinary technical fields. The main research work focuses on the study of biological prototypes, the establishment of theoretical models, the study of network models and algorithms, and the study of BP neural network application systems. It has been widely used in various scenarios, such as signal processing or pattern recognition, the construction of expert systems, and the production of robots. The same is true for basketball. This paper develops a set of visual movement tracking recognition and target detection models in response to the demand for automation of action assessment in basketball sports and uses BP neural network to realize basketball movement recognition.

3. BP Neural Network-Related Theoretical Methods

3.1. BP Neural Network Model

3.1.1. BP Neural Network Model Theory. Using BP network theory for basketball technical action, it can realize any nonlinear mapping from input to output of basketball technical action. First of all, for the strong ability of nonlinear mapping, the BP network can get the nonlinear mapping relationship between n-dimensional input and m-dimensional output by learning the actions of basketball training; secondly, the generalization ability is strong, and the generalization ability is through the BP network. The training extracts the nonlinear mapping relationship of basketball technical actions hidden in the sample mode and stores it in the weight matrix. When the nonsample data that has not appeared before is input into the trained BP network, the input and output can complete the ability of mapping correctly; finally, the fault tolerance is good. The process of extracting from a large number of basketball technical action samples and counting the law is the process of adjusting the weight matrix. The adjustment of the weight matrix does not depend on the error of individual samples. This is due to the error of all basketballs. The technical action samples contain correct rules, and when there are individual errors in the input samples, it will have little effect on the BP network.

Therefore, a typical BP neuron model is shown in Figure 1. It has n neuron model inputs. All the inputs of this layer are connected to the next layer by connecting weights w, and the final network output can be expressed as $o = f(wx + b)$.

If the transfer function from the hidden layer to the output layer in the BP neural network selects purlin type, any value can output the final output of the layer; and if the transfer function from the hidden layer to the output layer in the BP neural network selects the sigmoid type, the final output layer will be limited to a small range; when solving practical application problems, a detailed analysis of the problem is required to select a suitable transfer function. Generally speaking, the most common application is a single hidden layer network. The single hidden layer perceptron is usually called a three-layer perceptron; that is, the three layers include an input layer, a hidden layer, and an output layer. The schematic diagram of the three-layer BP network is shown in Figure 2 [5].

For the output layer,

$$O_k = f\left(\text{net}_k\right) = 1, 2, \ldots, l, \tag{1}$$

$$\text{net}_k = \sum_{j=0}^{m} W_{jk} y_j, \quad k = 1, 2, \ldots, l. \tag{2}$$

For the hidden layer,

$$y_j = f\left(\text{net}_j\right), \quad j = 1, 2, \ldots, m, \tag{3}$$

$$\text{net}_j = \sum_{i=0}^{n} v_{ij} x_i, \quad j = 1, 2, \ldots, m. \tag{4}$$

Generally, the transfer function $f(x)$ is a unipolar sigmoid function

$$f(x) = \frac{1}{1 + e^x}. \tag{5}$$

At this time, the transfer function $f(x)$ is continuous and derivable and satisfies
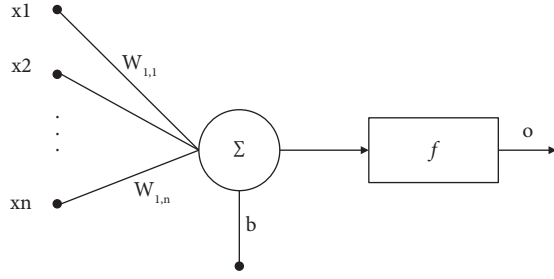
FIGURE 1: BP neuron model.
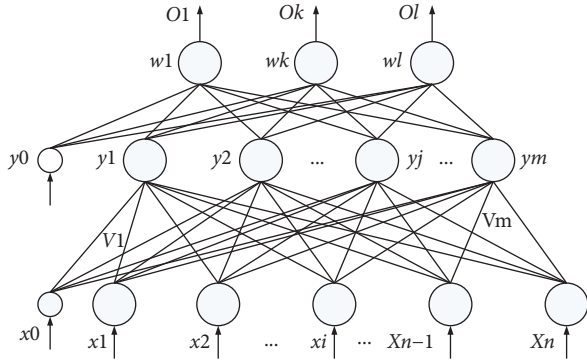
FIGURE 2: Three-layer BP neural network.

$$f(x) = f(x)[1 - f(x)]. \tag{6}$$

The main program implementation steps of BP algorithm are the following:

(1) initialization.

Initialize the weight matrix W and V with random values, set the training times counter q and the sample pattern counter p to 1, and set the error E to 0, the learning rate $\eta \in (0,1]$, and use a positive decimal as the network training requirement The achieved accuracy is E_{min}.

(2) Input the training sample pair and calculate the output of each layer.

Use the current samples X_p and d_p to assign values to the vector arrays X and d and calculate the components of Y and O.

(3) Calculate the output error of the network.

Assuming that there are P pairs of training samples, there are different errors for different sample networks $E^P = \sqrt{\sum_{k=1}^{l} (d_k^p - o_k^p)^2}$.

In this paper, the root mean square error $E_{RME} = \sqrt{1/P \sum_{P-1}^{P} (E^p)^2}$ is used as the total output error of the network.

(4) Calculate the error signal of each layer.

Use the above formula to calculate δ_k^o and δ_j^y.

(5) Modify the weights of each layer.

Use the above formula to calculate the components in W and V. Because the weight adjustment method

described above has low learning efficiency and slow convergence speed, it is easy to form a local minimum but cannot obtain the global optimum. It is learned during training. The new sample has many problems such as forgetting the trend of the old sample. Therefore, in this paper, we choose to use the increasing momentum term in the weight adjustment formula to solve this problem; that is, not only the adjustment of the error gradient descent direction at time t but also the adjustment of the error gradient descent direction before time t is considered.

Assume that the weight matrix of a certain layer is represented by W and the input vector of a certain layer is represented by X; when the momentum term is added, the expression of the weight adjustment vector is

$$\Delta W(t) = \eta \delta X + \alpha \Delta W(t-1). \tag{7}$$

α is called the momentum coefficient, generally $\alpha \in (0,1)$. In addition, the above problems can also be solved by adaptively adjusting the learning rate or introducing the jitter factor λ.

(6) Check whether all samples have completed a rotation training.

If $p < p$, then p and q increase by 1, and return to step (2); otherwise skip to step (7).

(7) Detect whether the total error of the network meets the accuracy requirements.

When $E_{RME} < E_{min}$, the training ends; otherwise, set E to 0 and p to 1 and return to Step (2). The implementation process of BP algorithm is shown in Figure 3.

3.2. Action Recognition.

Human body posture recognition is mainly to study the description of human body posture and predict human behavior. The recognition process refers to the process of recognizing human body actions based on changes in the positions of joints in the human body in a designated image or video screen. At present, many scholars have made the human action recognition of inertial sensors the research focus. Through the recognition of human actions by inertial sensors, the research results are applied to wearable devices. Human action recognition is mainly divided into data collection, preprocessing, data division, feature extraction, and model training. The data collection stage refers to the use of inertial sensors to collect the human body's movement signals and physiological signals and observe the human body's movement and physiological changes under the state of motion; the data preprocessing stage refers to the algorithm to denoise and normalize the messy original data through the algorithm. To make it meet the requirements of algorithm identification, the data division stage refers to the selection of individual actions for analysis in the data after the preprocessing is completed; the data feature extraction stage refers to extracting the relevant attribute features to become the data required for the sample location; the final model training stage refers to the
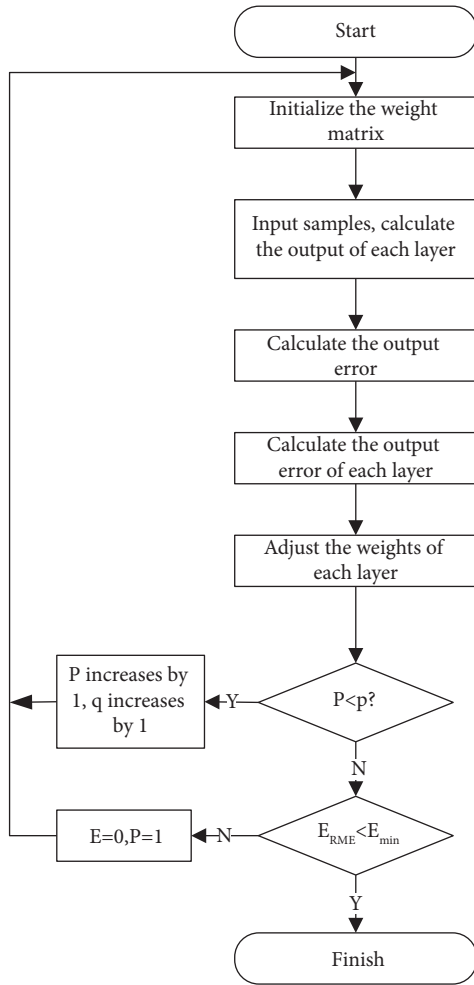
Figure 3: BP process algorithm.

algorithm model according to different classification principles to realize data recognition; the final model training stage refers to the realization of data recognition based on algorithm models with different classification principles, as shown in Figure 4.

3.2.1. Detection of Human Moving Targets.

In the process of classification and recognition of human walking features, accurately detecting moving targets is a very important step. Target detection refers to first detecting the changed area in the video or image sequence and then separating the detected changed target from the background image. Human motion recognition is used in various scenarios, such as robot navigation, security monitoring, medical image analysis, video image coding, and transmission. Detecting and separating moving targets from images provides great convenience for image processing and feature extraction.

3.2.2. Model and Statistics-Based Feature Extraction.

The feature extraction method based on the object model usually uses the geometric relationship of the object or the feature point of the object to estimate. The basic idea is to use a certain geometric model or structure to represent the structure and shape of the object and extract certain object features in the model and correspondence is established between the images. Then, the parameter information of these models is analyzed, the spatial posture of the object is estimated by geometric or other methods, and finally the image is matched with a template [6–11]. The random field model method is currently the mainstream method in feature extraction, and Markov random field and Gibbs random field are often selected.

At present, some computer vision and identity recognition researchers use model-based methods in the process of target feature extraction. In the experiment of extracting human walking characteristics, Han converted the human body contour data into a motion model, converted a two-dimensional plane into a three-dimensional model, and processed the data through the least square method. Lee et al. used an ellipse to represent the various parts of the human body and then used the ellipse parameters such as the ratio of long and short axes, the center of mass coordinates, and the body structure parameters as the characteristics of gait recognition [12–15]. The swinging process of the lower limbs of the human body is similar to the swinging of the pointer of a clock. Based on this idea, Cunado et al. constructed a pendulum model to extract the angle between the legs and used this as a feature for gait recognition.

4. Basketball Action Recognition Model Based on BP Neural Network

4.1. Feature Extraction of Basketball Technical Action.

The human body movements involved in basketball are more complicated. Figure 5 is an analysis of the composition of the basketball posture. The state of the root limbs divides the basketball movement posture into a static state and a movement state [16]. The static state refers to the state where the posture of the limbs remains unchanged, and the state of motion refers to the state of the limbs when performing basketball actions. Shooting, receiving, passing, dribbling and jumping, walking, and running of the lower limbs are defined as unit actions. In the state of motion, the action can be divided into instantaneous action and continuous action according to whether it has periodicity. For this reason, this paper proposes a division method based on unit action extraction.

4.1.1. Movement Status Division.

Dispersion is an index that can be used to reflect the degree of difference between observed variable values, and dispersion is an index that indicates the difference between sensor signal sample values. Taking angular velocity as an example, ω_n^x represents the nth angular velocity value in the x-axis direction, ω_{n-1}^x represents the n-1th angular velocity data in the x-axis direction, and d_n^x represents the sensor at the nth and previous moments. The angular velocity difference in the x-axis direction at a certain moment where there exists the dispersion index ω_{n-1}^x can be obtained by the following formula:

$$d_n^x = \left| \omega_n^x - \omega_{n-1}^x \right|. \tag{8}$$

Angular velocity and acceleration data are included in the motion data package. Comprehensive consideration of
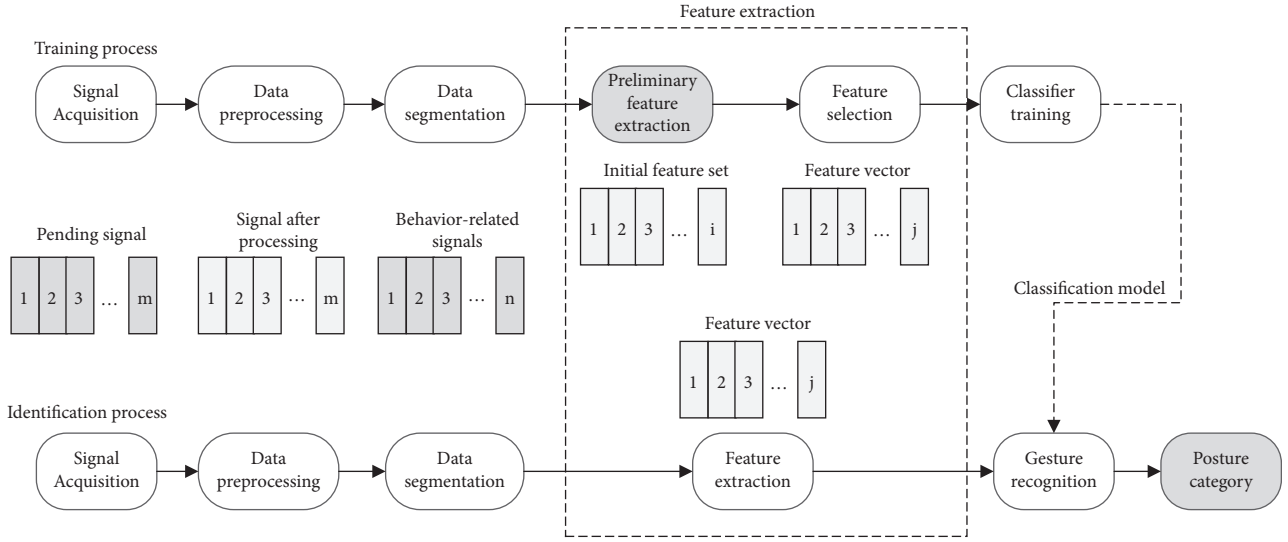
FIGURE 4: Flow chart of the human body gesture recognition method.

the data characteristics of each sensor is an important step to accurately divide the movement. d_n^x represents the dispersion of the acceleration sensor data at the nth time, d_n^g represents the dispersion of the angular velocity sensor data at the nth time, and $d_n^{a_x}$, $d_n^{a_y}$, $d_n^{a_z}$, $d_n^{g_x}$, $d_n^{g_y}$, and $d_n^{g_z}$, respectively, represent the dispersion of the acceleration and angular velocity of each axis; then, d_n^a and d_n^g can be obtained by the following formulas, respectively:

$$d_n^a = d_n^{a_x} + d_n^{a_y} + d_n^{a_z}. \tag{9}$$

$$d_n^g = d_n^{g_x} + d_n^{g_y} + d_n^{g_z}. \tag{10}$$

In the case of sports, the changes of the athletes' actions will be reflected in the sensor data in real time. Since the dispersion reflects the degree of difference between the sensor data, the division of the athlete's physical state can refer to the characteristics of the dispersion. In static conditions, the angular velocity and acceleration dispersion are kept below the threshold values λ_g and λ_a in sequence. The state of the athlete's limbs at the nth moment is represented by γ_n. When γ_n is 0, it indicates the static state, and when γ_n is 1, it indicates the motion state. The definition is shown in

$$\gamma_n = \begin{cases} 0, d_n^a < \lambda_a \text{ and } d_n^g < \lambda_g \\ 1, d_n^a \geq \lambda_a \text{ and } d_n^g \geq \lambda_g \end{cases} \tag{11}$$

The degree of data dispersion of different sensors is calculated, and different motion states can be distinguished and identified for different thresholds. At the same time, the instantaneous action and continuous action can be distinguished by whether the angular velocity of each sensor changes periodically. It can be known that the extraction of basketball motion state data can be achieved through motion state division.

4.1.2. Unit State Division. Studies have shown that the division of motion can be achieved with the help of the motion data of the legs and arms. In describing the angle change in the process of rigid body movement, the angular velocity has a very good performance. Therefore, the angular velocity can be used to divide the data. In reducing the influence of external noise, the Kalman filter algorithm can be used to fuse acceleration, magnetic field strength, and angular velocity data to effectively solve the problem.

4.1.3. Feature Extraction of Basketball Posture. After data division, the unit action data is obtained, which is composed of acceleration and angular velocity, a_n^x, a_n^y, and a_n^z represent the acceleration of the three axes of the nth sampling point, and g_n^x, g_n^y, and g_n^z indicate the angular velocity of the three axes at the nth sampling point. Use v and g_n to denote the acceleration vector sum and the angular velocity vector sum, respectively, which can be obtained by the following formulas:

$$a_n = \sqrt{(a_n^x)^2 + (a_n^y)^2 + (a_n^z)^2}. \tag{12}$$

$$g_n = \sqrt{(g_n^x)^2 + (g_n^y)^2 + (g_n^z)^2}. \tag{13}$$

An eight-dimensional vector can be constructed through the four parameters of three-axis acceleration, three-axis angular velocity, combined acceleration, and combined angular velocity, where the sampling point of each action is represented by N, and an $N \times 8$-dimensional matrix can form a sample, which can be based on the fact that each sample calculates the characteristics of each dimension. The time domain features in this experiment are composed of mean
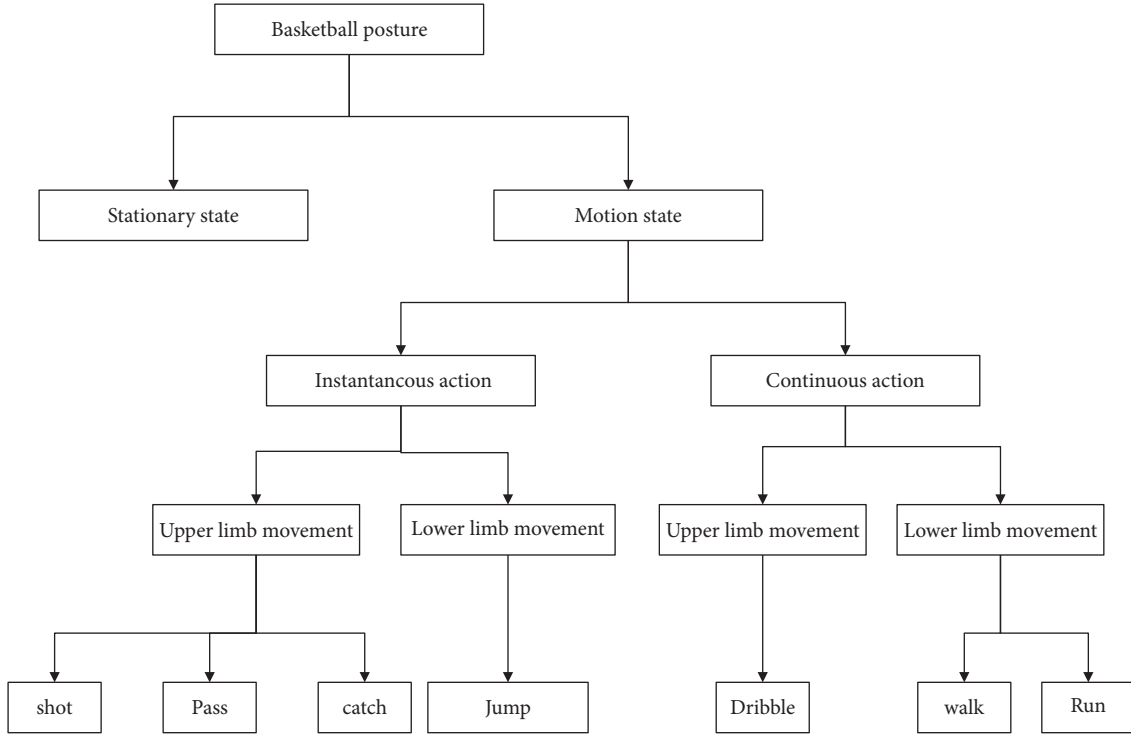
FIGURE 5: The composition of motion posture in basketball.

and variance. μ_a and δ^2 are used to represent the mean and variance of a certain component of acceleration in a unit action. It can be calculated by the following formulas, where a is the acceleration of a certain component:

$$\mu_a = E(a) = \frac{1}{N} \sum_{i=1}^{N} a_i. \tag{14}$$

$$\delta^2 = \frac{1}{N}(a_i - \mu_a)^2. \tag{15}$$

The peak value of the discrete Fourier transform and its corresponding frequency is the performance of frequency domain characteristics. The process of converting the signal from the time domain to the frequency domain can be realized by the discrete Fourier transform method. Among them, $S_{\text{DFE}}(n)$ represents the Fourier transform result of the nth sampling point, and j represents the imaginary part calculated by the following formula:

$$S_{\text{DFE}}(n) = \sum_{i=1}^{N-1} a_i e^{-j2\pi/N\text{in}}. \tag{16}$$

From the Fourier transform result, $S_{\text{DFE}}(K)$ represents the peak value, the sampling point K corresponds to the Fourier transform peak value, and the Fourier transform frequency f is obtained by the following calculation formula, where the sampling frequency is represented by f_s:

$$f = K \times \frac{f_s}{N}. \tag{17}$$

Through feature calculation, the time domain and frequency domain features of each dimension data in the sample can be obtained, thereby constructing a 32-dimensional feature vector, as shown in Tables 1 and 2.

In the experiment, sensor nodes are used to detect the movement information of different limbs to fix them on the forearms and calves of the experimenter. Due to the different placement of nodes, the division of upper- and lower-limb action sets of the data set is also different. Therefore, in order to further realize the movement division of the upper and lower limbs, it is necessary to establish different sample classifiers for the data set. Combining the results of the upper and lower limbs, the basketball posture of the current candidates can be obtained [17, 18].

4.2. BP Network Design

4.2.1. Preparation of Training Sample Set.
First, the input and output variables of the modeling system should be screened. Usually, the goal to be achieved is selected as the output variable [5, 19]. The variables that have a large impact on the output, easy to extract, and have little correlation should be selected as the input variables of the network. Use samples with or without a certain input variable to train the network and compare the effects to determine whether the variable is suitable as an input variable. After selecting the input and output variables, they must be represented. From the perspective of the nature of input and output, it can be divided into two types: language variables and numerical variables. Variables expressed in natural language are linguistic variables, and various attributes of things represented by natural language are their "linguistic values." The sample data that has undergone scale transformation and distribution transformation can also be used for

TABLE 1: Time domain features of data.

Feature	Feature ID	Feature description
Time domain characteristics	1 ~ 3	Acceleration sensors x, y, and z axis data mean value
	4	Mean value of total acceleration
	5 ~ 7	The mean value of the gyroscope's x, y, z axis data
	8	Mean value of closing angular velocity
	9 ~ 11	Acceleration sensors x, y, and z axis data variance
	12	Variance of total acceleration
	13 ~ 15	Variance of the total velocity: the variance of the gyroscope's x, y, and z axis data
	16	Variance of angular velocity

TABLE 2: Frequency domain features of data.

Feature	Feature ID	Feature description
Frequency domain analysis	17, 18	The DFT peak value of the acceleration sensor x data and the frequency of the corresponding peak value
	19, 20	The DFT peak value of the acceleration sensor y data and the frequency of the corresponding peak value
	21, 22	The DFT peak value of the acceleration sensor z data and the frequency of the corresponding peak value
	23, 24	The DFT peak value of the combined acceleration and the frequency of the corresponding peak value
	25, 26	The DFT peak value of the gyroscope x data and the frequency of the corresponding peak value
	27, 28	The DFT peak value of the gyroscope y data and the frequency of the corresponding peak value
	29, 30	The DFT peak value of the gyroscope z data and the frequency of the corresponding peak value
	31, 32	The DFT peak value of the gyroscope x data and the frequency of the corresponding peak value

network training. After transformation processing, the input and output data of the network are limited to the interval of [−1,1] or [1,0]. This process is called scale transformation, which can make the positions of various variables closer. The commonly used transformation formula is

$$\overline{x_i} = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}}. \tag{18}$$

$x_{\min}$ is the minimum value of the variable; $x_{\max}$ is the maximum value of the variable; x_i is input or output data

Normalization is a linear transformation, and when the distribution of the sample is not ideal, the distribution transformation is needed to change the sample. This distribution law generally uses logarithmic transformation, as well as square root and cube root transformation. The distribution transformation is a nonlinear transformation. The training sample set also needs to be prepared. The information capacity of the network may affect the classification ability of the BP neural network. The information capacity of the network is characterized by the total number of network weights and thresholds n_w. Generally, the matching relationship between the number of training samples P and the given training error ε is

$$P \approx \frac{n_w}{\varepsilon}. \tag{19}$$

There is a reasonable matching relationship between the information capacity and the number of training samples in the network. When solving the problem of less actual training sample data, it is difficult to meet the above requirements. For a certain number of samples, a suitable network parameter is required. Too few cannot express all the hidden laws, and too many samples will not be fully

trained. Based on experience, the number of training samples can be 5–10 times n_w. After determining the number of sample sets, the samples need to be selected and organized. The samples should be representative and balanced in categories; when organizing the samples, the input samples should be randomly selected from the training sample set, or samples of different categories should be cross-input.

4.2.2. Design of BP Network Structure. In the BP network structure design, it is necessary to design the number of nodes in the input layer and the output layer. At the same time, the design of the number of hidden layers and the number of hidden nodes of the network is also particularly important. The number of nodes in the input layer and output layer can be determined according to actual application requirements. The number of nodes in the input layer is usually obtained through the dimensions of the training sample vector; the number of nodes in the output layer is usually used as the dimension of the output space of the approximation function in the approximation network and is usually used as the number of categories in the classification network. Theoretical analysis shows that a network with an S-shaped hidden layer plus a linear input layer can approximate any rational function. Increasing the number of layers can improve accuracy and reduce errors, thereby making the network more complex. Therefore, in actual demand applications, usually consider designing a network with hidden layers.

The number of samples in training, the complexity of the hidden law, and the size of the noise can determine the number of hidden nodes. Since it is difficult to grasp the complexity of the sample rule and the size of the noise in the actual process, it is difficult to design the number of hidden
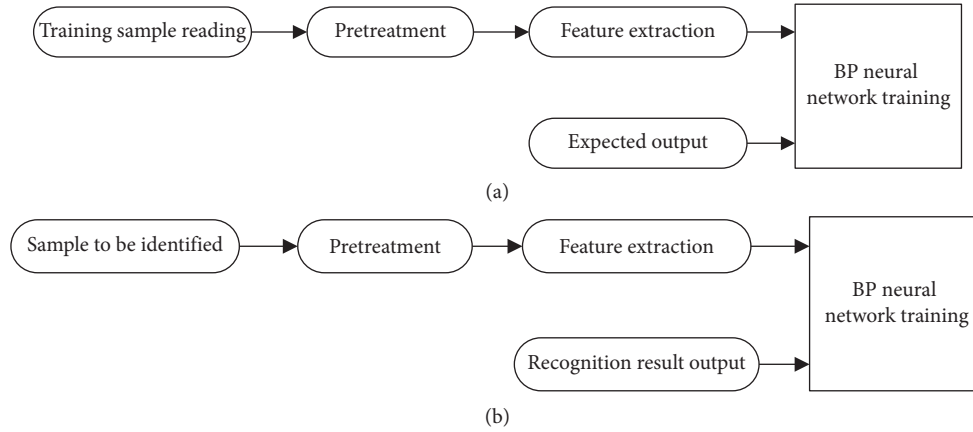
FIGURE 6: BP neural network recognition process. (a) BP neural network training process. (b) BP neural network recognition process.

nodes. In determining the optimal number of hidden nodes, the same sample set can be used to train networks with different numbers of hidden nodes, because the network error is the smallest at this time. The number of hidden nodes can be roughly estimated by the following formulas:

$$m = \sqrt{n + 1} + \alpha, \tag{20}$$

$$m = \log_2 n, \tag{21}$$

$$m = \sqrt{nI}, \tag{22}$$

$$m = \sqrt{0.43nl + 0.12l^2 + 0.51 + 2.54n + 0.77l + 0.35}, \tag{23}$$

M is the number of hidden nodes; n is number of output layer nodes; l is number of output layer nodes; α takes a constant between 1– and 10.

4.2.3. Design of Parameters.

Under normal circumstances, a small random value is selected to assign the initial weight, which not only ensures that the input value is small and can work in the area where the slope of the excitation function changes greatly, but also can avoid the unreasonable and unlimited increase of the absolute value of some weights after multiple continuous learning and training. The initial weight is generally a random number of (−1, 1). The learning rate is generally smaller in order to ensure the stability of the system. A large learning rate may make the system unstable, but a small learning rate makes the training time longer. In order to ensure that the error value of the network does not jump out of the bottom of the error surface and eventually tends to the minimum error, the learning rate is within the range of 0.01–1. The recognition process of BP network model basketball technical action recognition based on BP neural network mainly includes two aspects, namely, the training process of BP neural network and the basketball technical action recognition process of BP neural network. The basketball technical action recognition process of BP neural network is shown in Figure 6.

First, collect a lot of sample data amounts to train the designed BP neural network. After reading the training sample data, the sample data is preprocessed to extract the acceleration characteristic value of the gesture action, and the BP neural network is trained under the guidance of the expected output (teacher signal) to achieve the preset target accuracy. The trained BP neural network can then be used to recognize the samples to be recognized for predefined basketball technical actions. Of course, the sample to be recognized also needs to be preprocessed to extract the acceleration characteristic value of the gesture action, then use the trained BP neural network to recognize the technical action, and finally output the recognition result.

5. Experimental Design and Result Analysis

5.1. Experimental Design.

The entire data collection process is mainly to collect basketball's dribbling, running, standing dribbling, catching, shooting, passing, jumping, and other actions, as shown in Figure 7; each action is effective and repeated for 50 seconds per rate. A total of 5650 samples were collected, of which 3000 were upper-limb basketball movements and 2,650 were lower-limb basketball movements. The sampled movements were completed in strict accordance with regulations. The number of specific movements was recorded by the scorer, as shown in Table 3.

5.2. Result Analysis and Discussion.

From the results reflected in Tables 4 and 5, it can be clearly seen that the BP neural network has a good recognition effect in the recognition of upper-limb movements and lower-limb movements. Among them, the recognition accuracy rates of the entire upper and lower extremity basketball technical movements are as high as 93.2% and 99.2%, respectively, and the average recall rates of the upper and lower extremity basketball technical movements are as high as 93.2% and 99.2%, respectively. The recognition accuracy rates are as high as 93.2% and 99.2%, and the average recall rates of upper and lower limbs are 93.2% and 99.2%, respectively. Regardless of the algorithm, the accuracy of upper- and lower-limb basketball actions is, respectively, 97% and 84.9%. As for the recognition rate of upper-limb basketball action which is lower than that of lower-limb basketball action, the main reason is that upper-limb

Figure 7: Nine kinds of basketball moves.

Table 3: The number of different movements of the upper and lower limbs.

Upper and lower limbs	Behavior	Quantity
Arm movements	Dribble (standing)	500
	Dribble (walking)	500
	Dribble (running)	500
	Shot	500
	Pass	500
	Catch the ball	500
Leg movements	Run	500
	Jump	500
	Shot	500
	Dribble (running)	600
	Dribble (walking)	550

Table 4: Classification results of upper-limb movements by different classification algorithms.

Behavior	C4.5 Accuracy	C4.5 Recall rate	SVM Accuracy	SVM Recall rate	NB Accuracy	NB Recall rate	BP-ANN Accuracy	BP-ANN Recall rate
Catch the ball	0.937	0.949	0.964	0.977	0.899	0.928	0.964	0.979
Pass	0.909	0.931	0.978	0.989	0.935	0.954	0.982	0.979
Shot	0.924	0.910	0.986	0.992	0.956	0.918	0.980	0.984
Dribble (walking)	0.888	0.860	0.929	0.920	0.793	0.944	0.938	0.919
Dribble (standing)	0.788	0.770	0.819	0.570	0.780	0.704	0.910	0.881
Dribble (running)	0.781	0.805	0.693	0.863	0.805	0.760	0.881	0.898
Average	0.862	0.862	0.875	0.868	0.851	0.851	0.935	0.935

Table 5: Classification results of lower-limb movements by different classification algorithms.

Behavior	C4.5 Accuracy	C4.5 Recall rate	SVM Accuracy	SVM Recall rate	NB Accuracy	NB Recall rate	BP-ANN Accuracy	BP-ANN Recall rate
Jump	0.998	0.995	0.985	0.999	0.981	0.979	0.986	0.999
Run	0.984	0.982	0.999	0.980	0.960	0.979	0.998	0.986
Walk	0.974	0.974	0.990	0.995	0.982	0.960	0.994	0.995
Average	0.983	0.983	0.991	0.991	0.972	0.973	0.994	0.994

Table 6: Classification results of lower-limb movements by different classification algorithms.

Behavior	C4.5 Accuracy	C4.5 Recall rate	SVM Accuracy	SVM Recall rate	NB Accuracy	NB Recall rate	BP-ANN Accuracy	BP-ANN Recall rate
Catch the ball	0.945	0.947	0.969	0.975	0.865	0.930	0.986	0.985
Pass	0.920	0.940	0.978	0.991	0.945	0.960	0.991	0.991
Dribble	0.985	0.986	0.998	0.991	0.998	0.986	0.997	0.994
Shot	0.946	0.925	0.987	0.995	0.958	0.910	0.976	0.986
Average	0.964	0.964	0.990	0.990	0.965	0.965	0.991	0.991

movement is relatively more complicated and some actions have similarities, as shown in Table 6.

From the results reflected in Tables 4 and 5, it can be clearly seen that the BP neural network has a good recognition effect in the recognition of upper-limb movements and lower-limb movements. The recognition accuracy rates are as high as 93.2% and 99.2%, and the average recall rates of upper and lower limbs are 93.2% and 99.2%, respectively. Regardless of the algorithm, the accuracy of upper- and lower-limb basketball actions is, respectively, 97% and 84.9%. As for the recognition rate of upper-limb basketball action which is lower than that of lower-limb basketball action, the main reason is that upper-limb movement is relatively more complicated and some actions have similarities, as shown in Table 6.
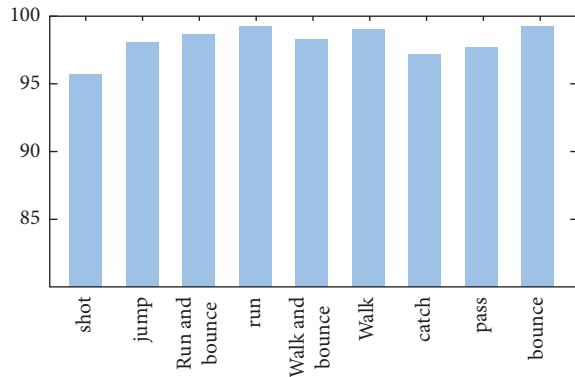
FIGURE 8: BP-ANN's recognition results of basketball gestures.

Experiments show that the best way to recognize upper- and lower-limb movements in basketball is the BP artificial neural network method. After the BP artificial neural network establishes the classifier of the basketball movement of the upper and lower limbs, it finally and effectively recognizes the basketball movement, as shown in Figure 8. It can be seen from the figure that the basketball movement recognition model established by the BP neural network is recognized. The accuracy rate can exceed 95%, and the recognition rate of the entire basketball action is as high as 98.85%.

6. Conclusion

On the basis of studying and summarizing the advantages and disadvantages of existing basketball action recognition methods, this paper designs a basketball action recognition method based on BP neural network. BP network has distributed storage and parallel processing of information, strong nonlinear mapping ability, and robustness. The advantages of good performance, fault tolerance, and strong generalization ability have successfully reduced the complexity of the traditional action recognition algorithm implementation process, improved the self-organization and self-adaptability of the algorithm implementation, and improved the recognition rate. On the basis of the above-mentioned research work, the feasibility of the recognition algorithm selected in this paper is verified through experiments.

Through analyzing the characteristics of basketball actions and aiming at the characteristics of basketball actions, it is proposed to divide the basketball action data into two stages. After the entire basketball action is fully recognized, each action unit is extracted for separate analysis, and finally test the characteristic data of 32 basketball technical movements, classify the samples according to the data obtained from the upper and lower limbs, construct a basketball movement classifier based on the technical movement data of the upper and lower limbs, identify the movements according to the classification algorithms commonly used by experts and scholars, and finally construct a most suitable action. Through a series of test actions, it is shown that the BP artificial neural network has the best recognition effect on the upper and lower limbs of basketball

actions, the overall accuracy rate is controlled above 99%, and the recognition accuracy of the movement posture in basketball is 98.85%. It shows that the effect of the basketball gesture recognition method used in this article is quite satisfactory. This research will be of great significance to the intelligent training and technical correction of basketball players in the future.

Data Availability

The dataset can be accessed upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

References

[1] D. Cunado, M. S. Nixon, and J. N. Carter, "Using gait as a biometric, via phase-weighted magnitude spectra//," in *Proceedings of the First International Conference on Audio and Video-Based Biometric Person Authentication*, pp. 93–102, Springer-Verlag, Crans-Montana, Switzerland, 12 March 1997.

[2] N. T. Trung, Y. Makihara, H. Nagahara, Y. Mukaigawa, and Y. Yagi, "Inertial-sensor-based walking action recognition using robust step detection and inter-class relationships[C]," in *Proceedings of the International Conference on Pattern Recognition (ICPR)*, pp. 3811–3814, Tsukuba International Congress Center, Tsukuba, Japan, 11 November 2012.

[3] S. J. Preece, J. Y. Goulermas, L. P. J. Kenney, and D. Howard, "A comparison of feature extraction methods for the classification of dynamic activities from accelerometer data," *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 3, pp. 871–879, 2009.

[4] J. D. Shutler, M. S. Nixon, and C. J. Harris, "Statistical gait recognition via velocity moments," in *Proceedings of the IEEE Colloquium on Visual Biometrics*, IEEE, London, UK, March 2000.

[5] L. Zhao, Y. Jiao, J. Chen, and R. Zhao, "Image style transfer based on generative adversarial network," in *Proceedings of the 2021 International Conference on Computer Network, Electronic and Automation (ICCNEA)*, pp. 191–195, IEEE, Xi'an, China, September 2020.

[6] L. A. Schwarz, A. Bigdelou, and N. Navab, "Learning gestures for customizable human-computer interaction in the operating room," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2011*, pp. 129–136, Springer, Berlin, Germany, 2011.

[7] C. K. Lim, Z. Luo, I.-M. Chen, and S. H. Yeo, "Wearable wireless sensing system for capturing human arm motion," *Sensors and Actuators A: Physical*, vol. 166, no. 1, pp. 125–132, 2011.

[8] A. V. Dowling, J. Favre, and T. P. Andriacchi, "Inertial sensor-based feedback can reduce key risk metrics for anterior cruciate ligament injury during jump landings," *The American Journal of Sports Medicine*, vol. 40, no. 5, pp. 1075–1083, 2012.

[9] F. Albinali, S. Intille, W. Haskell, and M. Rosenberger, "Using wearable activity type detection to improve physical activity energy expenditure estimation[C]," in *Proceedings of the 12th ACM international conference on Ubiquitous computing*, pp. 311–320, ACM, Copenhagen, Denmark, 26 September 2010.

[10] B. Bhanu and J. Han, "Individual recognition by kinematic-based gait analysis," *Proceedings - International Conference on Pattern Recognition*, vol. 3, pp. 343–346, 2012.

[11] S. Sarkar, P. J. Phillips, Z. Liu, I. R. Vega, P. Grother, and K. W. Bowyer, "The humanID gait challenge problem: data sets, performance, and analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 2, pp. 162–177, 2005.

[12] J. B. Hayfron-Acquah, M. S. Nixon, and J. N. Carter, "Automatic gait recognition by symmetry analysis," *Pattern Recognition Letters*, vol. 24, no. 13, pp. 2175–2183, 2003.

[13] A. Taniguchi, K. Watanabe, and Y. Kurihara, "Measurement and analyze of jump shoot motion in basketball using a 3-d acceleration and gyroscopic sensor[C]," in *Proceedings of the SICE annual conference (SICE)*, pp. 361–365, Akita, Japan, 20 August 2012.

[14] G. M. Paul, B. P. David, and L. M. Clare, "The physical and physiological demands of basketball training and competition [J]," *International Journal of Sports Physiology and Performance*, vol. 5, no. 1, pp. 75–86, 2010.

[15] M. Zhang and A. A. Sawchuk, "A feature selection-based framework for human activity recognition using wearable multimodal sensors[C]," in *Proceedings of the 6th International Conference on Body Area Networks*, pp. 92–98, ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), Beijing, China, 7 November 2011.

[16] L. N. N. Nguyen, D. Rodríguez-Martín, A. Català, C. Pérez-López, A. Samà, and A. Cavallaro, "Basketball activity recognition using wearable inertial measurement units[C]," in *Proceedings of the XVI International Conference on Human Computer Interaction*, pp. 60–67, Vilanova i la Geltrú, Vilanova i la Geltru Spain, 7 September 2015.

[17] A. D. Ignatov and V. V. Strijov, "Human activity recognition using quasiperiodic time series collected from a single tri-axial accelerometer," *Multimedia Tools and Applications*, vol. 75, no. 12, pp. 7257–7270, 2016.

[18] H.-C. Lin, S.-Y. Chiang, K. Lee, and Y.-C. Kan, "An activity recognition model using inertial sensor nodes in a wireless sensor network for frozen shoulder rehabilitation exercises," *Sensors*, vol. 15, no. 1, pp. 2181–2204, 2015.

[19] B. Longstaff, S. Reddy, and D. Estrin, "Improving activity classification for health applications on mobile devices using active and semi-supervised learning[C]," in *Proceedings of the Pervasive Computing Technologies for Healthcare (Pervasive-Health), 2010 4th International Conference on-NO PERMISSIONS*, pp. 1–7, IEEE, Munich, Germany, April 2010.

Research Article

Analysis of Behavioral Image Recognition of Pan-Entertainment of Contemporary College Students' Network

Hong Cui [1] **and Yuan Wang**[2]

[1]Xi'an Technological University, Shaanxi, Xi'an 710002, China
[2]Xi'an Polytechnic University, Shaanxi, Xi'an 710048, China

Correspondence should be addressed to Hong Cui; cuihong@xatu.edu.cn

With the continuous update and iteration of network technology and technological innovation, the handheld smart media of college students will become more and more sensitive. With the advancement of economic globalization, various ideologies and cultures in the world will rapidly invade, and the "pan-entertainment" of online media may intensify. Only through the government's supervision function and the self-discipline of the internet industry, we can strictly control and screen positive values. In order to better establish the correct employment value orientation of university students and further analyze the importance of the "pan-entertainment" behavior image recognition of college students, this study analyzes the related technology and basic theory of behavior recognition. After introducing several mainstream methods, the traditional dual-stream convolutional network method is improved, and the time information and spatial information extracted by the two channels are discussed for the weighted fusion of feature maps. Finally, using R(2 + 1)D structure and dual-stream network structure design, a deep learning-based spatiotemporal convolution behavior recognition algorithm is proposed. The proposed algorithm is tested and analyzed on the datasets UCF101 and HMDB51. The specific work content is as follows: (1) to summarize the widely used video behavior classification methods proposed so far and discuss the future development. Then, it mainly analyzes the existing technical bottlenecks of some methods based on deep learning methods and summarizes and explores an efficient, stable, and accurate spatiotemporal feature joint extraction and learning method theory. (2) The design of spatiotemporal convolutional network algorithm framework is proposed, the method of segmentation processing of long video is studied, the improvement of the dual-stream network decision-level fusion method is studied, and the R(2 + 1)D network is reorganized. The network algorithm is trained and tested on the UCF-101 dataset and HMDB-51 dataset under the condition of calling the pretrained model. Finally, the accuracy is compared with the existing classic algorithms to obtain better accuracy, which proves the effectiveness of the algorithm for the "pan-entertainment" behavioral image recognition of contemporary college students.

1. Introduction

At present, the "pan-entertainment" culture is increasingly permeating every corner of people's life, and college students are most active among the consumer groups of the "pan-entertainment" culture. Due to the immature development of ideology, psychology, and other aspects and the limited ability to recognize "pan-entertainment" culture, some college students are easily affected by the negative influence of "pan-entertainment" culture, but it is difficult to identify and analyze such behavior images. With the rapid development of computer science and the modern internet

world's demand for massive amounts of pictures and video information, contemporary college students' network "pan-entertainment" image behavior recognition, that is, the machine acquires the video taken by the camera, and then self-learning after preprocessing, combined with scene recognition, detect the actions of college students in the image. The machine is made smarter and more able to approach the characteristics of humans to detect images. From the 1970s to the present, some progress has been made in image recognition and analysis of abnormal motion analysis in this field. Nowadays, image behavior recognition has been successfully applied to life and can be seen

everywhere. Video recognition technology with a time dimension has become a hot and difficult point of current scientific research. For example, Wang et al. [1] adjusted and encoded DMM into pseudo-RGB images, converted their spatial and temporal behavior information into texture information, and fused three independent ConvNets networks for training and recognition. Rahmani and Mian [2] proposed a deep sequence learning view-invariant human behavior model. The method is to input each frame of a deep image into a specific convolutional neural network to learn advanced features and then transfer the human behavior in the unknown image to the model. Training and classification. Jin et al. designed a new type of RGB-D image recognition framework. The framework calculates the position deviation of 3D bone joint points and then uses the space independent nature of the joint points in the bag-of-words model to complete the vector offset and recognize human behavior. Wang et al. [3] constructed three different types of dynamic depth images, namely, dynamic depth images, dynamic depth regular images, and dynamic depth motion regular images to extract behavioral features in in-depth image sequences. Therefore, image behavior recognition technology has become one of the important contents of research and experimentation by scholars at home and abroad.

2. Behavior Image Recognition Technology Based on Deep Learning

2.1. Dual-Stream Convolutional Neural Network. Based on video processing, it can be naturally decomposed into two levels of temporal characteristics and spatial characteristics. Spatial features mainly involve the appearance description and environmental background of the subject in the "pan-entertainment" video actions of college students. Its essence is the "pan-entertainment" image recognition of static college students. The ability of spatial modeling should be strengthened to more efficiently obtain space. This deep convolutional neural network has achieved good results. The temporal feature is to clearly describe the motion feature between video frames by capturing the optical flow feature and the optical flow displacement field stacked between several consecutive frames. The network does not need to implicitly estimate motion, which reduces the difficulty of recognition to a certain extent. In 2014, Karen et al. proposed a dual-stream convolutional neural network that separately extracts and trains spatial and temporal features. Each stream of its architecture uses a convolutional neural network, and the final result is obtained after the average combination of the softmax layer scores of the two streams [4–7].

The overall framework of the dual-stream network is shown in Figure 1 below. The first network is a spatial stream network, which performs feature extraction and training on a single video frame after video preprocessing. This system is similar to image recognition. Because the action has important correlation information with the moving subject and the specific environment, the static "pan-entertainment" picture information of college students has become an important clue. The second network is the time stream, which extracts and trains the time information of the video. The optical flow diagram of multiple consecutive frames of the video is collected as input, and the motion information of the video is captured. The different information obtained between the two networks plays a complementary role [8], and it is also proved through experiments that the average result of the two convolutional neural networks is more accurate than the accuracy of the single network.

The biggest difference between dual-stream convolutional network and individual convolutional neural network training is the feature extraction of optical flow information. The optical flow feature is a set of dense optical flow, which is a set of displacement vector fields between adjacent frames. Let $d_t(u, v)$ be the displacement vector of the point (u, v) in the t-th frame. The optical flow is divided into two image channels, horizontal and vertical. d_t can be decomposed into a horizontal component d_x and a vertical component d_y. Frame t and frame $t + 1$ are obtained together. The optical flow information obtained from a set of continuous sequence frames L is stacked to obtain a total of $2L$ channels of optical flow. Assuming that the width of the video frame is W and the height is H, then the input of the time flow network is $I_\tau = \epsilon R^{W \times H \times 2L}$ and τ represents any video frame. There are two ways to calculate I_τ. One is a very well-understood optical flow field superposition method, which is to superimpose the calculated optical flow between adjacent frames [6, 9, 10]:

$$
\begin{aligned}
I_\tau(u, v, 2k - 1) &= d_{\tau+k-1}^x(u, v), \\
I_\tau(u, v, 2k) &= d_{\tau+k-1}^x(u, v), \\
u &= [1; w], v = [1; h], k = [1; L].
\end{aligned}
\tag{1}
$$

Another one is the optical flow superposition method that tracks the trajectory. Based on the inspiration of the trajectory descriptor, sampling the motion trajectory at the same position replaces the method of simple optical flow superimposition sampling at the same position in all adjacent frames. At the same time, this method also decomposes the optical flow into horizontal and vertical x and y vectors to generate $2L$ image channels. Starting from the position (u, v) of the t-th frame, let P_k be the k-th point traced along the trajectory, and then the following definitions are given:

$$
\begin{aligned}
I_\tau(u, v, 2k - 1) &= d_{\tau+k-1}^x(P_k), \\
I_\tau(u, v, 2k) &= d_{\tau+k-1}^x(P_k), \\
u &= [1; w], v = [1; h], k = [1; L], \\
P_1 &= [u; v], P_k = P_{k-1} + d_{\tau+k-2}^x(P_{k-1}), k > 1.
\end{aligned}
\tag{2}
$$

Both networks use the same structure and can be separately trained. This study chooses to pretrain on a larger dataset ImageNet. The spatial network stream only inputs a single RGB picture, changing the previous idea of selecting the center video frame, adopting a random selection method, and cropping the "pan-entertainment" pictures of college students to a size of 224×224. Since the optical flow network has dual vertical and horizontal channels instead of 3 RBG channels, the first
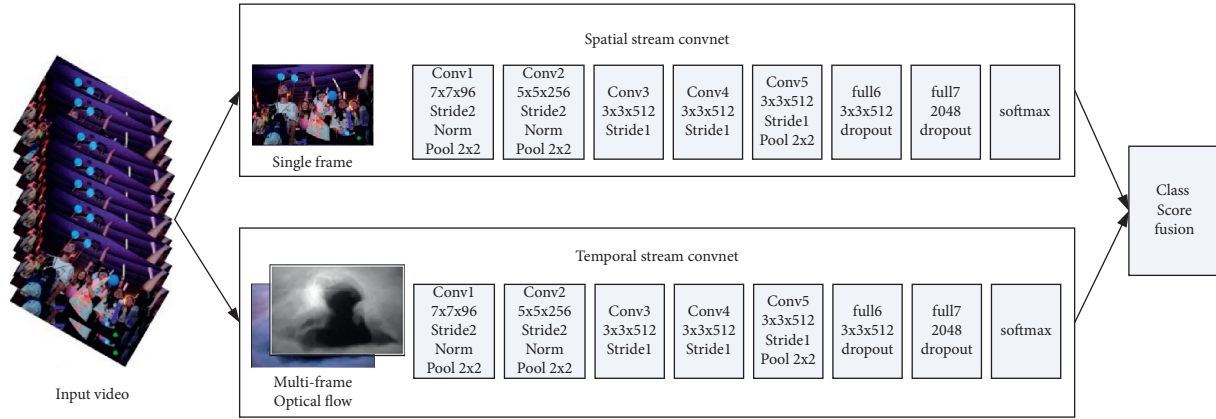
FIGURE 1: Dual-stream network structure diagram of video classification.
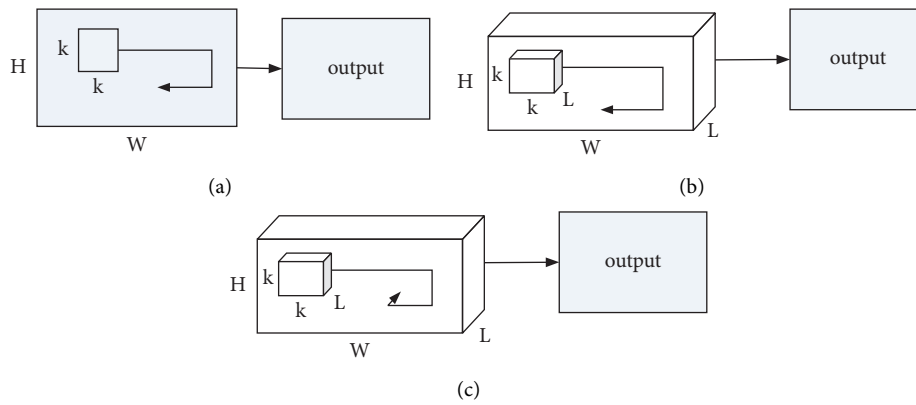
FIGURE 2: 2D convolution and 3D convolution.

convolutional layer is adjusted to a layer with $2L$ input channels. The previously calculated multiframe optical flow I is taken and a size of $224 \times 224 \times 2L$ from it is randomly cropped as input. After pretraining, the network was tested on UCF-101 and HMDB-51 and after multitask learning, and it was found that the method of training space and time separately is far better than the effect of single network training. At the same time, it is found through experiments that the convolutional network trained on dense optical flow still has good performance in a smaller dataset. In addition, the multitask learning used by the dual-stream network can simultaneously apply two different datasets to improve performance.

2.2. C3D Network. The theory that the dual-stream neural network separately extracts and trains temporal and spatial information opens up a new research direction for college students' "pan-entertainment" video behavior recognition, especially the extraction of optical flow features. However, the optical flow feature is also the biggest drawback of the dual-stream neural network. The optical flow vector needs to be extracted from adjacent frames, and the "pan-entertainment" video action of college students may require dozens of frames of input to accurately determine the action under special circumstances. It can be seen from this that the dual-stream network cannot extract the information of a

long sequence of videos, and it is even easy to lose more other spatiotemporal information [11–13].

As shown in Figure 2, (a) is the situation obtained by 2D convolution in a static image, and the output is a two-dimensional feature map; (b) is the situation when 2D convolution is used in video operations; at this time, the image becomes a multichannel image, and the output is still two-dimensional feature map; Figure c is the operation of 3D convolution in the video, and the output is a three-dimensional feature map. The video segment is set as $c \times l \times h \times w$, where c is the number of channels, l is the length (number) of the video frame, h and w are the width and height of the video frame; the size of the convolution kernel becomes three-dimensional $d \times k \times k$, where k is the size of the convolution kernel and d is the newly added time depth; then, the output value becomes $k \times l \times h \times w$. It is inferred from this that 3D convolution better completes the spatiotemporal modeling of video data [14], and the network can directly input video data without any preprocessing.

For college students' "pan-entertainment" video processing, an effective descriptor should have the following four characteristics: first, versatility and distinguishability; second, the descriptor should be compact. There are millions of videos, and the compact descriptors in the processing process can strengthen the retrieval task and storage; third,
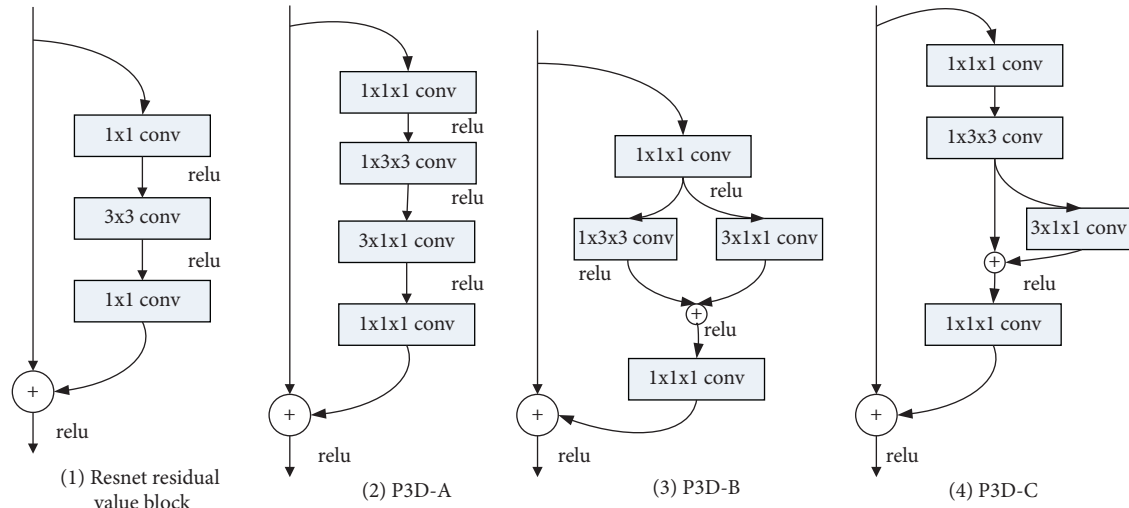
FIGURE 3: Three kinds of pseudo-3D block. (a) ResNet residual value block. (b) P3D-A. (c) P3D-B. (d) P3D-C.

the efficiency of calculation. Increasing the calculation speed is one of the goals we have been pursuing; fourth, simplicity. The features extracted by the 3D convolutional network contain other information such as the subject, background, and other information in the "pan-entertainment" video of college students, which can cope with various recognition requirements. The network architecture of C3D is shown in Figure 2. The size of the convolution kernel is $3 \times 3 \times 3$, and the stride is set to $1 \times 1 \times 1$. In order to keep the dimension unchanged, the size of the first pooling core is set to $1 \times 2 \times 2$, and the stride is $1 \times 2 \times 2$. The core size of the remaining pooling layer is $2 \times 2 \times 2$, and the stride is $2 \times 2 \times 2$. The network consists of 8 convolutional layers, 5 maximum pooling layers, and 2 fully connected layers and finally connected to the softmax layer. C3D chose the largest dataset Sports-1M for training. Compared with the UCF101 trained by the previous 2D convolutional network, it exceeded 5 times the category and 100 times the total number of videos. The input video is cropped into 16 112×112 segments, and the picture is horizontally flipped by 50% for data enhancement.
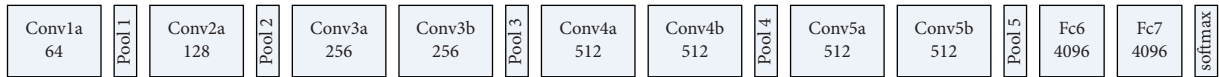
What are the differences in the learned features between the C3D network and the classic 2D network? This study uses the deconvolution method to detect what the 3D network has learned. After observation, it is found that C3D tends to track the subsequent development of motion information from the appearance information of the previous few frames. The C3D learning method can not only capture the appearance information of the moving subject well but also learn the variable motion information, which can be selectively learned. The proposal of the C3D network has achieved huge progress from the extraction of low-level semantics to the extraction of high-level abstract semantics, surpassing traditional deep learning methods and surpassing traditional manual methods. It is a simple and efficient model with good prospects. However, due to a large amount of calculation of the three-dimensional convolution operation, the large number of parameters generated in the C3D network has also become a problem to be solved.

2.3. R(2 + 1)D Network. After research and exploration, it is found that the 2D convolutional network is still the best choice in the analysis of action recognition. The 3D convolutional

network that introduces spatial and temporal dimensional features has also brought significant progress to the research of video recognition. Unfortunately, both are flawed. The 2D convolutional network can easily lose a lot of key information due to the inability to extract long-sequence video information; the 3D convolutional network is too computationally expensive, resulting in too much speed and too many parameters, which requires a lot of storage space and other problems. In 2017, Zhaofan Qiu et al. proposed Pseudo-3D Residual Net (P3D ResNet). Based on the ResNet network, the 3D convolution kernel undergoes a series of deformations, so that the 2D and 1D convolution kernels separately operate, which not only expresses the timing information well but can also greatly reduce the amount of calculation, making the network easier to optimize. On the basis that the residual idea of the ResNet network has achieved good results in the 2D convolutional network, the 3D convolution is split into 1D convolution about time information and 2D convolution about space information. The question that needs to be discussed is what kind of connection state is between 2D and 1D convolutions. Both series and parallel will affect the final effect. The three deformation methods are shown in Figure 3 below [15].

P3D has designed three convolution deformation methods: P3D-A decomposes $3 \times 3 \times 3$ convolution integrals into $1 \times 3 \times 3$ and $3 \times 1 \times 1$ and connects them in series. The output of a 2D convolution is used as the input of 1D convolution; P3D-B does the same decomposition. At this time, 2D convolution and 1D convolution are parallel, and there is no connection between the two convolutions. The final result is determined by the sum of the two convolutions; P3D-C is the combination of the two structures P3D-B and P3D-A. It needs to be pointed out that the bottleneck design in each residual block unit in the ResNet network structure makes that there are two convolutional layers of $1 \times 1 \times 1$ before and after, which are used to reduce the size of the input dimension and restore the output dimension. In this way, the effect of reducing computational complexity is achieved. In order to evaluate these three deformation methods, based on ResNet-50, the original residual unit is directly replaced with the three designed by P3D. It is found through experiments that the P3D-A series connection method achieves the best effect. Based on the above situation, in 2018 Du Tran et al. proposed the R(2 + 1)D convolution block and conducted related experiments on the kinetic dataset to demonstrate its usability and prove that the 3D convolution kernel is decomposed into separate space and time. Separate extraction instead improves accuracy. The proposal of the R(2 + 1)D network is also based on the changes made by R3D [8] that have applied ResNet to the 3D convolutional network, using the basic residual structure, without using the bottleneck design in P3D. Each residual block consists of two convolutional layers, and each layer has a ReLU activation function. Let x denote the input of $3 \times L \times H \times W$, L be the number of frames edited in the input, H and W denote the height and width of the video frame, respectively, and 3 denote the picture RGB channel. Let z be the tensor calculated by the i-th convolutional block in the residual network and the output of the i-th residual block, $\mathscr{F}(z_{i-1}; \theta_i)$ performs two convolutional layers through the weight θi synthesis and the role of the activation function ReLU [16].

$$Z_i = Z_{i-1} + \mathscr{F}(z_{i-1}; \theta_i). \tag{3}$$

In order to keep the R(2 + 1)D network and the R3D network roughly the same amount of parameters, the hyperparameter M_i is introduced.

The R(2 + 1)D convolution kernel is specifically composed of N $i - 1 \times 1 \times d \times d$ Mi 2D convolution kernel and $Mi \times t \times 1 \times 1$ Ni time 1D convolution kernel, instead of 3D Of N I N $i - 1 \times t \times d \times d$ convolution kernel, which has the following relationship.

$$N_{i-1} \times t \times d^2 \times N_i = N_{i-1} \times d^2 \times M_i + M_i \times t \times N_i,$$

$$N_{i-1} \times t \times d^2 \times N_i = \left(N_{i-1} \times d^2 + t \times N_i\right) \times M_i, \tag{4}$$

$$M_i = \left[\frac{t d^2 N_{i-1} N_i}{d^2 N_{i-1} + t N_i}\right].$$

Figure 4 shows the difference between (2 + 1)D convolution and 3D convolution. It is supposed that the input is single-channel spatiotemporal information. On the left is 3D
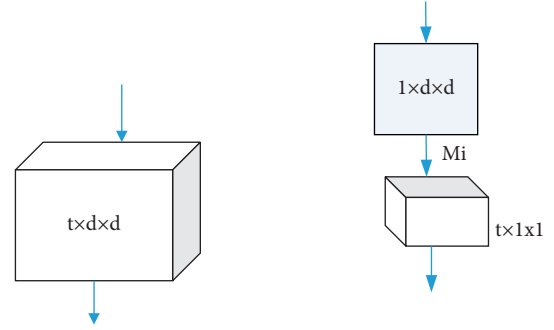
FIGURE 4: 3D convolution kernel and (2 + 1)D convolution kernel.

convolution, using a convolution kernel of size $t \times d \times d$, where t is time, and d represents the height and width of the space. On the right is the (2 + 1)D convolution block, which calculates spatial 2D convolution and time 1D convolution, respectively, and sets the number of 2D convolutions to M, so that the number of parameters in the (2 + 1)D block is the same as the entire 3D volume The number of parameters of the block matches [17, 18].

Compared with the convolution kernel of time convolution and spatial convolution, factoring into the 3D convolution kernel has two advantages. First, in the case of the same amount of parameters, the (2 + 1)D convolution kernel has doubled nonlinear mapping and has better network expression ability. Second, the (2 + 1)D network is easier to optimize, the loss rate of training and testing is lower, and the number of network layers is allowed to be deeper. This is confirmed by experiments on the large kinetic dataset. Whether it is the classic ResNet-3D network or the deformed network with the 2D convolutional layer and the 3D convolutional layer interlaced, the final result is not as good as R(2 + 1)D on the internet. One of the findings is that the effect of the ResNet-3D network, which independently separates spatial convolution from temporal convolution, is still stronger than that of hybrid 2D-3D convolutional networks. This further illustrates that separating space and time calculations is important to improve network performance factor. It is worth noting that the R(2 + 1)D network effect of the input optical flow graph still has a lot of room for improvement. Although the dense optical flow method using the Franeback algorithm improves the efficiency, the optical flow accuracy is not high, and it is worthy of in-depth study. In addition, in the experiment, the network obtained the best results after fusing the results of the input RGB image with the results of the input optical flow diagram. It can be seen that the timing characteristics and spatial characteristics are both very important.

3. "Pan-Entertainment" Behavioral Image Recognition of College Students Based on Spatiotemporal Convolutional Network

3.1. Dual-Stream Space-Time Integration Network Design. In this study, the algorithm chooses to model the spatial action and temporal information of the input relatively high-resolution "pan-entertainment" images of college students in the shallow network and uses dual-channel 2D Conv to

extract temporal and spatial features in parallel. In the later stage of the deep network part, 3D Conv is again used to perform space-time modeling. The spatiotemporal-r(2 + 1)d end-to-end model proposed in this study performs weight adjustment and fusion of the features extracted by the dual-stream neural network to obtain the middle-level semantic features, which are input into the R(2 + 1)D network for further learning and complete behavior recognition. At the same time, in order to be effective for long-term video, the video segmentation method is used in the input part of this study. The overall framework is shown in Figure 5.

The network framework is divided into three modules, namely, segmentation and RBG image preprocessing and optical flow image calculation of the input "pan-entertainment" video of college students, a weighted fusion of the dual-stream network part, and respatial modeling of the R(2 + 1)D network. First, the T_1, T_2, T_3, ..., Tn video is divided into K segments of equal length $\{s_1, s_2, s_3, ..., s_k\}$ where S_0 consists of multiple frames. Each video frame T_n is randomly sampled from S_0 and will be used as the input of the spatial network. The extracted "pan-entertainment" feature map is x_a, where a represents the "pan-entertainment" feature map extracted from the spatial domain. The time domain network corresponds to the input continuous optical flow image and the set t time as the corresponding time of the video frame T n; then, the position of the L continuous optical flow frame pictures corresponds to time t in the time domain. The time domain feature map obtained is x_b, where b represents the "pan-entertainment" feature map extracted from the time domain network. Then, the weighted sum fusion method is used to obtain consecutive spatiotemporal feature maps that are subsequently input to the R(2 + 1)D network.

$$M \in R^{H \times W \times D \times \lambda}\left(\lambda = \theta x^{ak} + \mu x^{bk}\right). \tag{5}$$

3.1.1. Data Preprocessing of Input Airspace Network. The

spatial network takes RBG pictures of static video frames as input. In the training process, the VGG-M-2048 model with pool1 and pool2 is selected, and the fully connected layer is removed for later feature fusion. This study deals with video segmentation. Among K segments of equal length, each segment randomly selects 1 frame of the image and crops it into 224×224 size. In order to enhance the generalization ability of the network, training samples are increased to prevent overfitting of the deep learning model, and a series of data enhancement operations are performed on the input images. This study adopts three processing methods for the input single-frame image:

(i) the picture is horizontally flipped, (ii) the angle of the picture is rotated, and (iii) the horizontal and vertical offset and shift transformation are carried out on the picture. After data expansion of the image data, the sample is increased by multiples to prevent overfitting to a certain extent.

3.1.2. Data Preprocessing of Input Time Domain Network.

In the time domain network input part of the dual-stream network, this study uses the TV-L1 method to calculate the optical flow, which is divided into horizontal and vertical optical flow diagrams. The optical flow feature is a set of dense optical flow, which is a set of displacement vector fields between adjacent frames, which can be used to extract the "pan-entertainment" information of college students and play an important role in video recognition. The optical flow image imported into the input of the network framework designed in this study contains the motion information of each static video frame image, which improves the correlation of spatiotemporal features on pixels and the robustness of processing video frame sampling. As shown in Figure 6, the optical flow diagram is a grayscale image calculated by decomposing the optical flow data into horizontal and hammer direction vectors, and there are $2L$ image channels.

In this study, $L = 10$ is set, and the horizontal and vertical optical flows of 10 consecutive frames are stacked to form 20 dense optical flow images as the input of the time domain network.

This study uses the VGG-M-2048 model pretrained on the ImageNet dataset. However, the RGB image with the number of channels as 3 in the first conv1 does not match the input data of the time domain network. For this problem, we use the cross-modality cross pretraining method. The weights of conv1 are averaged and copied into 20 copies as the weights of the time domain network conv1. Other weights remain the same to achieve the time domain network pretraining network parameter matching.

3.2. Improved Design of R(2 + 1)D Structure. The spatio-

temporal-r(2 + 1)d end-to-end model is the spatiotemporal feature map obtained after the feature fusion of the dual-stream network. Here, the spatiotemporal dual-stream network removes the fully connected layer because the output of the fully connected layer is high-level semantic features that will affect the image. The information on the time axis is not conducive to subsequent modeling. Then, the obtained spatiotemporal fusion feature map is input into the R(2 + 1)D network, which has a better effect than the C3D network. R(2 + 1)D uses the classic network ResNet. In the case of the fusion and series connection of the dual-stream network and the 3D convolutional network, the formed deep network is prone to the disappearance of the gradient, which makes the network effect worse. By inserting jump connections, ResNet directly propagates the gradient from the lossy layer at the end of the network to the early layer close to the input, which is to simplify the training structure of the deep structure. ResNet can solve the problem of gradient disappearance in deep networks to a certain extent.

Since the output part of the dual-stream network is a 28×28 feature map, this study removes the first three convolution blocks on R(2 + 1)D-34 and R(2 + 1)D-50, conv1_1, and conv2_1. Spatiotemporal downsampling is performed on the top and is implemented with a convolution step size of $2 \times 2 \times 2$. Finally, a 512-dimensional feature vector is an output, and the final result is output through the fully connected layer and the softmax layer. M is the number of 2D space convolution kernels after 3D convolution kernels are decomposed into (2 + 1)D
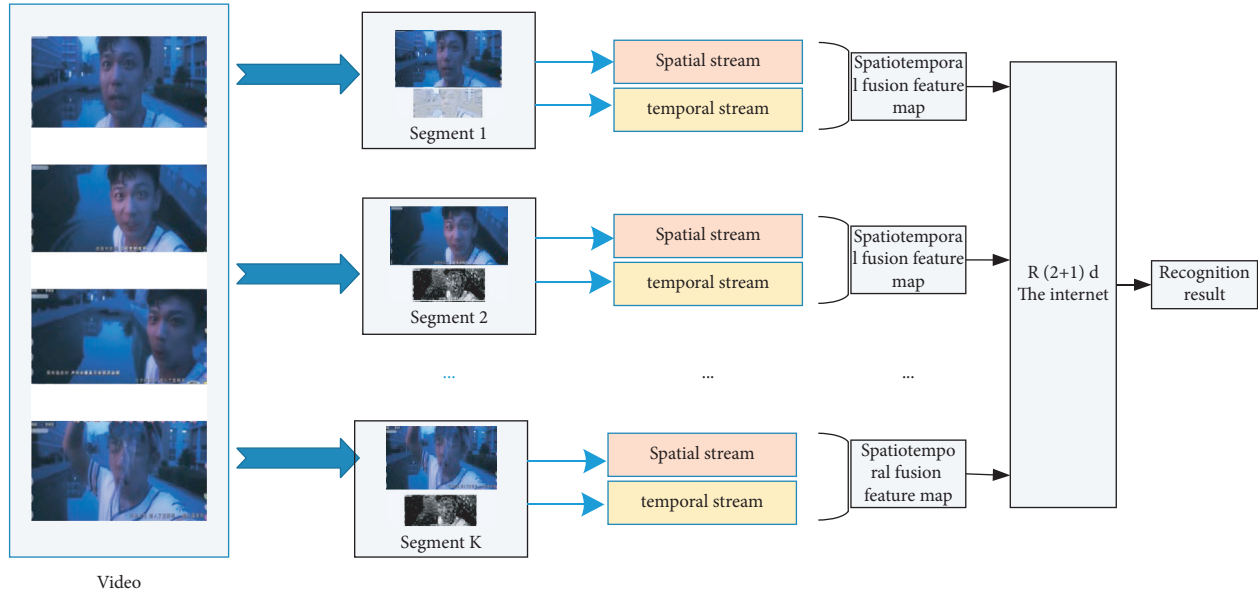
FIGURE 5: Spatiotemporal-r (2 + 1) d architecture design.

FIGURE 6: Optical flow diagram of continuous video frames.

convolution kernels. The network structure is shown in Table 1 below.

4. Experiment and Result Analysis

4.1. Experimental Data Set. In this study, UCF-101 and HMDB-51 are selected for evaluation experiments based on the most widely used datasets in the field of deep learning video behavior recognition. Next, we will introduce the HMDB-51 dataset. The HMDB-51 dataset comes from YouTube and Google videos collected from the internet. The pixels of the video frame are 320×240, and the average duration is 3.0 seconds. "Pan-entertainment" movements are divided into five types: (1) general facial movements; (2) facial movements with object manipulation; (3) whole-body movements; (4) body movements that interact with objects; and (5) body movements caused by human interaction.

Compared with the UCF-101 dataset, the dataset comes from real scene videos, and the background greatly changes. The small amount of data results in the limited training of the network and the existing algorithms generally have low recognition rates for it, making it one of the current challenging datasets. Figure 7 shows an example of some actions of HMDB-51.

Both UCF-101 and HMDB-51 have a standard three-segment evaluation protocol. This study evaluates the network performance based on the average recognition accuracy of all splits. This study divides the two datasets into the training set and test set in the experiment.

4.2. Parameter Setting. The dual-stream network module uses the VGG-M-2048 model pretrained on the ImageNet dataset, and the R(2 + 1)D network module uses the R(2 + 1)

TABLE 1: Improved network structure based on R(2 + 1)D-34/R(2 + 1)D-50.

Number of layers	Output size	R (2 + 1) D-34	R (2 + 1) D-50
Conv1_x	$\lambda \times 14 \times 14$	$\left[\begin{pmatrix} 1 \times 3 \times 3 & M_1 \\ 3 \times 1 \times 1 & 256 \\ 1 \times 3 \times 3 & M_1 \\ 3 \times 1 \times 1 & 256 \end{pmatrix}\right] \times 6$	$\left[\begin{array}{c} 1 \times 1 \times 1 \quad 256 \\ \begin{pmatrix} 1 \times 3 \times 3 & M_2 \\ 3 \times 1 \times 1 & 256 \end{pmatrix} \\ 1 \times 1 \times 1 \quad 1024 \end{array}\right] \times 6$
Conv2_x	$\lambda/2 \times 7 \times 7$	$\left[\begin{pmatrix} 1 \times 3 \times 3 & M_2 \\ 3 \times 1 \times 1 & 512 \\ 1 \times 3 \times 3 & M_2 \\ 3 \times 1 \times 1 & 512 \end{pmatrix}\right] \times 3$	$\left[\begin{array}{c} 1 \times 1 \times 1 \quad 256 \\ \begin{pmatrix} 1 \times 3 \times 3 & M_4 \\ 3 \times 1 \times 1 & 256 \end{pmatrix} \\ 1 \times 1 \times 1 \quad 2048 \end{array}\right] \times 3$
	$1 \times 1 \times 1$	Spatiotemporal pooling, fc layer with softmax	

FIGURE 7: Some examples of "pan-entertainment" actions of HMDB-51.

D-34 and R(2 + 1) D-50 model. Using small batch stochastic gradient descent (SGD), the loss function is cross-entropy loss, and the BN (batch normalization) layer is added to the network to accelerate the network convergence speed, to a large extent, to prevent overfitting and to improve the problem of gradient disappearance. Tables 2 and 3, respectively, list the detailed parameters on the two datasets.

The network is trained on two datasets. The input of the spatial network is an RGB image cropped to 224×224; the input of the time domain network is 20 optical flow images of 10 consecutive video frames, the size of which is also 224×224. On UCF-101, the batch size of the spatiotemporal-r(2 + 1)d-34 dataset is 128, the initial learning rate is set to 0.001, and the learning rate decays to 1/10 of the original for 5,000 iterations, with a total of 15,000 iterations. The batch size of the spatiotemporal-r(2 + 1)d-50 dataset is 64, the initial learning rate is 0.001, and the learning rate decays to 1/10 of the original for 30,000 iterations per 10,000 iterations. The batch size of the spatiotemporal-r(2 + 1)d-34/50 dataset on HMDB-51 is set to 64, the initial learning rate is 0.001, and the learning rate decays to 1/10 of the original every 3,000 times, with a total of 10,000 iterations. The network is based on the transfer learning method, using a pretraining model for training, the initial learning rate is set to a smaller value, and then, the weight of each network is fine-tuned.

4.3. Comparison of Fusion Feature Weights.
The sum of the fusion method of spatial convolutional network and time convolutional network forms a fusion method of different weights by setting the spatial weighting coefficient θ and the time domain weighting coefficient μ, and the final output feature map is obtained by adding the weights. The fifth convolutional layer of the dual-stream convolutional network module is chosen to experiment with different strategies on the fusion coefficient ratio. Based on the experiment, take the video segmentation into 3 groups and compare the average accuracy of 3 groups on the UCF-101 dataset and HMDB-51 dataset (all splits). At the same time, in order to compare the performance of the 34th layer and 50th layer of the improved r(2 + 1)d network structure, we use r(2 + 1)d-34/50 to separately experiment to discuss the weight ratio of the dual-stream fusion method. The results are shown in Tables 4 and 5. The weight ratios are taken in 7 different proportions. It can be seen that when the spatial feature map accounts for a large proportion, the recognition accuracy decreases; conversely, when the temporal feature map accounts for a larger proportion, the accuracy increases. It can be concluded that the time information extracted by the time domain network plays an important role in the overall network performance. In summary, it is found that whether it is using r(2 + 1)d-34 or r(2 + 1)d-50, $\theta: \mu = 4:6$, the network recognition performance is the best, reaching the highest accuracy.

4.4. Overall Network Performance Evaluation.
In order to prove that the network framework proposed in this study has certain advantages, it is compared with some existing classic algorithms on the public datasets UCF-101 and HMDB-51. Table 6 lists different comparison algorithms, including traditional hand-designed feature algorithms (IDT) and algorithms based on deep learning. It can be clearly seen that the accuracy of the framework based on deep learning algorithms (no. 3, 4, and 5) is greatly improved compared to the different feature encoding methods (no. 1 and 2) based on dense trajectories. Among them, the C3D algorithm has a relatively poor effect due to too many network parameters. However, the 3D Conv proposed by the C3D network opens the research direction of spatiotemporal convolutional neural networks based on the 3D convolution kernel, which is of great significance for video behavior recognition. At the same time, it can be seen that the accuracy of the most primitive dual-stream convolutional neural network algorithm has been improved after the LSTM recurrent neural network is added, indicating that a reasonable combination of the dual-stream network with other methods can improve the recognition effect. Analyzing the latest research results in recent years, the P3D network

TABLE 2: Parameter settings on UCF-101 dataset.

The Internet	cpoch	Batch_size	fropout	Initial learning rate	Momentum
Spatiotemporal-r(2 + 1)d-34	200	128	0.5	0.001	0.9
Spatiotemporal-r(2 + 1)d-50	200	64	0.5	0.001	0.9

TABLE 3: Parameter settings on HMDB-51 dataset.

The internet	cpoch	Batch_size	fropout	Initial learning rate	Momentum
Spatiotemporal-r(2 + 1)d-34	180	64	0.5	0.001	0.9
Spatiotemporal-r(2 + 1)d-50	180	64	0.5	0.001	0.9

TABLE 4: The average accuracy of r(2 + 1)d-34 fusion ratio of different spatiotemporal feature maps%.

$\theta_{spatial} : \mu_{temporal}$	UCF-101	HMDB-51
2 : 8	86.4	61.1
3 : 7	87.6	61.8
4 : 6	88.3	62.0
5 : 5	86.2	61.7
6 : 4	85.0	60.8
7 : 3	84.4	60.3
8 : 2	83.8	60.5

TABLE 5: The average accuracy of r(2 + 1)d-50 fusion ratio of different spatiotemporal feature maps%.

$\theta_{spatial} : \mu_{temporal}$	UCF-101	HMDB-51
2 : 8	90.5	65.3
3 : 7	90.7	65.7
4 : 6	92.1	66.1
5 : 5	91.8	65.8
6 : 4	90.4	64.2
7 : 3	89.2	62.5
8 : 2	88.5	59.8

TABLE 6: Comparison of accuracy rate of recognition of "pan-entertainment" image and video behavior of college students with different algorithms%.

Serial number	Method	UCF-101	HMDB-51
1	IDT + FV	85.9	59.1
2	IDT + bovw	87.9	60.9
3	Dual-stream(VGG-M)	88.0	59.4
4	Dual-stream+(LSTM)	88.6	—
5	TDD + IDT	91.5	65.9
6	C3D	85.2	—
7	P3D ResNet	88.6	—
8	MiCT-Net	88.9	63.8
9	Spatiotemporal-r(2 + 1)d	92.1	66.1

model using the ResNet network improved based on the traditional C3D network is an initial attempt to factorize the 3D convolution kernel. Compared with the traditional C3D, the effect obtained on the dataset UCF-101 greatly improved. Proposing a hybrid 2D and 3D convolutional deep network MiCT-Net compared to the traditional single 2D convolutional network and 3D convolutional network, the accuracy of the two datasets is also improved.

This study is based on the improvement of the dual-stream convolutional network algorithm. For long-term video spatiotemporal modeling, the recognition accuracy of 92.1% and 66.1% is achieved on UCF-101 and HMDB-51, respectively. Compared with the dual-stream method, the improvement is 4.1% and 6.7%, respectively. Compared with other classic algorithms, the method in this study also obtains higher accuracy. At the same time, the algorithm in this study also realizes the end-to-end network structure and realizes the effectiveness of the recognition task based on the "pan-entertainment" video behavior of college students.

5. Conclusions

This study mainly explores the recognition of "pan-entertainment" image and video behavior of college students based on the spatiotemporal convolutional neural network. The research work is as follows:

(1) The methods of "pan-entertainment" video behaviors proposed by college students so far are summarized and discussed, and the future development is discussed. By analyzing the technical bottlenecks of the current network structure in deep learning, an efficient, stable, and accurate method is summarized and explored.

(2) The dual-stream neural network is improved, and the classification decision fusion method is changed. The time information and spatial information separately extracted in the dual-channel 2D Conv are combined with "pan-entertainment" feature maps to form spatiotemporal feature maps and then behavior classification. The specific fusion method and fusion location are discussed, the influence of fusion location on network learning is explained through experiments, and the best fusion location and method are found.

(3) A spatiotemporal convolution algorithm is proposed, which connects the improved dual-stream network and the R(2 + 1)D network based on ResNet in series. The spatiotemporal feature map extracted by the dual-stream network is again input into R(2 + 1)D for spatiotemporal modeling. In order to achieve series connection, R(2 + 1)D-34 and R(2 + 1)D-50 and are used to reorganize them. This study uses the pretrained network on the ImageNet and kinetic datasets to perform experiments and fine-tunes the weights on the datasets UCF101 and HMDB51, respectively. Compared with other classic methods, the algorithm proposed in this study has achieved higher recognition accuracy.

(4) This study improves and combines several existing algorithm frameworks based on deep learning and proposes a spatiotemporal convolutional network algorithm framework for long-term videos that are difficult to process. Although comparing some classic algorithms with UCF101 and HMDB51, there is a significant improvement in accuracy, but there is still a lot of room for improvement. For example, designing the optimal network architecture is tried for 2D Conv in video recognition learning, and other datasets are added for experimentation.

Data Availability

The dataset can be accessed upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] P. Wang, W. Li, Z. Gao, C. Tang, J. Zhang, and P. Ogunbona, "Convnets-based action recognition from depth maps through virtual cameras and pseudocoloring," in *Proceedings of the 23rd ACM international conference on Multimedia*, Brisbane Australia, 2015.

[2] H. Rahmani and A. Mian, "3D action recognition from novel viewpoints," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1506–1515, Las Vegas, NV, USA, 2016.

[3] L. Jin, S. Gao, Z. Li, and J. Tang, "Hand-crafted features or machine learnt features? together they improve rgb-d object recognition," in *Proceedings of the 2014 IEEE International Symposium on Multimedia*, pp. 311–319, Taichung, Taiwan, 2014.

[4] P. Wang, W. Li, Z. Gao, Y. Zhang, C. Tang, and P. Ogunbona, "Scene flow to action map: a new representation for rgb-d based action recognition with convolutional neural networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 595–604, Honolulu, HI, USA, July 2017.

[5] D. Tran, L. Bourdev, R. Fergus, L. Torresani, and M. Paluri, "Learning spatiotemporal features with 3d convolutional networks," in *Proceedings of the International conference on computer vision*, December 2015.

[6] J. Yue-Hei Ng, M. Hauschnecht, S. Vijayanarasimhan, O. Vinyals, R. Monga, and G. Toderici, "Beyond short snippets: deep networks for video classification," in *Proceedings of the Conference on computer vision and pattern recognition*, June 2015.

[7] Y. Zhou, X. Sun, Z. J. Zha, and W. Zeng, "Mict: mixed 3d/2d convolutional tube for human action recognition," in *Proceedings of the Conference on computer vision and pattern recognition*, June 2018.

[8] H. Kuehne, H. Jhuang, E. Garrote, and P. T. Serre, "HMDB: a large video database for human motion recognition," in *Proceedings of the 2011 International Conference on Computer Vision*, November 2011.

[9] L. Bottou, "Stochastic gradient descent tricks," *Lecture Notes in Computer Science, Neural Networks: Tricks of the Trade*, Springer, Berlin, Heidelberg, pp. 421–436, 2012.

[10] X. Peng, L. Wang, X. Wang, and Y. Qiao, "Bag of visual words and fusion methods for action recognition: comprehensive study and good practice," *Computer Vision and Image Understanding*, vol. 150, pp. 109–125, 2016.

[11] M. Shu, "The value choice of the public in the era of pan-entertainment," *New Media Research*, vol. 34-35, 2018.

[12] Z. Shi, "Beware of pan-entertainment enslaving the self," *People's Forum*, vol. 44-46, 2018.

[13] W. Jia, "Pan-entertainmentism makes entertainment a fool," *People's Forum*, vol. 50-52, 2018.

[14] Q. Jiang, Y. Zhang, S. Tan, and Y. Yang, "Recognition of students' classroom behavior based on residual network," *Modern Computer*, vol. 20, pp. 23–27, 2019.

[15] X. Jin, *Research and Implementation of Face-To-Face Classroom Intelligent Management System Based on Seetaface Face Recognition Engine*, Jiangsu University, Zhenjiang, China, 2019.

[16] A. B. Dhivya and M. Sundaresan, "Tablet identification using support vector machine based text recognition and error correction by enhanced n - grams algorithm," *IET Image Processing*, vol. 14, no. 7, pp. 1366–1372, 2020.

[17] X. Yuan, B. Huang, Y. Wang, C. Yang, and W. Gui, "Deep learing-based feature representation and its application for soft sensor modeling with variable-wise weighted SAE," *IEEE Transformations on Industrial Infomatics*, vol. 14, no. 7, pp. 3235–3243, 2018.

[18] S. Wang, H. Chen, L. Wu, and J. Wang, "Novel smart meter data compression Method .via. staked. convolutional sparse auto-encoder," *Internate Journal of Elector Power and Energy Systems*, vol. 118, Article ID 105761, 2020.

Research Article

Deep Convolutional Nets Learning Classification for Artistic Style Transfer

R. Dinesh Kumar,[1] E. Golden Julie,[2] Y. Harold Robinson [ID],[3] S. Vimal,[4] Gaurav Dhiman [ID],[5] and Murugesh Veerasamy [ID][6]

[1]*CSE Department, Siddhartha Institute of Technology and Science, Hyderabad, India*
[2]*Department of Computer Science and Engineering, Anna University Regional Campus, Tirunelveli, India*
[3]*School of Information Technology and Engineering, Vellore Institute of Technology, Vellore, India*
[4]*Department of Artificial Intelligence and Data Science, Ramco Institute of Technology, Rajapalayam, India*
[5]*Department of Computer Science, Government Bikram College of Commerce, Patiala, Punjab, India*
[6]*Department of Computer Science, College of Computer Science, Bule Hora University, Blue Hora, Ethiopia*

Correspondence should be addressed to Murugesh Veerasamy; murugesh@bhu.edu.et

Humans have mastered the skill of creativity for many decades. The process of replicating this mechanism is introduced recently by using neural networks which replicate the functioning of human brain, where each unit in the neural network represents a neuron, which transmits the messages from one neuron to other, to perform subconscious tasks. Usually, there are methods to render an input image in the style of famous art works. This issue of generating art is normally called nonphotorealistic rendering. Previous approaches rely on directly manipulating the pixel representation of the image. While using deep neural networks which are constructed using image recognition, this paper carries out implementations in feature space representing the higher levels of the content image. Previously, deep neural networks are used for object recognition and style recognition to categorize the artworks consistent with the creation time. This paper uses Visual Geometry Group (VGG16) neural network to replicate this dormant task performed by humans. Here, the images are input where one is the content image which contains the features you want to retain in the output image and the style reference image which contains patterns or images of famous paintings and the input image which needs to be style and blend them together to produce a new image where the input image is transformed to look like the content image but "sketched" to look like the style image.

1. Introduction

A decade ago, when machine learning was an emerging application of artificial intelligence providing the ability to automate learning process from foregoing experiences without being explicitly programmed, the only limitation was assumed that a good computer program can never replace a human in creativity [1]. But as the exploration in the field grew, this gave rise to many other subfields like deep learning, which threw the limelight on the solution for replacing humans for their creativity or their process of recognizing objects or people [2]. One of such problems

which characterized human from a machine was art. Generating art has no rules which could be used to replicate a man's imagination. This paper renders an input image in the range of well-known art works [3]. This is conceptually close to texture transfer for style transfer. Today, there are few existing systems which replicate art from famous painters. One of these methods is style transfer using a neural network [4].

A neural network can be defined as a circuit of neurons which stimulates the behavior of a human brain. Mathematically, a neural network is a chain of functions which maps the inputs with their respective outputs based on their

weights, which defines the amount of influence one function has on the other (where each function represents a neuron) [5]. A neural network engages a larger amount of functions operating similar layers. One layer generates the input and every adjacent layer gathers the output from the previous layers. The final layer generates the output of the system, which is also known as visible layer. As the layers other than the input and output layers are not accessible, they are known as hidden layers [6].

The transformation of the style from one image to another image could be called the texture transfer issue. The aim of doing this is to manufacture an element from an input image and also gather the semantic component to the output image [7]. Several texture transfer methods used the texture synthesis functions of using dissimilar way to conserve the targeted output image. The features of the output image like intensity is computed for effective texture transfer. The frequency related texture data for preserving the target image combined the edge orientation data. A style transfer is utilized to extort the contents from the image for implementing the texture transfer. Hence, a normal procedure is used to analyze the image independent representation to observe the semantic contents and style related structures were demonstrated to control the subset of natural image characterization [8].

The computer vision is used to extract the semantic contents from the images. The labelled information for the concerned tasks like object recognition is to extract the quality image element in general feature extraction from the datasets and other visual elements [9]. The texture transformation algorithm comprises a texture formation technique with feature demonstration based convolutional neural networks (CNN). The style transfer technique minimizes and obtains solutions to the optimization issues in a neural network. The new images are constructed by implementing a feature extraction for the dataset images. The texture synthesis approach is used to increase the deep image representation [10].

The main contribution of our proposed work is as follows:

(i) Deep learning approach is applied to real-time images after implementing the object detection

(ii) The proposed methodology executes the style transfer efficiently in which the captured depth image is from the dataset

(iii) Validation is performed using the proposed training methodology which may produce the style transfer with reduced complexity

(iv) The visual object recognition is applied to the feature space with the usage of max-pooling layers

2. Related Work

The style images and the recombination of the images are separated by creating the artistic images with high quality using neural representations from the neural algorithm; this utilizes the artificial intelligence technique [11] which

presents a way to understand the artistic images algorithmically. Adversarial networks are created by generating the art based learning styles and conflicting from style standards; the system generates imagination based art using art distributions. It focuses on realistic texture rather than pixel accuracy [12]. The essential factors in neural style transfer demonstrate control over position and color data across spatial scale which improves by permitting high quality control over style and also assists in reducing normal faulty situation such as imaginary options.

A fast generator combined with feedforward network for initial style transfer has been proposed to get the styled output as an image with the produced unnoticed element with style image as the input throughput the inference period. The generator is constructed to perform the encoder and decoder by transferring the deep features. The spatial components are utilized to measure the level of styles that have been incorporated. The perceptual loss for every element has been reduced to identify the properties of the image styles from the artistic images. The multidomain related images are created using the mask element to implement the stabilization and evade the collapse in the real-time scenario. The selective stylized images are basically increasing the effectiveness of optimization related style transfer [13]. The neural techniques combined with patch related synthesis approach have been implemented to attain the stylization feature with high resolution. The neural methods have been trained to increase the stylization level in global standard and predict the output for relevant patch related synthesis at different level. The original artistic media is used in better way to improve the dependability of the image in stylish manner. Without incrementing the pixel size the generated image is in high quality. The visual quality is checked with the related style transfer method and the response is better with feature extraction metrics [14].

The deep learning based feature extraction has been implemented using the methods in which the identification of high level image features segregates the style and image contents. The image style adaptation methodology produces the style characteristics of several paintings and pertained styles for learning from the other images also. The incorporation of artificial intelligence concept with the art strengthens this method which has the improved one [15]. Neural learning techniques were implemented to be efficient for style transfer. A new image has been synthesized to keep the high level image with the low-level features for the style image. Moreover, the convolutional methods are used to implement the feature extraction for style images with semantic features identified from the images. The natural style elements are incorporated to improve the low-level features into the high level feature using machine learning approach [16]. Every painting may confine the integral enhancement for natural stylizations. The style transfer technique is combined with the direction based style transfer for improved texture extraction. An innovative field loss with direction is enhanced to the synthesized images. The loss function is used to identify if there is any loss in the time of style transfer and the incorporated method is constructed to reduce the loss itself without any specific methodology. A

simple interaction technique is developed to manage the generated fields and the direction of the texture in generated images. The texture enhancement method is used to implement the style transfer to the synthesized images [17].

The photographic related style transfer shows the improved output for spatial based semantic segmentation for input. The segmentation has been done within the region of the input image and the style referenced images are dissimilar in spatial contents. A spatial transformation technique is applied before implementing style transfer technique. The cross-correlation based feature maps are applied to compute the affine transformation to the generated image to implement the active style transfer for the semantically aligned regions. The pretrained CNN model is implemented to identify the reference images when the shadows are removed from the produced images [18]. The semantic related style transfer technique has been identified for getting the solution for the semantic based issues. The semantic matching is implemented to increase the style transfer quality. Every image is segregated into various regions with semantic values and improved painting [19]. The divided regions are further arranged related to the semantic elucidation. The source region is trained using the learning elements to produce an output in stylish manner. The semantic matches have been identified within the regions and the guaranteed semantic matching is ascertained within the source and the target outputs. The semantic gaps are identified whenever dividing the regions based on the semantic values for the real paintings and also the photographs. The domain adaptation method is developed to decrease semantic gap for regional segregation [20].

The synthesis images have been mostly used in the real-time applications with learning and training model developed to minimize the cost of the resources and also the human values. There are several differences identified for original images and synthesis images in real-time characteristics [21]. The style transfer is one of the solutions to minimize the gap within these kinds of images. The indoor synthesis images are converted into the improved images by minimizing the light influences. Hence, the content data of the real image is achieved by the style information. The convergence speed is converted to identify the real images in complex situations [22]. The visual artworks have been applied in a photo related style transfer method which may produce the realistic images. The outcome of this technique is used to assist the normal users to expand the style transfer images which are affected by the various real-time issues. The autonomous sky segmentation approach is used to segregate the input image with sky background. The background colors are characterized into the sky segmentation. The natural color transformation technique is adopted with the sky background and correction method to guarantee the quality output production [23].

The style transfer approach is also used in the field of medical images. The CNN and combined deep learning approaches are implemented in the computer vision tasks. The CNN model needs a huge amount of data sets that are very difficult in medical based image processing. For this problem, the image generation is used to increase the

computer vision [24]. A new engine is developed to exploit the network to confine the synthetic information. The style transfer method is utilized to enhance the visual realism in the public dataset combined to the semantic features using CNN methodology [25]. The optical images are applied in the deep learning concept under water image information to reduce the bottleneck. The detection of objects in real-time sonar images is utilized for the network based training [26]. The 3D artistic face modeling with controllable manner [27] enables the face geometry space and a face texture space based on 3D face dataset. The experiment is carried out in real time without GPU acceleration to achieve different cartoon characters. The deconstructed integer function [28] is applied to have different attributes as biomorphism, beauty, and symmetry. The random geometric graphs enable the creative artistic composition. The algorithm is introduced with new outline image [29] extracted from the content image. The variation regularization is applied to reduce the noise and to smoothen the boundary region and outline loss function applied on the outline image. The results experience the better design clothing shape which is reserved perfectly. The virtual space technology [30] enables the immersive experience which analyzes the multisensory and multitechnical spatial art style transformation form. The results show the better experience on art style transfer.

The neural style transfer has been constructed with several semantic representations with dual semantic loss which has been maintained with the particular values for the stylized outputs to every technique of the computed content images [31]. The color cast has been established according to the illumination modification and the temperature for increasing the productivity. The color calibration technique is used to transfer the exact color with semantic representation. The global attention functionality has removed the color cast from the input image for style transfer [32]. The neural style transfer technique has been implemented to convert the portrait image into the specific realistic image with some style and the pixel motion parameter with the color displacement from specific frames in semantic representation [33]. The optimization problems have the solutions through the metaheuristic techniques as the torus walk bat algorithm and modified bat algorithm are used to enhance the local search capacity ahead of utilizing the standard process [34, 35].

3. Proposed Method

The semantically useful style in a global level without user interaction and the low-level elements is removed and regaining the color related information is done to keep the fidelity without affecting the originality. The proposed system is constructed to diminish the computational complexity by producing the full resolution based images. The exploitation of previously used style transfer technique will enhance the professionalism. The image analogies have been minimized by using the neural related style transfer technique. The VGG16 is used to pretrain the convolutional neural networks in efficient manner. The dissimilar colors are mixed and produced the single region and are utilized to

perform the style transfer in semantically dissimilar regions. The max-pooling operation enhances the flow of the images in gradient level. An important element is implemented for producing art generation which is related with the creativity of art in real-time scenario.

VGG16, a pretrained neural network, is utilized to implement the content and style representations from its layers. Figure 1 represents the block diagram for art generation using neural algorithm using VGG-16 with 16 convolutional layers and 5 pooling layers. Whenever the input image and style image are passed through the network, input image is initialized to the content image. Style representation is extracted at the initial layers of the network as they extract the pixilated features. In the extraction, the layers extract the content in the image. Once the style and content representations are extracted, the output is generated by reducing the losses between them. The generated output image consists of the content representations of the input image and style representation. Convolution neural networks are used in image processing. They are made up of layers where every neuron delivers an input, computes a dot product, and pursues with nonlinearity optionally. Here, unlike any particular neural network, the neurons in the layers of CNN are prearranged in 3 dimensions: height, breadth, and depth, where depth refers to activation volume not the depth of the network.

It consists of layers of minute computational units which produce visual data randomly in a feedforward approach. A CNN is mainly built of convolutional based layers, these layers are the compilation of image filters that extract a particular feature from the input, where the feature map will be the targeted output. When CNN are trained for performing the image processing, they construct a representation of image which takes the object data with the processing functionality of the input image modified into the pixel based image representations for producing the quality images. The higher level contents and objects are arranged to execute the pixel values from the original image; the lower layers are reconstructed for the pixel values of the input image.

Convolution layer is the central unit of constructing the CNN that is mathematically operated to combine the group of data. It is implemented on the input data with convolution filter to create the feature map. Feature space is framed to confine the texture data and it is utilized to get the style of the image. The feature space is developed for delivering the filter responses in every layer. It contains the correlations within the dissimilar filter representations. By incorporating the feature correlations of various layers, the multiscale representation of the input image recognizes the texture components. The operation is computed by sliding the filter across the input image. At the particular location, there is a component related matrix formation and the resultant that creates the feature map. The common area of producing the convolution operation is performed by the standard filter which is shown in Figure 2.

Figure 3 demonstrates the generating feature map using convolutional input and filter. The filter across the aggregated input is used to produce the convolution results with
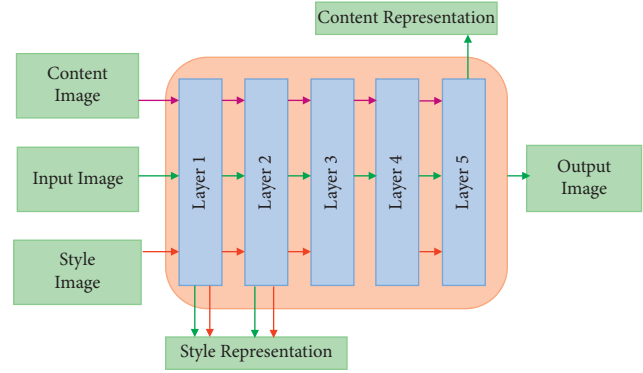
FIGURE 1: Block diagram of proposed system which takes input image.
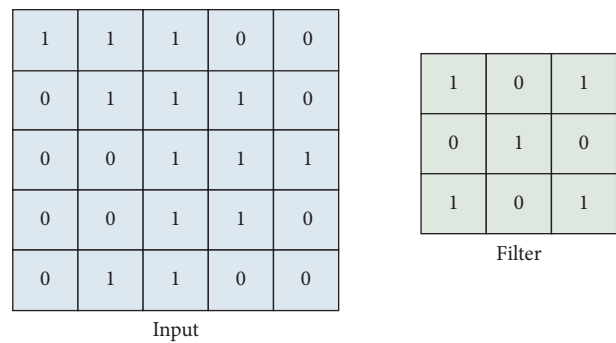
FIGURE 2: Images and filter.
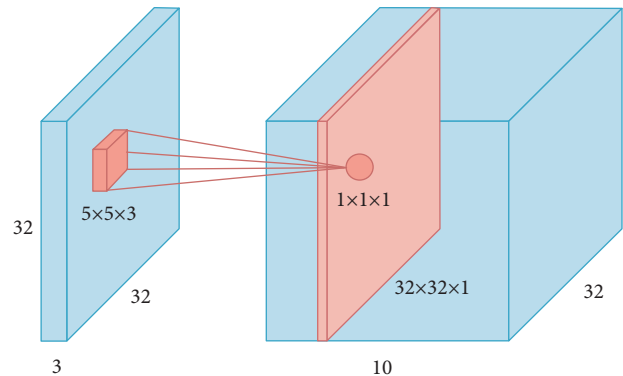
FIGURE 3: Generating feature map by convoluting input and filter.

feature map. The convolution operations are computed in 3 dimensions and the image is demonstrated as the 3-dimensional matrixes like depth, height, and width. A convolution filter holds the particular width and height as 5×5 and 3×3; it may cover the requirements of 3 dimensions. The feature maps are constructed to produce the final output with different kinds of filters and generated the output from the convolution layer. The convolution filter requires being determined in the input. If equal dimension requires maintenance, the padding concept is utilized to enclose the input with zeros.

In Figure 4, there is an input image with $32 \times 32 \times 3$ dimensions and a filter layer with $5 \times 5 \times 3$ dimensions. The filter has improved depth that matches the depth of the input

Input X Filter   Feature Map
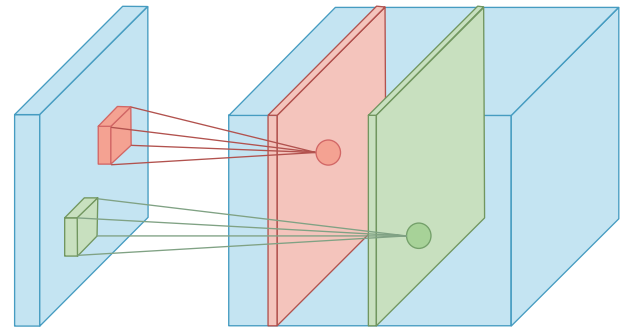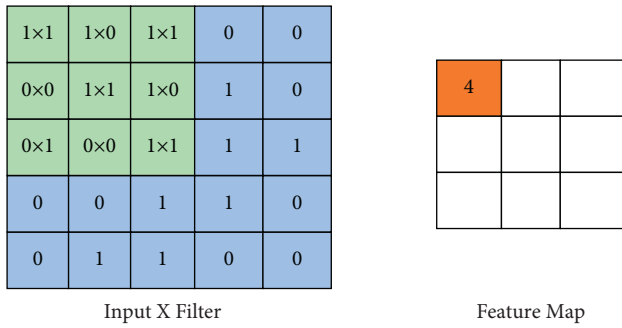
Figure 5: Generating a stack of feature maps using multiple filters.

Input X Filter   Feature Map

Figure 4: Generating a feature map with a single filter by convoluting the input layer with the filter.

Stride 1   Feature Map
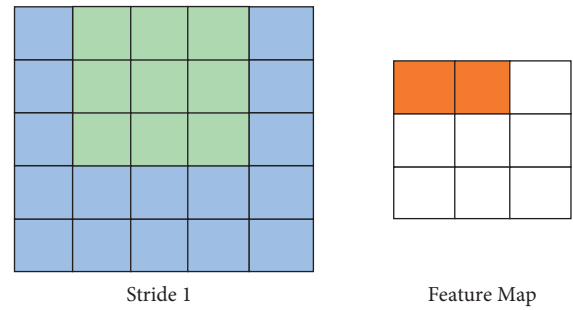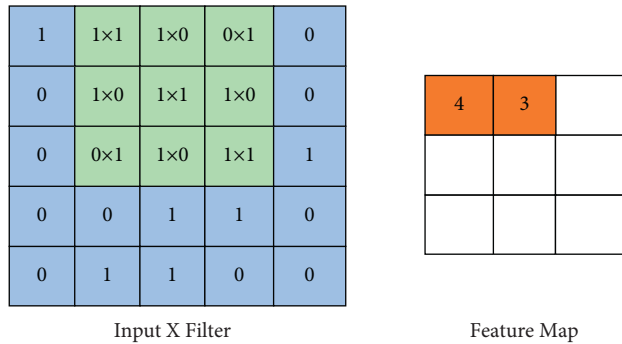
Figure 6: Stride value when it equals 1.

and the value for both is 3. Whenever the filter is located in a specific position, it wraps a volume of input and produces the output using the convolution operations. If 10 dissimilar filters are used, 10 different feature maps are used when computing them together combined with the dimension of depth for computing the size of $32 \times 32 \times 10$ convolution layers. The convolution operation for every filter is computed individually and results in a disjoint set. Nonlinearity could be achieved by producing the sum of weighted inputs along with the activation function; the convolution operation is passed along with the nonsaturating activation function called ReLu. The constructed final feature maps are the network with ReLu operation.

The sum of the matrix multiplication is performed in 3 dimensions as shown in Figure 5, but the result is scalar with the feature map of size $32 \times 32 \times 10$. The multiple filters are used to decrease the computational cost; it needs to utilize the specific filters within the particular time for learning purpose. The filters are mapped to the input image and learn every part of the image with the tiny filter sliding from one end to another end to produce the output image. The convolution process has the individual value in spite of using the style filter and the convolutions are joined to frame the output through the total amount of filters.

Stride demonstrates the convolution filter required to be moved at every step. Initially, the value of stride is 1 which is demonstrated in Figure 6. The stride is recommended to use the initial value 1. The padding could be frequently using layers in the convolution operation but not in pooling operation. The bigger strides are overlapped within the output fields that generate the output feature map of very minimum value.

From Figure 7, it can be observed that the size of the feature map is shorter than the input stride value increased to 2. After the convolution and pooling layers, few connected layers are included to conclude the CNN architecture. It is known that the output of the convolution layers and pooling layers is in 3 dimensions but fully connected layers expect a 1 dimension vector. Hence, the final output of the pooling layer is flattened to a vector which befalls the input to the fully connected layer, where destruction is organizing the 3-dimension volumes of numbers into a 1-dimension vector.

Figure 8 represents the padding in grey; the input can be embedded with zeros. The padding is the concept of merging the pixels into the image whenever the kernel processing padding value is 0. The testing process has the similar dimensions and the length to highest length will affect the accuracy. The dimensionality of the feature map equals the input value; then the padding is normally used to hold the feature map size that would process at every layer in the convolutional neural networks.

Pooling layers behind sample every feature map individually, diminishing the height and width maintenance with the depth together. Pooling is performed after the convolution operation to diminish the dimensionality, which enable us to reduce the amount of parameters which abbreviates the training time and overfitting. Pooling layer slides a window across its input and takes the highest value in the window where the window size and stride are in particular similar to a convolution. The most general type of pooling is max-pooling which receives the maximum assessment in the window as shown in Figure 9.

From Figure 10, the window and stride configurations segregate the feature map pooling size to minimum that the
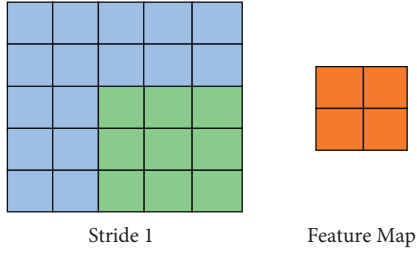
FIGURE 7: Stride value when it equals 2.
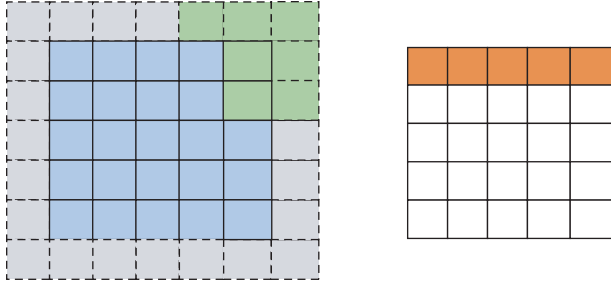
FIGURE 8: Padding.
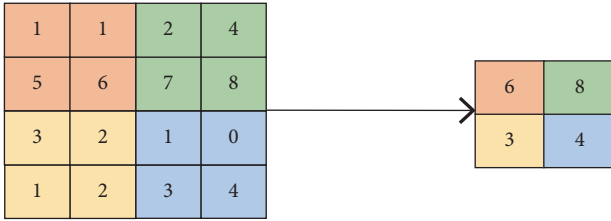
FIGURE 9: Max-pooling with 2×2 window and stride 2.
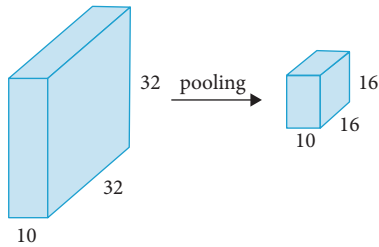
FIGURE 10: Reducing the feature map size by pooling.

feature map is maintaining the significant data. If the input to the pooling layer dimension is $32 \times 32 \times 10$, it could be reduced to $16 \times 16 \times 10$. There are 4 parameters of 3×3, 5×5, 7×7 filter sizes that can be used depending on application. It should be noted that filters have depth which is equivalent to the depth of its input. The filter count is a power of 2 within 32 and 1024. Utilizing additional filters might result in a dynamic model, but these risk overfitting according to the improved parameter value. Frequently, the network starts with minimum amount of filters at the arbitrary layers and progressively increases the value once the network increments. The width and the height are changed but the depth does not change because of the individuality of every layer for pooling.

The VGG network was trained to implement the object recognition with the feature space constructed using convolutional and max-pooling layers. The network is positioned to improve the scaling and weight of the activated filters within the images. With the arrangement of VGG network for producing the better output, the activation functions are generated for increasing the efficient feature maps. This model contains the fully connected layers for performing the pooling operating to increase the efficiency.

The CNN learns the styles through the parameters for training the restructuring loss Loss_R for the targeted output image (O) from an input image (I) computed using the following equation:

$$\mathrm{Loss}_R = ||O - I||^2. \tag{1}$$

The *perceptual* loss Loss_P within the style branch is computed using the following equation:

$$
\begin{aligned}
\mathrm{Loss}_P = {} & x \, \mathrm{Loss}_R \left(O_i, I_i\right) + y \, \mathrm{Loss}_{\mathrm{st}} \left(O_i, \, \mathrm{St}_i\right) + z \, \mathrm{Loss}_N \left(O_i\right) \\
& + p \, \mathrm{Loss}_{\mathrm{st}} \left(O_i, \, \mathrm{St}_i\right),
\end{aligned}
\tag{2}
$$

where St_i is the styled image, $\mathrm{Loss}_{\mathrm{st}}$ is restructuring loss for style, Loss_N is the restructuring loss for the normal, and $\mathrm{Loss}_{\mathrm{in}}$ is the restructuring loss for intensity. The restructuring loss for style $(\mathrm{Loss}_{\mathrm{st}})$ is computed using the following equation:

$$\mathrm{Loss}_{\mathrm{st}} \left(O_i, \, \mathrm{St}_i\right) = \sum_l \left\| \mathrm{GM}\left(\mathrm{FM}^l \left(O_i\right)\right) - \mathrm{GM}\left(\mathrm{FM}^l \left(\mathrm{St}_i\right)\right) \right\|^2. \tag{3}$$

The restructuring loss for intensity is computed using the following equation:

$$\mathrm{Loss}_{\mathrm{in}} \left(O_i, \, \mathrm{St}_i\right) = \frac{1}{\alpha} \sum_{\alpha=1}^{k} \left\| O_i^{\alpha} - \mathrm{St}_i^{\alpha} \right\|^2. \tag{4}$$

The complexity is analyzed in the network that uses the nonlinear values for fixing the layer position in the network. The input image $\vec{a}$ is prearranged in every layer of the CNN using filter reaction to the image. A layer is mapped with the individual filters for producing the feature maps of every matrix and identifies the position of the particular layer.

The style representation is obtained from an input image; the feature space is identified to identify the texture data. The feature space has been constructed to reply the filter in the layers. It contains the correlations within the dissimilar filter replies that are expected for spatial related feature maps. These correlations framed the generalized matrix (Ge_{ij}^l) using feature map Fe_{ik}^l in the following equation:

$$\mathrm{Ge}_{ij}^l = \sum_k \mathrm{Fe}_{ik}^l \, \mathrm{Fe}_{jk}^l. \tag{5}$$

4. Results and Discussion

The experiments have been done using the MATLAB software and the WikiArt data set has been utilized to implement the performance evaluation. The style classification is performed using the style ambiguity. The proposed
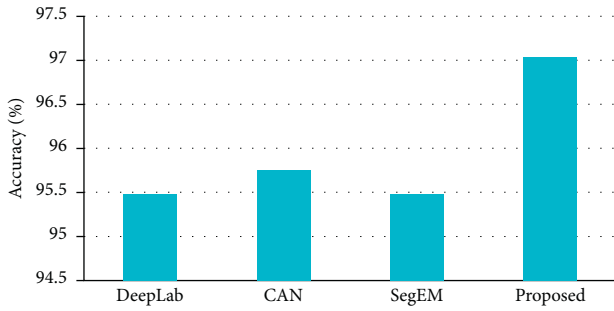
FIGURE 11: Accuracy %.

FIGURE 12: Total runtime.
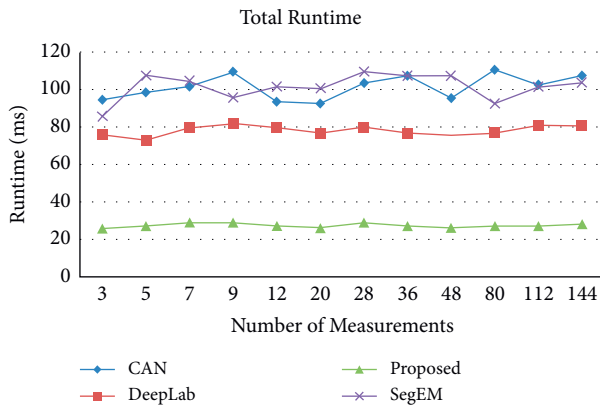
FIGURE 13: Restructuring loss %.

methodology is compared with the related methods of Deeplab [1], CAN [12], and SegEM [23], and the performance metrics used for these experiments are processing time, restructuring loss, and accuracy.

Figure 11 demonstrates the accuracy % for successfully performing style transfer; the proposed methodology has efficient functionality of improving the accuracy compared to the related methods.

Figure 12 demonstrates the runtime for the measurement of the style transfer and the results illustrated that the proposed methodology has diminished amount of runtime compared with the other methods. The restructuring loss is one of the primary parameters to increase the functionality of the proposed methodology.

Figure 13 illustrates that the proposed methodology has reduced amount of restructuring loss percent compared with the related methodologies. The curve is the plot for discovering the performance over time period which has been produced by the machine learning approach.

Our proposed method has produced significantly good results from the neural network based technique. The large-scale based artefacts are generated using this proposed approach. The produced result has high quality in resolution and it looks the same as the input image after successfully implementing the style transfer. Eliminating the style ambiguity loss is used to improve the accuracy of implementing the style transfer. Our proposed system has proved producing the new artefacts without affecting the resolution of the images.

5. Conclusion

The images generated through style transfer are dependent on iterations. The images generated with minimum iterations have more style features while the images generated with maximum iterations have more content features. This traces how humans produce and recognize artistic imagery which gives an algorithmic understanding. In conclusion, that pretrained neural network can be used for not only image recognition but also painting. The feature space demonstrates the larger levels of the input image through the VGG16 neural network for replicating the task while the input image has executed the style transfer to capture the depth image. The validation process is used for producing the style transfer with reduced amount of complexity and the max-pooling layers are involved into the feature space for visual object recognition.

Data Availability

Data will be made available upon request.

Disclosure

This research does not involve any human or animal participation.

Conflicts of Interest

The authors declare that they do not have any conflicts of interest.

Authors' Contributions

R. Dinesh Kumar contributed to writing original draft, writing review and editing, conceptualization, data curation, and validation. E. Golden Julie contributed to conceptualization, formal analysis, and supervision. Y. Harold Robinson contributed to conceptualization, formal analysis, writing original draft, writing review and editing, and supervision. S. Vimal contributed to conceptualization, writing original draft, formal analysis, and supervision. Gaurav Dhiman contributed to conceptualization, writing original

draft, formal analysis, and supervision. Murugesh Veerasamy contributed to conceptualization, formal analysis, writing original draft, writing review and editing, and supervision. All authors have checked and agreed the submission.

References

[1] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Deeplab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 2017, Article ID 2699184, 848 pages, 2017.

[2] T. Li, R. Qian, C. Dong et al., "Beautygan: instance-level facial makeup transfer with deep generative adversarial network," in *Proceedings of the 2018 ACM Multimedia Conference on Multimedia Conference*, pp. 645–653, ACM, Seoul, Korea, October 2018.

[3] W. Wu, Y. Zhang, C. Li, C. Qian, and C. C. Loy, "Reenactgan: learning to reenact faces via boundary transfer," in *Proceedings of the European Conference on Computer Vision*, pp. 622–638, Springer, Munich, Germany, September 2018.

[4] H. Chang, J. Lu, F. Yu, and A. Finkelstein, "Pairedcyclegan: asymmetric style transfer for applying and removing makeup," in *Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Salt Lake City, UT, USA, June 2018.

[5] Z. Ma, N. Wang, X. Gao, and J. Li, "From reality to perception: genre-based neural image style transfer," in *Proceedings of the IJCAI*, pp. 3491–3497, Stockholm, Sweden, July 2018.

[6] M. Lu, F. Xu, H. Zhao, A. Yao, Y. Chen, and L. Zhang, "Exemplar-based portrait style transfer," *IEEE Access*, vol. 6, pp. 58532–58542, 2018.

[7] N. Wang, X. Gao, and J. Li, "Random sampling for fast face sketch synthesis," *Pattern Recognition*, vol. 76, pp. 215–227, 2018.

[8] N. Wang, M. Zhu, J. Li, B. Song, and Z. Li, "Data-driven vs. model-driven: fast face sketch synthesis," *Neurocomputing*, vol. 257, pp. 214–221, 2017.

[9] F. Luan, S. Paris, E. Shechtman, and K. Bala, "Deep photo style transfer," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, July 2017.

[10] L. A. Gatys, A. S. Ecker, and M. Bethge, "Texture synthesis using convolutional neural networks," *Advances in Neural Information Processing Systems*, vol. 28, 2015.

[11] L. A. Gatsy, A. S. Ecker, and M. Bethge, "A neural algorithm of artistic style," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1–16, Boston, MA, USA, June 2015.

[12] E. Ahmed, B. Liu, M. Elhoseiny, and M. Mazzone, "CAN: creative adversarial networks generating 'art' by learning about styles and deviating from style norms," in *Proceedings of the ACC Eighth International Conference on Computational Creativity (ICCC)*, pp. 1–22, Atlanta, Georgia, USA, 2017.

[13] Z. Xu, M. Wilber, C. Fang, A. Hertzmann, and H. Jin, "Adversarial training for fast arbitrary style transfer," *Computers & Graphics*, vol. 87, 2020.

[14] O. Texler, D. Futschik, J. Fišer et al., "Arbitrary style transfer using neurally-guided patch-based synthesis," *Computers & Graphics*, vol. 87, 2020.

[15] L. Liu, Z. Xi, R. Ji, and W. Ma, "Advanced deep learning techniques for image style transfer: a survey," *Signal Processing: Image Communication*, vol. 78, pp. 465–470, 2019.

[16] H. Wu, Z. Sun, Y. Zhang, and Q. Li, "Direction-aware neural style transfer with texture enhancement," *Neurocomputing*, vol. 370, pp. 39–55, 2019.

[17] M. Guo and J. Jiang, "A robust deep style transfer for headshot portraits," *Neurocomputing*, vol. 361, pp. 164–172, 2019.

[18] J. H. Park, S. Park, and H. Shim, "Semantic-aware neural style transfer," *Image and Vision Computing*, vol. 87, pp. 13–23, 2019.

[19] T. Zhao, Y. Yan, I. S. Shehu, X. Fu, and H. Wang, "Purifying naturalistic images through a real-time style transfer semantics network," *Engineering Applications of Artificial Intelligence*, vol. 81, pp. 428–436, 2019.

[20] H. Li, L. Wan, S. Wang, and Z. Cui, "Photo2Silhouette: transferring silhouette photo style for images," *Optics & Laser Technology*, vol. 110, pp. 30–35, 2019.

[21] P. Andreini, S. Bonechi, M. Bianchini, A. Mecocci, and F. Scarselli, "Image generation by GAN and style transfer for agar plate image segmentation," *Computer Methods and Programs in Biomedicine*, vol. 184, Article ID 105268, 2020.

[22] S. Lee, B. Park, and A. Kim, "Deep learning based object detection via style-transferred underwater sonar images," *IFAC-PapersOnLine*, vol. 52, no. 21, pp. 152–155, 2019.

[23] M. Berning, K. M. Boergens, and M. Helmstaedter, "SegEM: SegEM: efficient image analysis for high-resolution connectomics," *Neuron*, vol. 87, no. 6, pp. 1193–1206, 2015.

[24] J. J. Virtusio, A. Talavera, D. S. Tan, K.-L. Hua, and A. Azcarraga, "Interactive style transfer: towards styling user-specified object," in *Proceedings of the 2018 IEEE Visual Communications and Image Processing (VCIP)*, December 2018.

[25] A.-D. Nguyen, S. Choi, W. Kim, and S. Lee, "A simple way of multimodal and arbitrary style transfer," in *Proceedings of the ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Brighton, UK, 2019.

[26] S. Chelaramani, A. Jha, and A. M. Namboodiri, "Cross-modal style transfer," in *Proceedings of the 2018 25th IEEE International Conference on Image Processing (ICIP)*, Athens, Greece, October 2018.

[27] T. Sucontphunt, "A practical approach for identity-embodied 3D artistic face modeling," *International Journal of Computer Games Technology*, vol. 2014, Article ID 781950, 10 pages, 2014.

[28] E. Estrada and P. Pereira-Ramos, "Spatial "artistic" networks: from deconstructing integer-functions to visual arts," *Complexity*, vol. 2018, Article ID 9893867, 8 pages, 2018.

[29] H. Wang, H. Xiong, and Y. Cai, "Image localized style transfer to design clothes based on CNN and interactive segmentation," *Computational Intelligence and Neuroscience*, vol. 2020, Article ID 8894309, 12 pages, 2020.

[30] L. Zeng and X. Dong, "Artistic style conversion based on 5G virtual reality and virtual reality visual space," *Mobile Information Systems*, vol. 2021, Article ID 9312425, 8 pages, 2021.

[31] W. Ye, X. Zhu, Z. Xu, Y. Liu, and C.-C. Chang, "A comprehensive framework of multiple semantics preservation in neural style transfer," *Journal of Visual Communication and Image Representation*, vol. 82, Article ID 103378, 2021.

[32] H. Huang, A. Yang, Yu Tang et al., "Deep color calibration for UAV imagery in crop monitoring using semantic style transfer with local to global attention," *International Journal*

of Applied Earth Observation and Geoinformation, vol. 104, 2021.

[33] J. Cui, Y. Q. Liu, H. J. Lu et al., "PortraitNET: photo-realistic portrait cartoon style transfer with self-supervised semantic supervision," *Neurocomputing*, vol. 465, 2021.

[34] H. B. Waqas, J. Ahmed, and T. R. Hafiz, "A modified bat algorithm with torus walk for solving global optimisation problems," *International Journal of Bio-Inspired Computation*, vol. 15, no. 1, pp. 1–13, 2020.

[35] W. Haider Bangyal, A. Hameed, J. Ahmad et al., D. Rawat, R. Etengu, D. B. Rawat, and R. Etengu, New modified controlled bat algorithm for numerical optimization problem," *Computers, Materials & Continua*, vol. 70, no. 2, pp. 2241–2259, 2022.

Research Article

Research on Engineering Geomechanics Characteristics and CFRP Reinforcement Technology Based on Machine Learning Algorithms

Baoqi Yan [iD],[1] Nuoya Zhang,[1] Ganggang Lu,[1,2] and Yue Hui[1]

[1]*Chang'an University, School of Civil Engineering, Xi'an, Shaanxi 710064, China*
[2]*China Northwest Architecture Design and Research Institute Co., Ltd, Xi'an, Shaanxi 710018, China*

Correspondence should be addressed to Baoqi Yan; 2018128055@chd.edu.cn

We have completed the design of an early warning and evaluation analysis module based on machine learning algorithms. Aiming at the prestressed CFRP-strengthened reinforced concrete bridges under natural exposure, we developed a theoretical model to analyze the long-term prestress loss of reinforced parts and the adhesion behavior of the CFRP-concrete interface under natural exposure conditions. The analysis deeply reveals the technical and engineering geomechanics characteristics of the D bridge. At the same time, through a series of experimental studies on the D bridge condition monitoring system, the data acquisition and transmission, processing and control of the D bridge condition monitoring system, and the bridge condition monitoring and evaluation software are provided. Regarding how to repair the engineering geomechanical characteristics of D bridge, we mentioned the prestressed CFRP reinforcement technology. The prestressed carbon fiber reinforced composite (CFRP) structure made of reinforced concrete (RC) makes better use of the high-strength characteristics of CFRP and changes. It strengthens the stress distribution of the components and improves the overall strength of the components. It is more supported by engineers in the civil engineering and transportation departments. However, most prestressed CFRP-reinforced RC structures are located in natural exposure environments, and the effect of natural exposure environments on the long-term mechanical properties of prestressed C FRP-reinforced RC components is still unclear. This article mainly uses the research on the engineering geomechanics characteristics and reinforcement technology of the bridge body, so that people have a deep understanding of its concept, and provides reasonable use methods and measures for the maintenance and protection of the bridge body in the future. This paper studies the characteristics of engineering geomechanics based on machine learning algorithms and applies them to the research of CFRP reinforcement technology, aiming to promote its better development.

1. Introduction

Since the 1990s, machine learning algorithms have gradually been applied to the study of geomechanical properties of bridge construction and related research on the properties of damaged parts of bridges [1]. The engineering geomechanical characteristic data of the bridge damage is sent to the machine, and the machine uses the learning algorithm calculation and pattern recognition to complete the work on the damage site [2]. The identification of the damage program can be traced back to the problem of functional adaptation [3]. The machine learning algorithm learns from

historical data to set the geoengineering features of different damage degree projects and adapts the appropriate feature plane to the special space where the feature data is sent to the machine learning algorithm [4]. The learning algorithm calculates the data on the surface of the corresponding feature and then performs interpolation to evaluate the degree of damage, and it plays a leading role in the repair of the bridge [5]. Due to the repeated action of the alternating load of the bridge, the maximum stress that the concrete bridge bears does not reach the allowable stress of the static strength design [6]. Fatigue cracks are generated at the local positions of the concrete members and propagate, and

finally, fatigue fracture will occur. It brings greater safety and social risks, and its maintenance also brings a huge economic burden, which brings great troubles to people's travel convenience [7]. Studies have shown that the cost of building a new bridge is relatively high, and the cost of strengthening a bridge is at least 10% of the cost of a new bridge. Relatively speaking, it is cheaper and more efficient to use reinforcement or repair instead of new construction [8]. Therefore, this article will focus on the research of carbon fiber reinforced composite material (CFRP) paste reinforcement technology to explore whether the engineering geomechanical properties of the damaged bridge can be restored [9]. At present, engineering practice and a number of research results show that prestressed CFRP reinforcement technology, as an active reinforcement technology, significantly improves the utilization efficiency of CFRP, provides a more convenient and efficient use plan for the application of CFPR, and makes better use of it [10]. The strength advantage lays a solid foundation for a wider range of applications. It can participate in the force (effective force) before the reinforced component is subjected to the second force, and the stress distribution of the reinforcing element can be changed to close the original crack or prevent its expansion [11]. Therefore, the prestressed CFRP reinforcement technology is more and more favored by traffic and civil engineers [12]. The use of prestressed CFRP reinforcement technology has become widespread in practice. Under actual operating conditions, the long-term mechanical properties of prestressed CFRP-reinforced concrete (RC) components have also attracted more and more attention, and they have been more and more used in various applications [13]. This kind of machinery played a greater role in the practice process.

2. Materials and Methods

2.1. Research Objects

2.1.1. Introduction to the Example Test. After the fatigue test of the reinforced concrete beam by the Structural Laboratory of Chang'an University, the fatigue test was analyzed by finite element simulation ABAQUS. Three typical long-span bridges are selected as representatives: Bridge A, Bridge B, and Bridge C. These were originally built in certain urban areas and are still in operation and are calculated by Midas Civil 2012 software. Then, based on the "Specifications for Design of Concrete Structures", the fatigue stress calculations of the three small and medium-sized bridges were confirmed, and the stress levels of the vertical steel bars under the design load of the three bridges were investigated to be 0.5 and 0.55, respectively [14]. The minimum stress level under load is 0.25.

2.1.2. Selection of Material Parameters. (1) According to concrete reference materials, all concrete beams in the fatigue simulation test of this study are made of C40 concrete. (2) In the rod fatigue simulation test, the strength grade of the rod is HRB400.

2.1.3. Design of Test Beam Conditions. It has 3 test strips, which can be divided into 2 groups. One group is the static load test, which measures the maximum load-bearing bending moment Mu of the concrete beam and determines the upper and lower limits of fatigue according to the stress level required by the fatigue test [15]. The other group is the test beam fatigue test.

2.2. Theoretical Basis

2.2.1. Early Warning Assessment of Long-Span Bridge Damage. In recent years, the overall structural condition monitoring technology based on vibration testing and modal analysis has been relatively mature in the machinery and aerospace industries. After the early warning method is used to determine the damage of the long-span bridge, the test mode analysis should be used to further diagnose the location and extent of the structural damage in order to monitor the long-span bridge structure [16] and provide basis and guidance for system identification, structure evaluation, and decision-making for maintenance and repair managers [17]. This content can be summarized as an early warning assessment of remote bridge damage. According to engineering data, D bridge's 3D benchmark finite element model has 11 elements and 10 nodes. Damage analysis allows us to select 10 nodes as damage objects. In the (x, y) quadrant along the x-direction, there are 1 to 10 nodes in turn. 10 nodes with different damage levels (5%–100%) can get 200 groups of damage conditions. The 200 damage conditions are divided into 20 condition groups, and each condition group has 26 damage conditions. Concerning the damage to 10 nodes numbered 1–10, the damage degree of each working condition group gradually increases. After a series of calculations, the influence of the location and damage degree of each node on the natural frequency of the bridge can be determined. This paper normalizes the natural frequencies of each damage condition group based on normal data.

$$\overline{f}_{ijk} = \frac{f_{iik}}{f_{i0}}, \quad i = 1, 2, \ldots, 10; \ j = 1, 2, \ldots, 20; \ k = 1, 2, \ldots, 10. \tag{1}$$

In the formula, f_{ijk} represents the i-th natural frequency of the j-th operating condition group and the k-th operating condition (dimension) normalized by data. The k-th working state is represented by f_{ijk}. There is an i-th natural vibration value (Hz) of the j-th working condition group before data normalization. f_{i0} represents the i-th natural vibration value (Hz) without damage.

2.2.2. Neural Network Method. Since the 1990s, neural networks have gradually been used in bridge research to identify structural damage. For some applications, only injury data from injury condition group 1 (5% injury) is used as the training set and sent to the neural network for training. We understand the distribution function between the damage location and the natural frequencies of each order. If a knot is damaged, even if the actual damage is

different from the training conditions, the damage will not have a significant impact on the above [18]. For the mentioned mapping feature and damage feature sensitive positioning factors, the neural network can identify the actual damage location according to the natural frequency distribution of each order in the multidimensional space. The specific process is as follows:

$$e_k(n) = d_k(n) - y_k(n). \tag{2}$$

We define the total energy function of the network as follows:

$$E_n(n) = \frac{1}{2} \sum_k e_k^2(n). \tag{3}$$

The network feedforward process is as follows:

$$
\begin{aligned}
v_j(n) &= \sum_i \omega_{ji}(n) y_i(n), \\
y_j(n) &= f\left(v_j(n)\right), \\
v_k(n) &= \sum_j \omega_{kj}(n) y_j(n), \\
y_k(n) &= f\left(v_k(n)\right).
\end{aligned}
\tag{4}
$$

The reverse calculation process of the network can be obtained:

$$\frac{\partial E(n)}{\partial \omega_{ji}(n)} = \left[-\sum_k \delta_k(n) \omega_{kj}(n) \right] f'\left(v_j(n)\right) y_i(n). \tag{5}$$

Then, the weight correction amount of the hidden layer unit is

$$
\begin{aligned}
\Delta w_{ji}(n) &= \eta \delta_j(n) y_i(n), \\
\omega_{ji}(n+1) &= \omega_{ji}(n) + \Delta \omega_{ji}(n) \\
&= \omega_{ji}(n) + \eta \delta_j(n) y_i(n).
\end{aligned}
\tag{6}
$$

2.3. Research Methods

2.3.1. Mechanical Analysis Corresponding to Rigid Body Displacement at Structural Level. The rigid body displacement of the structural plane usually includes the lateral displacement, the longitudinal displacement, the vertical displacement, and the rotation around the vertical, horizontal, and vertical axes of the entire structure. The rigid body displacement at the structural level is the same spatial change in the entire structure at the same time, so whether it is a statically indeterminate structure or a statically indeterminate structure, the rigid body displacement at the structural level will not generate force. The displacement of the rigid body at the structural level does not produce any force on the structure, so it is not always subject to force when the structure moves. In the case of rigid body displacement at the structural level when the bridge space changes, the force cannot be equal to the space deformation and the stress state cannot be derived from the bridge space deformation. If the structure-level rigid body displacement

is not distinguished, but the structure-level rigid body displacement and the component-level displacement are mixed and the mechanical state is analyzed, the force state can be obtained from the analysis of the mixed space. In this process, the shape change is inaccurate because it is assumed that the displacement of the rigid body can exert pressure on the structure at the structural level. Therefore, before the analysis of mechanical properties, structural rigid body displacement analysis is required to identify and eliminate the total rigid body displacement of the structure.

2.3.2. Mechanical Analysis Corresponding to Rigid Body Displacement at Component Level. The component-level rigid body displacement analysis is usually based on the structural-level rigid body displacement analysis. The rigid body displacement of the component usually includes the lateral displacement, vertical displacement, and deflection of the component. For statically indeterminate structures, the rigid body displacement at the component level will not produce a force on the component, but for statically indeterminate structures, the rigid body displacement of the component causes the component to receive minor internal forces. In the statically determinate structure of the rod, if there is no rigid body displacement at the rod level, the deformation shape includes the rigid body deformation of the entire rod, and the rigid body deformation of the entire rod will not affect the force of the member. Rigid body displacement analysis is used to identify and eliminate the total rigid body displacement of the component to obtain an accurate force state. For components with a statically indeterminate structure, the displacement of the rigid body at the component level will cause the component to generate secondary internal forces. The force state of the component can be calculated by adding the rigid body displacement obtained at the structural level to the force displacement, using the finite element model of the component.

2.3.3. Mechanical Analysis Corresponding to Changes in the Internal Space of Components. The internal deformation analysis of the component is usually based on the rigid body displacement analysis at the structural level and the rigid body displacement analysis at the component level. Deformation of the interior of a part usually includes vertical bending, horizontal bending, and twisting. From the change of the vertical bending moment of the member along the horizontal direction of the bridge, the change of the vertical bending moment of the member can be obtained, so as to determine the direction of the change of the vertical bending moment; the transverse bending change of the steel bar in the longitudinal direction of the bridge receives the change of the transverse bending moment. We determine the direction of lateral bending moment change. Changing the torsion of the rod changes the torque received by the rod and determines the direction of the torque.

2.4. Experimental Protocol. The experimental method to study the behavior of the FRP-concrete interface is the pull-

out test. There are three common types; single shear test, double-shear test, and bending test are all causal pull-out tests. The FRP termination is poorly peeled. The advantage of a simple shear test is that it is easy to handle, but it is very difficult to meet the final load level and may cause an eccentric load. Although the double-shear test can overcome the shortcomings of eccentric load, most concrete structures are in a bent state, and the double-shear test cannot fully simulate the actual situation of GRP reinforced concrete members. In this chapter, we choose the bending test as the experimental method to study the interface behavior of FRP concrete. In this experiment, a total of 6 samples with exposure times of 0, 180, and 360 days were tested. Table 1 shows the mechanical properties of the main materials in the experiment.

The specimen consists of two RC elements with a length of 300 mm and a cross-sectional dimension of 100 mm × 150 mm. The two concrete members are connected by two HRB hot-rolled ribbed steel bars with a length of 650 mm and a diameter of Φ12, leaving a gap of 10 mm between the two concrete members. Considering various effective bonding lengths, the maximum effective length of the test piece is 109 mm. Therefore, a GRP adhesive with a length of 150 mm is selected, and a GRP board with a size of 50 mm × 310 mm is attached from the outside on the underside of the sample. On the other side, a 200 mm × 600 mm FRP board is wrapped for shear reinforcement to ensure that peeling occurs first.

In this work, a 4-point bending test was carried out. Therefore, all samples were loaded into a hydraulic press with a capacity of 500 kN, and the path-controlled monotonic loading method was used to load the two-point loading system. The distance between the two load points and the two supports is 100 mm or 310 mm, and a monotonic load occurs at a constant speed of 0.2 mm/min. The force sensor is connected to the two loading heads and records the applied load. For each test, a linear variable differential sensor (LVDT) with a measuring range of 50 mm and an accuracy of 0.01 mm is placed in the center of the span to measure the deflection of the test piece. In order to record the strain of FRP, a total of 8 strain gauges were installed at a distance of 20 mm outside the entire length of FRP, and the static strain gauge automatically recorded the strain value of the entire test. Before the formal application, the test piece should be pretensioned to check the operating condition of the machine and reduce the gap between the load head and the test piece.

The results of the bending test of the test samples show the bonding behavior of the interface between the FRP and the concrete while considering the influence of the FRP type and exposure time, including the peeling load, the fracture mode, and the FRP load under various loads. As shown in Table 2, the final stresses of the specimens after exposure to the subtropical natural environment for 0 days, 180 days, and 360 days.

The way the sample breaks is to peel the FRP from the interface and place concrete in various thicknesses. When poor peeling occurs, there is a slight cracking sound in the middle of the span. As the load increases, the surface of the

TABLE 1: Main material parameters.

Material	Material parameters	
CFRP board	Tensile strength/MPa	4133
	Elastic modulus/GPa	230
	Elongation/%	1.88
	Thickness/mm	0.168
BFRP board	Tensile strength/MPa	2101
	Elastic modulus/GPa	90
	Elongation/%	2.5
	Thickness/mm	0.110
Adhesive	Tensile strength/MPa	49
	Elastic modulus/GPa	2.379
	Elongation/%	1.9
Concrete	Compressive strength/MPa	30.6

TABLE 2: Experimental results of test pieces.

Specimen number	FRP type	Exposure time/d	Peeling load/kN
CFRP-0		0	36.65
N-CFRP-180	CFRP	180	31.65
N-CFRP-360		360	32.97
BFRP-0		0	22.31
N-BFRP-180	BFRP	180	16.33
N-BFRP-360		360	16.63

concrete will begin to crack, producing a cracking sound, and eventually, the FRP will peel off the concrete. Compared with the control sample, the fracture mode of the CFRP sample is the shear fracture and delamination of the concrete, while the BFRP sample is the FRP directly peeled from the concrete. Generally, the damage of the specimen exposed to the natural environment is the peeling damage at the interface between FRP and concrete. In a naturally exposed environment, rainwater directly penetrates into the adhesive layer, and the adhesive layer is more sensitive to moisture, thereby reducing the shear strength of the adhesive layer interface. It can be seen that whether it is the CFRP-concrete interface or BFRP-concrete interface, their adhesion properties deteriorate with the increase of exposure time.

3. Results

3.1. Analysis of Dead Weight and Dead Load Effect of Engineering Geomechanics

3.1.1. The Dead Weight Effect. In the case of a straight beam bridge, the influence of its own weight is generally only vertical displacement, but in the case of a curved beam bridge, the center of gravity of its own weight moves to the outside of the curve, and there is an eccentric distance from the axis. This study uses the finite element model simulation method to calculate the displacement of the D bridge under its own weight and dead load, as shown in Figure 1. This study defines the following: the vertical displacement upward is positive, downward is negative, the lateral displacement of the curved bridge is positive, and the inward

(a)

(b)
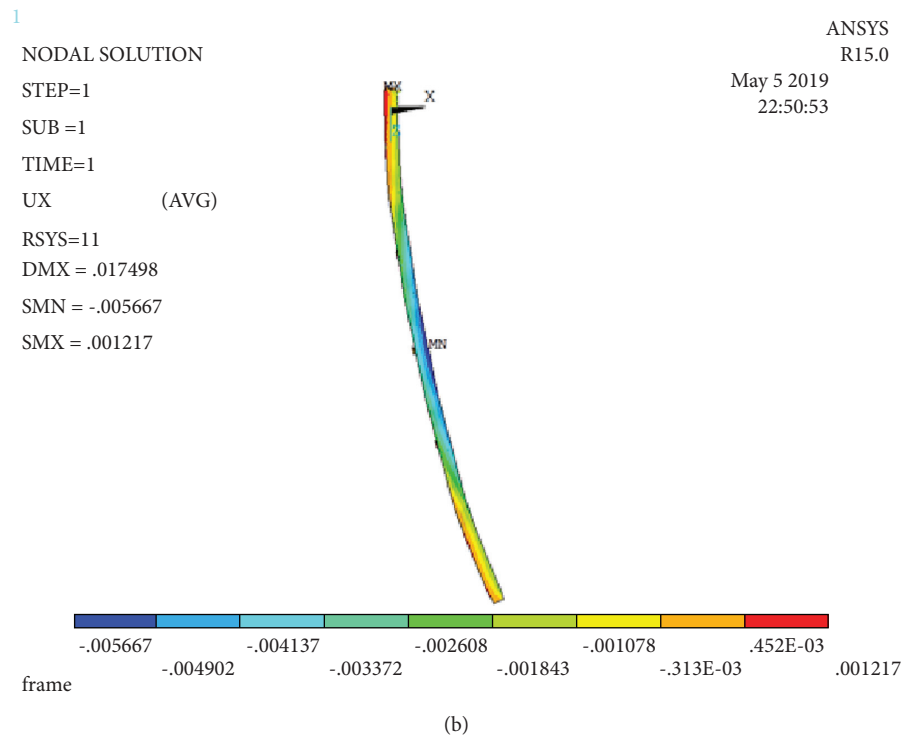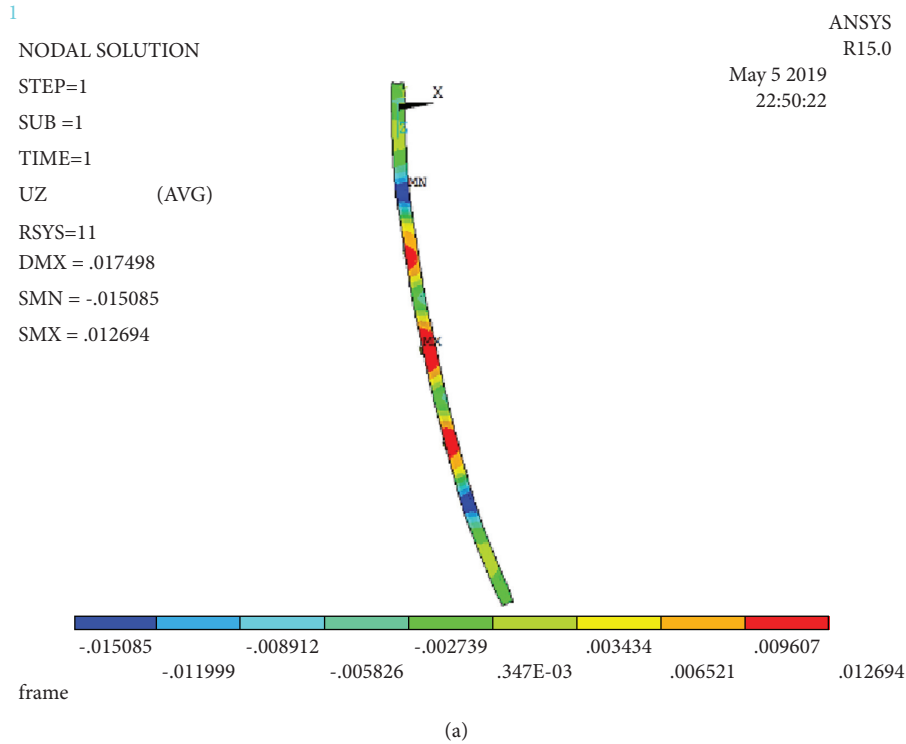
Figure 1: Displacement diagram of the bridge under dead load (vertical displacement diagram of node a and lateral displacement diagrams of b): (a) node vertical displacement graph and (b) node lateral displacement graph.

is negative, the outer member of the curved bridge rotates as positive, and the inward bending is negative.

From the impact vertical displacement diagram of bridge D under its own weight and dead load, it can be seen that the downward vertical displacement amplitude of the inner node is greater than that of the outer node, and the upward

vertical displacement amplitude is as follows. Therefore, every span of the bridge bends inward. As shown in Table 3, under its own weight and dead load, the pelvic support is located on the #1 column of the D bridge and is horizontal, so the corresponding main beam is clamped vertically on the #1 column to move vertically, and the lateral displacement is

TABLE 3: Displacement of 1#–4# pier top under dead load (unit: m).

Pier number	Vertical displacement	Lateral displacement
1	$2.30E-18$	$-1.62E-07$
2	$8.81E-04$	$-2.37E-02$
3	$1.02E-03$	$-3.61E-02$
4	$8.08E-04$	$-2.22E-02$

almost no exist. The bridge pier beam between 2#-4# is integrated with the main girder of the D bridge, and the main girder joints on the top of the column are all vertically and laterally displaced upward and laterally toward the inner side of the arch bridge.

3.1.2. Vertical Settlement of Pier Bottom.

In the static structure, the vertical settlement of the column will not produce any secondary stress, and the static structure bridge only undergoes spatial deformation without any internal force. In statically unsafe structures affected by the uneven vertical settlement of the columns, the bridge will not only produce spatial deformation but also generate secondary internal forces. Figure 2 shows the spatial deformation of bridge D when the piers are vertically settled.

It can be seen from Figure 2 that the vertical displacements of the inner and outer nodes of the main beam are consistent, and the vertical sinking of the column will not cause the main beam to twist. Under the action of vertical sinking, the corresponding main girder of D-shaped bridge 1# column top is supported vertically and horizontally, and the vertical and horizontal displacement is small.

3.2. Analysis of Temperature Effect of Engineering Geomechanics

3.2.1. Overall Temperature Change.

The ambient temperature of the bridge changes slowly and steadily over a long period of time. When the ambient temperature changes, the entire bridge structure will have the same temperature change. In the finite element model, the total temperature rises and falls, and the structure is usually used for simulation. This study is based on an ambient temperature of 18°C and considers the changes in the spatial shape of the D-shaped bridge when the main girder is raised or lowered by 12°C, as shown in Figure 3.

Figure 4 shows the displacement of the bridge under the influence of the overall cooling of the main girder.

It can be seen from Figures 3 and 4 that under the influence of the total heating and cooling of the main girder, the spatial change direction of the D bridge is opposite and the values are the same. As shown in Table 4, the main girder corresponding to the D 1# bridge top is subject to vertical and horizontal restrictions under the influence of the total temperature rise and decline of the main girder, and there is almost no lateral displacement in the vertical and horizontal restrictions. Under the influence of the total heating of the main girder, the 2#–4# beam reinforcement column has vertical downward displacement and lateral displacement on the corresponding main girder. Under the influence of the

overall cooling of the main girder, the column corresponds to the main beam and moves vertically upward and laterally inward.

3.3. CFRP-Reinforced Structure Analysis

3.3.1. Strain Changes.

Figure 5 shows the effect of natural exposure time on the bonding behavior of the FRP-concrete interface, showing the elongation of CFRP and BFRP along the bond length under different loads. The horizontal axis represents the distance from the strain gauge to the span center, and the vertical axis represents the strain value.

When the loading begins, only the first and second strain gauges show strain readings near the center of the span. In addition, the value of the first strain gauge closest to the center of the span is much higher than the value of the second strain gauge. As the stress increases, the FRP stress begins to appear at the far end, and the stress near the center increases sharply. When FRP peels off in the middle of the span, the interfacial tension is redistributed, and the elongation at the other end reaches its maximum value. At this time, the FRP has completely peeled off the concrete surface. When the applied load is close to the ultimate load, the strain value far from the center of the field is very small, indicating that the ultimate load of the specimen does not increase with the increase of the bonding length of FRP.

3.3.2. Bond Stress-Slip Relationship.

According to the data in Figure 5, the bilinear coupling stress-slip curve of the sample can be adjusted for different exposure times. As shown in Figure 6, as the exposure time increases, these points can be adjusted to the rising and falling levels of the model to obtain the bond stress-slip relationship between the CFRP-concrete interface and the BFRP-concrete interface.

3.4. The Influence of CFRP-Reinforced Structure Stress Distribution.

Figure 7 shows the comparison between the theoretical calculation results and the experimental results of the test piece N-CFRP-180. The results show that the interface shear stress distribution of the two results is similar at all stages. As the load increases, the interfacial tension undergoes a rising and falling phase, and the rate of change has a constant difference. Compared with the experimental results, the theoretical calculation results are more evenly distributed far away from the loading position, which is mainly due to the continuity of the theoretical model assumptions.

By substituting the bond stress-slip ratio of the specimen under different exposure times (0, 180, and 360 days) into the calculation process, the bond shear stress distribution along the bond longitudinal direction under different loads is obtained, as shown in Figure 8.

It can be seen from Figure 8 that, at the beginning of loading, the bond stress at the FRP-concrete interface near the center of the field is higher, and the shear stress in other areas is close to zero. This indicates that the interfacial shear tension is mainly concentrated in the center of the span and is not

(a)

(b)
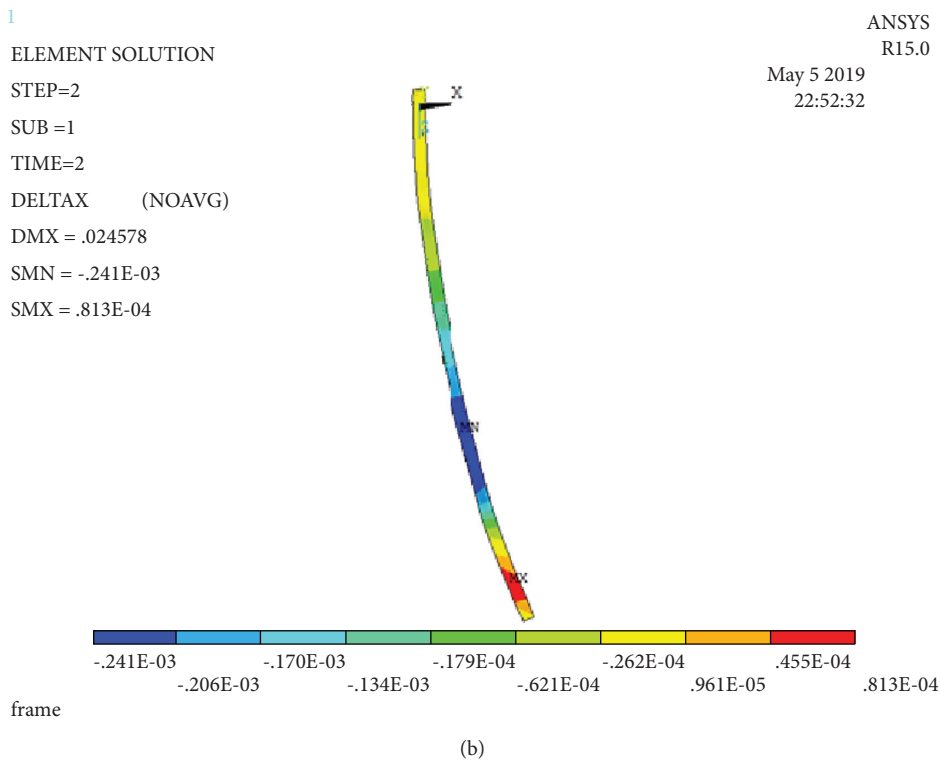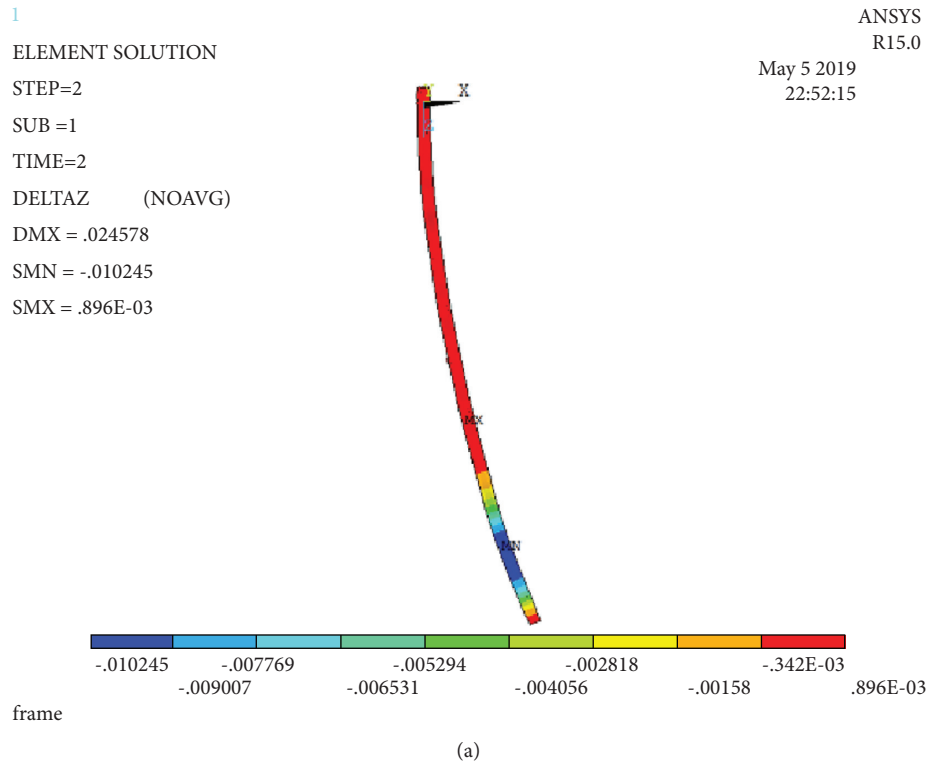
FIGURE 2: 2# Bridge displacement diagram under the action of pier settlement 0.01 m (vertical displacement diagram of node a lateral displacement diagram of node b): (a) node vertical displacement graph and (b) node lateral displacement graph.

transmitted to the distal end. As the load slowly increases, the position of maximum coupling stress moves backward, and shear stress begins to appear at the distal end. Under different natural load times, the nodal tension in the direction of the entire FRP node indicates that the more the maximum nodal load position moves backward, the longer the natural load time. This indicates that the behavior of the FRP-concrete interface is weakened by the effects of natural exposure.

(a)

(b)
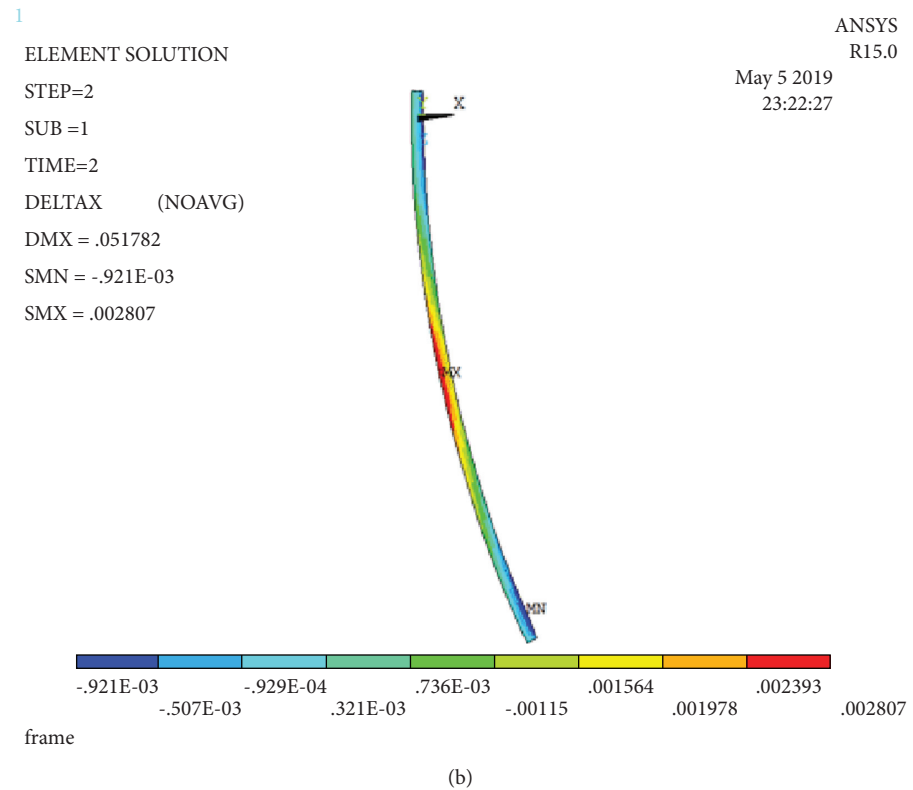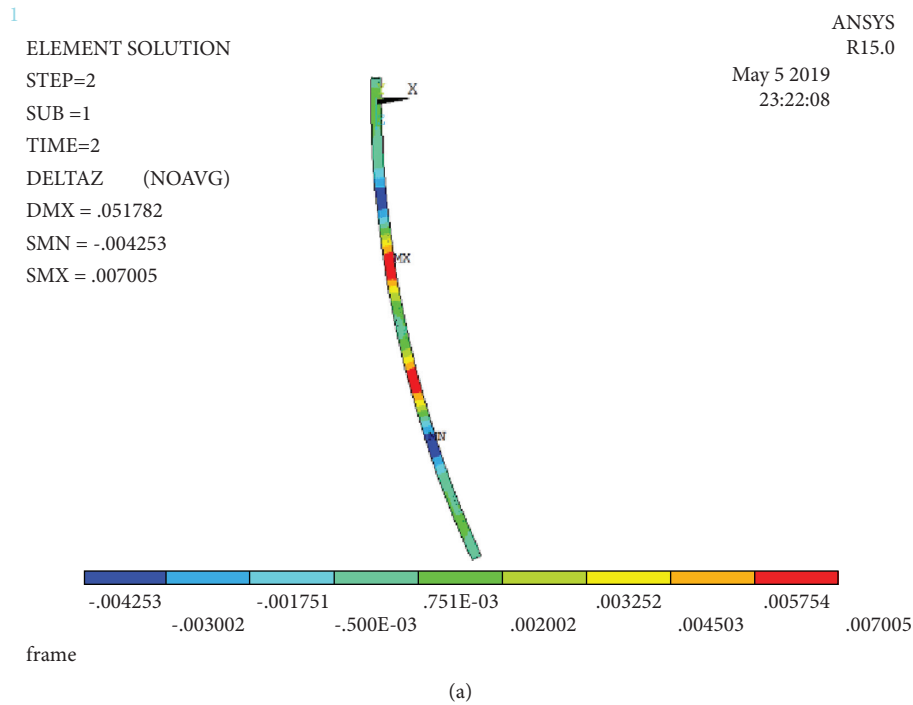
FIGURE 3: Bridge displacement diagram under the effect of the overall heating of the main girder (vertical displacement diagram of node a lateral displacement diagram of node b): (a) node vertical displacement graph and (b) node lateral displacement graph.

4. Discussion

4.1. Engineering Geomechanics Characteristics Analysis Technology. The ability of a bridge to deform under load is an important parameter for evaluating the structural

stiffness of the bridge during operation. Bridge deformation can be divided into two aspects: local deformation and overall deformation. Local deformation refers to the stress and strain of bridge components that are usually detected by strain sensors. The overall deformation reflects the overall
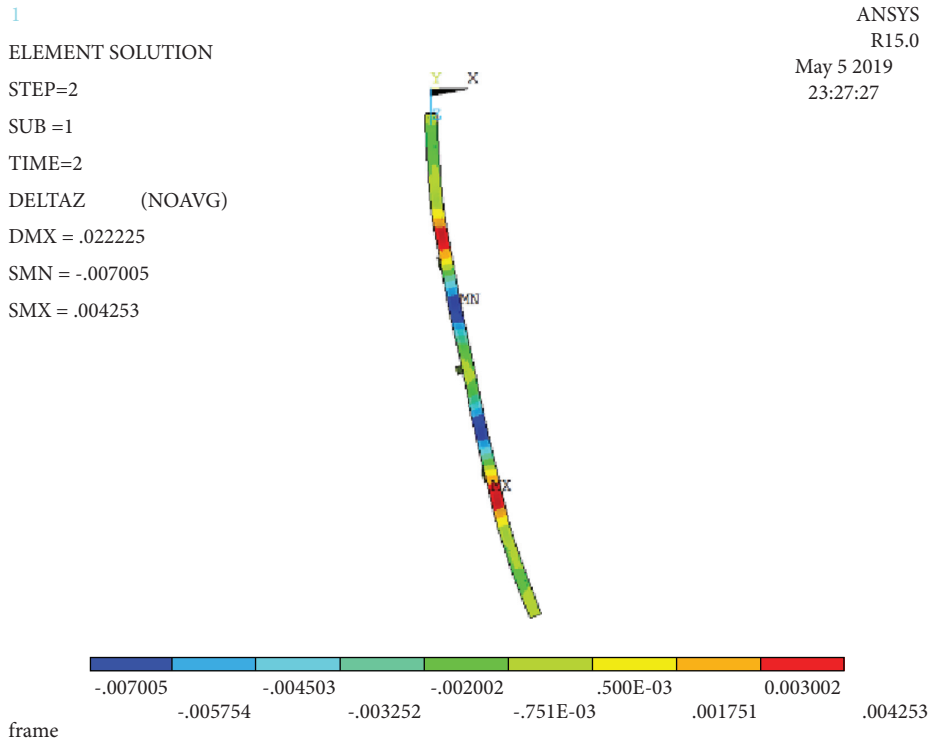
(a)
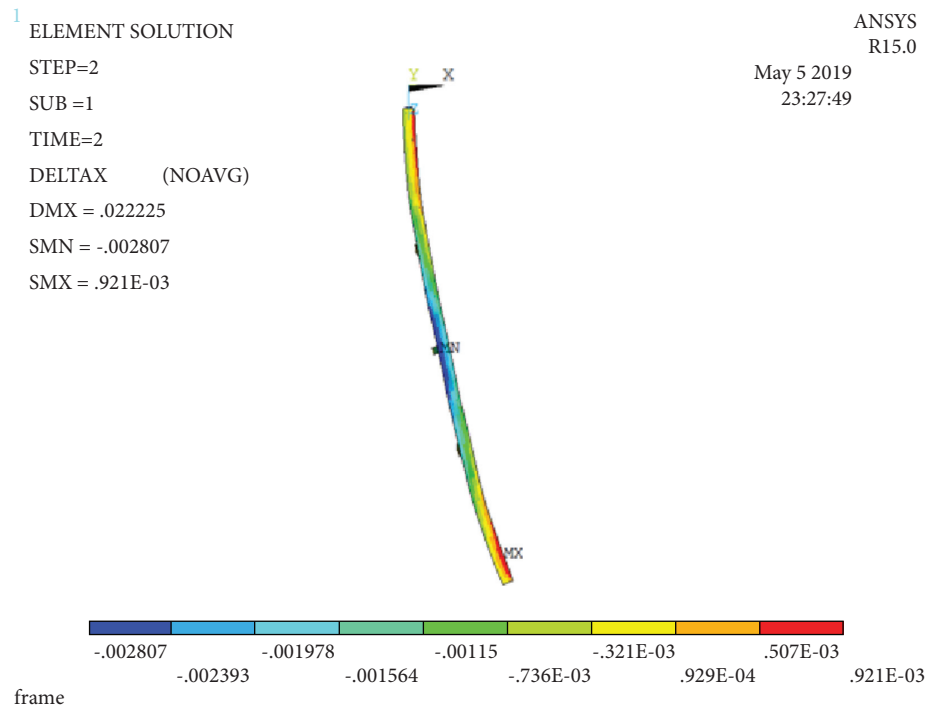
(b)

FIGURE 4: Bridge displacement diagram under the effect of overall cooling of the main girder (vertical displacement diagram of node a lateral displacement diagram of node b): (a) node vertical displacement graph and (b) node lateral displacement graph.

operating conditions of the bridge structure. The content of bridge deformation detection mainly includes main girder deflection, main girder vertical displacement, main girder lateral displacement, and main girder tip displacement.

In recent years, bridge deflection measurement methods based on digital imaging technology have developed rapidly. Compared with other deflection measurement methods, this technology does not require the installation of sensors. By

TABLE 4: Displacement of the top of pier 1#–4# under the effect of the overall temperature rise and fall of the main beam (unit: m).

Working condition	The overall temperature of the main beam is increased by 12°C		The overall temperature of the main beam is reduced by 12°C	
Pier number	Vertical displacement	Lateral displacement	Vertical displacement	Lateral displacement
1	$6.23E{-}20$	$-4.44E{-}07$	$-6.25E{-}18$	$4.42E{-}09$
2	$-1.33E{-}05$	$1.05E{-}02$	$1.35E{-}03$	$-1.03E{-}04$
3	$-1.16E{-}05$	$1.99E{-}02$	$1.18E{-}03$	$-1.97E{-}04$
4	$-1.31E{-}05$	$8.45E{-}03$	$1.33E{-}03$	$-8.43E{-}05$
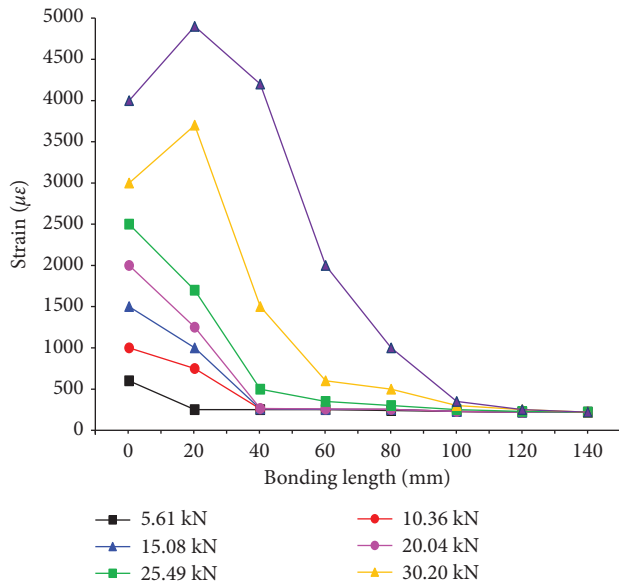
FIGURE 5: CFRP-reinforced control specimen.
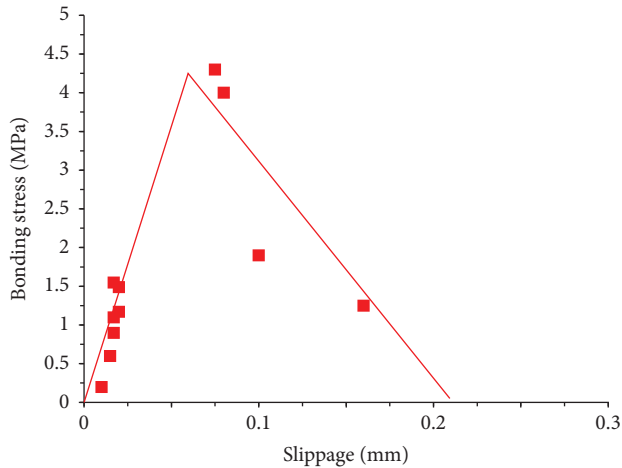
FIGURE 6: CFRP-reinforced control specimen.

comparing the digital images before and after the deformation of the structure, the structure displacement information is obtained to achieve long-distance noncontact measurement. It can be performed at high speed. It is the current bridge load test and the relative deflection of other buildings. There is a new method of change experiment. Digital imaging technology usually needs to first obtain 3D coordinate data based on image information and then perform the next step of the analysis. The 3D laser scanning technology can directly obtain the 3D coordinate data of the physical structure, which simplifies the data analysis process and reduces the error.

4.2. CFRP Reinforcement Model Analysis Principle

4.2.1. Selection of Concrete Failure Criteria. The actual performance and strength of concrete materials is a very complex issue. In the current concrete finite element analysis, a certain order of magnitude of concrete volume is generally regarded as a unit, as a continuous and uniform isotropic material. At this stage, scientists at home and abroad have proposed two 3D concrete failure standards. (1) The viewpoints used by the classical strength theory and the calculation formula standards of the classical strength theory are the Drucker-Prager strength standard and the Von Mises strength standard. (2) The more common multiparameter strength standards, are Bresler-Pister error standards, William-Warnke error standards, etc. This article uses ANSYS 2018 software for finite element analysis, so the Drucker-Prager strength standard is used.

4.2.2. Treatment of Concrete Cracks. The current finite element analysis can be divided into three commonly used concrete crack models: (1) individual crack models, (2) special unit models, and (3) scattered crack models. The advantage of the distributed crack model is that it does not increase the number of elements and the number of nodes, which is convenient for the realization of the finite element program. Therefore, the ANSYS finite element analysis adopts the diffusion crack model in this paper.

4.2.3. Bonding Treatment between Steel Bar and Concrete in Concrete Finite Element Analysis. In the existing finite element analysis, the two most commonly used methods for simulating the sliding of the steel bar and the concrete joint are as follows. (1) Method one is to add a fastener between the concrete and the steel bar to reflect the combined slip relationship. (2) Method two is to use material equation fitting to indirectly consider the effects of adhesion and slip. This method is simple and easy to use, without affecting the grid spacing, the calculation time convergence effect is good, and the result is high in accuracy, so this work uses this reinforced concrete pouring method.

(a)

(b)
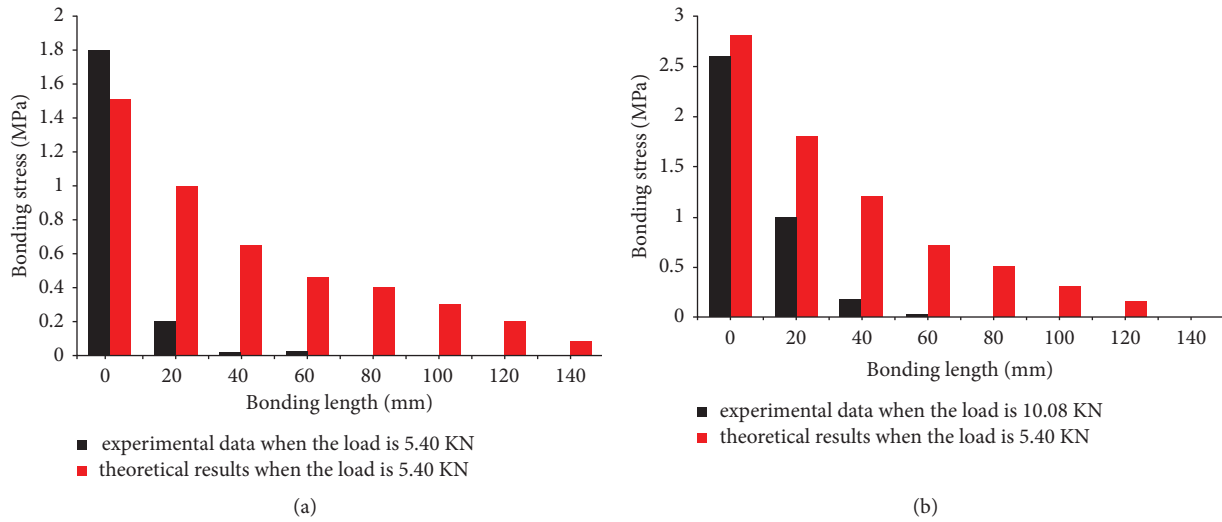
FIGURE 7: Comparison of theoretical results and experimental results of the N-CFRP-180 test piece: (a) $p = 5.4$ kN corresponding bond stress and (b) $p = 10.08$ kN corresponding bond stress.
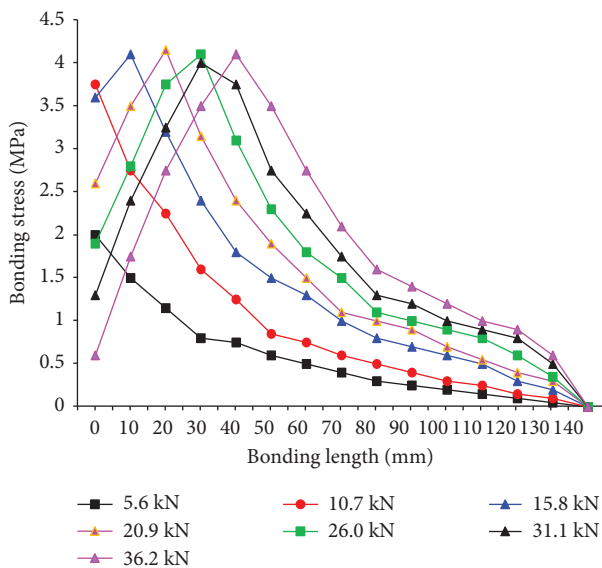
FIGURE 8: Theoretical results of the bond shear stress distribution at the interface of the specimen during the loading process.

4.3. CFRP Reinforcement Model Analysis Purpose.

It is very expensive to replace all hidden bridges, and the roads need to be closed when the bridges are replaced. Studies have shown that the cost of building a new bridge is relatively high, and the cost of strengthening a bridge is at least 10% of that of a new bridge. Relatively speaking, it is cheaper to use reinforcement and repair than to build a new one, and on the other hand, it is more efficient. Therefore, under the premise of minimizing bridge damage, aging, and improving bridge life, in a broad sense, strengthening and repairing bridges has important economic and social benefits.

Although CFRP paste reinforcement technology is common, the disadvantage of this passive reinforcement technology is that it does not have the high-strength performance of CFRP materials, and the reinforcement components only work after secondary stress. It is used in most scientific research and engineering technology and is recognized by the person in charge. In order to make up for the shortcomings of this technology, prestressed CFRP reinforcement technology was temporarily introduced, imitating the technical idea of prestressed steel bars/strands. It can participate in the force (effective force) before the second force is applied to the reinforced component and change the stress distribution of the reinforcing element to close the original crack or prevent its expansion. Therefore, the prestressed CFRP reinforcement technology is more and more favored by traffic and civil engineering engineers. The use of prestressed CFRP reinforcement technology has become widespread in practice. Under actual operating conditions, the long-term mechanical properties of prestressed CFRP-reinforced concrete (RC) components have also attracted more and more attention. RC bridge reinforcement structure is used for natural exposure environment. During operation, due to concrete shrinkage and creep, performance degradation, and CFRP-concrete interface mechanical performance degradation, the prestress is lost, resulting in the long-term mechanical performance of prestressed CFRP-reinforced RC structure deterioration. Under this engineering background, this article takes prestressed CFRP-strengthened RC beams as the research object to investigate the long-term mechanical properties under natural exposure environments and provides a scientific basis for the application of prestressed CFRP technology in engineering practice. The wide application of concrete structure in engineering has caused its safety and stability to attract much attention. At present, various countries in the world have carried out different degrees of research topics on material reinforcement methods. The research in this area in our country is still in its infancy, and the research has been less successful. The application in actual engineering has not been widely developed, and there are still many problems to be solved. Theoretical research needs to be further strengthened and effectively applied to actual projects. Starting from the reliability theory, feasible regulations and standards are formulated to guide practice.

5. Conclusion

Aiming at a typical long bridge, this paper evaluates the early warning of the damage of the long bridge through a machine learning algorithm designed for material parameter selection and test beam conditions. The research analysis used mainly includes three types of analysis: mechanical analysis corresponding to rigid body displacement at the structural level, mechanical analysis corresponding to rigid body displacement at the component level, and mechanical analysis corresponding to internal changes. Whether it is the CFRP-concrete interface or the BFRP-concrete interface, their adhesion properties deteriorate with the prolonged exposure time. By analyzing the engineering geomechanical characteristics under the influence of the overall heating and cooling of the main girder, the spatial direction of the D bridge is opposite and the value is the same. In the static structure, the vertical settlement of the column will not produce any secondary stress, and the static structure bridge only undergoes spatial deformation without any internal force. In statically unsafe structures affected by the uneven vertical settlement of the columns, the bridge will not only produce spatial deformation but also generate secondary internal forces. From the vertical displacement of the D bridge node under its own weight and dead load, it can be seen that the amplitude of the downward vertical displacement of the inner node is greater than the amplitude and the vertical amplitude of the outer node. The upward displacement is less than the displacement of the outer nodes, so all spans of the bridge are equal to inward bending. The wide application of concrete structure in engineering has caused its safety and stability to attract much attention. At present, various countries in the world have carried out different degrees of research topics on material reinforcement methods. The research in this area in our country is still in its infancy, and the research has been less successful. The application in actual engineering has not been widely developed, and there are still many problems to be solved. Theoretical research needs to be further strengthened and effectively applied to actual projects. Starting from the reliability theory, feasible regulations and standards are formulated to guide practice.

Data Availability

The dataset can be accessed upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

References

[1] I. S. Asaad and M. F. Omer, "Facies characterization and depositional environment of baluti formation (late triassic) from selected sections in the kurdistan region, northern Iraq," *Arabian Journal of Geosciences*, vol. 13, no. 23, p. 1253, 2020.

[2] I. Aarab, M. Derqaoui, A. Abidi et al., "Direct flotation of low-grade Moroccan phosphate ores: a preliminary micro-flotation study to develop new beneficiation routes," *Arabian Journal of Geosciences*, vol. 13, no. 23, p. 1252, 2020.

[3] H. Atapour and M. Moosavi, "Some effects of shearing velocity on the shear stress-deformation behaviour of hard-soft artificial material interfaces," *Geotechnical & Geological Engineering*, vol. 31, no. 5, pp. 1603–1615, 2013.

[4] M. Bahaaddini, P. C. Hagan, R. Mitra, and B. K. Hebblewhite, "Parametric study of smooth joint parameters on the shear behaviour of rock joints," *Rock Mechanics and Rock Engineering*, vol. 48, no. 3, pp. 923–940, 2015.

[5] M. Bahaaddini, P. C. Hagan, R. Mitra, and M. H. Khosravi, "Experimental and numerical study of asperity degradation in the direct shear test," *Engineering Geology*, vol. 204, pp. 41–52, 2016.

[6] J. Cheng, X. W. Xu, W. J. Gan, W. T. Ma, W. T. Chen, and Y. Zhang, "Block model and dynamic implication from the earthquake activities and crustal motion in the southeastern margin of Tibetan Plateau," *Chinese Journal of Geophysics*, vol. 55, no. 4, pp. 1198–1212, 2012.

[7] Y. Cui, "Effect of joint type on the shear behavior of synthetic rock," *Bulletin of Engineering Geology and the Environment*, vol. 78, no. 5, pp. 3395–3412, 2019.

[8] F. Duriyapong and K. Nakhapakorn, "Coastal vulnerability assessment: a case study of Samut Sakhon coastal zone," *Songklanakarin Journal of Science and Technology*, vol. 33, no. 4, pp. 469–476, 2011.

[9] C. Gehle and H. K. Kutter, "Breakage and shear behaviour of intermittent rock joints," *International Journal of Rock Mechanics and Mining Sciences*, vol. 40, no. 5, pp. 687–700, 2003.

[10] F. Hallouz, M. Meddi, G. Mahé, S. Alirahmani, and A. Keddar, "Modeling of discharge and sediment transport through the SWAT model in the basin of Harraza (Northwest of Algeria)," *Water Science*, vol. 32, no. 1, pp. 79–88, 2018.

[11] Z. M. He, Z. Y. Xiong, Q. G. Hu, and M. Yang, "Analytical and numerical solutions for shear mechanical behaviors of structural plane," *Journal of Central South University*, vol. 21, no. 7, pp. 2944–2949, 2014.

[12] H. Jahanian and M. H. Sadaghiani, "Experimental study on the shear strength of sandy clay infilled regular rough rock joints," *Rock Mechanics and Rock Engineering*, vol. 48, no. 3, pp. 907–922, 2015.

[13] E. Z. Lajtai, "Shear strength of weakness planes in rock," *International Journal of Rock Mechanics and Mining Science & Geomechanics Abstracts*, vol. 6, no. 5, pp. 499–515, 1969.

[14] E. Z. Lajtai, "Strength of discontinuous rocks in direct shear," *Géotechnique*, vol. 19, no. 2, pp. 218–233, 1969.

[15] M. A. Saleh, "Mangrove vegetation on abu minqar island of the red sea," *International Journal of Remote Sensing*, vol. 28, no. 23, pp. 5191–5194, 2007.

[16] T. S. Sankari, A. R. Chandramouli, K. Gokul, S. S. M. Surya, and J. Saravanavel, "Coastal vulnerability mapping using geospatial technologies in Cuddalore-Pichavaram coastal tract, Tamil Nadu, India," *Aquatic Procedia*, vol. 4, pp. 412–418, 2015.

[17] F. G. Xue, P. S. Julian, and M. H. Chris, "Direct shear test of sandstone-concrete joints," *International Journal of Geomechanics*, vol. 3, no. 1, pp. 21–33, 2003.

[18] J. B. Zhu, H. Li, and J. H. Deng, "A one-dimensional elastoplastic model for capturing the nonlinear shear behaviour of joints with triangular asperities based on direct shear tests," *Rock Mechanics and Rock Engineering*, vol. 52, no. 6, pp. 1671–1687, 2019.

Research Article

Straw Mulching with Minimum Tillage Is the Best Method Suitable for Straw Application under Mechanical Grain Harvesting

En Lei,[1,2] ChaoBo Wang,[2] Wen Xue Li,[2] Yue Dong Wang,[3] Yong Bing Yang,[3] Hua Bin Zheng,[1] and Qi Yuan Tang [iD][1]

[1]*College of Agronomy, Hunan Agricultural University, Hunan, Changsha 410128, China*
[2]*College of Biological and Agricultural Sciences, Honghe University, Mengzi, Yunnan, 661100, China*
[3]*Agricultural Maechinery Research Institute of Yunnan Honghe Hani and Yi Autonomous Prefecture, Mengzi, Yunnan 661100, China*

Correspondence should be addressed to Qi Yuan Tang; tlf3300@hunau.edu.cn

Mechanical grain harvesting is a crop production development direction. However, the residue management methods suitable for mechanical grain harvesting have been not established. In order to study the effect of residue management modes on maize yield formation and explore the best residue management methods for mechanical grain harvesting, four crop field surveys were carried out in Southwest China. Crops were mechanically harvested, and the residues were shredded and returned to the field using various straw application methods including straw deep burial with plowing (SDBP), straw shallow burial with rotary tillage (SSBRT), and straw mulching with minimum tillage (SMMT). The first-season rape residues were returned to the field, and the second-season maize yield under SDBP and SSBRT was significantly higher than that under SMMT. However, with the increase in rounds of residue application, compared with SDBP and SSBRT, SMMT continuously increased the soil moisture content in the 0–30 cm soil layer at the early stage of maize growth, increased the soil alkaline-hydrolyzed nitrogen content in the 0–20 cm and 40–60 cm layers, and reduced the soil compaction under 40 cm layer, which were more conducive to the root system growth. Maize yield with the SMMT increased by 5.4% compared with that of the previous season, while the yields with SDBP and SSBRT decreased by 16.7% and 12.7%, respectively, compared with those of the previous season. In conclusion, it is recommended to employ the SMMT method during crop mechanical harvesting, which is of great significance to improve soil quality and increase maize grain yield.

1. Introduction

Previous studies have pointed out that the long-term application of conventional farming methods and removing crop residues have severely disturbed and destroyed the natural structure of the soil, resulting to a continuous decline in the soil organic matter content, which restricts the increase of crop grain yield [1]. As a potential bioenergy, the crop residues after harvesting can be returned to the fields to combine with soil tillage. Straw application can improve soil tilth and fertility, maintain soil productivity, and increase crop grain yield [2]. Straw returns could continuously increase soil organic carbon in an interannual rotation cropping system [3].

Some studies have shown that continuous straw application significantly reduced the soil bulk density from the 4th year [4], which greatly improved the stability of soil aggregates [5]. Straw application can regulate soil moisture content and temperature, affect the vertical distribution of soil available nitrogen [6], and reduce ammonia volatilization from soils [7]. With straw application, the alkaline-hydrolyzed nitrogen (AH-N) content of the soil increased by 6%–14% and 8%–34%, respectively, compared to that of soils without crop residues [8]. After straw application for six consecutive years, the soil organic carbon and available nitrogen contents in each soil aggregate-size class increased by 27% and 12%, respectively [2], compared with those of

soils without crop residues. In addition, straw application significantly reduced the soil penetration resistance, leading to a 1.4-fold increase in water infiltration rate of the soil [9]. This effect is most significant in areas where the soils are of poor fertility [10]. Straw application can also increase the organic matter content and the accumulation of humus components in the soil, improve the structure of soil humus [11], increase the abundance of soil microbiota [5], and facilitate the activities of soil dehydrogenase and phosphatase [4, 12]. Straw returns improved the fertilizer effects on soil bulk density, soil porosity, maize root length, root surface area, root volume, and yield by 3.99%–7.27%, 3.89%–7.40%, 1.35%–71.01%, 19.16%–42.45%, 10.49%–22.73%, and 4.43%–7.05%, respectively [13]. Therefore, straw application is conducive to increase crop grain yield, nitrogen use efficiency [14], and economic benefit [15].

However, the undecomposed crop residues will hinder the emergence [16] and growth [17] of the crop in the coming season. Therefore, in areas with multiple cropping systems, additional nitrogen fertilizers are applied every year after straw application to promote the decomposition of crop residues [18, 19]. Straw application promotes the accumulation of nutrients in the surface soil layers; however, its effect on increasing crop grain yield is related to factors such as soil types [20] and soil tillage methods [21]. Soil tillage methods significantly influence the accumulation and transformation of nutrients in the soil [22]. A previous study has demonstrated that the plowing tillage technology during straw application can increase the organic carbon and total nitrogen contents of the surface soil layers, thereby increasing the maize grain yield and water use efficiency [23]. Other studies have shown that the carbon content of organic matter, humus, humic acid, and fulvic acid [11], as well as the total nitrogen release in the soil are significantly higher using the shallow burial method than those with deep burial treatment [24]. Straw shallow burial can significantly increase the degradation rate of straws, which may be related to the increase in the catabolic versatility of the soil microbiota [25] and the decrease in the ratio of G^+/G^- bacteria [26]. Compared with the straw deep burial, the comprehensive indices of soil nutrients can reach their maximums with the shallow burial method [27]. However, soil moisture content and temperature tend to decrease significantly whether using the deep burial with plowing tillage method or shallow burial method, which lead to a decrease in the emergence rate of the crop in the coming season [28].

In rain-fed arid areas, straw mulching with minimum tillage can significantly increase the organic carbon content [22], moisture content, and permeability [29] of the 0–100 cm soil layer, thereby increasing the crop grain yield and soil quality. However, in areas with irrigation systems, this method reduced the soil temperature and thus did not contribute to crop grain yield increase [30]. In other words, the effects of straw application with different soil tillage methods are significantly different among different regions.

In recent years, mechanical grain harvesting has continued to develop in the main maize-producing areas in China. Based on the limitation of harvesting machinery, when grain is harvested by machinery, the straw is crushed by the harvester and returned to the field. However, there was little research about the straw application under mechanical grain harvesting. The results of previous studies were mostly obtained under the condition of manual harvesting. Compared with the traditional way of artificial harvesting and straw recycling, crop residues are simultaneously shredded and returned to the field during mechanical grain harvesting, which poses new demand for studying straw application methods. This is also the challenge and superiority of the analysis program compared with previous studies.

In this study, we conducted field experiments in the red soil area of the Yunnan Plateau, which is the maize and rape double-cropping system production region in Southwest China. We investigated the influences of the straw application in combination with different soil tillage methods on soil tilth and maize grain yield and determined the most appropriate straw application method under mechanical grain harvesting. The findings from this study provide a theoretical basis and technical approach for improving the soil quality and increasing the maize grain yield.

2. Materials and Methods

2.1. Experimental Field. The field test was conducted between 2016 and 2018 in Jiale Village at the Town of Jinma, Honghe Hani and Yi Autonomous Prefecture, Yunnan Province, China (24°46′N, 103°30′E). The region is located at a typical production area of the maize and rape double-cropping system. The field has an altitude of 1,800 m, a subtropical monsoon climate, an annual average temperature of 15.2°C, and an annual precipitation of 979.7 mm. The soil type is red soil, with adoption of rain-fed farming and rotary tillage practices.

2.2. Straw Application Methods. The split plot experimentation scheme was adopted, maize cultivar was the main plot factor, and straw application was the subplot factor. Jinyu 99 (JY99), an early-maturing maize cultivar, and Baoyu 9 (BY9), a late-maturing maize cultivar, were used for cultivation. All mechanically harvested straw was shredded and applied to the field. Three straw application methods were used: straw deep burial with plowing (SDBP), straw shallow burial with rotary tillage (SSBRT), and straw mulching with minimum tillage (SMMT). Each experimental treatment was repeated in three blocks, and the experimental field was divided into 18 treatment plots (50 m × 2.4 m). In each plot, four rows of maize were planted at equal distances, with a total plot area of 2,200 m². Maize was cultivated alternatively with rape plants (Table 1).

Mechanical grain harvesting was conducted using a harvester (Kubota 4LZ-2.5, PRO688Q) with a semifeed header (Jiajiale 4 YG-4A). Plowing was conducted using a 1LH-438 moldboard plow towed with a tractor (LX1204). Rotary tillage was conducted with a rotary tiller towed with the same type of tractor. The detailed experimental procedure is summarized in Table 1.

TABLE 1: The experimental procedure and straw application methods.

Season and crop species	Cultivation method	Straw application method	Cumulative number of straw applications (CNSA)
Winter-seeded rape plant in 2016	Sowed on 10-25-2016. Mechanical furrowing was conducted at a depth of 8–12 cm. Then, a thin layer of soil was applied. Thinning was conducted at V2 and final thinning at V4 during intertillage and weeding. Fertilizers were applied at 276 kg/ha for N, 120 kg/ha for P_2O_5, and 150 kg/ha for K_2O. Nitrogen fertilizer was given at sowing, germination, and inflorescence emergence at a ratio of 3 : 3:4. Phosphorus and potassium fertilizers were applied at sowing. Top dressings were applied to the furrows between two rows. Crop maintenance was conducted according to the standard practice for high-yield crops.	Mechanical grain harvesting was carried out, and straw was shredded on 4-20-2017. SDBP: straw was buried at 30–40 cm during plowing. Then, rotary tillage was carried out. SSBRT: straw was buried at 15–20 cm during rotary tillage. SMMT: straw was directly applied to the surface of the soil.	Cumulative zero rounds of straw application before the rape plant was sown in 2016 (0 CNSA)
Summer-seeded maize in 2017	Sowed on 4-28-2017 at a density of 67,500 plants/ha. Mechanical furrowing was conducted at a depth of 10–15 cm. The fertilizer was applied at 180 kg/ha for N, 112.5 kg/ha for P_2O_5, and 135 kg/ha for K_2O. Nitrogen and potassium fertilizers were given at sowing and inflorescence emergence at a ratio of 5 : 5. Phosphorus fertilizer was applied at sowing. Fertilizers applied during the period of inflorescence emergence were applied into furrows, and top dressings were applied into deep holes punched between every two crops. Crop maintenance was conducted according to the standard practice for high-yield crops, and each plot was treated the same.	Mechanical grain harvesting was carried out, and straw was shredded on 10-15-2017. Straw application was carried out as per the three methods of the winter-seeded rape plant in 2016.	Cumulative one round of straw application before maize was sown in 2017 (1 CNSA)
Winter-seeded rape plant in 2017	Sowed on 10-28-2017. The same method as winter-seeded rape plant in 2016.	Mechanical grain harvesting was carried out, and straw was shredded on 4-25-2018. Straw application was carried out as per the three methods of winter-seeded rape plant in 2016.	Cumulative two rounds of straw application before the rape plant was sown in 2017 (2 CNSA)
Summer-seeded maize in 2018	Sowed on 5-1-2018. The same method as summer-seeded maize in 2017.	Mechanical grain harvesting was carried out on 10-22-2018.	Cumulative three rounds of straw application before maize was sown in 2018 (3 CNSA)

2.3. Soil Samples and Collection.
Soil profiles ($120 \times 60 \times 60$ cm) and root samples ($45 \times 30 \times 60$ cm) were manually collected in the maize ear initiation stage (V12) and silking stage (R1) in representative areas in each plot (Figure 1). Soil penetration resistance and water content were tested outside of the four corners of the profile collection area. The alkaline-hydrolyzed nitrogen (AH-N) content and root volume were tested at each depth range (0–10, 10–20, 20–30, 30–40, 40–50, and 50–60 cm).

2.4. Soil Penetration Resistance, Water Content, and AH-N Content.
Soil penetration resistance was tested with a penetrometer (TJSD-750), and water content was tested with a soil moisture sensor (TZS-1K). Then, a soil drill was used to collect soil samples. Four sampling points in the same depth range were mixed. Roots and shedding were discarded. Each sample was sieved with a 0.18 mm filter. The AH-N content was determined with the alkaline hydrolysis diffusion method.

2.5. Root Volume.
A stainless-steel blade was used to collect the three sub-blocks of root samples. Samples were transferred to nylon mesh bags before the soil, and other impurities were washed off. All visible maize roots were collected into Ziplock bags and scanned using a high-resolution scanner (Epson Perfection V700). Images were analyzed with the plant root measurement and analysis system (WinRHIZO Pro-2016). The contour map of the root system was drawn by a type of drawing software (Surfer 8.0). The discrete data of root volume were selected, and the interpolation method in Surfer 8.0 was used to grid the data, the regular grid file was obtained, and then the contour map is drawn.

2.6. Determination of Leaf Area Index, Dry Matter Accumulation, and Chlorophyll Content.
For each plot, 4 plant samples were collected in the V12 stage and R1 stage and 10 were collected in the physiological maturity (R6) stage. Organs were separated in each plant. For leaf area indices,
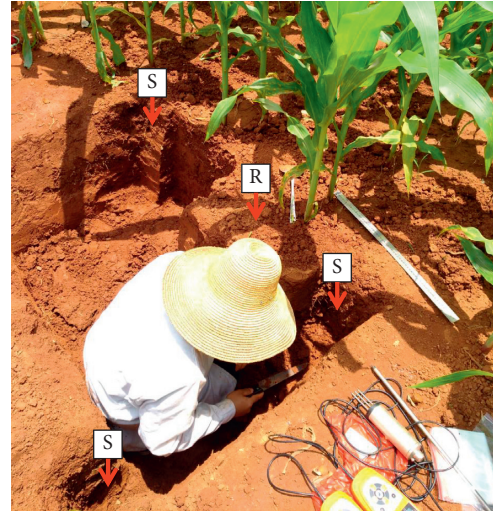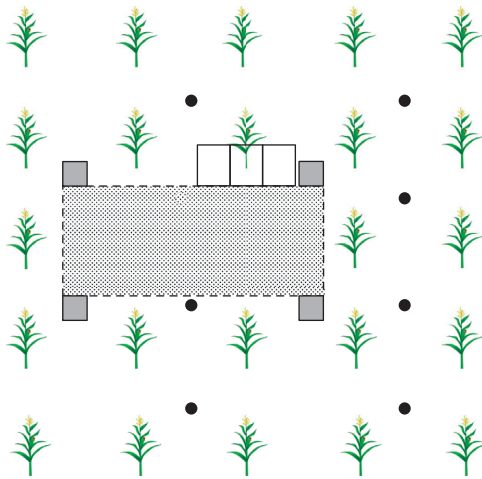
FIGURE 1: Soil samples and collection. maize plant. locations of top-dressing application. soil profile collection region ($120 \times 60 \times 60$ cm). sampling points for soil penetration resistance and water content test. maize root sampling area ($45 \times 30 \times 60$ cm), including three sub-blocks at the plant, left to the plant, and right to the plant. Tests on soil and root samples were carried out at 0–10, 10–20, 20–30, 30–40, 40–50, and 50–60 cm in depth. The capital R stands for maize root sampling area and the S stands for soil sampling points.

the lengths and widths of all leaves were measured and the leaf area index was calculated as length × width × coefficient, where coefficient = 0.50 when the leaves were not fully unrolled and 0.75 otherwise. For dry matter accumulation, each sample was baked at 105°C and then dried at 75°C till the weight reached constant before weighing. For samples collected at the physiological maturity stage, biomasses and harvest indices were calculated based on the measured dry matter accumulation values. SPAD values were measured with a chlorophyll meter (SPAD-502). In the V12 stage, the top fully unrolled leaves were measured. At R1, ear leaves were measured. Measurements were taken at a spot slightly deviant from the center of the leaves. Fifteen consecutive plants were measured in each plot.

2.7. Grain Yield. The total weight of the harvested maize grains in each plot was measured (total weight). Then, 3 random samples (>1 kg) were collected and weighed (gross weight). Impurities from these samples were discarded before the samples were weighed again (net weight). Impurity ratio was calculated: (gross weight–net weight) ÷ gross weight.

3 random grain samples (100 kernels) were collected and weighed (wet weight). Then, grains were dried at 75°C till weight became constant before weighing (dry weight). The grain water content was calculated as (wet weight–dry weight) ÷ net weight. The yield assuming a 14% water content was calculated as total weight × (1 – impurity ratio) × (1 – water content) ÷ (1 – 14%).

2.8. Statistical Analysis. Data were organized with Microsoft Excel 2016. The data-processing system of SPSS statistics 8.0 software was used for analysis of variance. Analysis of variance (ANOVA) was followed by LSD tests at $P = 0.05$ where appropriate. The least significant difference was adopted for the multiple comparisons in the analysis of variance of experimental treatment, and the significance level was 0.05. The LSD algorithm between different levels of different varieties and straw application is $LSD_{0.05} = t_{0.05(df)} \sqrt{MS_e/ar}$, where $t_{0.05(df)}$ is the critical t-value of factor degree freedom under the F test, MS_e is the error mean square, a is the factor level number, and r is the experimental treatment repetition times. The LSD algorithm of variety × straw application interaction is $LSD_{0.05} = t_{0.05(df_e)} \sqrt{2MS_e/r}$, where df_e is the test error degree freedom.

3. Results

3.1. Yield and Harvest Index. The maize yield in each season was significantly affected by the straw application methods (Table 2). After the first straw application (from the rape plant harvest, 1 CNSA), the maize yields using the SDBP and SSBRT were higher than that using the SMMT by 26.8% and 17.7%, respectively. However, after three seasons of straw application (rape, maize, and rape, 3 CNSA), the yields and biomasses using the SDBP, SSBRT, and SMMT were not significantly different ($P > 0.05$). Furthermore, compared to the first maize harvest, with the increase in rounds of mechanical harvesting and straw application, the yields with the SDBP and SSBRT were reduced by 16.7% and 12.7%, respectively, whereas the yield using the SMMT was increased by 5.4%. The increase in yield was mainly a result of the increase in biomass, i.e., the harvest indices were barely changed. The yield (average of 2017 and 2018 harvests) of late-maturing cultivar BY9 was significantly higher than the early-maturing cultivar JY99 (8.4% and 7.7%, respectively), mainly due to the higher biomasses and harvest indices of BY9.

3.2. Dry Matter Accumulation before Silking, Leaf Area Index, and Leaf Chlorophyll Content. Our results indicate that straw application methods strongly influenced dry matter

TABLE 2: Maize yields and harvest indices under different practices of crop straw application.

CNSA	Treatments		Kernel yield (t/ha)	Biomass (t/ha)	Harvest index (%)	Theoretical yield (t/ha)
	Straw returns	SDBP	11.46 ± 0.50a	21.12 ± 0.86a	52.96 ± 0.68a	12.76 ± 0.61a
		SSBRT	10.64 ± 0.47b	20.07 ± 0.82b	53.65 ± 1.43a	12.30 ± 0.62a
		SMMT	9.04 ± 0.27c	18.11 ± 0.74c	54.31 ± 1.02a	11.22 ± 0.56b
	Cultivars	BY9	10.80 ± 0.43a	20.56 ± 0.87a	54.72 ± 1.12a	12.81 ± 0.66a
		JY99	9.96 ± 0.39b	18.97 ± 0.74a	52.55 ± 0.96b	11.37 ± 0.53b
1		BY9×SDBP	11.53 ± 0.56a	21.58 ± 0.87a	53.03 ± 0.95bc	13.05 ± 0.65a
		BY9×SSBRT	11.16 ± 0.49a	21.47 ± 0.87a	54.72 ± 1.64ab	13.40 ± 0.76a
	Interaction	BY9×SMMT	9.72 ± 0.25b	18.62 ± 0.87bc	56.41 ± 0.77a	11.97 ± 0.57bc
		JY99×SDBP	11.39 ± 0.43a	20.66 ± 0.85ab	52.89 ± 0.40bc	12.46 ± 0.56ab
		JY99×SSBRT	10.12 ± 0.44b	18.67 ± 0.76bc	52.57 ± 1.21c	11.19 ± 0.47c
		JY99×SMMT	8.36 ± 0.29c	17.59 ± 0.60c	52.20 ± 1.27c	10.47 ± 0.55c
	Straw returns	SDBP	9.55 ± 0.29a	17.38 ± 0.56b	54.11 ± 1.28a	10.74 ± 0.39b
		SSBRT	9.29 ± 0.35a	16.79 ± 0.72b	52.64 ± 1.24b	10.09 ± 0.60c
		SMMT	9.53 ± 0.28a	18.50 ± 0.60a	54.06 ± 1.08a	11.43 ± 0.31a
	Cultivars	BY9	9.80 ± 0.32a	18.58 ± 0.66a	55.41 ± 1.25a	11.74 ± 0.34a
		JY99	9.10 ± 0.28b	16.53 ± 0.59b	51.79 ± 1.14b	9.76 ± 0.52b
3		BY9×SDBP	9.83 ± 0.30ab	18.25 ± 0.50b	55.78 ± 1.17a	11.60 ± 0.20b
		BY9×SSBRT	9.52 ± 0.36ab	17.47 ± 0.78bc	54.59 ± 1.28a	10.87 ± 0.63bc
	Interaction	BY9×SMMT	10.06 ± 0.30a	20.03 ± 0.71a	55.85 ± 1.31a	12.75 ± 0.20a
		JY99×SDBP	9.26 ± 0.27bc	16.50 ± 0.61bc	52.44 ± 1.39ab	9.87 ± 0.58d
		JY99×SSBRT	9.05 ± 0.33c	16.11 ± 0.66c	50.68 ± 1.20b	9.31 ± 0.56d
		JY99×SMMT	9.00 ± 0.25c	16.97 ± 0.49bc	52.26 ± 0.84ab	10.11 ± 0.42 cd

Groups labeled with different letters were significantly different ($P < 0.05$). The ± sign is followed by the standard deviation. CNSA: cumulative number of straw returns.

accumulation before silking, leaf area index, and SPAD value (Table 3). After the first straw application, dry matter accumulation, leaf area index, and SPAD during V12 and R1 stages were significantly higher while using the SDBP than with SSBRT and SMMT, whereas after three consecutive straw applications, the SPAD value during the V12 stage and all three index values during the R1 stage were higher with SMMT. Furthermore, the three index values had a trend of reduction after three rounds of straw application with the SDBP and SSBRT, whereas for the SMMT, these numbers either increased or did not significantly change.

By comparing between the two cultivars, we found that straw application methods only affected SPAD values during the first season, whereas dry matter accumulation before silking and leaf area index were not significantly affected. Considering BY9 had a higher yield, we deduce that the dry matter accumulation after silking, compared to before silking, contributes more to the high yield of the late-maturing cultivar.

3.3. Volume and Distribution of the Root System. Figure 2 illustrates the root volume and distribution in each straw application method (average of the two cultivars). We found that, after one straw application, the SDBP and SSBRT methods showed higher average root volume in 0−60 cm depth range compared with the SMMT. In contrast, after three consecutive straw applications, the root volumes using the SDBP and SSBRT were not notably higher than using the SSMT. Particularly, the average root volume of the 0−40 depth range with the SDBP was less than that using the SMMT by 29.6% at V12 and 7.5% at R1, after three consecutive straw applications.

In addition, we found that even after one round of straw application, the root volume in the 40−60 cm soil layer was not significantly different among the SMMT, SDBP, and SSBRT methods. As the root volume at deeper soil layers is positively correlated with yield, these data suggest that the immediate increase of root volumes with the SDBP and SSBRT may not have a strong benefit on the increase of yield. By comparing root volumes between one straw application and three straw applications, we found that in the R1 stage, 0−60 cm root volume with the SMMT was increased by 18.2%, whereas that using SDBP was decreased by 28.0% and that by SSBRT was not significantly changed. These results suggest that the SMMT may have a delayed but long-term advantage in enhancing maize root volume.

3.4. Soil Penetration Resistance. Figure 3 illustrates soil penetration resistance at different depths during V12 and R1 stages with different straw application time periods. By comparing SDBP and SMMT, we found that soil penetration resistance with the SDBP was lower than that with SMMT at 0−30 cm range by an average of 32.8% (on average between 1 CNSA and 3 CNSA; the same below). However, soil penetration resistance was higher with the SDBP at 40−60 cm range by over 60.0%. We also compared SSBRT and SMMT, and we found that while soil penetration resistance at 0−20 cm was not significantly different, the SSBRT method had a 24.8% higher soil penetration resistance than SMMT method at 20−60 cm. Furthermore, we also found that the deep-layer soil penetration resistance (50−60 cm) increased after three straw applications in both the SSBRT (791.9 KPa) and SDBP (427.3 KPa) methods.

TABLE 3: Dry matter accumulation, leaf area index, and SPAD value of maize under different practices of crop straw application.

CNSA	Treatments	Maize ear initiation stage (V12)			Silking stage (R1)		
		Dry matter weight (t/ha)	Leaf area index (m/m)	Leaf SPAD values	Dry matter weight (t/ha)	Leaf area index (m/m)	Leaf SPAD values
1	Straw returns						
	SDBP	0.39 ± 0.04a	4.28 ± 0.14a	51.88 ± 2.11a	9.67 ± 0.33a	5.19 ± 0.18a	56.40 ± 0.83a
	SSBRT	0.26 ± 0.02b	3.15 ± 0.11b	47.37 ± 1.26b	6.71 ± 0.23b	4.37 ± 0.13b	53.15 ± 1.28b
	SMMT	0.25 ± 0.01b	3.12 ± 0.18b	47.53 ± 1.12b	6.34 ± 0.21c	4.29 ± 0.18b	47.38 ± 1.10c
	Cultivars						
	BY9	0.29 ± 0.02a	3.42 ± 0.11a	50.05 ± 1.96a	7.33 ± 0.25a	4.58 ± 0.17a	51.51 ± 1.12b
	JY99	0.30 ± 0.02a	3.61 ± 0.18a	47.79 ± 1.03b	7.80 ± 0.27a	4.64 ± 0.15a	53.11 ± 1.01a
	Interaction						
	BY9×SDBP	0.35 ± 0.03b	3.92 ± 0.09b	53.43 ± 2.01a	9.57 ± 0.30a	5.25 ± 0.16a	55.73 ± 0.75ab
	BY9×SSBRT	0.26 ± 0.02c	3.14 ± 0.07c	49.39 ± 2.16b	6.37 ± 0.24 cd	4.17 ± 0.15c	52.47 ± 1.48c
	BY9×SMMT	0.27 ± 0.01c	3.21 ± 0.16c	47.34 ± 1.70bc	6.06 ± 0.20d	4.33 ± 0.20bc	46.33 ± 1.14d
	JY99×SDBP	0.42 ± 0.04a	4.64 ± 0.19a	50.32 ± 2.20ab	9.76 ± 0.36a	5.12 ± 0.20a	57.07 ± 0.91a
	JY99×SSBRT	0.26 ± 0.01c	3.15 ± 0.15c	45.35 ± 0.35c	7.04 ± 0.23b	4.57 ± 0.10b	53.83 ± 1.07bc
	JY99×SMMT	0.23 ± 0.02c	3.03 ± 0.20c	47.71 ± 0.54bc	6.61 ± 0.22bc	4.24 ± 0.15c	48.43 ± 1.05d
3	Straw returns						
	SDBP	0.17 ± 0.01a	2.53 ± 0.15a	42.78 ± 1.00b	5.12 ± 0.18ab	4.42 ± 0.20b	53.43 ± 1.07b
	SSBRT	0.18 ± 0.01a	2.56 ± 0.15a	44.27 ± 0.87ab	4.80 ± 0.17b	4.24 ± 0.17c	53.67 ± 1.04ab
	SMMT	0.17 ± 0.01a	2.45 ± 0.11a	45.95 ± 1.10a	5.30 ± 0.12a	4.61 ± 0.19a	54.65 ± 0.83a
	Cultivars						
	BY9	0.18 ± 0.01a	2.60 ± 0.11a	44.63 ± 0.95a	5.20 ± 0.17a	4.49 ± 0.18a	52.94 ± 0.87b
	JY99	0.16 ± 0.01a	2.43 ± 0.16b	44.03 ± 1.03a	4.94 ± 0.13a	4.35 ± 0.18a	54.89 ± 1.09a
	Interaction						
	BY9×SDBP	0.19 ± 0.01a	2.86 ± 0.13a	43.33 ± 0.74 cd	5.53 ± 0.20a	4.76 ± 0.22a	52.13 ± 0.67d
	BY9×SSBRT	0.17 ± 0.01bc	2.45 ± 0.12bc	44.33 ± 0.92bc	4.65 ± 0.15c	4.12 ± 0.17 cd	53.10 ± 1.23 cd
	BY9×SMMT	0.17 ± 0.01bc	2.48 ± 0.08bc	46.23 ± 1.19a	5.41 ± 0.17a	4.60 ± 0.16ab	53.60 ± 0.71bc
	JY99×SDBP	0.14 ± 0.01d	2.20 ± 0.16c	42.23 ± 1.26d	4.70 ± 0.15c	4.07 ± 0.17d	54.73 ± 1.46ab
	JY99×SSBRT	0.18 ± 0.00ab	2.66 ± 0.18ab	44.20 ± 0.82bc	4.95 ± 0.18bc	4.36 ± 0.17bc	54.23 ± 0.85ab
	JY99×SMMT	0.16 ± 0.01c	2.42 ± 0.14bc	45.67 ± 1.01ab	5.18 ± 0.07ab	4.62 ± 0.21a	55.70 ± 0.95a

Groups labeled with different letters were significantly different ($P < 0.05$). The ± sign is followed by the standard deviation. CNSA: cumulative number of straw returns.

Together, these results suggest that, although SDBP and SSBRT methods can effectively decrease soil penetration resistance in the shallow and medium soil layers, they can also increase the deep-layer soil penetration resistance. In contrast, SMMT cannot effectively decrease soil penetration resistance in the shallow layer, but it will not have a serious adverse impact on the deeper soil structure.

3.5. Soil Water Content. We found that the straw application methods significantly affected the soil water content, as shown in Figure 4. During the V12 stage (averaged between 1 CNSA and 3 CNSA; the same below), the SMMT method produced higher water content in the 0−30 cm soil layer than SDBP and SSBRT methods (by 15.2% and 18.0%, respectively). Similar differences were observed in the 30−60 cm layer as well (11.3% and 7.3%, respectively). Importantly, in the superficial soil layer (0−10 cm), which is essential for maize germination, the SMMT still produced the highest water content, followed by SDBP (21.3%) and SSBRT (19.0%). During the R1 stage, the water content by the SMMT remained significantly higher than that by the SDBP (by 19.4% in the 0−30 cm layer and 6.9% in the 30−60 cm layer), while not significantly different from that by SSBRT.

3.6. Alkaline-Hydrolyzed Nitrogen Content. Figure 5 shows that straw application methods strongly impacted the AH-N content in the soil. In the V12 stage, the SMMT showed significantly more AH-N content in the 0−20 cm and 40−60 cm layers, compared with SDBP and SSBRT (15.9%

and 22.7% for 0−20 cm and 7.9% and 48.6% for 40−60 cm, respectively, calculated with the averages between 1 CNSA and 3 CNSA). In comparison, the SDBP and the SSBRT only showed higher content in the 20−40 cm layer (21.5% and 16.7% higher than SMMT, respectively). During the R1 stage, in the 0−20 cm layer, AH-N content by the SMMT was significantly higher than that by the SDBP method (by 9.7%), but it was not significantly different from that by the SSBRT; in the 20−40 cm layer, AH-N content with the SMMT was higher than that with the SSBRT but lower than with SDBP (by 12.8% and 5.5%, respectively); in the 40−60 cm layer, AH-N content with the SMMT was significantly higher than with SDBP and SSBRT (by 73.8% and 71.9%, respectively).

In addition, by comparing the data from one and three straw applications, we found that during V12, AH-N content with the SMMT in the 0−20 cm and 40−60 cm layers increased by 3.1% and 16.9%, respectively, and the number decreased in the 20−40 cm layer by 12.9%, but the overall (0−60 cm) AH-N content increased by 2.6%. These results demonstrate that SMMT is overall advantageous in the continued improvement of AH-N content, particularly in the deep soil layers.

4. Discussion

The soil tillage and straw burial depth during straw application are determinants affecting soil tilth and maize grain yield [31]. Previous studies have demonstrated that compared with straw mulching with minimum tillage, straw deep burial with plowing and straw shallow burial with rotary tillage can improve the soil structure and root growth
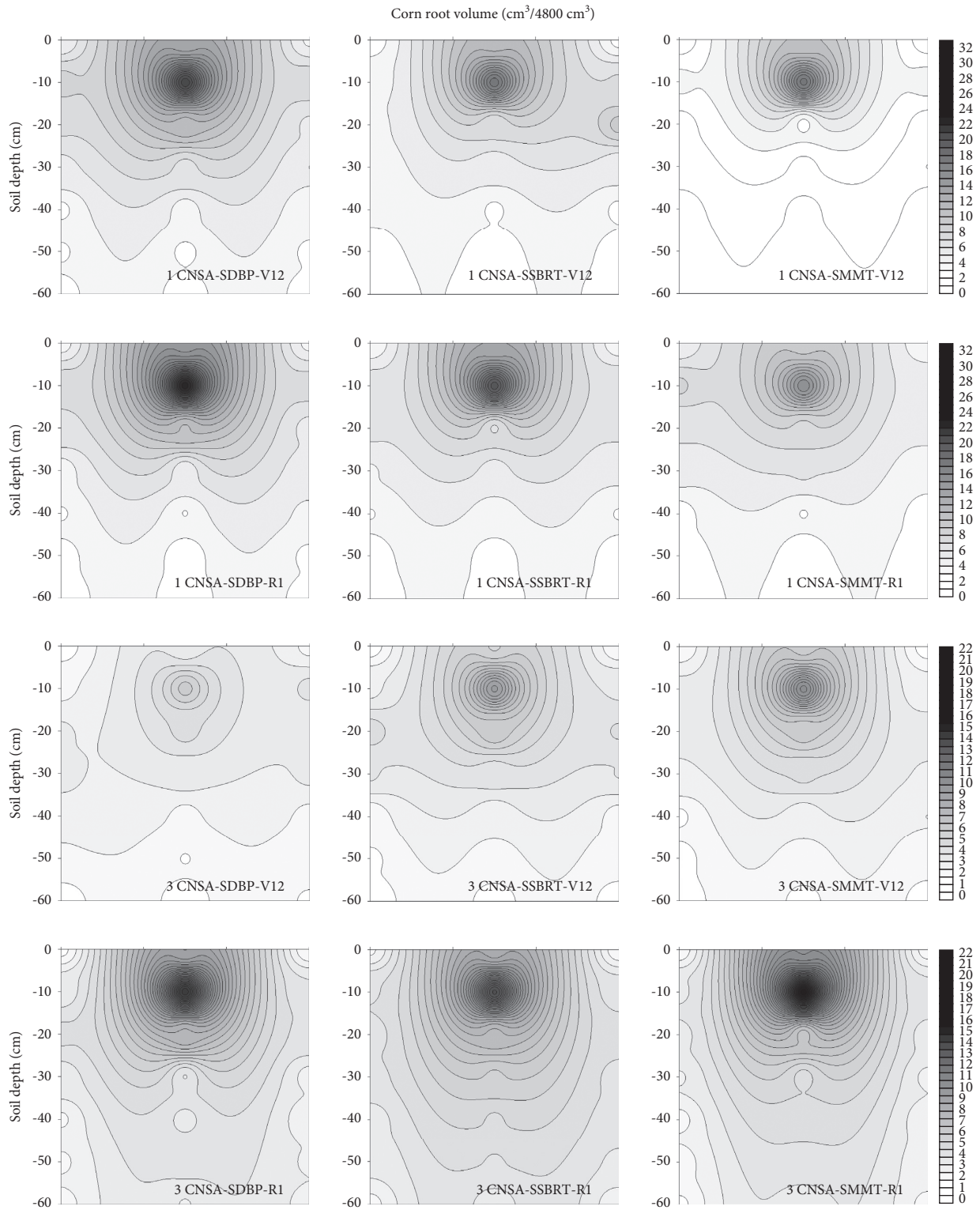
Figure 2: Maize root distribution under different practices of crop straw application.

environment, reduce the soil bulk density of the 0–20 cm arable layer, and increase the permeability and water retention performance of the soil, which promote root expansion and improve the capacity of roots to absorb water and nutrients [32]. Straw application with plowing tillage can restrict the horizontal water flow, promote straw decomposition, and increase soil moisture content and accumulation of available nutrients, thereby increasing the utilization of nitrogen fertilizer and maize grain yield [33]. In addition, the maize grain yield is higher with straw deep
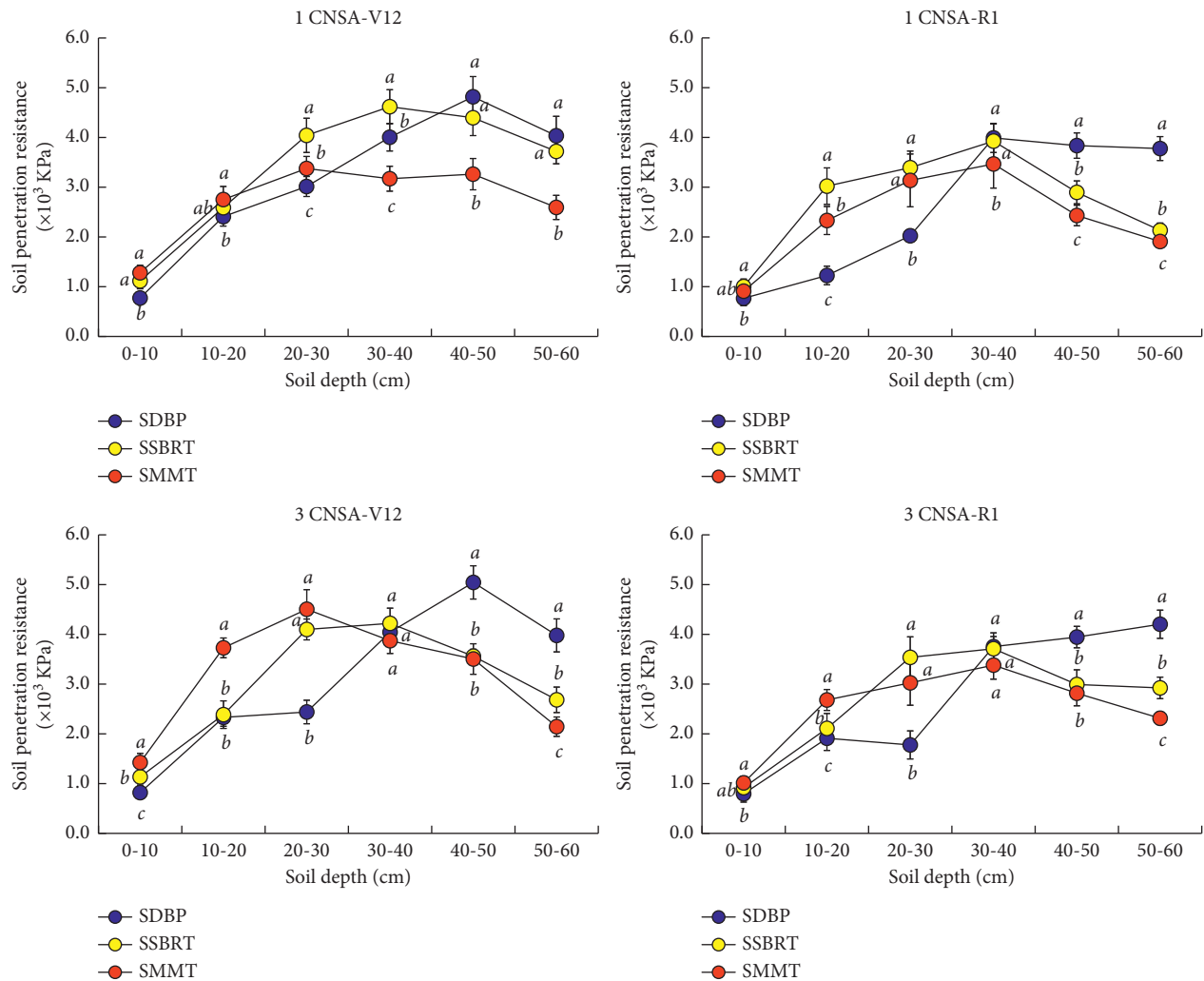
FIGURE 3: Soil penetration resistance in the maize-growing season under different practices of crop straw application.

burial with plowing compared to that with straw shallow burial [34].

However, soil macroaggregates are easily broken to form microaggregates or sand-to-clay-sized particles due to the strong disturbance of the arable layer by plowing or rotary tillage [35, 36], resulting in a decrease in the organic carbon content [37], which affects the growth and development of maize root systems [38]. Previous studies have indicated that compared with conventional soil rotary tillage, straw mulching with minimum tillage can provide a suitable soil environment for crop growth. This method can significantly increase the soil organic carbon [39–42] and moisture contents [43] and enhance the activities of urease, dehydrogenase, alkaline phosphatase, and other enzymes [44], which ultimately increases maize kernel weight [45] and grain yield [39, 40]. Compared with continuous straw application with plowing tillage, the straw mulching with minimum tillage method can significantly increase the content and mean mass diameter of water-stable aggregates in the 20–50 cm soil layer, which effectively reduces the damage of soil aggregate structure and increases the water storage capacity of the 0–200 cm soil layer. The crop grain

yield and water use efficiency by this method were increased by 15.1% and 27.5%, respectively, compared with continuous straw application with plowing tillage [46]. Some studies also implied that the crop grain yield by straw mulching with conventional tillage is higher than that by straw mulching with minimum tillage; however, the operability and economic benefits of the latter are greater than those of the former [47, 48].

Our results show that compared with straw mulching with minimum tillage, straw deep burial with plowing and straw shallow burial with rotary tillage methods can reduce the density of surface soil layers and increase the root system volume after the first straw application, which plays a leading role in increasing the maize grain yield. However, after consecutive rounds of mechanical harvesting and straw application, soil density of 0–40 cm layer was significantly increased with straw deep burial with plowing and straw shallow burial with rotary tillage. The interaction between soil tillage and straw application is the main factor affecting soil porosity. Specifically, the mixture of straws and soils increased the soil porosity [49] and reduced moisture content of the 0–30 cm soil layer and AH-N content of the 0–20 and
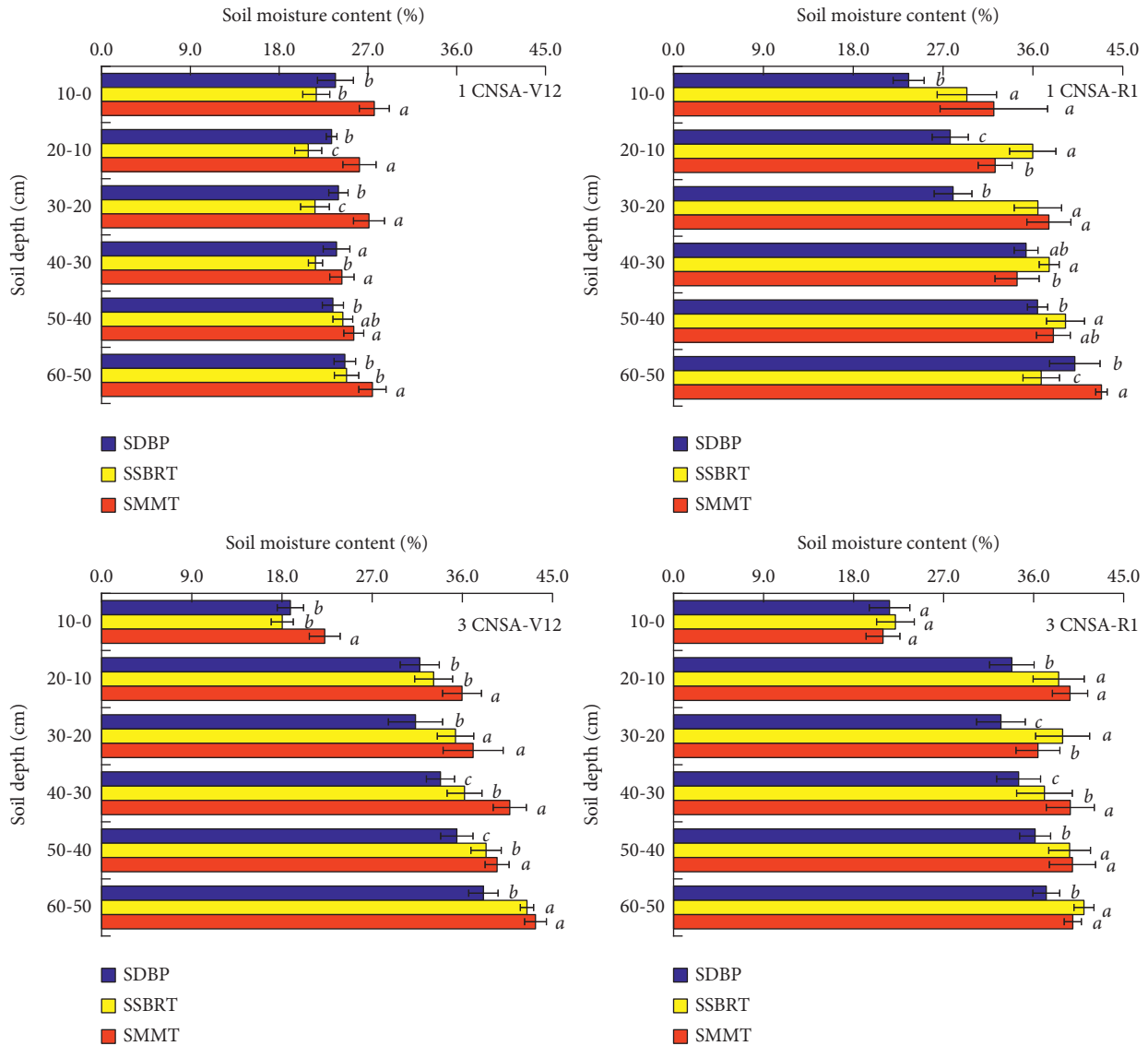
FIGURE 4: Soil water content in the maize-growing season under different practices of crop straw application.

40–60 cm soil layers at the early stage of maize growth. As a result, the growth and development of the maize root system are suppressed, leading to a significant lower maize grain yield compared with that of the previous season. In the Yunnan Plateau in Southwest China, considering frequent seasonal droughts during maize planting and poor soil fertility of the arable layer, the employment of straw mulching with minimum tillage method can increase soil moisture and nutrient contents, which is of great significance in improving the maize emergence rate and grain yield.

Nevertheless, there are some issues associated with the straw mulching with minimum tillage during maize seeding process. Either in mechanical ditching with manual seeding or mechanical seeding, the removal of straw covers by ditching a shovel or seed meter will increase water loss in the soil, which, in turn, leads to a decrease in maize emergence rate [50, 51]. Therefore, the utilization of appropriate equipment for seeding and increase in the seeding depth are required to provide a suitable soil environment for seed germination and emergence.
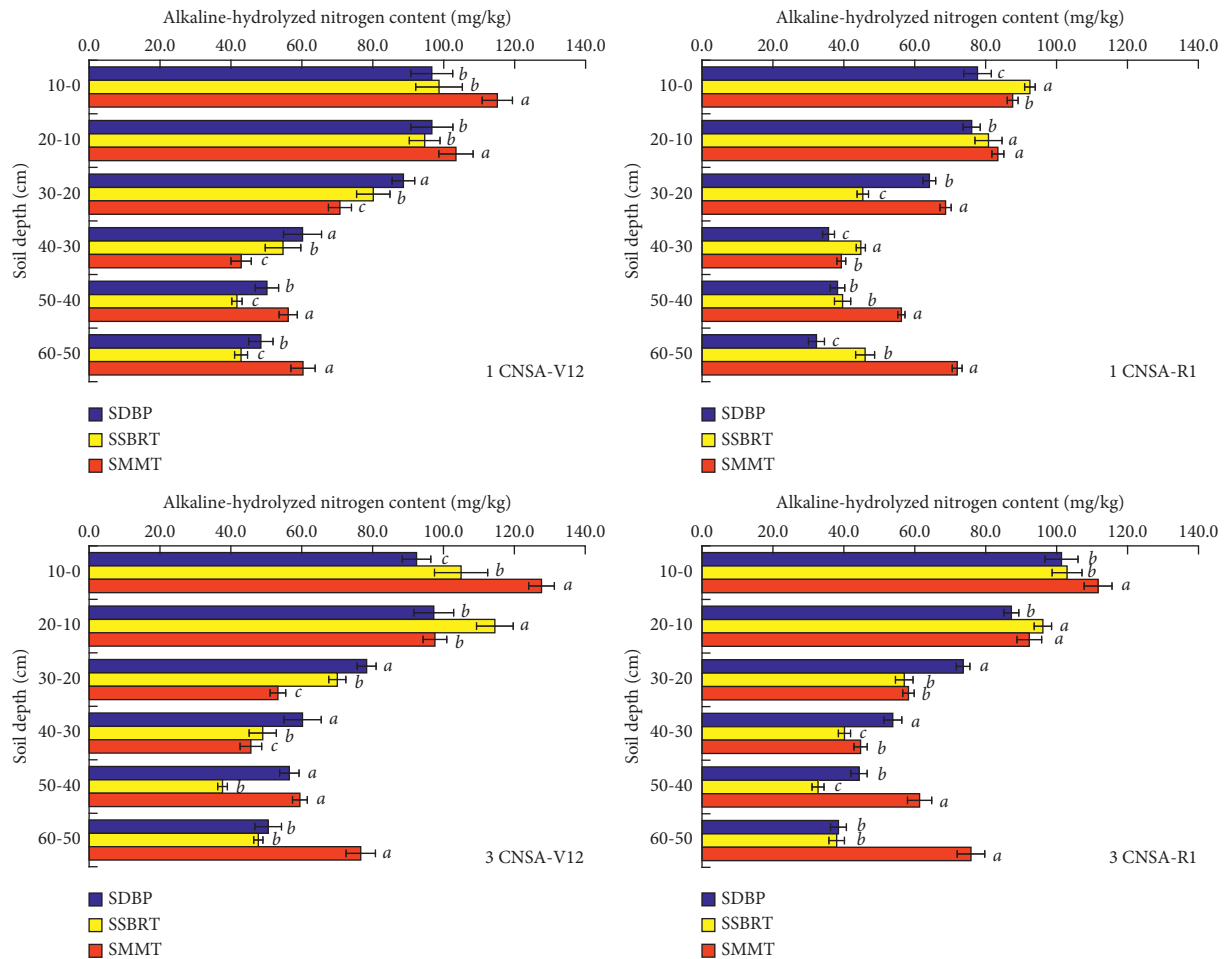
FIGURE 5: Alkaline-hydrolyzed nitrogen content in the maize-growing season under different practices of crop straw application.

5. Conclusion

During mechanical harvesting, the maize grain yield significantly decreases with repeated straw applications using straw deep burial with plowing and straw shallow burial with rotary tillage methods. In contrast, the straw mulching with minimum tillage method can increase the soil moisture content of the arable layer, increase the alkaline-hydrolyzed nitrogen content of shallow or deep soil layers, and promote the growth and development of the maize root system, thus increasing the maize grain yield.

Data Availability

The experimental data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest regarding the present study.

Acknowledgments

References

[1] I. A. Dahri, A. A. Tagar, J. Adamowski, N. Leghari, A. R. Shah, and S. A. Soomro, "Influence of straw incorporation-to-planting interval on soil physical properties and maize performance," *International Agrophysics*, vol. 32, no. 3, pp. 341–347, 2018.

[2] H. Zhao, A. G. Shar, S. Li et al., "Effect of straw return mode on soil aggregation and aggregate carbon content in an annual maize-wheat double cropping system," *Soil and Tillage Research*, vol. 175, pp. 178–186, 2018.

[3] X. Hao, X. Han, S. Wang, and L.-J. Li, "Dynamics and composition of soil organic carbon in response to 15 years of straw return in a Mollisol," *Soil and Tillage Research*, vol. 215, Article ID 105221, 2022.

[4] X. Wang, Z. Jia, L. Liang et al., "Changes in soil characteristics and maize yield under straw returning system in dryland farming," *Field Crops Research*, vol. 218, pp. 11–17, 2018.

[5] M. T. Rahman, Z. C. Guo, Z. B. Zhang, H. Zhou, and X. H. Peng, "Wetting and drying cycles improving aggregation and associated C stabilization differently after straw or biochar incorporated into a Vertisol," *Soil and Tillage Research*, vol. 175, pp. 28–36, 2018.

[6] D. Yu, Z. Wen, X. Li, X. Song, H. Wu, and P. Yang, "Effects of straw return on bacterial communities in a wheat-maize rotation system in the North China Plain," *PLoS One*, vol. 13, no. 6, Article ID e0198087, 2018.

[7] L. Yan, Z. Zhang, Y. Chen, Q. Gao, W. Lu, and A. M. Abdelrahman, "Effect of water and temperature on ammonia volatilization of maize straw returning," *Toxicological and Environmental Chemistry*, vol. 98, no. 5-6, pp. 638–647, 2016.

[8] S. W. Mei, S. Wang, and X. W. Yang, "Straw returning and nitrogen application on winter wheat quality and yield under wheat-maize multiple cropping system," *Ekoloji*, vol. 28, pp. 2183–2192, 2019.

[9] A. S. Sidhu and H. S. Sur, "Effect of Incorporation of legume straw on soil properties and crop yield in a maize-wheat sequence," *Tropical Agriculture*, vol. 70, pp. 226–229, 1993.

[10] J. Bakht, M. Shafi, M. T. Jan, and Z. Shah, "Influence of crop residue management, cropping system and N fertilizer on soil N and C dynamics and sustainable wheat (*Triticum aestivum* L.) production," *Soil and Tillage Research*, vol. 104, no. 2, pp. 233–240, 2009.

[11] W. Fan, J. Wu, J. Li, and J. Hu, "Comparative effects of different maize straw returning modes on soil humus composition and humic acid structural characteristics in Northeast China," *Chemistry and Ecology*, vol. 34, no. 4, pp. 355–370, 2018.

[12] J. Paul, A. K. Choudhary, V. K. Suri, A. K. Sharma, V. Kumar, and Shobhna, "Bioresource nutrient recycling and its relationship with biofertility indicators of soil health and nutrient dynamics in rice-wheat cropping system," *Communications in Soil Science and Plant Analysis*, vol. 45, no. 7, pp. 912–924, 2014.

[13] Y. Fan, J. Gao, J. Sun et al., "Effects of straw returning and potassium fertilizer application on root characteristics and yield of spring maize in China inner Mongolia," *Agronomy Journal*, vol. 113, no. 5, pp. 4369–4385, 2021.

[14] T. Pandiaraj, S. Selvaraj, and N. Ramu, "Effects of crop residue management and nitrogen fertilizer on soil nitrogen and carbon content and productivity of wheat (*Triticum aestivum* L.) in two cropping systems," *Journal of Agricultural Science and Technology A*, vol. 17, pp. 249–260, 2015.

[15] C. Jiang and W. Yu, "Maize production and field CO2 emission under different straw return rates in Northeast China," *Plant Soil and Environment*, vol. 65, no. 4, pp. 198–204, 2019.

[16] L. Bechini, C. Costamagna, L. Zavattaro, C. Grignani, J. Bijttebier, and G. Ruysschaert, "Barriers and drivers towards the incorporation of crop residue in the soil. Analysis of Italian farmers' opinion with the theory of planned behaviour," *Italian Journal of Agronomy*, vol. 10, no. 4, pp. 178–184, 2015.

[17] C. B. Henriksen, J. Rasmussen, and C. Søgaard, "Ridging in autumn as an alternative to mouldboard ploughing in a humid-temperate region," *Soil and Tillage Research*, vol. 85, no. 1-2, pp. 27–37, 2006.

[18] K. Manevski, C. D. Børgesen, X. Li et al., "Optimising crop production and nitrate leaching in China: measured and simulated effects of straw incorporation and nitrogen fertilisation," *European Journal of Agronomy*, vol. 80, pp. 32–44, 2016.

[19] J. M. Carefoot, H. H. Janzen, and C. W. Lindwall, "Crop residue management for irrigated cereals on the semi-arid Canadian prairies," *Soil and Tillage Research*, vol. 32, no. 1, pp. 1–20, 1994.

[20] X. Zhang, N. Sun, L. Wu, M. Xu, I. J. Bingham, and Z. Li, "Effects of enhancing soil organic carbon sequestration in the topsoil by fertilization on crop productivity and stability: evidence from long-term experiments with wheat-maize cropping systems in China," *The Science of the Total Environment*, vol. 562, pp. 247–259, 2016.

[21] K. Saffih-Hdadi and B. Mary, "Modeling consequences of straw residues export on soil organic carbon," *Soil Biology and Biochemistry*, vol. 40, no. 3, pp. 594–607, 2008.

[22] M. A. Bughio, P. Wang, F. Meng, Q. Chen, J. Li, and T. A. Shaikh, "Neoformation of pedogenic carbonate and conservation of lithogenic carbonate by farming practices and their contribution to carbon sequestration in soil," *Journal of Plant Nutrition and Soil Science*, vol. 180, no. 4, pp. 454–463, 2017.

[23] Y. Li, D. Song, P. Dang, L. Wei, X. Qin, and K. H. M. Siddique, "Combined ditch buried straw return technology in a ridge-furrow plastic film mulch system: implications for crop yield and soil organic matter dynamics," *Soil and Tillage Research*, vol. 199, Article ID 104596, 2020.

[24] P. Tian, P. Sui, H. Lian et al., "Maize straw returning approaches affected straw decomposition and soil carbon and nitrogen storage in Northeast China," *Agronomy*, vol. 9, no. 12, p. 818, 2019.

[25] S. Sharma, A. Rangger, M. von Lützow, and H. Insam, "Functional diversity of soil bacterial communities increases after maize litter amendment," *European Journal of Soil Biology*, vol. 34, no. 2, pp. 53–60, 1998.

[26] X. Xia, P. Zhang, L. He et al., "Effects of tillage managements and maize straw returning on soil microbiome using 16S rDNA sequencing," *Journal of Integrative Plant Biology*, vol. 61, no. 6, pp. 765–777, 2019.

[27] D. Zhou, Y. Su, Y. Ning et al., "Estimation of the effects of maize straw return on soil carbon and nutrients using response surface methodology," *Pedosphere*, vol. 28, no. 3, pp. 411–421, 2018.

[28] J. L. Zhao, Y. Lu, H. L. Tian, H. L. Jia, and M. Z. Guo, "Effects of straw returning and residue cleaner on the soil moisture content, soil temperature, and maize emergence rate in China's three major maize producing areas," *Sustainability*, vol. 11, no. 20, p. 5796, 2019.

[29] P. Sharma, V. Abrol, G. R. M. Sankar, and B. Singh, "Influence of tillage practices and mulching options on productivity, economics and soil physical properties of maize (*Zea mays*)-wheat (*Triticum aestivum*) system," *Indian Journal of Agricultural Sciences*, vol. 79, pp. 865–870, 2009.

[30] H. Ram, Y. Singh, K. S. Saini, D. S. Kler, J. Timsina, and E. J. Humphreys, "Agronomic and economic evaluation of permanent raised beds, no tillage and straw mulching for an irrigated maize-wheat system in Northwest India," *Experimental Agriculture*, vol. 48, no. 1, pp. 21–38, 2012.

[31] M. Bahrani, M. Raufat, and H. Ghadiri, "Influence of wheat residue management on irrigated corn grain production in a reduced tillage system," *Soil and Tillage Research*, vol. 94, no. 2, pp. 305–309, 2007.

[32] K. L. Ding and M. J. Hann, "Effects of soil management on soil properties and crop yield," *Transactions of the CSAE*, vol. 16, pp. 28–31, 2000.

[33] E. Nafi, H. Webber, I. Danso, J. B. Naab, M. Frei, and T. Gaiser, "Soil tillage, residue management and site interactions affecting nitrogen use efficiency in maize and cotton in the Sudan Savanna of Africa," *Field Crops Research*, vol. 244, Article ID 107629, 2019.

[34] B. Prochazkova, J. Malek, and J. Dovrtel, "Effect of different straw management practices on yields of continuous spring barley," *Rostlinna Vyroba*, vol. 48, pp. 27–32, 2002.

[35] E. T. Elliott, "Aggregate structure and carbon, nitrogen, and phosphorus in native and cultivated soils," *Soil Science Society of America Journal*, vol. 50, no. 3, pp. 627–633, 1986.

[36] G. P. S. Sodhi, V. Beri, and D. K. Benbi, "Soil aggregation and distribution of carbon and nitrogen in different fractions under long-term application of compost in rice-wheat system," *Soil and Tillage Research*, vol. 103, no. 2, pp. 412–418, 2009.

[37] D. A. Angers, A. N'dayegamiye, and D. Côté, "Tillage-induced differences in organic matter of particle-size fractions and microbial biomass," *Soil Science Society of America Journal*, vol. 57, no. 2, pp. 512–516, 1993.

[38] J. M. Tisdall and J. M. Oades, "Organic matter and water-stable aggregates in soils," *Journal of Soil Science*, vol. 33, no. 2, pp. 141–163, 1982.

[39] K. S. Gangwar, K. K. Singh, S. K. Sharma, and O. K. Tomar, "Alternative tillage and crop residue management in wheat after rice in sandy loam soils of Indo-Gangetic plains," *Soil and Tillage Research*, vol. 88, no. 1-2, pp. 242–252, 2006.

[40] J. Xu, H. Han, T. Ning, Z. Li, and R. Lal, "Long-term effects of tillage and straw management on soil organic carbon, crop yield, and yield stability in a wheat-maize system," *Field Crops Research*, vol. 233, pp. 33–40, 2019.

[41] R. Saha and P. K. Ghosh, "Soil organic carbon stock, moisture availability and crop yield as influenced by residue management and tillage practices in maize-mustard cropping system under hill agro-ecosystem," *National Academy Science Letters*, vol. 36, no. 5, pp. 461–468, 2013.

[42] C. E. P. Lima, M. R. Fontenelle, N. R. Madeira et al., "Compartimentos de carbono orgânico em Latossolo cultivado com hortaliças sob diferentes manejos," *Pesquisa Agropecuária Brasileira*, vol. 51, no. 4, pp. 378–387, 2016.

[43] M. Mtyobile, L. Muzangwa, and P. N. S. Mnkeni, "Tillage and crop rotation effects on soil carbon and selected soil physical properties in a Haplic Cambisol in Eastern Cape, South Africa," *Soil and Water Research*, vol. 15, pp. 47–54, 2020.

[44] D. A. Bossio, J. A. Fleck, K. M. Scow, and R. Fujii, "Alteration of soil microbial communities and water quality in restored wetlands," *Soil Biology and Biochemistry*, vol. 38, no. 6, pp. 1223–1233, 2006.

[45] F. A. Chandio, Y. M. Li, S. A. Shaikh et al., "Effect of straw incorporation by tillage implements combination on physicochemical properties of soil and maize productivity in field condition. Fresen," *Fresenius Environmental Bulletin*, vol. 27, pp. 7527–7535, 2018.

[46] D. Liu, X. Zhang, J. Li, and X. D. Wang, "Effects of different tillage patterns on soil properties, maize yield and water use efficiency in Weibei Highland, China," *Journal of Applied Ecology*, vol. 29, pp. 573–582, 2018.

[47] Z. Li, X. Lai, Q. Yang, X. Yang, S. Cui, and Y. Shen, "In search of long-term sustainable tillage and straw mulching practices for a maize-winter wheat-soybean rotation system in the Loess Plateau of China," *Field Crops Research*, vol. 217, pp. 199–210, 2018.

[48] X. Lu and X. Lu, "Tillage and crop residue effects on the energy consumption, input-output costs and greenhouse gas emissions of maize crops," *Nutrient Cycling in Agroecosystems*, vol. 108, no. 3, pp. 323–337, 2017.

[49] N. Tangyuan, H. Bin, J. Nianyuan, T. Shenzhong, and L. Zengjia, "Effects of conservation tillage on soil porosity in maize-wheat cropping system," *Plant Soil and Environment*, vol. 55, no. 8, pp. 327–333, 2009.

[50] E. M. Brandelero, A. G. D. Araújo, and R. Ralisch, "Coverage mobilization in the sowing line and its influence on temperature and water content and on maize emergence," *Engenharia Agrícola*, vol. 35, no. 1, pp. 98–108, 2015.

[51] M.-P. Hiel, S. Barbieux, J. Pierreux et al., "Impact of crop residue management on crop production and soil chemistry after seven years of crop rotation in temperate climate, loamy soils," *PeerJ*, vol. 6, p. e4836, 2018.

Research Article

Standardized Judgment Method of Shooting Training Action Based on Digital Video Technology

Guo Qing [ORCID] **and HuBao Hui**

Physical Education Department of Gansu Agricultural University, Gansu 730070, Lanzhou, China

Correspondence should be addressed to Guo Qing; guoq@gsau.edu.cn

Aiming at the difficulty of standardizing the action of basketball shooting training, a new method of standardizing the action of basketball shooting training is proposed based on digital video technology. The digital video signal representation, video sequence coding data structure, and video sequence compression coding method are analyzed, and the pixels of basketball shooting training action position space are sampled to collect basketball shooting training images. The time difference method is used to extract the movement target of basketball shooting training from a digital video sequence. Based on digital video technology, the initial background image is estimated, and the update rate is introduced to update the background estimation image. According to the pixel value sequence of the basketball shooting training image, the pixel model of the basketball shooting training image is defined and modified. By judging whether the defined pixel value matches the background parameter model, the standardization of shooting training can be realized. The experimental results show that the proposed method has good stability, high precision, and short time in determining the standardization of shooting movement, can correct the wrong shooting movement in real time, and can effectively guide basketball shooting training.

1. Introduction

Basketball is quite different from other sports. It is a high-intensity and comprehensive sport [1, 2]. Basketball belongs to the same field antagonistic event group dominated by technical and tactical ability, and the technical and tactical level is the decisive factor for the competitive level of basketball [3]. In the actual competition process, basketball players need to have diversified basketball qualities to ensure the victory of the competition. Among them, the more important point is the coordination and stability of athletes' physical functions. In the development of basketball, shooting is its key offensive technology. The essence of a basketball game is a shooting game, which also shows that the stability of shooting has an important relationship with the outcome of the game [4]. Among them, the factors affecting the standardization of athletes' shooting action mainly include athletes' bodies, technology, and psychology. In the process of competition, athletes need to ensure that they master the nature, time, and score of the competition [5]. In the actual basketball game, the attacking team needs to use different techniques or tactics to create more shooting opportunities and ensure

shooting scores [6, 7]. The defensive team should actively defend and prevent the other team from scoring. The accuracy and standard of shooting in basketball are directly related to the score. Therefore, it is of great significance to reasonably test and judge the shooting action of basketball.

At present, scholars in related fields have carried on the research on action judgment and obtained some research results. Reference [8] proposed a motion similarity judgment method based on motion primitives. Based on the computational model of kinematics, the similarity of motion and its performance to the human similarity judgment of the same motion are determined. By performing the action similarity task and comparing it with the computational model solving the same task, the action similarity judgment was realized by classifying the actions based on the learned kinematics primitive. The method has high reliability and provides necessary basis for human action classification. Reference [9] proposed a basketball motion image target detection method based on an improved Gaussian mixture model. Edge detection, gray processing, target capture, target recognition, image detection, and other technologies

are integrated into basketball sports video, and Gaussian probability density mixing is used to select the appropriate number of continuously updated parameters and each pixel area to achieve basketball sports image detection. The method is effective to some extent. However, the above methods are difficult to determine the standardization of basketball shooting training movements.

In view of the above problems, a method to judge the movement standardization of shooting training based on digital video technology is proposed. The innovation of the research method is to use the time difference method to extract the movement target of basketball shooting training. Based on digital video technology, the initial background image is estimated and the update rate is introduced to update the image. According to the pixel value sequence of the basketball shooting training image, the pixel model of the basketball shooting training image is defined and modified. By discriminating the matching relation between pixel value and background parameter model, the standardization judgment of shooting training movement can be realized. Compared with the previous research results, the method designed based on digital video technology has better stability, high accuracy, and short time and can correct the wrong shooting action in real time, which can effectively guide basketball shooting training.

2. Digital Video Technology

Digital video technology is to first use video capture equipment such as cameras to convert the color and brightness information of external images into electrical signals, and then record them into storage media [10–12]. Digital video is video recorded in digital form, as opposed to analog video. Digital video has different production methods, storage methods, and broadcast methods. For example, digital video signals are generated directly through digital cameras and stored on digital tape, P2 card, blue disc, or disk, so as to obtain different formats of digital video, which is then played on a PC, a specific player, etc.

2.1. Representation of Digital Video Signal. Video is described as a group of continuous images, and each image is regarded as a two-dimensional pixel array. The color representation of each pixel includes three components: red R, green G, and blue B, which is called the RGB space representation of the image. The color coordinates used for the three digital TV systems are different. For digital video capture and display, all three digital TV systems use RGB primary colors, but the definition of each primary color spectrum is slightly different. For the transmission of digital video signal, in order to reduce the required bandwidth and be compatible with monochrome digital TV system, the brightness/chroma coordinate system is adopted [13, 14]. The color coordinates used in the NTSC, PAL, and SECAM systems are all derived from the YUV coordinates used for PAL, and YUV is derived from the XYZ coordinates. According to the relationship between the RGB primary color and the YUV primary color, the value of the luminance component Y can be determined by the value of RGB. The

two chromaticity values U and V are proportional to the color differences $B - Y$ and $R - Y$, respectively, and are adjusted to the desired range. The classic conversion relationship between the YUV coordinate system and the RGB primary color value is

$$\begin{bmatrix} Y \\ U \\ V \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ -0.147 & -0.289 & 0.436 \\ 0.615 & -0.515 & -0.100 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}. \tag{1}$$

The conversion of two color spaces is based on the characteristics of a human visual system: in RGB space, if one of the three signals R, G, and B changes, the color of the total image will change, and the human eye can easily detect this change. However, human eyes have different responses to the changes of Y U and V signals. Among them, they are sensitive to the changes of luminance signals, but not very sensitive to the changes of chrominance signals. In this way, we can consider more luminance signals and adopt some processing methods for chrominance signals to improve the compression ratio.

2.2. Data Structure of Digital Video Sequence Coding. In the coding scheme, the video sequence is divided and multiplexed by multiple layers to establish such a data structure.

(1) Sequence: the video sequence starts with the sequence header, including several image groups, and ends with a sequence end code.

(2) Group of pictures (GOP): GOP is a head followed by a series of images, which allows fast random access to the sequence, fast search, and editing. It is the smallest coding unit that can be decoded independently in the sequence [15]. The first image in the GOP is an intracoded image (I frame), followed by a forward prediction coded image (P frame) and a bidirectional prediction image (B frame). Each GOP has only one I frame, and this I frame is used as the first frame to start coding. The P frame is encoded by motion-compensated prediction relative to the previous I frame or P frame, and the P frame can be used as a reference frame for other P frame or B frame coding. B frame is encoded by motion compensation prediction of two frames, one is the past frame, and the other is the future frame. The frame arrangement of the GOP is shown in Figure 1.

The standard does not specify the number of P and B frames in a GOP, nor their specific sequence, except for the first frame, and only one frame is I frame. Any sequence and frame number can be used to design the encoding scheme. The prediction of each P and B frame is based on the previous reference prediction frame. Too many frames in the group layer will affect the quality and compression ratio of coding. Therefore, 10–15 B frames are generally selected in each group layer, and 2–3 B frames are separated between the two P frames.

(3) Image: image is the basic coding unit of video sequence [16]. The image is composed of three rectangular matrices representing the luminance Y and two
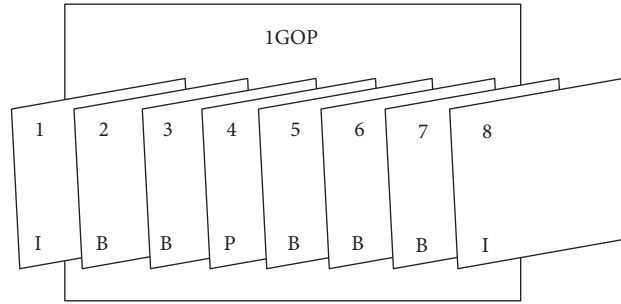
FIGURE 1: Frame arrangement of GOP.

chrominance C_b and C_r values. Each video standard divides the image into macroblock groups. H.261 and H.263 use a fixed macroblock structure, whereas MPEG1/2 allows a flexible structure, and MPEG4 arranges a variable number of macroblocks into a group.

(4) Group of blocks (GOB): H.261 and H.263 divide the image into GOBs. Each GOB includes three macroblock lines and 11 macroblocks in each GOB line, and the GOB header defines the position of the GOB in the image.

(5) Slice: a slice consists of several successive macroblocks into a unit. The size of the slice can be changed. The slice layer provides anti-interference ability against data errors.

(6) Macroblock (MB): Macro block (MB): MB is a basic concept in video coding technology. It can divide the image into many blocks of different sizes and implement different compression strategies in different locations. A coded image is usually composed of several macroblocks. A macroblock is composed of a luminance pixel block and two additional chrominance pixel blocks. In general, the brightness block is 16×16 pixel block, chroma block is 8×8-size pixel block; several macroblocks in each image are arranged in the form of slices. The video coding algorithm takes macroblocks as units, encodes

macroblocks one by one, and organizes them into a continuous video code stream.

(7) Block: block is the smallest coding unit in the standardized video coding algorithm. It consists of 8×8 pixel composition [17].

2.3. Digital Video Sequence Compression Coding Method

2.3.1. Transform Coding. Transform coding does not directly encode the spatial image signal, but first maps and transforms the spatial signal to another orthogonal vector space to generate a batch of transform coefficients, and then encodes these transform coefficients [18–20]. In the digital video sequence image compression and coding technology, the compression performance and error of discrete cosine transform (DCT) are very close to those of K-L transform, and DCT has the characteristics of moderate computational complexity, separability, and fast algorithm. Therefore, there are many schemes using DCT coding in image data compression [21–23]. At present, DCT is used in almost all transform-based image encoders.

Assuming that the range of the spatial variables is $x = 0, 1, \ldots, N - 1$ and $y = 0, 1, \ldots, N - 1$, and the range of the frequency domain variables is $u = 0, 1, \ldots, N - 1$, $v = 0, 1, \ldots, N - 1$, the two-dimensional discrete cosine sine transform formula is

$$F(u, v) = \frac{2}{N} E(u) E(v) \left[\sum_{x=0}^{N-1} \sum_{y=0}^{N-1} f(x, y) \cos\left(\frac{2x + 1}{2N} u\pi\right) \cos\left(\frac{2y + 1}{2N} v\pi\right) \right]. \tag{2}$$

In formula (2), when $u = 0, v = 0$, $E(u), E(v) = 1/\sqrt{2}$. The two-dimensional inverse discrete cosine transform formula is

$$f(x, y) = \frac{2}{N} \left[\sum_{x=0}^{N-1} \sum_{y=0}^{N-1} E(u) E(v) F(u, v) \cos\left(\frac{2x + 1}{2N} u\pi\right) \cos\left(\frac{2y + 1}{2N} v\pi\right) \right]. \tag{3}$$

In formula (3), when $u = 0, v = 0$, $E(u), E(v) = 1/\sqrt{2}$. The core of the two-dimensional discrete cosine transform is separable, so both the forward and inverse transforms can

decompose the two-dimensional transform into a series of one-dimensional transforms (rows, columns) for calculation [24, 25].

In the MPEG series, since the basic unit of DCT transformation is a luminance block or a chrominance block, the size is 8×8, so $N = 8$ can be used in formula (2), so that

$$F(u, v) = \frac{1}{4} E(u) E(v) \left[\sum_{x=0}^{7} \sum_{y=0}^{7} f(x, y) \cos\left(\frac{2x+1}{16} u\pi\right) \cos\left(\frac{2y+1}{16} v\pi\right) \right]. \tag{4}$$

In practical application, considering the characteristics of separable variables in formula (4), rewriting the latter part of formula (4) can obtain

$$\sum_{x=0}^{7} \sum_{y=0}^{7} f(x, y) \cos\left(\frac{2x+1}{16} u\pi\right) \cos\left(\frac{2y+1}{16} v\pi\right) = \sum_{x=0}^{7} \cos\frac{2x+1}{16} u\pi \left(\sum_{y=0}^{7} f(x, y) \cos\frac{2y+1}{16} v\pi \right). \tag{5}$$

Set an intermediate variable $F(i, v)$; then,

$$F(i, v) = \sum_{j=0}^{7} f(i, j) \cos\frac{2j+1}{16} v\pi. \tag{6}$$

The coefficient can be written as

$$F(u, v) = \frac{1}{4} E(u) E(v) \left[\sum_{x=0}^{7} F(i, v) \cos\frac{2x+1}{16} u\pi \right]. \tag{7}$$

It can be seen that after such processing, the two-dimensional DCT transform is decomposed into one row DCT and one column DCT transform, which is easy to be realized by computer. For the inverse transformation, the variables can still be separated to facilitate the implementation of the algorithm.

2.3.2. Predictive Coding. Predictive coding is a technique to improve compression performance through statistical redundancy. Based on the previously encoded pixel values, the encoder can estimate and predict the pixel values to be encoded and decoded [26–28]. For a large number of static or slowly varying regions in the sequence image, the conditional patching method can be used to store the first frame image in the reference frame and send it to the other party. Then, the predicted value $\widehat{I}_k(z)$ of the pixel sampling value $I_k(z)$ of the k frame image at $z = (x, y)$ is the restoration value $I_{k-1}'(z)$ of the pixel value at the same position of the $k-1$ frame image. The interframe difference is expressed as

$$FD_k(z) = I_k(z) - I_{k-1}'(z). \tag{8}$$

For sequence images with relatively moderate amount of motion, encode the frame difference $FD_k(z)$ of the transmitted subblock, where k is the subscript of the subblock, and use the following formula to restore the subblock:

$$I_k'(z) = I_{k-1}'(z) + FD_k(z). \tag{9}$$

3. Standardized Judgment Method of Shooting Training Action

The standardized judgment method of shooting training action based on digital video technology is mainly to collect the basketball shooting training image by sampling and characteristic analysis of the pixels in the position space of basketball shooting training action. Using the time difference method, the basketball shooting training target is found and extracted in real time in the digital video sequence. Based on digital video technology, the initial background estimation image is introduced, and the update rate is introduced to update the background estimation image. According to the pixel value sequence of the basketball shooting training image, the pixel model of the basketball shooting training image is defined, and the defined pixel value model parameters are modified. The standardized judgment of shooting training action is realized by judging whether the defined pixel value matches the background parameter model.

3.1. Collecting Basketball Shooting Training Images. Assuming that the Gaussian mixture model labels the spatial position rotation of basketball shooting action, at multiple points in the basketball shooting space, the shape coordinate of the basketball shooting action under the initial deformation is X, the width of the entire characteristic image of the basketball court is W, and the height is H. The three-dimensional spatial feature image I of basketball shooting is divided into several subblocks by using the grid model. The matching coordinate of the central point of the matching point along the gradient direction on the grid model is calculated as $X\prime$, and then, the spherical grid model of basketball in the hands of players is calculated. The triangular partition pheromone of single-frame basketball shooting action is obtained at the j manual calibration point (x_{ij}, y_{ij}):

$$P(i, j) = \left[X \text{int}\left(\frac{W}{2}\right) - 1 \right]\left[X' \text{int}\left(\frac{H}{2}\right) - 1 \right]. \tag{10}$$

The basketball shooting action sampling image has 8×8 pixels in the grid surface. The sampling point density feature is extracted, and the mean square error between the standardized feature points (x'_{ij}, y'_{ij}) of the shooting action is

$$Err_{ij} = \frac{1}{N} \sum_{i=1}^{N} \sqrt{\left(x'_{ij} - x_{ij}\right)^2 + \left(y'_{ij} - y_{ij}\right)^2}. \tag{11}$$

In the above formula, N is the total number of uniformly distributed grids of the image. Considering all the pixel feature points of n spatial positions, the difference error vector of basketball players in shooting and lifting the ball is obtained as

$$ERR = \frac{1}{n} \sum_{j=0}^{n-1} Err_{ij} = \frac{1}{n} \times \frac{1}{N} \sum_{j=0}^{n-1} \sum_{i=1}^{N} \sqrt{\left(x'_{ij} - x_{ij}\right)^2 + \left(y'_{ij} - y_{ij}\right)^2}. \tag{12}$$

$$D_k(x, y) = \begin{cases} f_k(x, y) - f_{k-1}(x, y), & \text{when} \quad f_k(x, y) - f_{k-1}(x, y) > 0 \\ 0, & \text{when} \quad f_k(x, y) - f_{k-1}(x, y) \le 0 \end{cases} \tag{13}$$

Negative difference:

$$D_k(x, y) = \begin{cases} \left| f_k(x, y) - f_{k-1}(x, y) \right|, & \text{when} \quad f_k(x, y) - f_{k-1}(x, y) > 0 \\ 0, & \text{when} \quad f_k(x, y) - f_{k-1}(x, y) \ge 0 \end{cases} \tag{14}$$

Full difference:

$$D_k(x, y) = \left| f_k(x, y) - f_{k-1}(x, y) \right|. \tag{15}$$

(2) Binarize the image D_k after the difference to obtain

$$R_k(x, y) = \begin{cases} 1 & \text{foreground,} \quad D_k(x, y) < \text{thresholding} \\ 0 & \text{background,} \quad D_k(x, y) \le \text{thresholding} \end{cases}. \tag{16}$$

In the above formula, D_k contains the change of the scene between two consecutive frames of images. This change is composed of many factors. It can be considered that the change of the moving target is obvious. Given a threshold, when the difference of a pixel value in the differential image is greater than a given threshold, the pixel is considered to be a foreground pixel, possibly a point on the target; otherwise, it is considered a background pixel [32–34].

(3) Postprocessing the image R_k to obtain R'_k, where the area of the moving target should be greater than the given threshold. Morphological filtering and noise removal can be used to eliminate noise in small areas.

Thus, the pixel sampling and feature analysis of three main position spaces of basketball shooting training action are realized, and the image acquisition of basketball shooting training is completed.

3.2. Extracting the Goal of Basketball Shooting Training.
The time difference method mainly uses the difference of two or several consecutive frames in the digital video sequence to extract the moving target of basketball shooting training [29–31]. The basic process of the time difference method is shown in Figure 2.

(1) Calculate the difference image D_k between the k frame image f_k and the $k - 1$ frame image f_{k-1}; according to the following three methods, the difference image obtained is expressed as follows.

Positive difference:

(4) Judge the postprocessing result R'_k, mark the area larger than the given threshold as the target, and obtain its complete location information. Through the above steps, the basketball shooting training sports goal is extracted.

3.3. Judging the Standardization of Basketball Shooting Training Action

(1) Initial background estimation image: the single Gaussian distribution background model is suitable for single-modal background situations. It establishes a model $\eta(x, \mu_t, \sigma_t^2)$ represented by a single Gaussian distribution for the color of each image point, where t represents time. Let the current color value of the basketball shooting training image point be f_i; calculate the average brightness μ_0 of each pixel and the variance δ_0^2 of pixel brightness of the basketball shooting training image in the digital video sequence, and take the image B_0 with Gaussian distribution composed of μ_0 and δ_0^2 as the initial background estimation image, which is expressed as
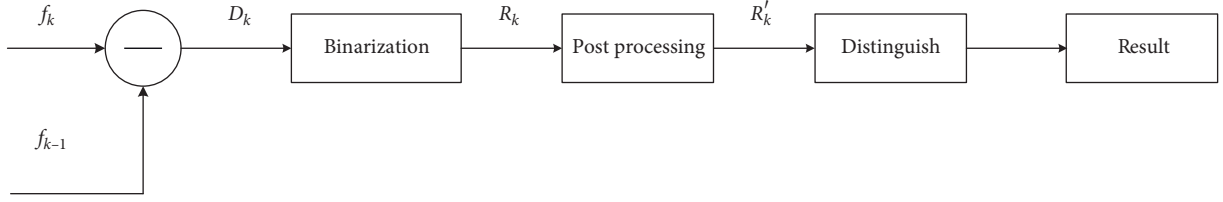
FIGURE 2: Basic process of the time difference method.

$$B_0 = \left[\mu_0, \delta_0^2\right]. \qquad (17)$$

In formula (17), $\mu_0 = 1/T\sum_{i=0}^{T-1} f_i$, $\delta_0^2 = 1/T\sum_{i=0}^{T-1}(f_i - \mu_0)^2$.

(2) Update the background estimation image: the update of the single Gaussian distribution background model refers to the update of the Gaussian distribution parameter of the basketball shooting training image. The constant update rate α that represents the update speed is introduced, and the update of the Gaussian distribution parameter of this point can be expressed as

$$B_t = \left[\mu_t, \delta_t^2\right]. \qquad (18)$$

In the above formula, $\mu_t = (1-\alpha)\mu_{t-1} + \alpha f_i$, $\delta_t^2 = (1-\alpha)\delta_{t-1}^2 + \alpha(f_i - \mu_t)^2$.

(3) Define the pixel model: define the distribution model for each basketball shooting training image pixel. Set the pixel value sequence of the basketball shooting training image as $\{x_{t-k}, x_{t-k+1}, \ldots, x_t\}$, and on this basis, define a set of multiple single models:

$$P_t(x, y) = \left\{p_{i,t} | i = 1, 2, \ldots, K\right\}. \qquad (19)$$

In the above formula, $p_{i,t} = [w_{i,t}, m_{i,t}, l_{i,t}]$ is each single model, which consists of three parameters, where $w_{i,t}$ is the weight of this single model, and its size reflects the current reliability of the pixel value represented by this model; $m_{i,t}$ is the mean value of this single model, which reflects the center of each single peak distribution; and $l_{i,t}$ is the width of the unimodal distribution of this single model, and its size reflects the degree of instability of the pixel value, and its role is equivalent to that of the aforementioned single model. K is the number of single models, which reflects the number of peaks in the multipeak distribution of pixel values. Its selection depends on the pixel value distribution and also on the computing power of the system. The usual value is between 3 and 5. In order to keep the model close to the current distribution of pixel values, it is necessary to update the parameters of this model for each defined pixel value [35].

(4) Correct the pixel value model parameters: the parameter correction steps are as follows:

Step 1: for each new pixel value, first check whether it matches the model. The detection method is

$$\begin{array}{c} For\ i = 1 \quad to \quad K \\ if\ |x_{i,t} - m_{i,t}| < \alpha l_{i,t} \quad then \quad matched \qquad (20) \\ else \quad unmatched. \end{array}$$

Step 2: after the detection in step 1, the weight of the single model matching the defined pixel value is corrected as

$$w_{i,t} = w_{i,t-1} + \beta w_{i,t-1}. \qquad (21)$$

The parameters of the single model that matches the defined pixel values are corrected as follows:

$$m_{i,t} = \frac{1}{k+1} \sum_{j=0}^{k} x_{t-j}$$

$$l_{i,t} = \frac{1}{k+1} \sum_{j=0}^{k} \left(x_{t-j} - m_{i,t}\right)^2. \qquad (22)$$

Step 3: after completing the above correction, it is necessary to normalize the weight of single model in the model as follows:

$$w_{i,t} = \frac{w_{i,t}}{\sum_{j=t}^{K} w_{j,t}}. \qquad (23)$$

(5) Establishment of background pixel model: the above model is used to define the pixel value of the basketball shooting training image; that is, it is necessary to judge whether the defined pixel value is target pixel or background pixel, so as to realize the standardized judgment of shooting training action. Calculate each single model $w_{i,t}/l_{i,t}$, arrange each model in descending order of the single model, consider the previous model to be a background model, and obtain the model of background pixels [36] expressed as

$$B = \text{argmin}_N \sum_{k=1}^{N} w_{i,t} > 1 - \alpha w_l. \qquad (24)$$

Through the above steps, the pixel value of the basketball shooting training image is defined by using the background parameter model, and the standardized judgment of shooting training action is realized by judging whether the defined pixel value matches the background parameter model.

4. Experimental Analysis

4.1. Experimental Environment and Data. In order to verify the effectiveness of the method for determining the standardization of shooting training movements based on digital video technology, the experiments were conducted on a computer with Intel Core i7-6800K, 3.4 GHz, Nvidia GeForce GTX1080 (8G) graphics card and 24G memory. The operating system is Win10, and the software platform is Anaconda3 and Visual Studio 2015. The resolution of basketball shooting action visual image sampling is 320×240. A group of basketball shooting action visual image simulation data express a basketball shooting action. There are 100 test sample image sets in each shooting action mode and a total of 1024×1000 test sets in basketball shooting action visual image database. The reference background neighborhood is 5×5 image blocks, that is, 20×20 pixels, and the model update parameters are $\beta_0 = 0.95, \beta_1 = 0.99, \beta_2 = 0.90$, and the threshold t_{IBSCI} is determined after many experiments. In order to ensure the absolute fairness of the experimental results, the ball selection processing in the whole experimental process is completed by the artificial intelligence robot, and the relevant participants only serve as the detection and verification personnel to supervise and investigate the ball selection operation of the robot. According to the above parameters, SolidWorks is used to establish a simplified visual analysis model of basketball shooting action, import the analysis data into ADAMS software for image processing and analysis, and make standardized judgment on basketball shooting action. The standardized action mode of basketball shooting is shown in Figure 3.

Save the basketball shooting standardized action data shown in Figure 3 as . TXT text data, load it into the image data processing software, conduct computer vision analysis, guide the actual shooting action, collect the basketball shooting training image, and obtain the original basketball shooting information. The collection results are shown in Figure 4.

Figure 4 shows the original basketball shooting information collection results. In order to realize the standardized judgment of shooting training action, it is necessary to extract the training target from the collected original basketball shooting information. The proposed standardized judgment method of shooting training action based on digital video technology is used to extract the collected original basketball shooting training target and judge the standardization of shooting training action. The results are shown in Figure 5.

It can be seen from the analysis of Figure 5 that the standardized judgment method of shooting training action based on digital video technology can effectively realize the extraction and detection of moving targets in basketball shooting training, correct shooting errors in real time, and effectively guide basketball shooting training.

In order to evaluate and compare the proposed standardized judgment method of shooting training action, the accuracy rate and recognition rate are used as evaluation indexes, and the calculation formulas of accuracy rate and recognition rate are as follows:
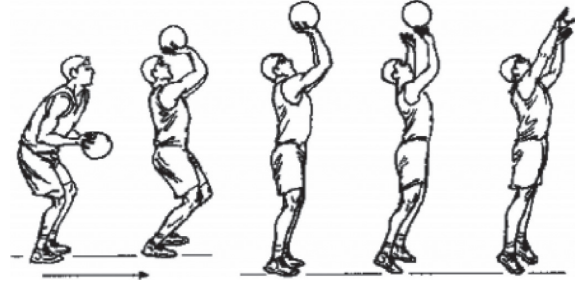
FIGURE 3: Standardized action mode of basketball shooting.

$$P = \frac{TP}{TP + FP}$$
$$R = \frac{TP}{TP + FN} \tag{25}$$

where TP is the number of foreground pixels that are correctly detected, FP is the number of pixels whose background is misjudged as foreground, and FN is the number of pixels whose foreground is misjudged as background.

4.2. Comparison of Standardized Judgment Accuracy of Shooting Training Action. In order to further verify the judgment accuracy of the proposed method, the accuracy is taken as the evaluation index. The higher the accuracy, the higher the judgment accuracy. By comparing the method of reference [8] and the method of reference [9], the standardized judgment accuracy of shooting training action of different methods is obtained, and the comparison results are shown in Figure 6.

According to the analysis of Figure 6, when the number of experiments is 30, the average standardized judgment accuracy of the shooting training action of the method of reference [8] is 84.6%, the average standardized judgment accuracy of the shooting training action of the method of reference [9] is 70.3%, and the average standardized judgment accuracy of shooting training action of the proposed method is as high as 95.2%. Therefore, compared with the method of reference [8] and the method of reference [9], the proposed method has a higher accuracy of the standardized judgment of shooting training action and can effectively improve the accuracy of standardized judgment of shooting training action.

4.3. Comparison of Standardized Judgment and Stability of Shooting Training Action. Further verify the judgment stability of the proposed method, and take the recognition rate as the evaluation index. The higher the recognition rate, the better the judgment stability of the method. By comparing the method of reference [8], the method of reference [9], and the proposed methods, we get the comparison results of the standardized judgment stability of shooting training actions of different methods, as shown in Figure 7.

According to the analysis of Figure 7, when the number of experiments is 30, the average standard judgment
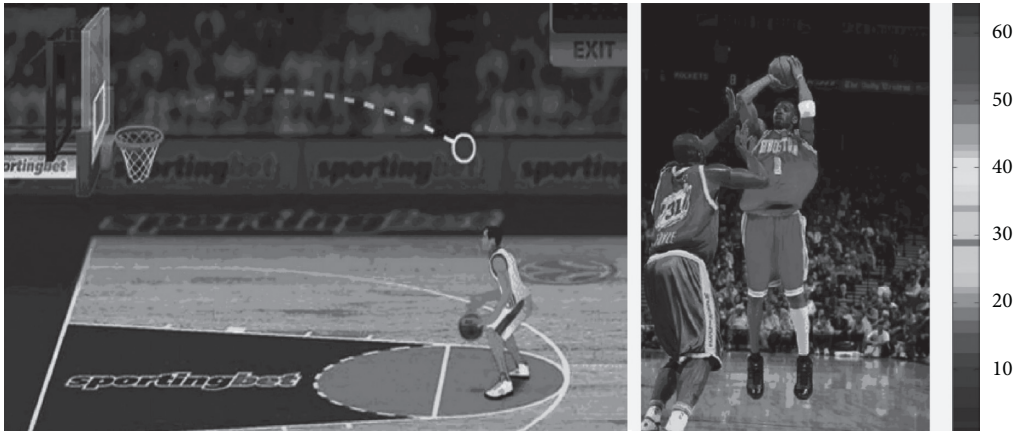
Figure 4: Original basketball shooting information collection results.

Figure 5: Extract the target results of basketball shooting training.
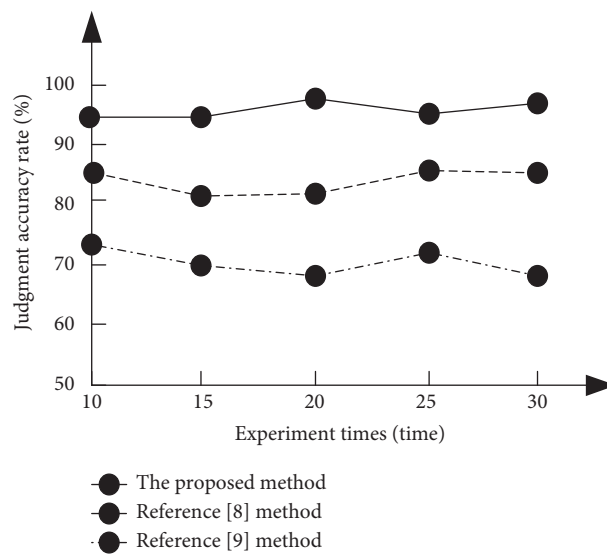
Figure 6: Comparison results of the standardized judgment accuracy of shooting training actions of different methods.
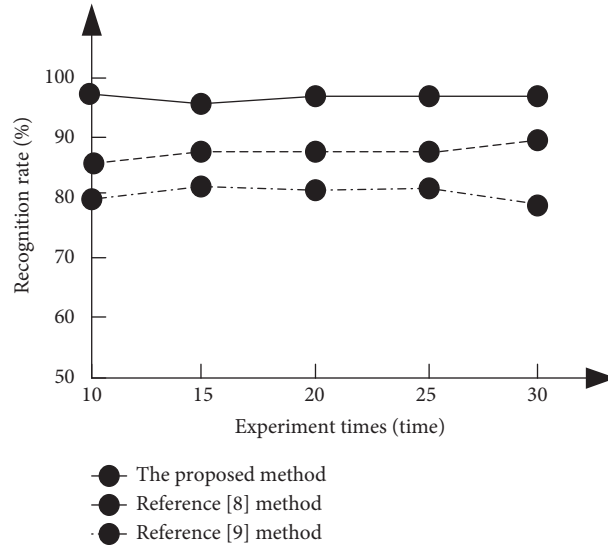
FIGURE 7: Comparison results of the standardized judgment stability of shooting training actions of different methods.

TABLE 1: Comparison results of the standardized judgment time of the shooting training action of different methods.

Number of experiments	The proposed method	The method of reference [8]	The method of reference [9]
10	4.36	7.98	9.76
15	7.12	11.8	13.8
20	9.23	15.7	17.6
25	12.2	19.5	23.9
30	15.3	23.9	26.5

recognition rate of shooting training action of the method of reference [8] is 88.4%, the average standard judgment recognition rate of shooting training action of the method of reference [9] is 80.2%, and the average standard judgment recognition rate of the shooting training action of the proposed method is as high as 96%. It can be seen that compared with the method of reference [8] and the method of reference [9], the proposed method has better stability in judging the standardization of shooting training action.

4.4. Comparison of Standardized Judgment Time of Shooting Training Action. On this basis, the judgment time of the proposed method is verified, and the method of reference [8], the method of reference [9], and the proposed method are compared. The standardized judgment time of the shooting training action of different methods is compared, and the comparison results are shown in Table 1.

According to the data in Table 1, with the increase in the number of experiments, the standardized judgment time of shooting training actions of different methods increases. When the number of experiments reaches 30, the standardized judgment time of the shooting training action of the method of reference [8] is 23.9 s, the standardized judgment time of the shooting training action of the method of reference [9] is 26.5 s, whereas the standardized judgment time of the shooting training action of the proposed method is only 15.3 s. Therefore, compared with the method of

reference [8] and the method of reference [9], the standardized judgment time of shooting training action of the proposed method is shorter.

5. Conclusion

(1) The proposed standardized judgment method of shooting training action based on digital video technology gives full play to the advantages of digital video technology

(2) The standardized judgment of shooting training action is high, which can effectively shorten the judgment time and has good judgment stability

(3) Correct shooting mistakes in real time, and effectively guide basketball shooting training

However, in the process of standardized judgment of shooting training action, dimension reduction is not considered to deal with the characteristics of shooting training action, so as to reduce the amount of calculation. Therefore, in the next research, the dimension of shooting training action characteristics is reduced to further reduce the judgment time.

Data Availability

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Conflicts of Interest

The authors declare that they have no conflicts of interest regarding this work.

Acknowledgments

References

[1] Z. Pan and C. Li, "Robust basketball sports recognition by leveraging motion block estimation," *Signal Processing: Image Communication*, vol. 83, no. 10, Article ID 115784, 2020.

[2] H. Dong, "Evaluation of the value of basketball players based on wireless network and improved Bayesian algorithm," *EURASIP Journal on Wireless Communications and Networking*, vol. 2020, no. 1, pp. 1–11, 2020.

[3] J. Li and Y. Yang, "Study on the influential factors of ordinary college elite men's basketball teams tactic execution," *Bulletin of Sport Science & Technology*, vol. 27, no. 2, pp. 125–127, 2019.

[4] Y. Chen, Y. Qiu, and W. Ren, "A normalized score-based weighted PageRank algorithm on ranking prediction of basketball games," *Modern Physics Letters B*, vol. 35, no. 18, Article ID 2150302, 2021.

[5] D. Castillo, J. Raya-González, A. T. Scanlan, S. Sánchez-Díaz, and Y. Javier, "The influence of physical fitness attributes on external demands during simulated basketball matches in youth players according to age category," *Physiology & Behavior*, vol. 233, no. 1, Article ID 113354, 2021.

[6] J. Xu and C. Yi, "The scoring mechanism of players after game based on cluster regression analysis model," *Mathematical Problems in Engineering*, vol. 2021, no. 3, pp. 1–7, 2021.

[7] J. Vera, R. Molina, D. Cárdenas, B. Redondo, and R. Jiménez, "Basketball free-throws performance depends on the integrity of binocular vision," *European Journal of Sport Science*, vol. 20, no. 3, pp. 407–414, 2020.

[8] V. Nair, P. Hemeren, A. Vignolo, N. Noceti, and G. Sandini, "Action similarity judgment based on kinematic primitives," *Robotics*, vol. 30, no. 8, Article ID 13176, 2020.

[9] H. Lv and X. Dong, "Target detection algorithm for basketball moving images based on improved Gaussian mixture model," *Microprocessors and Microsystems*, vol. 83, no. 6, Article ID 104010, 2021.

[10] N. A. Shelke and S. S. Kasana, "Multiple forgeries identification in digital video based on correlation consistency between entropy coded frames," *Multimedia Systems*, vol. 24, no. 7, pp. 1–14, 2021.

[11] W. E. Bruehs and D. Stout, "Quantifying and ranking quality for acquired recordings on digital video recorders," *Journal of Forensic Sciences*, vol. 65, no. 4, pp. 1155–1168, 2020.

[12] L. O'Donnell, R. Mander, M. Denton et al., "Portable digital video camera configured for remote image acquisition control and viewing," *Driving ip forward*, vol. 2, no. 20, Article ID 10356304, 2019.

[13] S. Dinmore, "Beyond lecture capture: creating digital video content for online learning – a case study," *Journal of University Teaching and Learning Practice*, vol. 16, no. 1, pp. 1–10, 2019.

[14] D. Chi and J. Zhou, "Deep learning-based luma and chroma fractional interpolation in video coding," *IEEE Access*, vol. 7, no. 8, pp. 112535–112543, 2019.

[15] S. Sowmyayani, V. Murugan, and J. Kavitha, "Fall detection in elderly care system based on group of pictures," *Vietnam Journal of Computer Science*, vol. 8, no. 2, pp. 199–214, 2020.

[16] A. M. Atto, A. Benoit, and P. Lambert, "Timed-image based deep learning for action recognition in video sequences," *Pattern Recognition*, vol. 104, no. 11, Article ID 107353, 2020.

[17] M. Wang, J. Lin, J. Zhang, and W. Xie, "Fine-grained region adaptive loop filter for super-block video coding," *IEEE Access*, vol. 8, no. 12, pp. 445–454, 2020.

[18] A. Nakagawa and K. Kato, "Quantitative understanding of VAE by interpreting ELBO as rate distortion cost of transform coding," *Machine Learning*, vol. 30, no. 7, Article ID 15190, 2020.

[19] N. Li, Y. Zhang, and C. Kuo, "Explainable machine learning based transform coding for high efficiency intra prediction," *Image and Video Processing*, vol. 21, no. 12, Article ID 11152, 2020.

[20] S. Milani, E. Polo, and S. Limuti, "A transform coding strategy for dynamic point clouds," *IEEE Transactions on Image Processing*, vol. 29, no. 3, pp. 8213–8225, 2020.

[21] S. Kansal and R. K. Tripathi, "Adaptive geometric filtering based on average brightness of the image and discrete cosine transform coefficient adjustment for gray and color image enhancement," *Arabian Journal for Science and Engineering*, vol. 45, no. 3, pp. 1655–1668, 2020.

[22] L. Sun, S. Liang, P. Chen, and Y. Chen, "Encrypted digital watermarking algorithm for quick response code using discrete cosine transform and singular value decomposition," *Multimedia Tools and Applications*, vol. 80, no. 2, pp. 1–16, 2020.

[23] K. Ramadan, M. I. Dessouky, and F. El-Samie, "Equalization and blind CFO estimation for performance enhancement of OFDM communication systems using discrete cosine transform," *International Journal of Communication Systems*, vol. 33, no. 3, Article ID e3984, 2020.

[24] S. Agha, U. A. Gulzari, F. Shaheen, and F. Jan, "A high throughput two-dimensional discrete cosine transform and MPEG4 motion estimation using vector coprocessor," *Journal of Real-Time Image Processing*, vol. 17, no. 5, pp. 1319–1330, 2020.

[25] Z. Yuan, D. Liu, X. Zhang, and Q. Su, "New image blind watermarking method based on two-dimensional Discrete Cosine Transform," *Optik - International Journal for Light and Electron Optics*, vol. 204, no. 2, Article ID 164152, 2019.

[26] S. Hovsepyan, I. Olasagasti, and A.-L. Giraud, "Combining predictive coding and neural oscillations enables online syllable recognition in natural speech," *Nature Communications*, vol. 11, no. 1, pp. 3117–3124, 2020.

[27] L. Annabi, A. Pitti, and M. Quoy, "Bidirectional interaction between visual and motor generative models using Predictive Coding and Active Inference," *Neural Networks*, vol. 143, no. 11, pp. 638–656, 2021.

[28] T. Hueber, E. Tatulli, L. Girin, and J. L. Schwartz, "Evaluating the potential gain of auditory and audiovisual speech-predictive coding using deep learning," *Neural Computation*, vol. 32, no. 3, pp. 1–30, 2020.

[29] L. Zhang, T. Zhang, H. S. Shin, and X. Xu, "An efficient underwater acoustical localization method based on time difference and bearing measurements," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, no. 12, Article ID 8501316, 2020.

[30] B. Yazdanshenasshad and M. S. Safizadeh, "Reducing the additional error caused by the time-difference method in transit-time UFMs," *Science, Measurement & Technology, IET*, vol. 51, no. 6, pp. 1049–1057, 2019.

[31] B. Zhang, W. Bu, and A. Xiao, "Efficient difference method for time-space fractional diffusion equation with Robin fractional derivative boundary condition," *Numerical Algorithms*, vol. 244, no. 4, pp. 1–24, 2021.

[32] Y. Q. Chen, Z. L. Sun, and K. Lam, "An effective sub-superpixel-based approach for background subtraction," *IEEE Transactions on Industrial Electronics*, vol. 67, no. 1, pp. 601–609, 2019.

[33] S. M. Roy and A. Ghosh, "Foreground segmentation using adaptive 3 phase background model," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 6, pp. 2287–2296, 2019.

[34] A. Kushwaha, A. Khare, O. Prakash, and M. Khare, "Dense optical flow based background subtraction technique for object segmentation in moving camera environment," *IET Image Processing*, vol. 14, no. 5, pp. 2532–2538, 2020.

[35] S. Wang, "Simulation of visual saliency evaluation method for multimedia human-computer interaction interface," *Computer Simulation*, vol. 37, no. 03, pp. 161–164, 2020.

[36] P. Guo, X. Zhu, H. Zhang, and X. Zhang, "Adaptive background mixture model with spatio-temporal samples," *Optik*, vol. 183, no. 4, pp. 433–440, 2019.

Research Article

Automatic Repair Method for D2D Communication Routing Buffer Overflow Vulnerability in Cellular Network

BaoPing Yang and Kun Jiang [iD]

Huanggang Normal University of Physics and Telecommunications, Huanggang Normal University, Huanggang 438000, Hubei, China

Correspondence should be addressed to Kun Jiang; jiangkun@hgnu.edu.cn

Repairing D2D communication routing buffer overflow in a cellular network is of great significance in improving communication quality and security. Due to the increase of user usage, the communication data are easy to exceed the boundary of the buffer, resulting in the reduction of covered data information. The traditional repair methods mainly repair through the characteristics of covered data information, ignoring the impact of network topology information transmission delay and packet loss during calculation, resulting in the problem of low communication security. A cellular network routing buffer overflow repair algorithm based on the homomorphic analysis of node residual energy is proposed; the cellular network D2D communication routing protocol is designed; the cellular network D2D communication protocol path index is determined; then, the cellular network D2D communication routing protocol is designed by analyzing node residual energy; and the cellular network D2D communication network routing optimization method based on AHP is designed. Big constructs the energy model of cellular network D2D communication network, solves and sets the routing optimization objective function, realizes the control of network routing, and repairs the buffer overflow. The experiment results show that the improved method can effectively reduce the packet loss rate of communication data, improve the anti-interference ability of the system, and ensure the security of network communication.

1. Introduction

Cellular network D2D communication usually takes electromagnetic wave as the carrier to receive and transmit network signals. A cellular network electromagnetic wave communication system plays an important role in implementing long-range precision strike and target precision guidance [1]. In the electromagnetic wave communication system of the cellular network, because the routing protocol evolves with the mobile location algorithm, the routing location cannot meet the needs of useful QoS, resulting in the buffer overflow problem, which needs to be repaired [2, 3]. Therefore, it is of great significance to study the automatic repair method of D2D communication routing buffer overflow vulnerability in the cellular network.

In literature [4], due to the lack of software execution information, static buffer overflow detection technology often reports too many false positives. Manually checking all static warnings is time-consuming. Bovinspector, a framework for automatically verifying static buffer overflow warnings, is proposed, and suggestions are provided for C programs to automatically repair real buffer overflow warnings. Given the program source code and static buffer overflow warning, Bovinspector first performs warning reachability analysis. The Bovinspector then symbolically executes the source code under the guidance of reachability warnings. Each reachable warning is verified and classified by checking whether all path conditions and buffer overflow constraints can be met at the same time. For each verified real warning, Bovinspector will provide suggestions and automatically repair it with 11 repair strategies. However, this method does not consider the characteristics of cellular network D2D communication, so the repair performance needs to be further improved. In [5], in case of earthquake,

debris flow, and other disasters, the communication infrastructure may fail, resulting in blind areas and inconvenient communication for residents. Therefore, a new scheme to connect these infrastructure fault areas, namely, energy-aware device-to-device communication scheme (need), is proposed. The proposed scheme uses cluster technology to connect users in the infrastructure failure area that often cannot directly access the cellular network. Compared with traditional clustering methods, it increases the process of determining candidate cluster heads before determining the final cluster heads. Based on the location and residual energy, the final CHS is selected from the candidate CHS, and the dual CHS in the cluster operates alternately to share the communication cost. In addition, in order to improve the routing efficiency, an improved ant colony algorithm (Maca) is proposed. The experiment results show the effectiveness of the proposed demand scheme in energy consumption and energy balance, and show that the scheme significantly prolongs the lifetime of the whole network. However, this method has the problem of large amount of calculation. Reference [6] studies multihop device-to-device (D2D) communication for cell coverage expansion. Considering the inband underlay D2D mode, the purpose is to meet the signal-to-noise ratio requirements of the pre-allocated resource block (RB) on the downlink connection, the signal-to-noise ratio requirements of the pre-allocated RB on each D2D side chain connection, and the maximum allowable interference at the base station receiver on all uplink RBS. While meeting these requirements, power control and routing are performed to minimize user equipment energy consumed in the system. An optimization problem is proposed, which is a mixed integer nonlinear programming and solved by Generalized Benders Decomposition (GBD). GBD decomposes the formula into main subproblem, auxiliary subproblem, and feasibility subproblem. However, this method is to repair vulnerabilities, and the repair performance needs to be further analyzed. In reference [7], network security is one of the main issues to be considered when designing a smart grid communication network (SGCN). However, due to the openness and unpredictability of wireless network, it is often vulnerable to different antagonistic attacks. It uses vulnerabilities to launch cross-layer attacks in the process of data transmission. In order to solve this problem, a new trust-based routing framework is proposed. The framework uses Bayesian reasoning to calculate direct trust and, combined with the evidence of reliable neighbors, uses the D-S theory to calculate indirect trust. Analytic hierarchy process (AHP) calculates the credibility and trust of nodes by using cross-layer indicators such as transmission rate, buffer capacity, and received signal strength. In addition, combining the fuzzy theory with BDS-AHP, fairness is considered when cross-layer metrics are used to calculate link trust to achieve reliable routing. Through a large number of experiments, the performance of the proposed fuzzy-based trusted routing algorithm (FBDS-AHP) when malicious nodes initiate packets is evaluated. However, when calculating the link trust, this method is vulnerable to interference nodes, resulting in the decline of repair performance.

It can be seen from the analysis that the traditional method uses the electromagnetic wave diversity reception predictive control algorithm to repair the buffer overflow [8]. Because the diversity equalizer in QoS space shares the same channel with the diversity receiver, the edge space gain is very large. For QoS requirements, a cellular network routing buffer overflow recovery algorithm based on node residual energy homomorphism analysis is proposed to improve the algorithm. The performance test by experiment shows the superiority of the algorithm in this article.

2. D2D Communication Routing Buffer Overflow in Cellular Network

Establishing network topology, solving robust coefficient, and updating routing are the whole process of graph routing buffer overflow repair. In the cellular network D2D communication system, there are many routing hole space branches, resulting in poor repair effect of D2D communication routing buffer overflow. First, according to the cellular network D2D communication structure, the initial network topology is obtained, as shown in Figure 1.

In order to ensure the safe operation of cellular network D2D communication, each link index in the path must be optimized by calculating the path index of the cellular network D2D communication protocol [9]. The D2D communication protocol path index of the cellular network selected in this article includes the reliability index and the stability index. Among them, the reliability index can judge the quality of cellular network D2D communication, and the stability index can judge the data transmission stability of cellular network D2D communication.

The reliability index can be subdivided into link quality and path quality, and these two indexes can be expressed by calculation. If the expression of link quality reliability index is L, the following formula can be obtained:

$$L = \frac{N}{X},\tag{1}$$

where N refers to the number of communication information correctly received by the link and X is the total number of messages sent by the sender. Through formula (1), the link quality reliability is obtained to reflect the periodic propagation correct probability of cellular network D2D communication. On this basis, let the expression of path quality reliability index be P, and the following formula can be obtained:

$$P = \prod L^K,\tag{2}$$

where K refers to the first transmission success rate of cellular network D2D communication. The D2D communication reliability index of the cellular network can be calculated by formula (1) and formula (2). For the calculation of stability index, this article adopts the link failure estimation model, sets up the evaluation node in the cellular network D2D communication, and takes the node as the center to establish the spatial three-dimensional coordinate axis. The link failure time is determined by calculating the
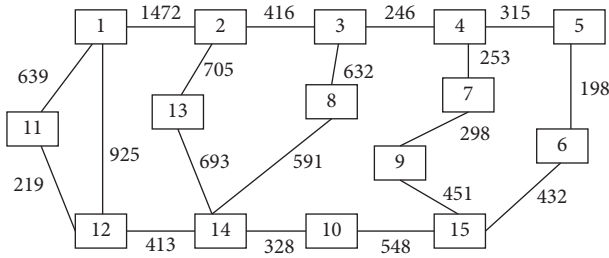
FIGURE 1: Communication network topology.

distance between the node and the communication transmission speed vector of cellular network D2D [10]. Using the cosine theorem, this process can be expressed by calculation. If the objective function is T, the following formula can be obtained:

$$T = \frac{d \cos \theta + \sqrt{R^2 - (d \sin \theta)^2}}{|v|},$$ (3)

where d refers to the communication range radius of cellular network D2D; R refers to the coordinates of the evaluation node; v refers to the transmission speed of cellular network D2D communication; and θ is the angle of transmission motion. According to formula (3), when θ is greater than 90°, the distance between the node and the communication transmission speed vector of cellular network D2D is far, and the link failure time is short; when θ is less than 90°, the distance between the node and the communication transmission speed vector of cellular network D2D is close, and the link failure time is long. It can be seen that the greater the T value, the higher the communication stability of cellular network D2D; On the contrary, the smaller the T value, the lower the communication stability of cellular network D2D. The correlation index of cellular network D2D communication refers to the correlation degree before reliability and stability. Generally, it can be expressed by intersecting path and disjoint path. The intersecting path indicates that the cellular network D2D communication is greatly disturbed by the outside world and there is a risk of chain breaking, resulting in low reliability and stability of cellular network D2D communication [11]. Disjoint path means that the cellular network D2D communication is less disturbed by the outside world and there is no risk of chain breaking, resulting in high reliability and stability of cellular network D2D communication.

A program is usually composed of multiple subroutines (modules). The larger the program size, the more the modules. After compiling a program, it is divided into three areas in the memory to store the program code area, data area, and stack area. Among them, various variables and buffer areas defined in the program are stored in the data area, and the return address of the calling subroutine is stored in the stack area. When the control returns from the subroutine, return to the main control program according to the address indicated at the top of the stack.

When writing a program in the C language, the input function is generally used to obtain the name entered by the user. When the user tries to input more name characters that can be processed by the buffer into the buffer, if there is no buffer cross-border check mechanism in the program, the buffer overflow will occur when an ultralong string is input. For example, if char name is defined in the program written by the programmer, when a string longer than 15 bytes is input, the character buffer will overflow. By judging the reliability index and stability index, it is judged that there is a vulnerability in the cellular network D2D communication, and then, the repair is completed according to the node residual energy analysis.

3. Homomorphism Analysis of Node Residual Energy

3.1. User Transmit Power Adjustment. In the D2D communication system, the communication link between adjacent users is allowed to be established directly, so as to complete the transmission of information. In the process of establishing a direct link, although it does not pass through the relay forwarding of the base station [12], the base station needs to control the whole process through the center or auxiliary control. In terms of network control, the D2D communication network architecture can be divided into network center control network architecture and network auxiliary control network architecture. The D2D network communication architecture is given in Figure 2.

3.1.1. Network Center Control Network Architecture. In the network center control network architecture, the whole process of D2D communication is controlled by the network management center (base station). According to the user's environment, the base station can quickly complete the user discovery process and judge whether adjacent users establish a direct communication link. Before determining to establish a direct communication link, the base station also needs to allocate communication resources to D2D users and configure network parameters. In addition, under the centralized control mode, the base station can efficiently manage the interference between users [13, 14], which is conducive to improving the network performance. However, this central control mode requires a lot of signaling overhead, which is difficult to obtain in the actual wireless network.

3.1.2. Network Auxiliary Control Network Architecture. In the network-aided control network architecture, firstly, for D2D users, it needs to be able to perceive the relevant information of other users and the use of network resources according to its own business needs, so as to independently complete device discovery and dynamically adjust its transmission power. In addition, the base station can master the information in the whole network [13], assist D2D users in resource allocation, identity authentication and other functions, and then establish a direct communication link between users. In this way, D2D autonomy has distributed characteristics, and its communication process can be completed with a small amount of signaling overhead.

(a)   (b)
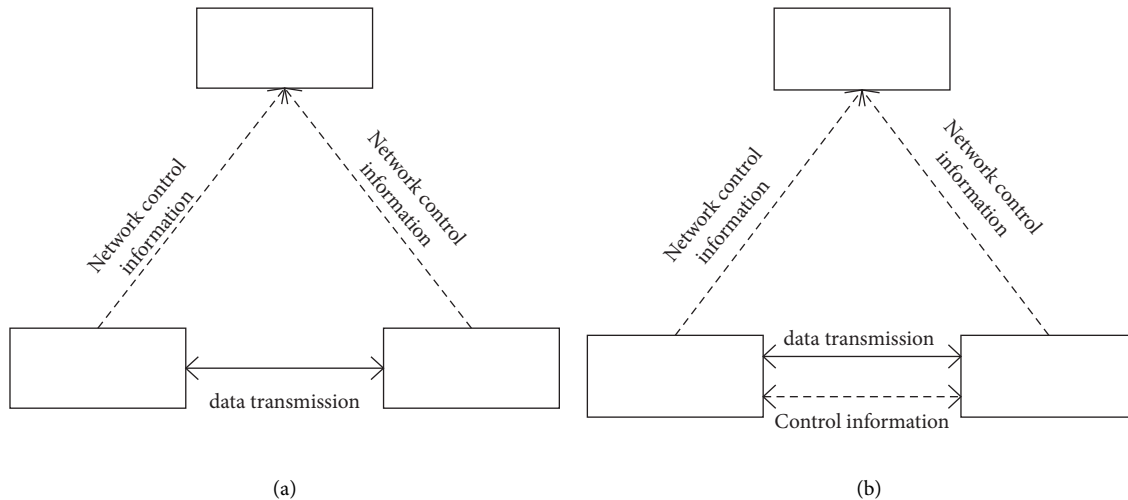
FIGURE 2: D2D communication network architecture. (a) Network center control network architecture. (b) Network auxiliary control network architecture.

Moreover, through the auxiliary role of the base station, the uncontrollability caused by the complete autonomy of D2D communication is avoided. However, compared with the network center control network architecture, this network architecture is not efficient enough.

There are three working modes of D2D communication network:

(1) Cellular Mode. The cellular mode here is the same as the traditional cellular mode. Because users do not have the conditions to establish a direct communication link, users can only complete data transmission through the base station. Compared with the D2D communication mode [15], the cellular mode avoids interference between users, but the spectrum resource utilization is low. The specific structure is shown in Figure 3.

(2) Private Mode. For the dedicated mode, two adjacent users who are qualified to establish D2D communication links can directly establish communication links to complete the transmission of information. However, the reserved spectrum resources need to be allocated to D2D users, and these reserved frequency bands, including uplink and downlink frequency bands, are orthogonal to each other, which can avoid mutual interference between users. Moreover, compared with the cellular mode [16], the dedicated mode only needs one downlink spectrum or one uplink spectrum to complete D2D communication, hence high resource utilization.

(3) Sharing Mode. In the sharing mode, D2D users multiplex the authorized frequency band in the cellular network for data transmission. Moreover, D2D users can multiplex the spectrum resources of downlink users and uplink users. Compared with cellular mode and dedicated mode, shared mode has the advantage of high-frequency spectral efficiency. However, D2D multiplexing the spectrum resources of cellular users will cause interference between two types of
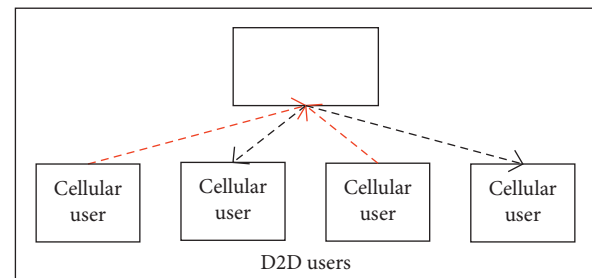
FIGURE 3: Cellular mode.

users. Effective resource allocation algorithm and interference management scheme are needed to avoid or reduce these interferences [17, 18].

In a complete D2D communication process, it mainly includes three steps: (1) device discovery, (2) communication establishment, and (3) data transmission. A cellular user will find the users communicating with it autonomously or through the base station according to their own business needs and the relevant information of other users, then complete the establishment of communication link under the control of the base station, and finally realize the data transmission of two adjacent users. The specific implementation process is shown in Figure 4. Taking two adjacent users as an example, the whole implementation process of D2D communication is analyzed and studied.

(1) User A prepares to communicate with user B according to his own needs. At this time, user A sends a request to establish a communication link to the base station in his cell, which includes some information of user B [19].

(2) After receiving the request from user A, the base station starts to collect the relevant information of user B and find the location of user B, and then, the base station sends D2D communication measurement signaling to two users [20].
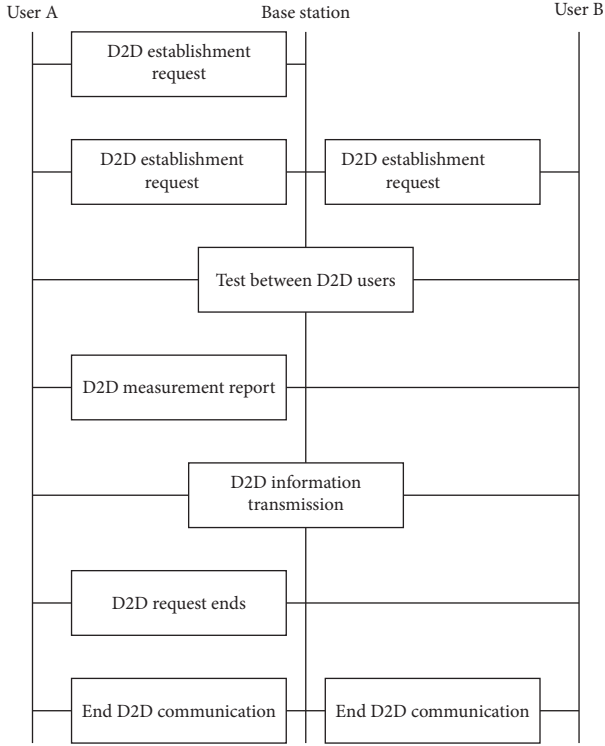
Figure 4: D2D communication implementation flowchart.

(3) After receiving the measurement signaling, user A and user B start to measure the channel state information between their two users, which is used by the base station to judge whether the conditions for establishing D2D communication link can be met between the two users, send the test results to the base station through user A or user B, and record the test results.

(4) After receiving the channel state information between the two users, the base station judges whether user A and user B can establish a D2D communication link according to the predetermined conditions for establishing a D2D communication link. If the conditions are true, the base station sends the judgment result to the two users.

(5) After receiving the notification, user A and user B establish a direct communication link to complete data transmission.

(6) After user A and user B complete data transmission, the user sends a request to the base station to end D2D communication.

(7) After receiving the request to end D2D communication, the base station will change the network parameter configuration, withdraw the determination information of D2D communication between user A and user B, and end D2D communication between user A and user B.

In the D2D communication system, compared with unicast communication mode, multicast communication mode is that multiple D2D users are combined together, and

they share the same content and can transfer information to multiple D2D users at the same time. In addition, due to the introduction of multicast technology, D2D users can form multicast communication with multiple users and transmit the same data resources to other users at the same time, which can improve the utilization efficiency of network resources and reduce the time of establishing communication between D2D users. As shown in Figure 5, multiple D2D users form a multicast communication group. In each multicast communication group, one user can transmit data to one of them or send data to multiple D2D users at the same time.

(4) Broadcast Communication Mode. D2D broadcast communication means that when a D2D user transmits information, other users away from the user can receive the information.

With the development of communication technology and the improvement of people's living standards, the number of users in the communication network is increasing rapidly, and the load of the base station is increasing. As a result, the base station cannot guarantee to provide good service quality for users within the coverage. As one of the key technologies of LTE and future 5G, D2D communication allows adjacent devices to communicate directly under the control of the cellular system. The data between users does not go through the base station routing, which can effectively reduce the base station load [21] and shorten the end-to-end delay. In addition, the D2D communication under the cellular network allows D2D users to reuse channel resources of cellular users, thereby significantly improving high-frequency spectrum utilization and system capacity. D2D communication technology not only brings many benefits, but also brings new problems, that is, CO frequency interference caused by spectrum sharing. Therefore, effective interference management and resource allocation schemes must be developed to solve the above interference problems [22] and maximize the throughput of D2D links in the system on the premise of resource constraints.

3.2. Design of Electromagnetic Wave Multiple Routing Protocol in Cellular Network. The adopted channel model not only considers the path loss, but also considers the fast fading caused by multipath effect and the slow fading caused by shadow effect. Therefore, the channel gain between the cellular user and the D2D user receiver can be expressed as

$$h_{cd} = k \cdot \delta_{\text{cld}} \cdot \xi_{\text{cld}} \cdot d_{\text{cld}}^{-\alpha}, \tag{4}$$

where k represents the path loss coefficient; α represents path loss index; δ_{cld} represents the fast fading factor obeying exponential distribution; ξ_{cld} represents the slow fading factor obeying log positive attitude distribution; and $d_{\text{cld}}^{-\alpha}$ represents the distance between the cellular user and the D2D user receiving end.

It is assumed that there are D pairs of D2D users in the system, and C cellular users use mutually orthogonal channels for communication. When D2D user d multiplexes
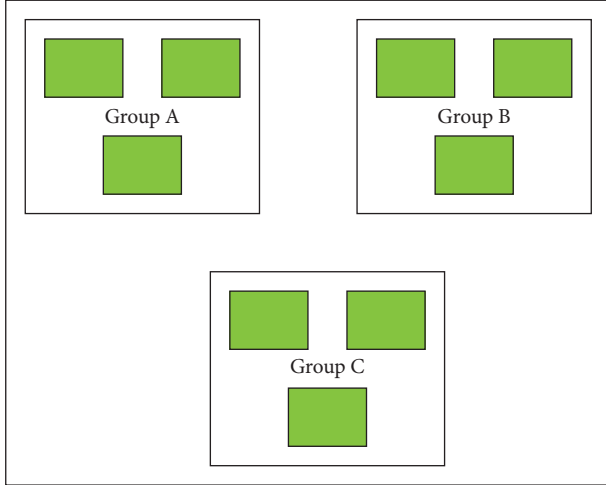
FIGURE 5: D2D multicast communication mode.

the channel resources of cellular user i, the signal-to-interference noise ratio and rate of cellular user i are, respectively,

$$\gamma_i = \frac{P^i h_{cb}}{P_d^i h_{dc} + N_0},$$

$$r^i = \log_2 (1 + \gamma_i). \tag{5}$$

The SINR and rate of D2D user d are

$$\gamma_d^i = \frac{P_d^i h_{dd}}{P^i \log_2 (1 + \gamma_d^i)},$$

$$r_d^i = \log_2 (1 + \gamma_d^i). \tag{6}$$

When D2D users and cellular users share the same channel, cellular users and D2D users will produce common channel interference, and the closer the distance, the greater the interference. Generally, the number of users performing D2D communication at the same time in the cell is less than the number of cellular users. Therefore, there may be multiple cellular channel resources as potential multiplexing objects for D2D users. In order to ensure the QoS requirements of D2D users, the base station allows D2D users to reuse the channel resources only when the interference of cellular users to D2D users is less than a certain threshold. This section defines the areas where cellular users are restricted from multiplexing:

$$P_{c,\max} h_{cd} \leq I_{c,d}. \tag{7}$$

Given the channel resource set multiplexed by D2D users, the transmission power of D2D users is optimized through power control to maximize the throughput of D2D link under resource constraints [23]. From the previous step, the reusable potential cellular channel of any D2D user D can be determined, and it can be obtained that the set of
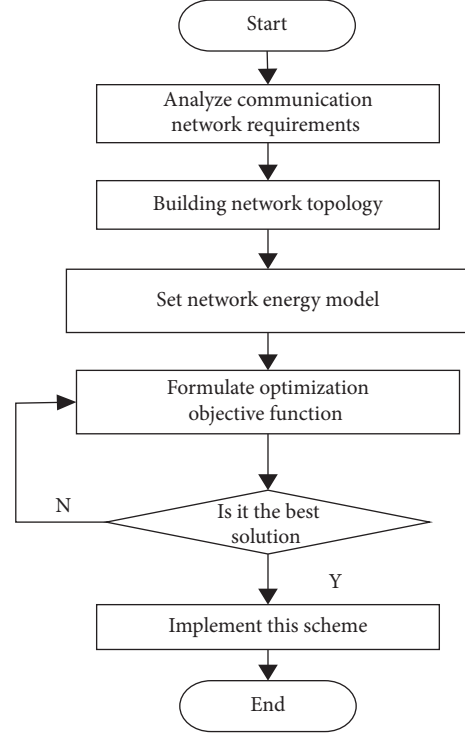
FIGURE 6: Route optimization process of the cellular network D2D communication network.

resources D_d can reuse is S_d. Therefore, the optimization objective can be expressed as

$$\begin{cases} \max u_d^k = \sum_{i \in S_{d,k}} \log_2 \left(1 + \frac{P_d^i h_{dd}}{P^i \log_2 (1 + \gamma_d^i)} \right) & 0 \leq P_d^i \leq P_{c,\max}, \\ \sum_{i \in s_{d,k}} \leq P_{c,\max}. \end{cases} \tag{8}$$

According to the optimization objectives and constraints, the water injection algorithm is used to allocate power to D2D users. According to the Karush–Kuhn–Tucker (KKT) condition,

$$L(P_d^i, \alpha, \beta) = \sum_{i \in S_{d,k}} \log_2 \left(1 + \frac{P_d^i h_{dd}}{P^i \log_2 (1 + \gamma_d^i)} \right)$$

$$- \alpha (P_d^i - P_{c,\max}) - \beta \left(\sum_{i \in S_{d,k}} P_d^i - P_{c,\max} \right). \tag{9}$$

To calculate the partial derivative, there are

$$P_d^i = \left[\delta - \frac{1}{H} \right]^+, \tag{10}$$

of which

$$[\varphi]^+ = \max\{0, \varphi\},$$

$$H = \frac{h_{dd}}{P^i \log_2\left(1 + \gamma_d^i\right)}. \tag{11}$$

Through the power control in the previous step, we can know the maximum throughput and corresponding transmission power of any resource set multiplexed by D2D users at a set time. At the same time, in the process of channel allocation, the power adjustment of users on different channels will be completed through the scheme of power control part.

3.3. Design of Routing Optimization Method of Cellular Network D2D Communication Network Based on AHP.

The traditional daily cellular network D2D communication network routing often has the problem of excessive energy consumption in use, which affects the stability of the cellular network D2D communication network. Therefore, set the cellular network D2D communication network routing optimization method based on AHP to deal with it. The specific routing optimization process is shown in Figure 6.

The routing in the communication network is processed in detail according to the steps set in 6. In order to ensure the reliability of the optimization results, the calculation accuracy is set in the calculation process.

3.3.1. Building a Cellular D2D Communication Network Energy Model.

In order to control the energy consumption of the communication network, a cellular network D2D communication network energy model is constructed. Before building the energy model, complete the setting of the network model. The nodes in this network model have the following characteristics: the node number is single and random [24, 25]. All nodes have corresponding fixed positions, and node energy is limited. The location of the base station is fixed and is not considered in the energy model. All nodes in the network are isomorphic nodes with the same initial energy and communication radius. All nodes have the ability to directly realize data transmission with the base station. Nodes have certain data integration ability, and their current residual energy can be obtained through calculation [26, 27]. The data transmission capacity and power of all nodes can be adjusted, and the data fusion technology can be used to reduce the amount of data transmission. The nodes can determine the transmission distance according to the strength of signal reception.

The first-order radio model is used as the basis of model design, and it is set as the form of free space model according to the characteristics of the cellular network D2D communication network [28]. Its network architecture has the following characteristics: when the distance between data transmitting node r_1 and data receiving node r_2 is less than d_1, the energy consumption of data transmission power amplification is directly proportional to d_1^2. Set the transmitting node to send the data of o bits to the receiving node with distance d_2. The energy consumption consists of

transmission power consumption and power amplification consumption, which can be expressed as follows:

$$W_i(h, d) = \begin{cases} hW_{\text{elec}} + hW_f d^2, \\ hW_{\text{elec}} + hW_{\text{amp}} d^4, \end{cases} \tag{12}$$

where W_{elec} represents the energy consumption of the transmitting circuit or receiving circuit when transmitting or receiving 1 bit data information and the free space signal amplification factor is W_f [29,30]. When there are too many channels in the communication network, the signal amplification factor is W_{amp}, d represents the distance between the transmitting node and the receiving node, and the data length in the communication network is set to h bit, and then d_1 is

$$d_1 = \sqrt{\frac{\alpha_f}{\alpha_{mp}}}, \tag{13}$$

where α represents the straight-line distance of the node. The energy consumption formula of the node receiving h bit data can be obtained from the above formula:

$$W_i(h) = W_{\text{re-elec}}(h) = hW_{\text{elec}}. \tag{14}$$

The node in the communication network transmits the collected data information to the cluster head node, performs data fusion through this node, and finally sends the fused data to the base station [31]. The process of data information fusion also needs to consume a lot of energy, so the energy consumption model of the D2D communication network is established as the following formula :

$$W_{DA}(h) = h \sum_i W_i(h), \tag{15}$$

where W_{DA} represents the energy consumption of fused h bits data and DA represents the data fusion coefficient. So far, the construction of the energy consumption model of cellular network D2D communication network is completed, and the optimization of network routing is realized on the basis of this model.

3.3.2. Setting Routing Optimization Objective Function.

Since the structure of the communication network itself is difficult to change in practical application, in order to reduce the impact of link interruption on network operation, corresponding optimization strategies are proposed [32, 33]. Set the objective function in the optimization process as

$$\min\left(\beta \overline{T} + \chi GJ\right), \tag{16}$$

where $\overline{T}$ represents the average transmission time of data information, T_i represents the transmission delay of data packet i, H represents the number of data packets, GJ is set as the service balance of the whole network, and the average data transmission delay of the communication network is

$$\overline{T} = \sum_{i=1}^{H} \frac{T}{H_i}, \tag{17}$$

where H_i represents the number of services on the i-th link, m represents the number of links in the communication network, $T_{\max}$ represents the longest delay in data transmission in the communication network, $H_{\max}$ represents the maximum number of services carried by the communication network link, and β and χ are constant terms in the calculation process. The service balance degree of the communication network can be obtained by the following formula :

$$GJ = \sqrt{\sum_{i=1}^{m} \frac{\left(H_i - \sum_{i=1}^{H} H_i\right)^2}{m}}. \tag{18}$$

The energy consumption and data transmission delay are set as the routing optimization objectives. It is applied to the model set above to remove the interrupted link x and obtain a new network topology. The shortest path algorithm is used to find the shortest transmission path for packet A [34, 35], then the optional path set of packet can be expressed as $P(p_i, \quad i = 1, 2, \ldots, n)$, and the fuzzy analytic hierarchy process is used to obtain the optimal communication network routing combination P_s. On the basis of the above calculation, the adjacent nodes with the shortest route are introduced to generate the subshortest route path and finally form the shortest route set as an alternative to route optimization. In the process of communication, when the signals are separated, the high-order Bessel function detection and control technology are introduced to make the network branch signals add in phase during communication, offset the phase offset of signal fading in the process of network communication, and finally achieve the purpose of routing buffer overflow repair.

4. Experiment and Performance Test

In order to verify the effect of this algorithm in repairing routing buffer overflow in the cellular network, experiment verification is carried out. The TOSSIM simulator of the TinyOS system is used to construct the cellular network experiment model, and the delay, energy efficiency, and the number of contracts are used to test the superiority of the method proposed in this article. In the experiment parameter setting, the routing Grace period is 2 ms, and the mac-pr deployment is adjusted. The network routing frequency is 7.8 Hz, and only 5.5 s is relayed each time. The mac-pr can accommodate 400 nodes. The network routes under different bandwidths, CLS lengths, and power losses are selected. See Table 1 for specific parameters.

4.1. Node Residual Energy. The test results of node residual energy of the automatic repair method for D2D communication routing buffer overflow vulnerability in the cellular network designed in literature [4], literature [7], and this article are shown in Figure 7.

According to the two test results shown in Figure 7, it can be seen that in the method of reference [4], due to the failure of the inner ring node, the distance of the outer ring node sending information is too large, and its energy is

TABLE 1: Network routing design parameters.

Paragraph A (ECA test indicators)		Paragraph B (ECA test indicators)	
Parameter	Index	Parameter	Index
Frequency	7.8 Hz	Frequency	7.8 Hz
Bandwidth	130 Mhz	Bandwidth	80 Mhz
CLS length	1.9 ms	CLS length	1.1 ms
Breakpoint loss	11%	Breakpoint loss	13%
WSN receiving antenna	LC	WSN receiving antenna	LA
Receiving rate	93%	Receiving rate	96%

exhausted first. The method in reference [7] does not cluster in the aggregation area, so the overhead is controlled and the protocol energy consumption is improved. This protocol uses the ant algorithm to search the communication route and find the global optimal transmission path.

4.2. Routing Buffer Overflow Repair Results under Improved Algorithm. Based on the above experiment environment and parameter design, the QoS oriented experiment of magnetic wave communication and routing buffer overflow repair in cellular network is carried out. The algorithm in this article is used to repair the overflow of D2D communication routing buffer in a slow cellular network. The repair results are shown in Figure 8.

By analyzing Figure 8, it can be seen that the amplitude corresponding gain is the standard to measure the result of routing buffer overflow repair. After using this algorithm to repair the routing buffer overflow in the cellular network, the amplitude corresponding gain is improved compared with that before repair, which shows that the repair result is ideal, and proves the effectiveness of the routing buffer overflow repair algorithm proposed in this article.

Under the condition of the same number of iterations, the network communication packet loss rate before and after repair is used as the test index to obtain the comparison results of the network communication packet loss rate before and after communication network routing buffer overflow repair, as shown in Table 2.

Summarize the data in Table 2 and draw a broken line diagram of the network communication packet loss rate before and after communication network routing buffer overflow repair, as shown in Figure 9.

Analyzing Table 2 and Figure 9, with the increase in the number of iterations, the packet loss rate of cellular network D2D communication before and after routing buffer overflow repair also gradually decreases. However, under the same number of iterations, the packet loss rate of network communication after routing buffer overflow repair using this algorithm is always lower than that without repair. The average packet loss rate after routing buffer overflow repair using this algorithm is 19%, whereas the average packet loss rate without routing buffer overflow repair is 30%. Overall, it shows that the optimization of D2D communication routing buffer overflow repair in the cellular network can effectively reduce the packet loss
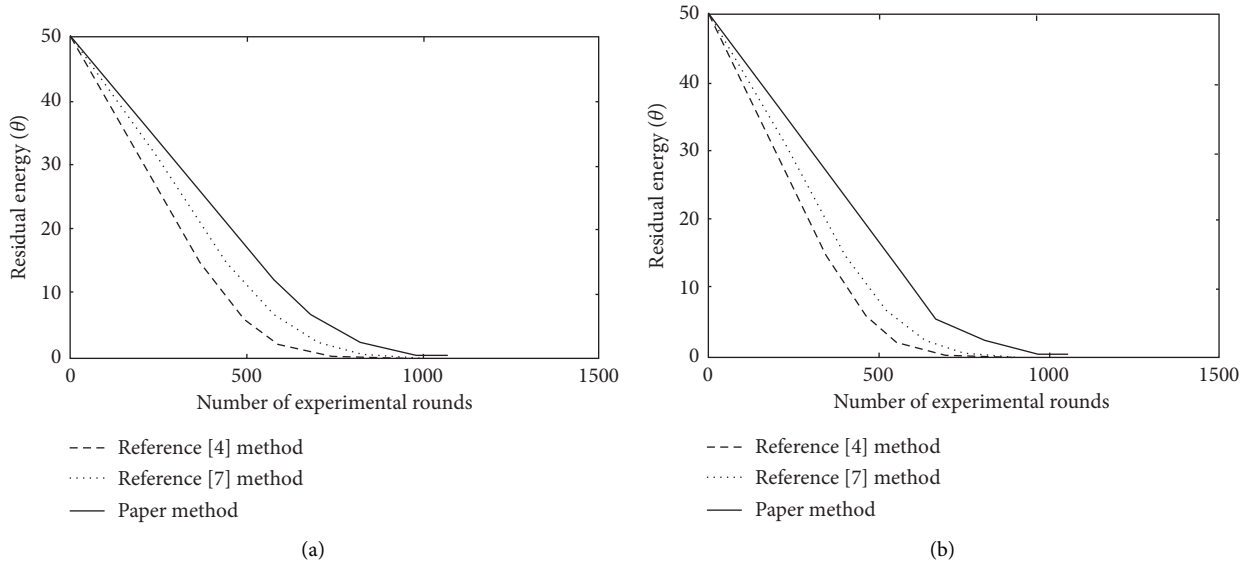
Figure 7: Comparison of θ value of node residual energy. (a) First test. (b) Second test.
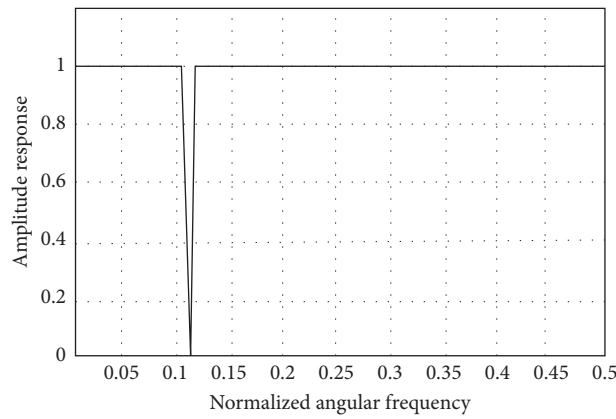
Figure 8: Routing buffer overflow repair results.

Table 2: Comparison results of the network communication packet rate before and after repair.

Number of iterations	Using buffer repair (%)	Buffer repair not used (%)
0	52	85
10	43	55
20	33	42
30	20	37
40	18	32
50	15	20
60	12	18
70	5	15
80	3	12
90	2	10
100	1	3

rate of communication data, improve the anti-interference ability of the system, ensure the network communication security, and meet the QoS standards and requirements.

4.3. Communication Performance Comparison under Overflow Repair. Under different experimental times, the communication performance of cellular network D2D before
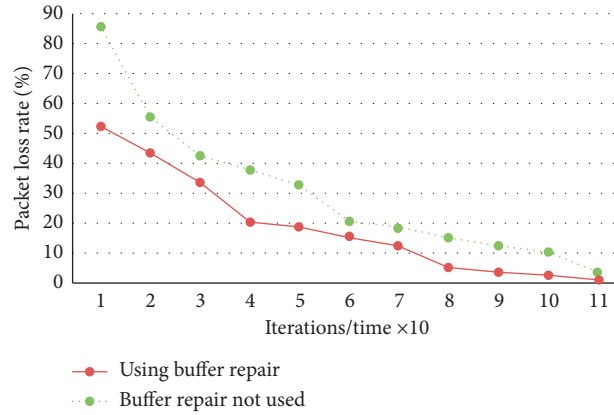
Figure 9: Broken line diagram of network communication packet loss rate comparison.

Table 3: Communication performance after the repair of routing buffer overflow in the communication network.

Number of experiments (times)	Blocking rate (%)	Equilibrium (%)	Stability (%)
10	0.5	97	96
20	0.8	98	97
30	0.6	96	95
40	0.4	95	96
50	0.7	97	97
60	0.6	98	98
70	0.5	96	94
80	0.3	98	96

Table 4: Communication performance of the communication network before routing buffer overflow repair.

Number of experiments (times)	Blocking rate (%)	Equilibrium (%)	Stability (%)
10	4	69	71
20	3	68	72
30	5	65	74
40	2	63	73
50	6	67	71
60	4	69	72
70	5	68	74
80	3	66	75

and after routing buffer overflow repair is compared, and the communication blocking rate, communication route data distribution balance of cellular network D2D, and communication route data stability of cellular network D2D are compared. The results are shown in Tables 3 and 4.

By analyzing Tables 3 and 4, it can be seen that in terms of communication blocking rate, the communication blocking rate before routing buffer overflow repair is large, and the average communication blocking rate is 4%, while the blocking rate after routing buffer overflow repair is significantly reduced, and the average communication blocking rate is only 0.55%, which is 86.5% lower than that before repair; In terms of routing data allocation balance, the routing data allocation balance before routing buffer overflow repair is small, and the average routing data allocation balance is 66.9%, while the routing data allocation balance after routing buffer overflow repair is significantly increased, and the average routing data allocation balance is 96.9%,

which is 44.8% higher than that before repair. In terms of routing data allocation stability, the routing data allocation stability before routing buffer overflow repair is small, and the average routing data allocation stability is 72.6%, which is poor. After routing buffer overflow repair, the routing data allocation stability has been significantly improved, and the average routing data allocation stability is 96.1%, which is 32.3% higher than that before repair. It shows that the stability of routing data distribution after repair is good, because the cellular network routing buffer overflow repair algorithm based on node residual energy homomorphism analysis is proposed, the cellular network routing buffer overflow repair algorithm based on homomorphic analysis of node residual energy is proposed, the cellular network D2D communication routing protocol is designed, the cellular network D2D communication protocol path index is determined, and then the cellular network D2D communication protocol is designed by analyzing node residual

energy, and the cellular network D2D communication network energy model is constructed. Solve and set the routing optimization objective function to control the network routing and repair the buffer overflow.

5. Conclusion

In the cellular electromagnetic wave communication system, because the routing protocol evolves with the mobile location algorithm, the location routing cannot effectively meet the QoS requirements, resulting in the buffer overflow problem, which needs to be repaired. The traditional method uses the predictive control algorithm of electromagnetic wave diversity reception to repair the buffer overflow. Because the diversity equalizer in QoS space shares the same channel with the diversity receiver, the edge space gain is very large. For QoS requirements, a cellular network routing buffer overflow repair algorithm based on node residual energy homomorphism analysis is proposed to determine the path index of cellular network D2D communication protocol, then design the cellular network D2D communication routing protocol, build the cellular network D2D communication network energy model, solve and set the routing optimization objective function, and realize the control of network routing, Fix buffer overflow. The experiment results show that the proposed method can effectively reduce the packet loss rate of communication data, improve the communication performance and anti-interference ability of cellular network D2D communication, and ensure the communication security of the cellular network, and has a good application value.

Data Availability

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Conflicts of Interest

The authors declare that they have no conflicts of interest regarding this work.

Acknowledgments

References

[1] N. Anjum, Z. Yang, H. Saki, M. Kiran, and M. Shikh-Bahaei, "Device-to-device (d2d) communication as a bootstrapping system in a wireless cellular network," *IEEE Access*, vol. 7, no. 99, pp. 6661–6678, 2019.

[2] G. L. Shen, "Vulnerability of vulnerability defense control simulation of network system resource data," *Computer Simulation*, vol. 37, no. 4, pp. 308–311, 2020.

[3] F. T. Zuhra, K. Bakar, A. Ahmed, K. M. Almustafa, and N. Islam, "Lltp-qos: low latency traffic prioritization and qos-aware routing in wireless body sensor networks," *IEEE Access*, vol. 7, no. 99, pp. 152777–152787, 2019.

[4] F.-J. Gao, Y. Wang, L.-Z. Wang, Z. Yang, and X.-D. Li, "Automatic buffer overflow warning validation," *Journal of Computer Science and Technology*, vol. 35, no. 6, pp. 1406–1427, 2020.

[5] H. Rong, Z. Wang, H. Jiang, Z. Xiao, and F. Zeng, "Energy-aware clustering and routing in infrastructure failure areas with d2d communication," *IEEE Internet of Things Journal*, vol. 6, no. 5, pp. 8645–8657, 2019.

[6] A. Ibrahim, T. Ngatched, and O. A. Dobre, "Using bender's decomposition for optimal power control and routing in multihop d2d cellular systems," *IEEE Transactions on Wireless Communications*, vol. 18, no. 99, p. 1, 2019.

[7] D. Velusamy, G. Pugalendhi, and K. Ramasamy, "A cross-layer trust evaluation protocol for secured routing in communication network of smart grid," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 1, pp. 193–204, 2020.

[8] G. V. Abgaryan, "On the resonant passage of electromagnetic wave through waveguide with diaphragms," *Lobachevskii Journal of Mathematics*, vol. 41, no. 7, pp. 1315–1319, 2020.

[9] I. Singh and N. P. Singh, "Coverage probability analysis of device-to-device communication underlaid cellular networks in uplink over $\kappa - \mu$ κ - μ/$\eta - \mu$ η - μ fading channels," *International Journal of Wireless Information Networks*, vol. 26, no. 1, pp. 39–47, 2019.

[10] O. A. Amodu, M. Othman, N. K. Noordin, and I. Ahmad, "Transmission capacity analysis of relay-assisted d2d cellular networks with interference cancellation," *Ad Hoc Networks*, vol. 117, no. 11, Article ID 102400, 2021.

[11] M. H. Khoshafa, T. Ngatched, M. H. Ahmed, and A. Ibrahim, "Improving physical layer security of cellular networks using full-duplex jamming relay-aided d2d communications," *IEEE Access*, vol. 8, no. 99, p. 1, 2020.

[12] A. N. Rao, B. R. Naik, and L. N. Devi, "On the relay node placement in wsns for lifetime maximization through metaheuristics," *Materials Today: SAVE Proceedings*, vol. 4, 2020.

[13] K. D. Rahi and Y. M. Hasan, "Utilizing device to device communication in cellular networks with the distance-based model selection method," *IOP Conference Series: Materials Science and Engineering*, vol. 870, Article ID 012116, 2020.

[14] A. Alnoman and A. S. Anpalagan, "Computing-aware base station sleeping mechanism in h-cran-cloud-edge networks," *IEEE Transactions on Cloud Computing*, vol. 9, no. 3, pp. 958–967, 2019.

[15] J.-A. He, L. Jia, L. Xu, and W. Chen, "Throughput maximization for multiple d2d group communications underlaying cellular networks," *Wireless Communications and Mobile Computing*, vol. 2020, Article ID 8833579, 10 pages, 2020.

[16] H. H. Su, W. B. Qu, and Y. Peng, "Uplink and downlink throughput optimization scheme for millimeter wave d2d communication," *Procedia Computer Science*, vol. 166, pp. 551–556, 2020.

[17] V. Sathya, S. M. Kala, S. Bhupeshraj, and B. R. Tamma, "Raptap: a socio-inspired approach to resource allocation and interference management in dense small cells," *Wireless Networks*, vol. 27, no. 1, pp. 441–464, 2021.

[18] S. K. Kasi, I. H. Naqvi, M. K. Kasi, and F. Yaseen, "Interference management in dense inband D2D network using spectral

clustering & dynamic resource allocation," *Wireless Networks*, vol. 25, no. 7, pp. 4431–4441, 2019.

[19] L. B. Kumar and P. Krishnan, "Multi-hop convergent fso-uwoc system to establish a reliable communication link between the islands - sciencedirect," *Optics Communications*, vol. 474, 2020.

[20] D. T. Do, M. Nguyen, T. A. Hoang, and B. M. Lee, "Exploiting joint base station equipped multiple antenna and full-duplex d2d users in power domain division based multiple access networks," *Sensors*, vol. 19, no. 11, 2019.

[21] M. Aditya, H. Pancholi, P. Priyanka, and G. S. Kasbekar, "Beyond the vcg mechanism: truthful reverse auctions for relay selection with high data rates, high base station utility and low interference in d2d networks," *Wireless Networks*, vol. 26, no. 6, 2020.

[22] I. Budhiraja, N. Kumar, and S. Tyagi, "Energy-delay tradeoff scheme for NOMA-based D2D groups with WPCNs," *IEEE Systems Journal*, vol. 9, no. 99, pp. 1–12, 2020.

[23] Y. Sun, M. Miao, Z. Wang, and Z. Liu, "Resource allocation based on hierarchical game for d2d underlaying communication cellular networks," *Wireless Personal Communications*, vol. 117, no. 2, pp. 281–291, 2021.

[24] M. Deypir and F. Gouya, "Multi-node d2d communications for wireless video delivery over cellular networks," *AEU - International Journal of Electronics and Communications*, vol. 110, Article ID 152863, 2019.

[25] M. Susanto, A. Abadi, Herlinawati, and A. Trisanto, "Uplink power control based on sinr for d2d enabled in cellular communication network," *Journal of Physics: Conference Series*, vol. 1376, Article ID 012019, 2019.

[26] R. A. Kumar and K. S. Prasad, "Performance analysis of gfdm modulation in heterogeneous network for 5g nr," *Wireless Personal Communications*, vol. 439, 2020.

[27] H. V. Vu and T. Le-Ngoc, "Performance analysis of underlaid full-duplex d2d cellular networks," *IEEE Access*, vol. 7, pp. 176233–176247, 2019.

[28] A. Yx, A. Hz, and B. Pr, "Unified scheduling for predictable communication reliability in cellular networks with d2d links," *Computer Communications*, vol. 167, pp. 1–14, 2021.

[29] S. Kubota, H. Kawaki, B. Perbal, K. Kawata, T. Hattori, and T. Nishida, "Cellular communication network factor 3 in cartilage development and maintenance," *Journal of Cell Communication and Signaling*, pp. 1–11, 2021.

[30] S. Wang and M. Di Renzo, "On the mean interference-to-signal ratio in spatially correlated cellular networks," *IEEE Wireless Communications Letters*, vol. 9, no. 3, pp. 358–362, 2020.

[31] W. Santipach and K. Jiravanstit, "On selecting transmission mode for d2d transmitter in underlay cellular network with multi-antenna base station," *Digital Communications and Networks*, vol. 3, no. 3, 2021.

[32] W. Baohua, F. Shuyi, W. Cheng, and W. Qiang, "Optimization strategy of anti-interference performance based on bpso," *Journal of Physics: Conference Series*, vol. 1693, Article ID 012061, 2020.

[33] K. Kopsidas and M. Abogaleela, "Utilizing demand response to improve network reliability and ageing resilience," *IEEE Transactions on Power Systems*, vol. 34, no. 3, pp. 2216–2227, 2019.

[34] A. Iguider, O. Elissati, A. En-Nouaary, and M. Chami, "Shortest path method for hardware/software partitioning problems," *International Journal of Information Systems and Social Change*, vol. 12, no. 3, pp. 40–57, 2021.

[35] J. Gokulraj, J. Senthilkumar, Y. Suresh, and V. Mohanraj, "Data prediction approaches for efficient data transmission using optimized leibler distance matrix-based data aggregation in wireless sensor network," *Journal of Ambient Intelligence and Humanized Computing*, vol. 3, no. 26, pp. 1–8, 2021.

Research Article

High Spatial Resolution Remote Sensing Data Classification Method Based on Spectrum Sharing

Meimei Duan [1] **and Lijuan Duan**[2]

[1]*School of Electronic and Control Engineering, North China Aerospace Industry Institute, Langfang, Hebei Province 065000, China*
[2]*Institute of Energy Internet Innovation, Tsinghua University, Beijing 100084, China*

Correspondence should be addressed to Meimei Duan; duanmm@nciae.edu.cn

Existing remote sensing data classification methods cannot achieve the sharing of remote sensing image spectrum, leading to poor fusion and classification of remote sensing data. Therefore, a high spatial resolution remote sensing data classification method based on spectrum sharing is proposed. A page frame recovery algorithm (PFRA) is introduced to allocate the wireless spectrum resources in low-frequency band, and a dynamic spectrum sharing mechanism is designed between the primary and secondary users of remote sensing images. Based on this, D-S evidence theory is used to fuse high spatial resolution remote sensing data and correct the pixel brightness of the fused multispectral image. The initial data are normalized, the feature of spectral image is extracted, the convolution neural network classification model is constructed, and the remote sensing image is segmented. Experimental results show that the proposed method takes shorter time and has higher accuracy for high spatial resolution image segmentation. High spatial resolution remote sensing data classification is more efficient, and the accuracy of data classification and remote sensing image fusion are more ideal.

1. Introduction

The spatial resolution of remote sensing images with high spatial resolution is greatly improved, which is fully reflected in the obvious internal differentiation of features, increased texture, rich details, and prominent edges [1]. Remote sensing images, especially high-resolution remote sensing images, have broad application prospects in land use and land cover change [2]. However, due to the uncertainty in the acquisition and processing of high spatial resolution remote sensing information, the classification accuracy of remote sensing data is difficult to meet the needs of land cover change, environmental monitoring, and thematic information extraction.

In order to improve this problem, scholars in relevant fields at home and abroad have also put forward some research results. Reference [3] proposed a remote sensing image classification algorithm combining IFCM (improved

FCM) clustering and variational inference. In the feature extraction stage, the spatial pixel template method is used to extract the pixel feature points, and the posterior distribution of parameters is approximated based on the variational inference method in Bayesian statistics to obtain the image classification results. In reference [4], a classification fusion algorithm based on machine learning is proposed. The classification fusion results of ranking level and measurement level are output, and the typical areas of Landsat 8 remote sensing images in Beijing are used for classification prediction. A new object-oriented classification method is proposed in reference [5], which uses the segmentation algorithm to perform the initial over segmentation of the original image. The segmentation unit with good homogeneity is obtained. The segmentation unit is taken as the object to be processed; Gravitational Self-Organizing Map (G SOM) is used to cluster the segmented objects, the clustering results are obtained, and the consistency function

is used to integrate the diverse clustering results at the least cost, so as to realize fast and automatic decision classification. In reference [6], a multisensor classification strategy based on deep learning integration process and decision fusion framework is studied. Random feature selection is used to generate two independent CNN-SVM integrated systems, one for LiDAR and VIS and the other for HS data to overcome similarity and over matching.

However, none of the above studies can realize the spectrum sharing of remote sensing images, resulting in the poor effect of remote sensing data fusion and classification. Therefore, a high spatial resolution remote sensing data classification method based on spectrum sharing is proposed. First, share the low-frequency wireless spectrum resources of remote sensing images. D-S evidence theory fusion method is used to realize high spatial resolution remote sensing data fusion. The convolution neural network classification model is constructed to realize the classification of high spatial resolution remote sensing data based on spectrum sharing. The experimental results show that the image segmentation time of the proposed high spatial resolution remote sensing classification method is shorter, the remote sensing image can be segmented accurately, and the data fusion effect of remote sensing image is better. The above experimental results show that the proposed high spatial resolution remote sensing data classification method based on spectrum sharing is practical and can provide a reliable theoretical basis for this field.

2. Method

2.1. Remote Sensing Image Dynamic Spectrum Sharing Method. PFRA-based radio spectrum resource allocation optimization method is used for remote sensing image in low-frequency band [7, 8] to improve the problem of spectrum resource shortage. If the number of available channels is M, the channel gain of subtransmitter S_{T_1}, subreceiver S_{R_2}, subtransmitter S_T, main receiver P_R, main transmitter P_T, and subreceiver S_R in channel m is described as $g_{ss}(m, S_{T_1}, S_{R_2})$, $g_{sp}(m, S_T, P_R)$, and $g_{ps}(m, P_T, S_R)$. Among them, channel gain is composed of large-scale attenuation and small-scale attenuation. In channel m, the transmitting power of secondary transmitter S_T and main transmitter P_T is $P_s(m, S_T)$, $P_p(m, P_T)$, and $S_{L,x}$ is the secondary remote sensing link, respectively. Each remote sensing link $S_{L,x}$ has channel request $C_{S_{L,x}}$, so the channel allocation matrix shall meet the request convergence conditions of the sub-remote sensing link channel, and each sub-remote sensing link channel request shall be set to the same fixed value, namely,

$$\sum_{m=1}^{M} a(m, S_{L,x}) = C_{S_L}. \tag{1}$$

In the process of wireless spectrum sharing of remote sensing images, the same spectrum can be shared with the primary user only when the disturbance caused by the secondary user to the primary user is less than a fixed limit

[9]. Meanwhile, the secondary transmitter has the maximum transmitting power, and the sum of the transmitting power of the secondary transmitter within each frequency band shall not exceed the maximum transmitting power, namely,

$$\sum_{S_T \in I} P_s(m, S_T) g_{sp}(m, S_T, P_R) \leq I_{P_R}^m,$$

$$\sum_{m \in W_{S_{Tx}}} P_s(m, S_{Tx}) \leq P_s^{\max}. \tag{2}$$

In the allocation of wireless spectrum resources in the low-frequency band of remote sensing image, the network income should be considered. The dry ratio of received signals of subreceiver S_{Rx} in channel m is described as follows:

$$\gamma_s(m, S_{L,x}) = \frac{S_s}{(N_s + N_p + P_{N_0})}. \tag{3}$$

where S_s represents the signal available in the sub-remote sensing link, N_s represents the disturbance originating in the remaining sub-remote sensing links, N_p represents the disturbance originating in the main remote sensing link, and P_{N_0} represents the noise power. Therefore, the reachable speed of the comparison is

$$r_s(m, S_{Tx}) = lb(1 + \gamma_s(m, S_{Tx})). \tag{4}$$

There is a certain correlation between the reachability of sub-remote sensing link in channel m and the channel allocation result, so the channel allocation matrix can be expressed as

$$A_{M \times S} = \{a(m, S_{L,x}) | a(m, S_{L,x}) \in \{0, 1\}\}_{M \times S}. \tag{5}$$

In the formula, $a(m, S_{L,x}) = 1$ represents the assignment of channel m to sub-remote sensing link $S_{L,x}$ and $a(m, S_{L,x}) = 0$ represents the assignment of channel m to sub-remote sensing link $S_{L,x}$. Only when channel m is assigned to the sub-remote sensing link, the relative reachable speed can be obtained. Therefore, the sum of all RS link speeds is expressed as

$$R_s^{\text{total}} = \sum_{x=1}^{s} R_s(S_{L,x})$$

$$= \sum_{x=1}^{S} \sum_{m=1}^{M} r_s(m, S_{L,x}) a(m, S_{L,x}), \tag{6}$$

$$\max R_s^{\text{total}} = \sum_{x=1}^{s} R_s(S_{L,x}). \tag{7}$$

The greater the speed sum of sub-remote sensing links, the better the spectrum utilization.

The problem of wireless spectrum sharing of remote sensing image in low-frequency band is to solve channel assignment matrix $A_{M \times N}$. The sum of the perturbation elements in each channel is taken as the perturbation coefficient of the channel [10] and expressed as follows:

$$F_m = \sum_{x_i=1}^{S} \sum_{x_j=x_i+1}^{S} a(m, S_{L,x_i}) a(m, S_{T,x_j}) f(S_{L,x_i}, m, S_{T,x_j}), \tag{8}$$

$$\min F_{\text{mean}} = \frac{1}{M} \sum_{m=1}^{M} F_m. \tag{9}$$

Formula (9) ensures that the mean of the disturbance coefficient is minimized and that the disturbance between each secondary user assigned to the same spectrum is minimized. In addition, in the allocation of remote sensing image spectrum resources, the disturbance to the primary user should be minimized. By minimizing the disturbance, all the primary users can be controlled within the disturbance. Therefore, the disturbance obtained by primary user P_R in spectrum m may be expressed as

$$I_{P_{R_p}}^{m} = \sum_{x=1}^{S} I(P_{R,p}, m, S_{L,x}), \tag{10}$$

$$\min I_P = \max \left(\sum_{x=1}^{S} I(P_{R,p}, m, S_{L,x}) \right). \tag{11}$$

In formula (11), $I(P_{R,p}, m, S_{L,x})$ indicates the magnitude of the disturbance to primary user $P_{R,p}$ caused by the secondary transmitter S_{Tx} in the spectrum m.

Since the spectrum of remotely sensed images is dynamically changing, the problem of dynamic spectrum sharing of remotely sensed images can be described as follows: based on the limited spectrum resources available, a dynamic spectrum sharing mechanism between the primary and secondary users of remotely sensed images can be modelled [11] to share the free frequency bands to the secondary users to optimize utilization.

Figure 1 depicts the relationship between the dynamic spectrum sharing links of remote sensing images.

Spectrum estimation value of remote sensing image is as follows:

$$r_{ji} = b_i \lg \left(1 + \frac{1}{l_{ji}^2} \times \delta \right). \tag{12}$$

where b_i is used to describe the channel bandwidth of spectrum y_i, l_{ji} is used to indicate the demand for idle channels by secondary user j, and δ is a constant, which is affected by factors such as transmission power, noise, and antenna gain, so the channel value estimation matrix is expressed in $R = \{r_{ji}\}_{Q \times Z}$.

Suppose a channel is only available to one secondary user, and all secondary users can select only one free channel at a time. $P = \{p_{ji}\}_{Q \times Z}$ is used to describe the sharing matrix; in the case of element $p_{ji} = 1$, the spectrum holder y_i will share the free channel to the second level user j, and in the case of $p_{ji} = 0$, the spectrum holder y_i will not share the free channel to the second level user j. Therefore, a reasonable sharing matrix must meet the following conditions:

$$\sum_{j=1}^{Z} p_{ji} \times j \leq q_i, \quad \forall i = 1, 2, \ldots, Q,$$

$$\sum_{i=1}^{Q} p_{ji} \times j \leq 1, \quad \forall j = 1, 2, \ldots, Z. \tag{13}$$

In formula (13), q_i represents that each spectrum holder can share free channels to a secondary user at most. Optimal dynamic spectrum sharing is to maximize spectrum efficiency, which can be obtained by using the following linear optimization equation:

$$P = \arg \max_{P} \sum_{j=1}^{Z} \sum_{i=1}^{Q} p_{ji} \cdot (r_{ji} \sim w_i). \tag{14}$$

Constraints are expressed as

$$\begin{cases} \sum_{j=1}^{Z} p_{ji} \times j \leq q_i, & \forall i = 1, 2, \ldots, Q, \\ \sum_{j=1}^{Q} p_{ji} \times j \leq 1, & \forall j = 1, 2, \ldots, Z, \\ p_{ji} \in \{0, 1\}, & \begin{array}{l} \forall i = 1, 2, \ldots, Z, \\ \forall j = 1, 2, \ldots, Z. \end{array} \end{cases} \tag{15}$$

2.2. Normalization of High Spatial Resolution Remote Sensing Data.

In order to achieve the ideal image processing effect, it is necessary to normalize the initial data in classification. In the classification model based on convolution neural network [12, 13], using standardized method to input data into different classification space, the operation process will make a big difference between different image classification results. The spectral reflectance difference per single pixel of the initial image is large, and the numerical span is relatively large in the separation process and will increase the computational difficulty. Firstly, each trajectory segment of high spatial resolution remote sensing data is normalized. This operation can make the trajectory of each single pixel's spectral curve more obvious and easier to judge and increase the variation of trajectory and reduce the complexity so as to improve the speed and accuracy of classification training. Assuming all pixels as column vector x_i, the formula is as follows:

$$X_i = \frac{X_i - \mu}{\lambda} - 1. \tag{16}$$

where μ represents the pixel average value of the initial image and λ represents the pixel standard deviation of the image in the i-th curve band.

2.3. Classification Method of High Spatial Resolution Remote Sensing Data Based on Spectrum Sharing.

High spatial resolution remote sensing monitoring refers to the use of high spatial resolution remote sensing technology for target monitoring in order to achieve quantitative analysis and
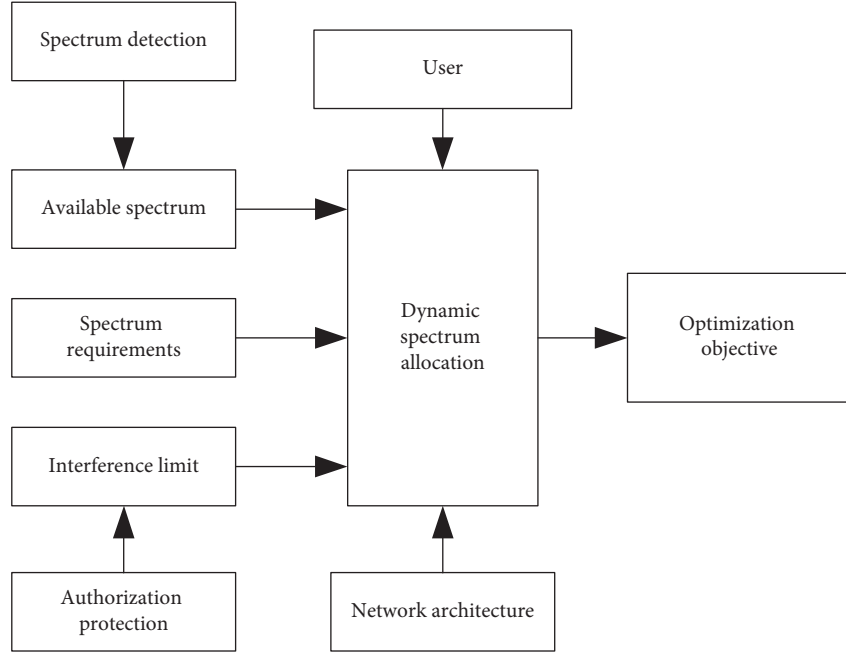
Figure 1: Relationship between dynamic spectrum sharing links.

determination of the characteristics and processes of surface change from monitoring data [14]. So far, high spatial resolution remote sensing technology has been widely used in meteorology, land, ocean, agriculture, geology, military, and other fields.

Classification is one of the main objectives of high spatial resolution remote sensing monitoring, which is a method of dividing each pixel or region into a certain type of terrain based on features collected by airborne LiDAR and hyperspectral technology [15]. The basic principle is as follows: because of different kinds of objects and different responses to electromagnetic waves, the high spatial resolution remote sensing data collected by airborne LiDAR and hyperspectral high spatial resolution remote sensing technology are different, which leads to different feature parameters. Data classification is achieved by using this feature to distinguish the target object from other objects [16].

Based on the above description, the classification of high spatial resolution remote sensing monitoring data is generally divided into six steps, as shown in Figure 2.

As can be seen from Figure 2, the classification of high spatial resolution remote sensing monitoring data includes data acquisition; preprocessing of high spatial resolution remote sensing monitoring data to improve the quality of data; feature extraction to select and reflect the characteristics of the target object; classification realization to be based on features, using classification algorithms to achieve classification; evaluation and analysis to analyze the effectiveness and feasibility of the method through simulation experiments; and output to show the test results in the form of images, statistical tables, etc. [17, 18].

2.3.1. Image Fusion. D-S evidence theory fusion method is used to realize high spatial resolution remote sensing data fusion [19]. The D-S theory of evidence constructs a trust structure as follows.

Basic probability distribution function: set U to represent a finite set and m to represent the probability distribution function on set U, which satisfies the following conditions [20]: $m(\phi) = 0$; for any $A \subset U$, $0 \leq m(A) < 1$; $\sum_{A \subset U} m(A) = 1$.

The basic probability function m: $2^U \longrightarrow [0, 1]$ mentioned above has a certain degree of confidence, which is usually subjectively defined.

Trust function is given by

$$Bel: 2^U \longrightarrow [0, 1] \cdot Bel(A) = \sum_{B \subseteq A} m(B). \tag{17}$$

The trust function of any subset A in set U is the sum of the basic probability functions of all subsets, and the trust function describes the global trust degree of evidence to A is true.

The expression of likelihood function [21] is

$$Pl(A) = 1 - Bel(A'). \tag{18}$$

In this expression, A' stands for the complement of A. Evidence theory can combine different kinds of evidences, which are related to each other, so as to fuse different kinds of evidences and get the final conclusion. The fusion rules are as follows: assuming that $m_1(A)$ and $(A \in 2^U)$ are the basic probability functions under different evidence U, the following combination rules of evidence theory may be applied:
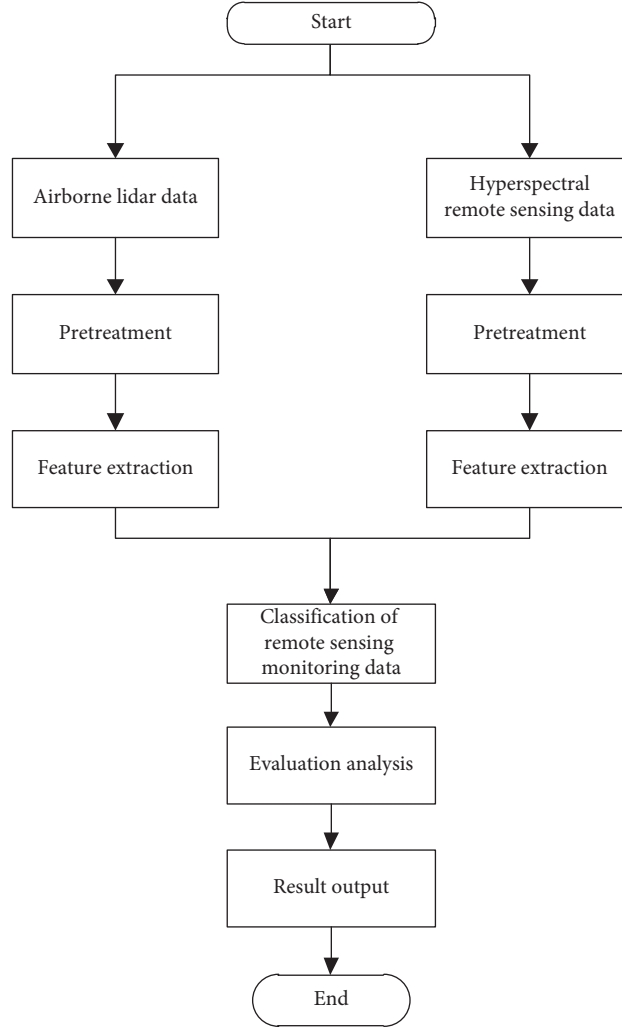
FIGURE 2: Basic process of high spatial resolution remote sensing monitoring data classification.

$$m(A) = \begin{cases} 0, & A = \phi, \\ K \sum_{B \cap C = A} m_1(B) m_2(C), & A \neq \phi. \end{cases} \quad (19)$$

In the process of fusion, if the trust degree is 0, $\sum_{B \cap C = \phi} m_1(B) m_2(C)$ will be discarded, and if no compensation is made after discarding, the total trust value will be less than 1, and parameter K will be introduced, that is, the part that can be compensated is discarded, and the total trust degree after fusion will be 1. The expression after fusion is m_1, m_2 orthogonal and $m = m_1 \oplus m_2$; similarly, assuming that there are multiple related evidence parallel fusion, then $m = m_1 \oplus m_2 \oplus \cdots \oplus m_n$.

In the operation of D-S evidence theory, the evidence accumulation can reduce the hypothesis set continuously, the complexity of time and information is low, and it has very good effect in dealing with the unstable factors caused by fuzzy generation.

Despite the preliminary data processing, however, the amount of data received by the coordinator for each data acquisition is still very large, including the environmental parameters transmitted from the acquisition nodes [22]. D-S

evidence theory is used to realize the fusion analysis of various environmental parameters, and the support degree of each set of data to various hypotheses is given to guide the control decision. The data fusion method is shown in Figure 3.

2.3.2. Pixel Brightness Value Correction.

Because of the saturation problem of pixel brightness in different images in HS-2 satellite dataset, it is necessary to correct the pixel brightness of fused multispectral images [23, 24]. Taking the upper limit of the cumulative pixel brightness value within the range of [1, 63] as the standard and taking the standard pixel brightness value of the multispectral camera as the reference data, a univariate quadratic model is constructed:

$$D_q = \delta \times D^2 + \gamma \times D + \lambda. \quad (20)$$

In (19), D_q represents the brightness value of corrected pixels, D represents the brightness value of pixels before correction, and δ, γ, and λ are all regression parameters. The quadratic model of one variable described in formula (19) is
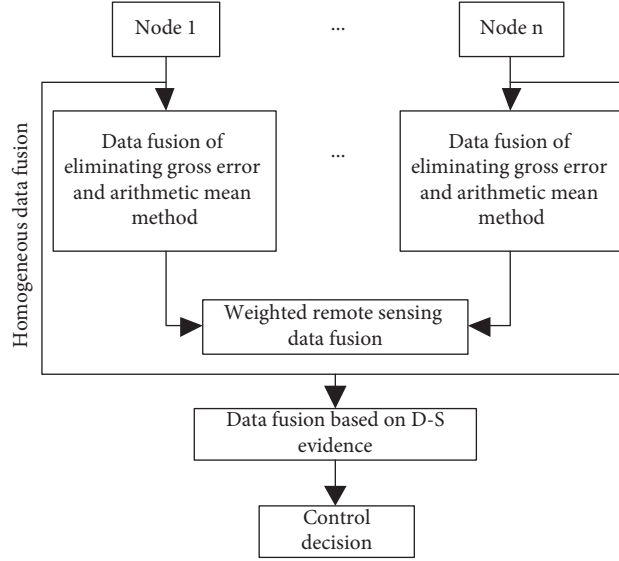
FIGURE 3: Remote sensing data fusion process.

used to eliminate unstable pixels from high spatial resolution remote sensing images of HS-2 [25, 26].

3. Classification Algorithm Design of High Spatial Resolution Remote Sensing Data

3.1. Spectral Image Feature Extraction. The feature extraction of high spatial resolution remote sensing data is mainly through the sinusoidal two-dimensional transformation function modulated by Gaussian function, and the expression formula is [27–29]

$$g(x, y; \vartheta, \theta, \varphi, \sigma, \gamma) = \exp\left(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right)\exp\left(j\left(2\pi\frac{x'}{\delta\vartheta} + \varphi\right)\right),$$

$$x' = x\cos\theta + y\sin\theta,$$

$$y' = x\cos\theta - y\sin\theta. \tag{21}$$

In the formula (20), x and y represent the values of the transverse coordinate system, ϑ the sine of the wavelength of the factor, θ the angle between directions of the specified parallel function, γ the spatial phase ratio of the ellipticity of the function, σ the standard deviation of the sine 2D transformation function determined by space and wavelength, and φ the composite function with the default value and the imaginary part transformation.

In order to improve the accuracy of spatial resolution remote sensing data classification, a spectral feature extraction model is proposed. The calculation formula is given by

$$W\left(T_{ij}'\middle\|T_{ij}\right) = \left(Z_{ij} \times T_{ij}\right). \tag{22}$$

In the formula, the key point of feature extraction is to capture the location of the central pixel, design and extract the tunnel to focus the position of the central pixel Z_{ij} on the

T_{ij}, convert the spectral feature vectors after the input and convolution data output to one-dimensional vectors, and then take r as the radius input data, then take the pixel as the center point and $Z_{ij} = R^{k \times k} (k = 2R + 1)$ as the initial data input to the designed feature extraction tunnel, and then realize the spectral feature extraction in the region with the central pixel Z_{ij} as the target [30]. The design formula is as follows:

$$F\left(T_{ij}, T_{ij}'\right) = f\left(W\left(T_{ij}'\middle\|T_{ij}\right) + b\right). \tag{23}$$

When the gradient of 2D transform function reaches saturation, when $x \rangle 0$, the gradient of the function is 1, so the problem of gradient dispersion is alleviated in the process of extracting. Therefore, the tunnel extracting method based on 2D transform function is not only accurate but also fast.

3.2. Construction of Convolution Neural Network Classification Model. Different classification models of convolution neural networks can be constructed by different training methods [31, 32]. Figure 4 shows the classification model structure. Set the training sample coefficient to ξ, so that $\xi = \{\xi_i | i = 1, 2, 3, 4\}$, and ξ_i is the equity coefficient between layers i and $(i - 1)$.

A high spatial resolution remote sensing data matrix is established, and each pixel sample in the matrix is combined into one, the number of columns is set to 1, and the number of rows is the number of fragments of each high spatial resolution remote sensing data, so as to construct the classification processor model of convolution neural network. Therefore, input layer A_1 is $(n_1, 1)$, and n_1 is the number of fragments of high spatial resolution remote sensing data, hidden convolution layer B_2 is composed of 20 convolution cores whose size is $(k_1 \times 1)$, including $(20 \times n_2 \times 1)$ nodes and $n_2 = n_1 - k_1 + 1$, so there are $[20 \times (k_1 + 1)]$ training coefficients between the convolution layer and the input layer, and pool C_3 is the second layer in the hidden convolution layer, whose function size is $(k_2, 1)$, and there are $(20 \times n_3 \times 1)$ nodes without coefficient.
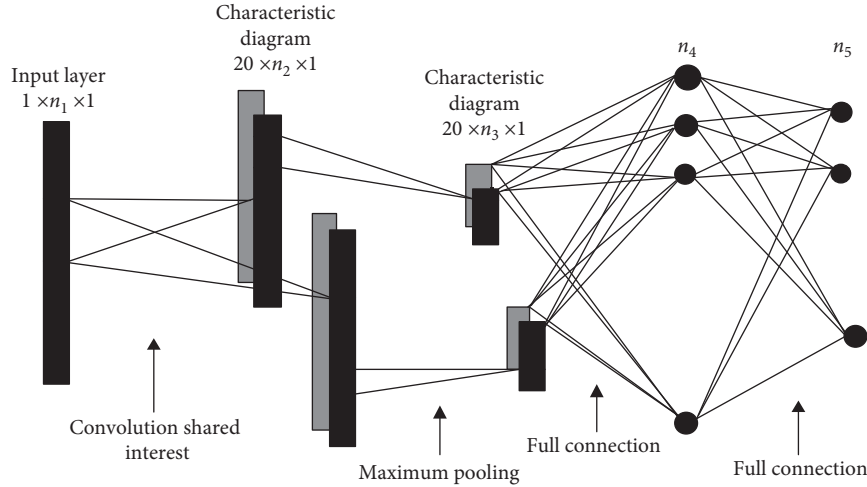
Figure 4: Structure of convolutional neural network classification model.

3.3. Remote Sensing Image Segmentation. After preprocessing the high spatial resolution remote sensing image, the high spatial resolution remote sensing image is segmented according to the graph cut theory to obtain a large number of remote sensing image blocks [33, 34]. In high spatial resolution remote sensing image, the boundary pixel and color fluctuation are the key features of high spatial resolution remote sensing image. Based on this feature, the target boundary in high spatial resolution remote sensing image can be described by energy function. The energy function is mapped into s-t network, and the boundary of high spatial resolution remote sensing image can be divided according to the minimum cost. The RGB distance S_{ij} between the nodes with the above two features can be obtained through the following formula:

$$S_{ij} = \left\| X_i - X_j \right\|^2. \tag{24}$$

where i and j, respectively, represent any two nodes in the high spatial resolution remote sensing image, and their corresponding pixel RGB values are X_i and X_j, respectively.

The edge weight k_{ij} of S-T network can be determined by the following formula:

$$\begin{cases} k_{ij} = A^{-S_{ij}/2\theta^2}, \\ \\ \theta^2 = \dfrac{1}{|A|} \displaystyle\sum_{(i,j)\in A} S_{ij}, \end{cases} \tag{25}$$

where A and θ^2 represent the S-T network boundary and node segmentation cost, respectively.

When the initial annular region contains the land class boundary of high spatial resolution remote sensing image, in order to minimize the cost of image cutting, it is necessary to ensure that the active contour of S-T network includes the land class boundary of remote sensing image.

Wine $\cos t(f)$ can be expressed as $\cos t(f)$, and the curve with the smallest $\cos t(f)$ is the optimal segmentation curve of the terrestrial boundary of the high spatial resolution remote sensing image [35]:

$$f = \arg\min_f \cos t(f) = \arg\min_f \sum_{(i,j)\in f} k_{i,j}. \tag{26}$$

In the S-T network, due to the influence of boundary thickness on the size of k_{ij}, the corresponding pixel RGB value of nodes on both sides of the boundary of the high spatial resolution remote sensing image will fluctuate significantly, and the smaller the k_{ij} value, the smaller the cumulative weight value of the pixels on the boundary of the high spatial resolution remote sensing image [36]. The S-T network is segmented to obtain the ring line f, which contains the cumulative weights of the edges as its segmentation cost.

4. Analysis of Experimental Results

4.1. Comparison of Remote Sensing Image Fusion Effects. In order to accurately judge the changes of spectral information and spatial details before and after fusion, statistical quantitative indicators are used to evaluate the effect of remote sensing image fusion. During the experiment, the correlation coefficient method is used to evaluate the effect of high spatial resolution remote sensing image fusion. The calculation process of correlation coefficient is as follows:

$$\rho(G, B) = \frac{\sum_{j=1}^{N} \sum_{i=1}^{M} (G(i,j) - \overline{G})(H(i,j) - \overline{H})}{\sqrt{\sum_{j=1}^{N} \sum_{i=1}^{M} (G(i,j) - \overline{G})^2 \sum_{j=1}^{N} \sum_{i=1}^{M} (H(i,j) - \overline{H})^2}}. \tag{27}$$

In the above formula, G and H represent multispectral image and panchromatic image respectively, M represents the number of rows of image, and N represents the number of columns of image. The calculation formula is as follows:

$$\overline{G} = \sum_{j=1}^{N} \sum_{i=1}^{M} \frac{G(i,j)}{(M,N)},$$

$$\overline{H} = \sum_{j=1}^{N} \sum_{i=1}^{M} \frac{H(i,j)}{(M,N)}. \tag{28}$$

Due to the large vegetation area in the study area, when using this method to extract office information, the panchromatic band image and the 2–4 band of multispectral image in the remote sensing data of the study area are selected. In order to verify the effect and feasibility of this method and increase the feasibility of experimental results, the remote sensing image classification method based on IFCM clustering and variational inference proposed in reference [3] and the remote sensing image classification method based on heterogeneous machine learning algorithm fusion proposed in reference [4] are used as experimental comparison methods. The remote sensing image fusion results of the three methods are shown in Figure 5.

According to the analysis of Figure 5, compared with the two experimental comparison methods of remote sensing image classification method based on IFCM clustering and variational inference proposed in reference [3] and remote sensing image classification method based on heterogeneous machine learning algorithm fusion proposed in reference [4], the remote sensing image fused by this method has high definition and can effectively enhance the detailed features in the space of remote sensing image.

The remote sensing image classification method based on IFCM clustering and variational inference proposed in reference [3] is compared with the remote sensing image classification method based on heterogeneous machine learning algorithm fusion proposed in reference [4] and the image fusion effect of this method. The correlation coefficient results of different bands of multispectral images are shown in Table 1.

By analyzing Table 1, compared with the two experimental comparison methods of remote sensing image classification method based on IFCM clustering and variational inference proposed in reference [3] and remote sensing image classification method based on heterogeneous machine learning algorithm fusion proposed in reference [4], the correlation coefficient of remote sensing image fused by this method is high, indicating that this method has the strongest ability to retain spectral information.

4.2. Comparison of Accuracy and Kappa Coefficient.
Taking the overall accuracy P_{OA} and kappa coefficient as test indicators, the higher the two coefficients, the better the urban environmental layout effect. The calculation formulas of overall accuracy P_{OA} and kappa coefficient are as follows:

$$P_{OA} = \frac{\sum_{i=1}^{n} X_{ii}}{M},$$

$$K = \frac{M \sum_{i=1}^{n} X_{ii} - \sum_{i,j=1} \left(X_{i+} X_{+j} \right)}{M^2 - \sum_{i,j=1} X_{i+} X_{+j}}, \tag{29}$$

where M represents the number of landscapes used for staggered zone layout; X_{ii} represents the number of landscapes in the error matrix; and X_{i+} and X_{+j} represent the number of landscapes in row i and column j, respectively.

The overall accuracy and kappa coefficient test results of different methods are shown in Figures 6 and 7, respectively.

By analyzing the data in Figures 6 and 7, it can be seen that the overall accuracy and kappa coefficient obtained by the proposed method in multiple iterations are higher than those obtained by the remote sensing image classification method based on IFCM clustering and variational inference and the remote sensing image classification method based on the fusion of heterogeneous machine learning algorithms because the proposed method introduces the page frame recovery algorithm (PFRA). The allocation of low-frequency wireless spectrum resources is completed, a dynamic spectrum sharing mechanism between primary and secondary users of remote sensing images is designed, the overall accuracy and kappa coefficient are improved, and a good layout effect is obtained.

4.3. Comparison of Time Coefficient Indicators.
Taking the time coefficient as the test index, the proposed method, the remote sensing image classification method based on IFCM clustering and variational inference, and the remote sensing image classification method based on the fusion of heterogeneous machine learning algorithms are used for testing. The larger the time coefficient, the longer the time consumption of the method. Figure 8 shows the test results of time coefficient.

As can be seen from Figure 8, the time coefficients obtained by the proposed method are lower than those obtained by IFCM clustering and variational inference and by heterogeneous machine learning algorithm because the proposed method uses D-S evidence theory to fuse high spatial resolution remote sensing data and correct the pixel brightness of the fused multispectral image. The initial data are normalized to extract spectral image features, which reduces the time used and improves the classification efficiency of remote sensing data.

4.4. Comparison of Time Consumption and Accuracy of Remote Sensing Image Segmentation.
The segmentation results of the proposed method, the classification method based on IFCM clustering and variational inference, and the classification method based on heterogeneous machine learning algorithm are shown in Table 2.

Table 2 shows that the average accuracy rate of remote sensing image segmentation is 99.67%, 2.24% higher than that of remote sensing image classification based on IFCM clustering and variational inference and 11.37% higher than that of remote sensing image classification based on heterogeneous machine learning algorithm fusion. At the same time, the average time consumption of remote sensing image segmentation is 0.38 s, 0.56 s lower than that of remote sensing image classification based on IFCM clustering and variational inference and 0.83 s lower than that based on heterogeneous machine learning algorithm fusion. The

(a)

(b)

(c)

(d)
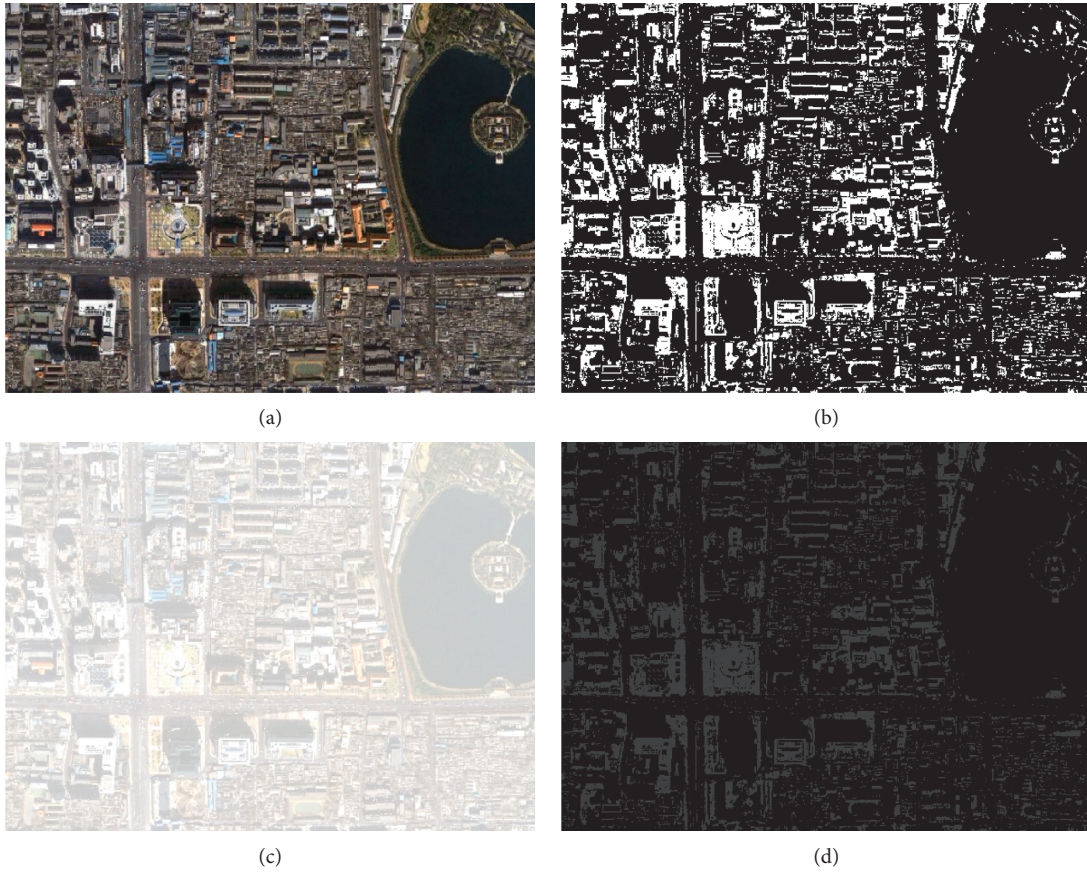
Figure 5: Remote sensing image fusion results. (a) Original multispectral image. (b) Fusion results of this method. (c) Remote sensing image classification method based on IFCM clustering and variational inference. (d) Remote sensing image classification method based on heterogeneous machine learning algorithm fusion.

Table 1: Image fusion effects of different methods.

Different methods	Band 1	Band 2	Band 3
Paper method	0.9986	0.9994	0.9984
Remote sensing image classification method based on IFCM clustering and variational inference	0.8503	0.8588	0.8060
Remote sensing image classification method based on heterogeneous machine learning algorithm fusion	0.8592	0.8610	0.8577

-•- Method proposed in this paper

-♦- Remote sensing image classification method based on IFCM clustering and variational inference

-▲- Remote sensing image classification method based on heterogeneous machine learning algorithm fusion
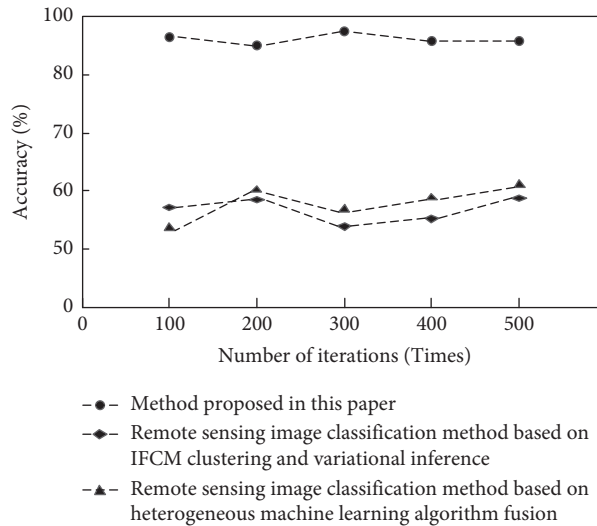
Figure 6: Accuracy of different methods.

- -●- Method proposed in this paper
- -◆- Remote sensing image classification method based on IFCM clustering and variational inference
- -▲- Remote sensing image classification method based on heterogeneous machine learning algorithm fusion
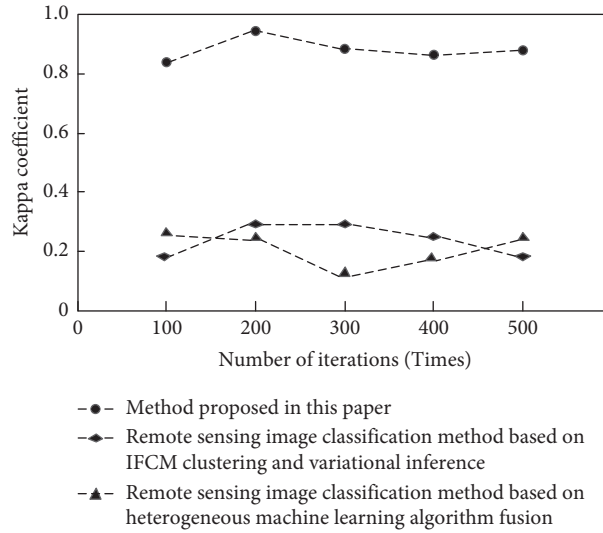
FIGURE 7: Kappa coefficient of different methods.

- -●- Method proposed in this paper
- -◆- Remote sensing image classification method based on IFCM clustering and variational inference
- -▲- Remote sensing image classification method based on heterogeneous machine learning algorithm fusion
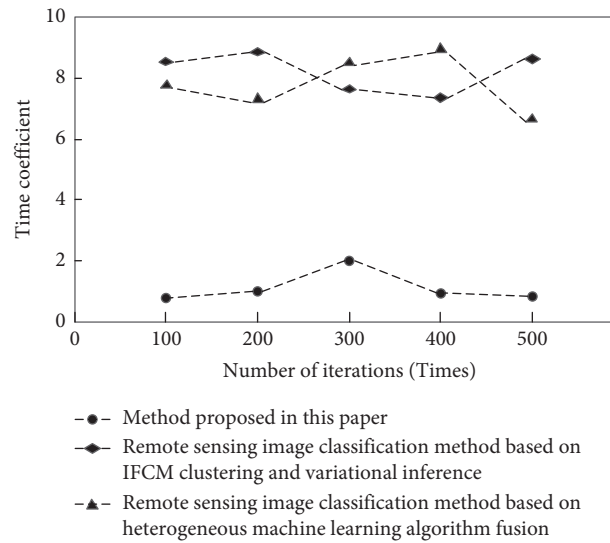
FIGURE 8: Time coefficients of different methods.

TABLE 2: Segmentation effect of remote sensing image.

Land type river surface class	Paper method		Remote sensing image classification method based on IFCM clustering and variational inference		Remote sensing image classification method based on heterogeneous machine learning algorithm fusion	
	Segmentation time (s)	Accuracy (%)	Segmentation time (s)	Accuracy (%)	Segmentation time (s)	Accuracy (%)
Cultivated land	0.17	99.27	0.55	86.94	1.80	86.65
Construction land	0.14	99.67	0.89	87.12	1.85	87.34
Highway	0.17	97.69	0.79	85.86	1.93	85.60
Groundbreaking	0.19	98.90	1.16	82.36	0.90	87.55
Garden class	0.11	98.19	0.62	84.21	0.91	89.26
Forest land	0.12	92.04	1.83	88.84	1.38	86.98
Land type	0.27	92.01	1.99	84.55	1.68	86.47

above data show that this method can reduce the time consumption of remote sensing image segmentation on the basis of ensuring high segmentation accuracy.

5. Conclusion

In order to optimize the data fusion and classification accuracy of traditional remote sensing data, a high spatial resolution remote sensing data classification method based on spectrum sharing is proposed. The following conclusions are drawn:

(1) The remote sensing image fused by this method has high definition and can effectively enhance the spatial detail features of remote sensing image. The correlation coefficient of the fused remote sensing image is high, and the high spatial resolution remote sensing data classification method based on spectrum sharing has the strongest ability to retain spectral information.

(2) The overall accuracy obtained by this method in multiple iterations is high. According to the analysis results, the layout optimization model is constructed to realize the layout optimization of urban ecotone, improve the overall accuracy and kappa coefficient, and obtain a good layout effect.

(3) The time coefficient obtained in the test process of this method is low, which provides relevant information for the layout optimization of urban ecotone, reduces the optimization time, and improves the efficiency of optimizing the layout of urban ecotone.

(4) The average accuracy of remote sensing image segmentation in this method is as high as 95.78%, and the average time consumption of remote sensing image segmentation in this method is 0.38 s, which can reduce the time consumption of remote sensing image segmentation on the basis of ensuring high segmentation accuracy.

In the future research work, the research will focus on the following two aspects:

(1) In the future, we can study the construction of deep learning network automation. How to automatically analyze and make decisions for remote sensing image analysis tasks, build a deep learning network suitable for the current task, and adaptively adjust the network structure and network learning parameters is a very practical research work, which provides a solid foundation for reducing the difficulty of remote sensing image analysis and processing tasks and improving the utilization rate and value of remote sensing images in the future.

(2) How to jointly use multisource remote sensing data to improve the classification effect of high score remote sensing images is an important and difficult problem at present. Therefore, based on the existing research foundation, constructing a high spatial resolution remote sensing data classification method based on spectrum sharing has high research value and application value.

Data Availability

The raw data supporting the conclusions of this article will be made available from the authors, without undue reservation.

Conflicts of Interest

The authors declare that they have no conflicts of interest regarding this work.

Acknowledgments

References

[1] W. Messaoudi, M. Farah, and I. R. Farah, "Fuzzy spatio-spectro-temporal ontology for remote sensing image annotation and interpretation: application to natural risks assessment," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 27, no. 5, pp. 815–840, 2019.

[2] M. Borhani, "Consecutive spatial-spectral framework for remote sensing image classification," *Earth Science India*, vol. 13, no. 2, pp. 271–285, 2020.

[3] Z. J. Xiang, L. Huang, and H. Chu, "Remote sensing image classification based on IFCM clustering and variational inference," *Computer Engineering and Design*, vol. 40, no. 7, pp. 2059–2063, 2019.

[4] Z. K. Tian, Y. Y. Fu, and S. H. Liu, "Remote sensing image classification based on heterogeneous machine learning algorithm fusion," *Computer Science*, vol. 46, no. 5, pp. 242–247, 2019.

[5] H. Huang, X. W. Zheng, and G. Y. Sun, Y. Hao, A. Zhang, J. Rong, and H. Ma, "Seismic image classification based on gravitational self-organizing map," *Remote Sensing For Land & Resources*, vol. 31, no. 3, pp. 95–103, 2019.

[6] B. Bigdeli, P. Pahlavani, and H. A. Amirkolaee, "An ensemble deep learning method as data fusion system for remote sensing multisensor classification," *Applied Soft Computing*, vol. 110, no. 1, Article ID 107563, 2021.

[7] V. P. Lakhin and E. A. Sorokina, "Low-frequency continuous MHD spectrum of toroidally rotating tokamak plasmas with anisotropic pressure," *Plasma Physics Reports*, vol. 45, no. 3, pp. 179–194, 2019.

[8] Y. P. Llerena and P. R. L. Gondim, "Social-aware spectrum sharing for D2D communication by artificial bee colony optimization," *Computer Networks*, vol. 183, no. 1, Article ID 107581, 2020.

[9] J. S. Hyde, R. A. Strangeway, and J. W. Sidabras, "Dispersion EPR: considerations for low-frequency experiments," *Applied Magnetic Resonance*, vol. 29, no. 5, pp. 1–14, 2021.

[10] A. M. Mamadou, J. Toussaint, and G. Chalhoub, "Survey on wireless networks coexistence: resource sharing in the 5G era," *Mobile Networks and Applications*, vol. 25, no. 3, pp. 1749–1764, 2020.

[11] R. Mondal, A. K. Shaw, I. T. Iliev et al., "Predictions for measuring the 21-cm multifrequency angular power spectrum using SKA-Low," *Monthly Notices of the Royal Astronomical Society*, vol. 494, no. 3, pp. 4043–4056, 2020.

[12] A. M. Hassan, A. A. Belal, M. A. Hassan, F. M. Farag, and E. S. Mohamed, "Potential of thermal remote sensing techniques in monitoring waterlogged area based on surface soil moisture retrieval," *Journal of African Earth Sciences*, vol. 155, no. 7, pp. 64–74, 2019.

[13] A. Hovi, E. Lindberg, M. Lang et al., "Seasonal dynamics of albedo across European boreal forests: analysis of MODIS albedo and structural metrics from airborne LiDAR," *Remote Sensing of Environment*, vol. 224, pp. 365–381, 2019.

[14] D. Carless, D. J. Luscombe, N. Gatis, K. Anderson, and R. E. Brazier, "Mapping landscape-scale peatland degradation using airborne lidar and multispectral data," *Landscape Ecology*, vol. 34, no. 6, pp. 1329–1345, 2019.

[15] M. Boell, H. R. Alves, M. Volpato, and D. D. Ferreira, "Exploiting feature extraction techniques for remote sensing image classification," *IEEE Latin America Transactions*, vol. 16, no. 10, pp. 2657–2664, 2019.

[16] R. Nijhawan, J. Das, and B. Raman, "A hybrid of deep learning and hand-crafted features based approach for snow cover mapping," *International Journal of Remote Sensing*, vol. 40, no. 1, pp. 759–773, 2019.

[17] F. L. A. Conceiç ao, F. L. C. Pádua, A. Lacerda, A. C. Machado, and D. H. Dalip, "Multimodal data fusion framework based on autoencoders for top-N recommender systems," *Applied Intelligence*, vol. 49, no. 9, pp. 3267–3282, 2019.

[18] S. Ma, B. Jia, J. Wu, Y. Yuan, Y. Jiang, and W. Li, "Multi-vibration information fusion for detection of HVCB faults using CART and D–S evidence theory," *ISA Transactions*, vol. 113, no. 12, pp. 210–221, 2021.

[19] R. Guillaume and D. Dubois, "A min-max regret approach to maximum likelihood inference under incomplete data," *International Journal of Approximate Reasoning*, vol. 121, no. 6, pp. 135–149, 2020.

[20] Y. Sun, J. Luo, T. Wu et al., "Synchronous response analysis of features for remote sensing crop classification based on optical and SAR time-series data," *Sensors*, vol. 19, no. 19, p. 4227, 2019.

[21] T. Kattenborn, J. Lopatin, M. Förster, A. C. Braun, and F. E. Fassnacht, "UAV data as alternative to field sampling to map woody invasive species based on combined sentinel-1 and sentinel-2 data," *Remote Sensing of Environment*, vol. 227, no. 6, pp. 61–73, 2019.

[22] R. K. Vemuri, P. Reddy, B. P. Kumar, and J. Ravi, "Deep learning based remote sensing technique for environmental parameter retrieval and data fusion from physical models," *Arabian Journal of Geosciences*, vol. 14, no. 13, pp. 1–10, 2021.

[23] M. P. Mariana, A. Gonzalez-Sanchez, S. I. Jimenez-Jimenez, and G. S. Alberto, *International Journal of Remote Sensing*, vol. 40, no. 1, pp. 420–438, 2019.

[24] A. T. Erdogan, R. Walker, N. Finlayson et al., "A CMOS SPAD line sensor with per-pixel histogramming TDC for time-resolved multispectral imaging," *IEEE Journal of Solid-State Circuits*, vol. 54, no. 6, pp. 1705–1719, 2019.

[25] M. Salah, "SRTM DEM correction over dense urban areas using inverse probability weighted interpolation and Sentinel-2 multispectral imagery," *Arabian Journal of Geosciences*, vol. 14, no. 9, pp. 1–16, 2021.

[26] Z. Mushtaq and S.-F. Su, "Environmental sound classification using a regularized deep convolutional neural network with data augmentation," *Applied Acoustics*, vol. 167, no. 10, Article ID 107389, 2020.

[27] D. Ruiz, B. Bacca, and E. Caicedo, "Hyperspectral images classification based on inception network and kernel PCA," *IEEE Latin America Transactions*, vol. 17, no. 12, pp. 1995–2004, 2019.

[28] N. Kaur, P. S. Tiwari, H. Pande, and S. Agrawal, "Utilizing advance texture features for rapid damage detection of built heritage using high-resolution space borne data: a case study of unesco heritage site at bagan, Myanmar," *Journal of the Indian Society of Remote Sensing*, vol. 48, no. 11, pp. 1627–1638, 2020.

[29] A. E. Maxwell and T. A. Warner, "Is high spatial resolution DEM data necessary for mapping palustrine wetlands?" *International Journal of Remote Sensing*, vol. 40, no. 1, pp. 118–137, 2019.

[30] Y. Li, X. Wang, Q. Huang, X. Hu, and W. Xie, "Robust multi-view representation for spatial-spectral domain in application of hyperspectral image classification," *IET Computer Vision*, vol. 13, no. 2, pp. 90–96, 2019.

[31] B. Hha, L. A. Ying, and B. Xz, "Convolutional neural network learning for generic data classification," *Information Sciences*, vol. 477, no. 11, pp. 448–465, 2019.

[32] I. Mitiche, M. D. Jenkins, P. Boreham, A. Nesbitt, and G. Morison, "An expert system for EMI data classification based on complex bispectrum representation and deep learning methods," *Expert Systems with Applications*, vol. 171, no. 1, Article ID 114568, 2021.

[33] H. He, T. Chen, M. Chen, D. Li, and P. Cheng, "Remote sensing image super-resolution using deep-shallow cascaded convolutional neural networks," *Sensor Review*, vol. 39, no. 5, pp. 629–635, 2019.

[34] W. Wang, Y. Fu, F. Dong, and F. Li, "Semantic segmentation of remote sensing ship image via a convolutional neural networks model," *IET Image Processing*, vol. 13, no. 6, pp. 1016–1022, 2019.

[35] A. Rsnen and T. Virtanen, "Data and resolution requirements in mapping vegetation in spatially heterogeneous landscapes," *Remote Sensing of Environment*, vol. 230, no. 9, Article ID 111207, 2019.

[36] X.-Y. Tong, G.-S. Xia, Q. Lu et al., "Land-cover classification with high-resolution remote sensing images using transferable deep models," *Remote Sensing of Environment*, vol. 237, no. 6, Article ID 111322, 2020.

Research Article

Diagnosis and Classification Decision Analysis of Overheating Defects of Substation Equipment Based on Infrared Detection Technology

Zhigang Shi, Yunlong Zhao ⬤, Zhanshuang Liu, Yanan Zhang, and Le Ma

State Grid Qinghai Maintenance Company, Xining 810021, China

Correspondence should be addressed to Yunlong Zhao; ylzhao.sgqh@yahoo.com

Substation equipment is not only the main part of the power grid but also the essential part to ensure the development of the national economy and People's Daily life of one of the important infrastructure. How to ensure its normal operation and find the sudden failure has become a hot issue to be solved urgently. For thermal fault diagnosis needs to classify and identify different power equipment first, this paper designed an SVM infrared image classifier, which can effectively identify three types of common power equipment. The classifier extracts HOG features from the infrared images of power equipment processed by the above segmentation and combines them with SVM multiclassification to achieve the purpose of improving the recognition accuracy. The experiment uses the classifier to identify three kinds of equipment, and the results show that the comprehensive recognition accuracy of the classifier is more than 95.3%, which is better than the traditional classification method and meets the demand for classification accuracy. In this paper, the traditional method of relative temperature difference is improved by using the temperature data of the infrared image, which can automatically judge the thermal failure level of electric power equipment. Experiments show that the diagnosis system designed in this paper can classify faults and give treatment suggestions while judging whether there are thermal faults for three types of power equipment, which verifies the feasibility and effectiveness of the substation infrared diagnosis technology designed in this paper.

1. Introduction

With the rapid development of the social economy, the requirements for the safety and reliability of substation equipment are constantly improving. Whether the electrical equipment can operate safely and reliably has become the key to the normal operation of the power system. Substation equipment in a long-term working state will inevitably have faults, and the thermal fault is the most common fault of electrical equipment, which is manifested in abnormal temperature inside the equipment: such as poor contact heating caused by accumulated pollution on the surface of electrical equipment, insufficient clamping force of isolating switch finger that leads to overheating of the switch knife-edge, and so on. Therefore, detecting and diagnosing the temperature change of substation equipment plays an important role in the normal operation of the power grid [1].

Infrared thermal image detection technology has the advantages of no contact, high-temperature measurement efficiency, wide detection range, and so on, which can effectively improve the reliability of results and effectively improve the safety factor. In current substations, infrared thermal image detection technology has been widely used for fault detection of power equipment. Infrared thermal imaging technology can judge whether there is a thermal fault in the current running equipment in time when the power equipment is running, In order to quickly locate potential dangerous areas, it is convenient to take control measures before equipment failure and operation interruption, reduce the possibility of users' power interruption caused by substation power outage detection and troubleshooting, prevent premature failure, delay the service life of power equipment, and reduce expensive power outage and downtime [2]. The infrared thermal imaging technology currently applied in the

field of substation equipment diagnosis; whether there are thermal defects in power equipment can be judged by manual analysis after a professional holds a thermal infrared imager to collect equipment images. This method will consume a lot of time when it is urgent to troubleshoot a large number of equipment and will also lead to diagnostic errors due to inexperience or fatigue diagnosis of staff. In view of the shortcomings of the above-mentioned manual analysis, it is necessary to use an intelligent identification and diagnosis method to judge whether there is a thermal fault in substation equipment so as to realize a rapid diagnosis and investigate the power accidents caused by potential thermal faults. Through the processing of infrared image denoising and segmentation of substation equipment, the thermal fault judgment method and infrared power equipment image recognition method are introduced, which lays a foundation for automatic recognition and thermal fault diagnosis of power equipment based on infrared image [3].

Overheating of electrical equipment is a common phenomenon. If the overheating area can be found accurately and directly, the fault range will be directly reduced. Therefore, we propose a new overheated area detection algorithm. This algorithm uses Ostu algorithm to segment the general area of equipment, then refines the target area by edge, and uses the FCM clustering algorithm to accurately divide the overheated part. After verification, the algorithm has practical value [4]. The demand for electricity has always been huge, and the country is constantly expanding the scale of the power grid to meet the transmission capacity. Obviously, the expansion of the scale also increases the load of power equipment; the frequency of faults also increases; and the workload of troubleshooting is increasing day by day. We combine patrol robot, support vector machine (SVM), and MATLAB software for simulation analysis and compare them with information entropy method. Finally, we apply the two methods to a project for periodic observation at the same time and find that the patrol robot with support vector machine has high efficiency [5]. Pressure monitoring technology can be used to monitor the status of oil-free power equipment because the insulation overheating of oil-free power equipment will lead to pressure change. Taking the change trend of pressure bearing index as an example, the pressure rises from 1 kPa to 3 kPa within 50 hours, and the pressure rises obviously, which shows that the pressure monitoring technology can monitor local overheated insulation defects online, and this technology is a hot spot for this problem at present [6].

Parallel conductors in AC systems often have the problem of current imbalance, which will lead to overheating and aging, and promote the occurrence of faults. Then, the fault diagnosis system is necessary. Based on a large section aluminum conductor, a spacer, and two terminal substation connectors, we have a Hall effect sensor system, which has the real-time and contactless characteristics. Through experiments, it can accurately detect the fault and current at the same time, which shows its effect and feasibility [7]. The heat dissipation problem of transformer in substation building must be solved under the condition of unsatisfactory ventilation. We use CFD to simulate the heat

value around the equipment, and the model should consider the expansion fin to enhance the heat dissipation. We put forward three schemes, all of which consider the air outlet position. Finally, it is explained that choosing the appropriate air inlet and outlet position in the substation plays a key role in avoiding overheating of the transformer [8].

2. Infrared Image Segmentation of Substation Equipment

2.1. Intuitionistic Fuzzy Clustering Algorithm. Intuitionistic fuzzy set theory is widely used to deal with fuzzy and uncertain data and ambiguous relationships and has been applied in many fields. There is a normalized relationship among three members in an intuitionistic fuzzy set. As long as the values of any two members in the three are known, the values of the remaining members can be obtained by using the normalized relationship. Common intuitionistic fuzzy sets are usually represented by a membership function and intuitionistic index, and nonmembership can be generated by specific functions. The clustering algorithm is also used in the power system in infrared image segmentation, in which K-means clustering is a commonly used feature with low complexity and fast convergence [9–12]. However, the noise processing effect is not ideal, and it is widely used in infrared images.

The segmentation method based on fuzzy clustering divides image data into several classes by a certain criterion, which makes the similarity between the same classes maximum and the similarity between different classes minimum. In the actual clustering process, the segmentation method based on fuzzy clustering uses a membership function to classify image data, which represents the degree of the pixel belonging to a certain region in image segmentation. Fuzzy C-means clustering algorithm is an unsupervised algorithm. Its core idea is to update the clustering center and membership function continuously until the best clustering center cut-off is obtained. The core idea of this algorithm in image segmentation is to get the best image classification result through the best clustering center. In the process of image segmentation with fuzzy C-means clustering algorithm, human interference is not needed, and the inherent fuzziness of the image can be effectively processed. An intuitionistic fuzzy clustering method based on local information is proposed [9].

2.2. Fuzzy C-Means Clustering. Fuzzy C-means clustering (FCM) is a process of dividing the elements of a set into several self-similar classes, which divides a set of elements into several classes, and the elements in each class are similar [13, 14]. Elements in the same class have great similarities, while elements in different classes have great differences. FCM proposed by Dunn in 1973 is generalized by Bezdek J.C. Its objective function is defined as follows:

$$J_m = (U, X) = \sum_{j=1}^{c} \sum_{k=1}^{n} u_{jk}^m d^2\left(x_k, v_j\right), \tag{1}$$

where U represents fuzzy membership matrix, c represents the number of clustering centers, n represents the number of pixels in the image, x_k denotes the k-th data point, v_j denotes the j-th cluster, u_{jk} represents the fuzzy membership degree of the k-th pixel sample to the j-th cluster center, and $d^2(x_k, v_j)$ is expressed as the distortion degree by Euclidean distance as follows:

The following restrictions are met:

$$\sum_{j=1}^{c} u_{jk} = 1, \tag{2}$$

where $d^2(x_k, v_j)$ is represented as the distortion by Euclidean distance as follows:

$$d^2(x_k, v_j) = x_k - v_j^2. \tag{3}$$

When the function reaches the extreme value, the algorithm has the best effect, so the minimum values of $J_m(U, X)$ calculated by Lagrange number multiplication are as follows:

$$u_{jk} = \left[\sum_{i=1}^{c} \frac{x_k - v_j}{x_k - v_i} \right]^{-2/m-1},$$

$$v_j = \frac{\sum_k^n u_{jk}^m x_k}{\sum_k^n u_{jk}^m} \tag{4}$$

2.3. Intuitionistic Fuzzy C-Means Clustering (IFCM).
Main ideas of FCM algorithm [15, 16] described in the characteristic function (whose value can only take 0 or 1) to any membership function in the range of [0, 1]. However, the FCM algorithm cannot describe the membership degree between right and wrong. In order to solve this problem, an intuitionistic FCM algorithm is proposed.

In intuitionistic fuzzy set theory, "hesitation degree" is put forward, that is, it represents the status of neutral attributes, and depicts the fuzzy information in objective reality more completely and accurately. The n-dimensional data set X is represented by the following formula:

$$A = \{(x, u_A(X), v_A(X)), x \in X\}, \tag{5}$$

where $u_A(X)$ is the membership degree, $v_A(X)$ is a non-membership degree, and the corresponding function is shown in the following formula:

$$u_A(X): x \longrightarrow [0, 1]; v_A(X): x \longrightarrow [0, 1]. \tag{6}$$

If $x \in X$, x does not belong to $u_A(X)$ and $v_A(X)$. Then, when the condition $0 \le u_A(X), +(X) \le 1$ is satisfied, the hesitation degree is described by the following formula:

$$\pi_A(x) = 1 - u_A(X) - v_A(X). \tag{7}$$

I represents the image, X represents the image pixels, and the gray value is [0, 1]. The image is normalized as shown in the following formula:

$$u_A(X) = \frac{x - I_{\min}}{I_{\max} - I_{\min}}, \tag{8}$$

where $I_{\max}$ and $I_{\min}$ represent the maximum gray value and the minimum gray value in the pixel, respectively. Sugeno calculates the nonmembership degree of intuitionistic fuzzy sets by using negative function as follows:

$$v_A(X) = \frac{1 - u_A(x)}{1 + \lambda u_A(x)}, \tag{9}$$

where λ is a normal number; this ensures that the sum of membership degree and nonmembership degree will not be greater than 1.

Furthermore, let d: $\mathrm{IFS}(X) \times \mathrm{IFS}(X) \longrightarrow [0, 1]$. If d is a distance measure, the following attributes should be satisfied:

(1) $0 \le d(A, B) \le 1$

(2) If $A = B$ then $d(A, B) = 0$

(3) $d(A, B) = d(B, A)$

(4) If $A \subseteq B \subseteq C$ and $A, B, C \in \mathrm{IFS}s(X)$, then $d(A, B) \le d(A, C)$ and $d(B, C) \le d(A, C)$

The normalized distances of A and B are shown in the following formula:

$$d_{IFS}(A, B) = \left\{ \frac{1}{2n} \sum_{i=1}^{n} \left[(u_A(x_i) - u_B(x_i))^2 + (v_A(x_i) - v_B(x_i))^2 + (\pi_A(x_i) - \pi_B(x_i))^2 \right] \right\}^{1/2}. \tag{10}$$

Formula (10) is improved to obtain the following formula:

$$\tilde{d}(A_1, A_2) = \left\{ \frac{1}{2n} \sum_{i=1}^{n} \left[p_{12}(u_{A1}(x_i) - u_{A2}(x_i))^2 + q_{12}(v_{A1}(x_i) - v_{A2}(x_i))^2 + \rho(A_1, A_2)(\pi_{A1}(x_i) - \pi_{A2}(x_i))^2 \right] \right\}^{1/2}. \tag{11}$$

The weight of p_{12}, q_{12}, and $\rho(A_1, A_2)$ can be adjusted according to the application data. The IFCM function definition is shown in the following formula:

$$J_{\mathrm{IFCM}} = \sum_{J=1}^{C} \sum_{k=1}^{n} u_{jk}^m d_{\mathrm{IFS}}^2 (x_k, v_j), \qquad (12)$$

where d_{IFS}^2 is the intuitionistic ambiguity distance. The iterative formula of membership degree and clustering centroid based on the Lagrange multiplier method is as follows:

$$u_{jk} = \left[\sum_{i=1}^{c} \frac{d_{\mathrm{IFS}}(x_k, v_j)}{d_{\mathrm{IFS}}(x_k, v_i)} \right]^{-2/m-1}, \qquad (13)$$

$$u_A(v_j) = \frac{\sum_{k=1}^{n} u_{jk}^m u_A(x_k)}{\sum_{k=1}^{n} u_{jk}^m}, \qquad (14)$$

$$v_A(v_j) = \frac{\sum_{k=1}^{n} u_{jk}^m v_A(x_k)}{\sum_{k=1}^{n} u_{jk}^m}, \qquad (15)$$

$$\pi_A(v_j) = \frac{\sum_{k=1}^{n} u_{jk}^m \pi_A(x_k)}{\sum_{k=1}^{n} u_{jk}^m}. \qquad (16)$$

In the IFCM algorithm, the cluster center v_j is expressed as $v_j = u_A(v_j) v_A(v_j) \pi_A(v_j)$. The convergence of the IFCM algorithm is proved by formulas (13)–(16), and the local minimum of the objective function is proved by formula (13).

3. Automatic Classification of Power Equipment Based on the Infrared Image

3.1. Feature Extraction of the Infrared Image. Feature is a collection of the essence and characteristics of a certain object relative to other objects, so feature extraction can be regarded as using a mapping relationship to describe high-dimensional image features [17] from low-dimensional spatial features. Generally speaking, images have their own characteristics that can be distinguished from other images, such as color, edge information, texture features, and so on. Common image features are as follows:

(1) Color characteristics

Color feature, as a global feature with high efficiency and good effect, expresses the surface characteristics of the target object. Half of it is represented as a three-dimensional space, and the other is represented by three color parameters, such as RGB, XYZ, and other color spaces. Color features are based on pixels and insensitive to the change of image direction and size, so they cannot describe the local changes of target objects in the image.

(2) Texture features

Texture feature, as a common statistical feature in image recognition, is generated by color spatial transformation. Although there is no completely unified theoretical concept at present, it is similar to

color feature, and it is also an external feature to express the target object in the image. Texture feature extraction, which is widely used in practice, generally includes statistical method and model method. In the actual texture extraction, the resolution will change, and there will be a large error at this time.

(3) Spatial relationship characteristics

Spatial relation feature, as a feature index in image space, means that there is a certain relationship between the position or direction of the segmented object in image space. Spatial relation features are generally used when it is necessary to enhance the ability of the object of the expressed image. However, it is easily affected by the position translation, rotation change, and size change of the image or a certain area of the image in actual operation, which will make the spatial relationship feature unable to accurately express the scene information of the image. Therefore, it will be combined with other feature extraction methods in practice to improve the accuracy of the expression of the image scene.

(4) Shape characteristics

Shape features are generally divided into contour and region features. The contour feature describes the outer boundary of the object in the image, and the region feature covers all the object regions in the image. Shape feature extraction is not affected by image position transformation or scale transformation. Common shape feature extraction, such as moment invariants, wavelet description, and so on.

For the infrared image studied in this paper, color information maps the temperature data of the target object, instead of the color characteristics of the object itself. Therefore, its contrast is not high. The texture of the target object cannot be clearly displayed. It is also important to note that in the actual operation of collecting infrared images of power equipment, the image will inevitably change in spatial position, so using the above three features, that is, color, texture, and spatial relationship features, to describe the target equipment in infrared images is not in line with the actual feature extraction requirements. Compared with other features, shape features can better identify the power equipment in infrared images. Therefore, the Hu moment of the invariant moment is used to recognize the shape features of power equipment in the infrared image.

3.2. Infrared Image Recognition. Image recognition is one of the practical applications of pattern recognition in digital image processing. Generally, image segmentation and feature extraction are carried out, and after image analysis, a certain feature in the selected image is recognized and then classified. The specific flow is shown in Figure 1. In this paper, image acquisition refers to the use of infrared thermal imager for power equipment image acquisition and import into the computer for the next processing operation. Image segmentation is to segment the target image, that is, the faulty power equipment from the image background.
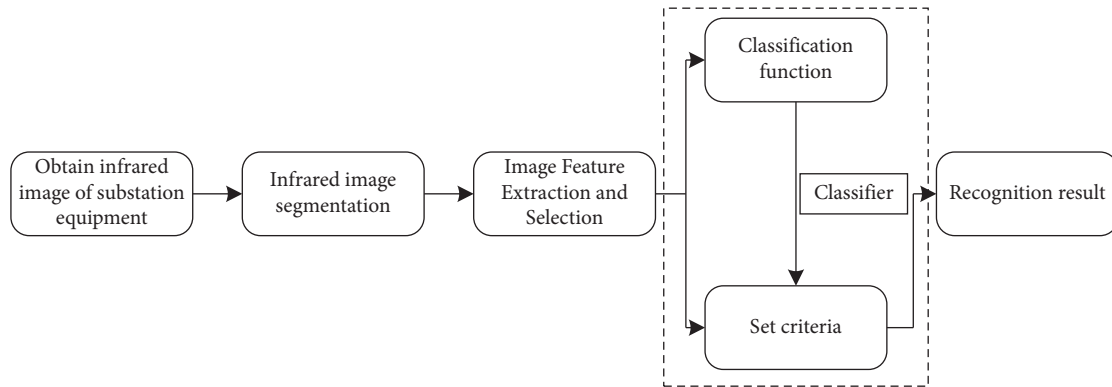
Figure 1: Flow chart of image recognition.

Feature extraction and selection are to process the image information and then extract the features that can describe the essence of the target object segmented from the image. The classifier uses the features extracted in the previous step to classify and recognize the image features through appropriate functions and evaluation criteria.

Up to now, many algorithms for image recognition have been widely used, such as artificial neural network recognition, fuzzy recognition, and template matching. In this paper, a support vector machine algorithm based on statistical theory is used to classify and recognize infrared images of substation equipment. The classification function mainly classifies images according to the features extracted from images. Generally, minimum distance classification is adopted, and there are machine learning methods, such as SVM, KNN, SVM, BPNN, CNN, and transfer learning. According to the set standard, the image features are classified so as to realize the classification function.

3.3. Feature Extraction Based on Directional Gradient Histogram.

The feature based on directional gradient histogram is a feature descriptor for object recognition, which uses directional gradient histogram of the local area of the statistical image to form features. In the field of image classification, using HOG features combined with SVM classification has been widely used [18].

HOG feature extraction is a good way to describe the image or shape of the local region of the image through the distribution of gradient or edge direction density. The process is shown in Figure 2.

The specific process is described by Figure 2:

(1) Firstly, the infrared image of power equipment is normalized by gamma ray, and the image is divided into smaller units, which are called cell units. In order to unify the overall standard, images are generally converted into gray images, and the gray values are set in the range of [0, 1].

(2) The gradient value is calculated. The gradient operator $[-1, 0, 1]$ is used to convolve the image, and the gradient component in the horizontal x-axis direction is obtained. Similarly, the gradient components perpendicular to the y-axis direction are

obtained by using this operator, and the gradient size and direction of each pixel in the digital image are calculated by using these two gradient components.

(3) Secondly, the directional gradient histogram is constructed. Each cell is divided into nine uniform directional blocks; then each pixel is mapped to a fixed angle range; and the gradient directional histogram of the cell is obtained, which contains the corresponding nine-dimensional feature vector of the cell.

(4) In block normalized gradient histgram, the cells are combined into an interval, and the HOG features of all cells in this interval are obtained by connecting the eigenvectors in series. The overlapping of different intervals means that the different results of each cell feature will be expressed in the final feature vector, so the feature vector in the block should be normalized.

(5) Finally, the HOG features are collected. All overlapping blocks are collected by HOG features, and the final combined feature vectors are used for classification.

3.4. Substation Equipment Classification Based on SVM.

In practical applications, two types of classifiers widely used are artificial neural networks and support vector machines. Both of them use the training set to obtain the recognition system model to classify the data set to be recognized. The difference is that the former is based on statistical theory and simulates the complex regular network built by the human brain neural network. The BP neural network model [19] shown in Figure 3 is most commonly used in practical applications.

In the latter, the input space is changed into a high-dimensional space by a specific function so as to transform the complex nonlinear relationship into a linearized relationship. When using a neural network to summarize complex laws, the recognition accuracy of data to be recognized is generally improved by adding a large number of known result data sets. However, although the recognition error rate decreases, the corresponding network scale and complexity will gradually increase, that is to say, a small
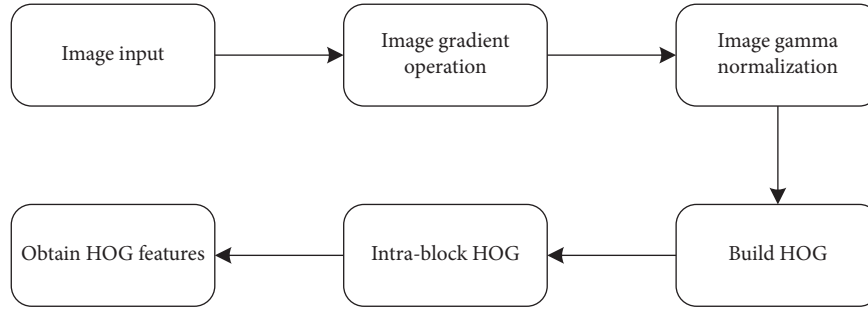
FIGURE 2: Flowchart of HOG feature extraction.
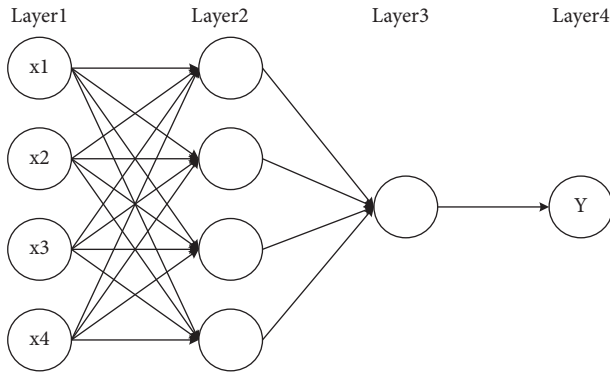
FIGURE 3: Basic model of the neural network.
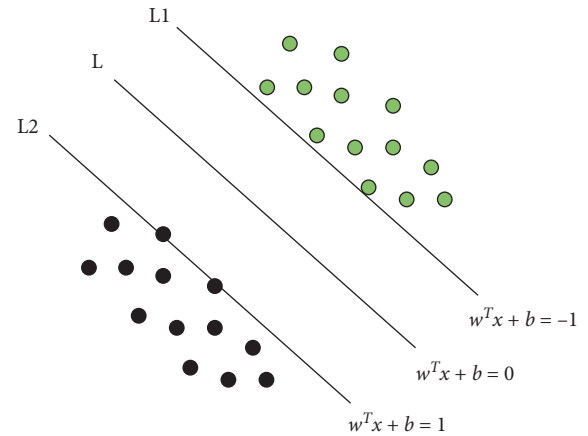
FIGURE 4: SVM linear classification model.

number of training sets cannot ensure a high recognition rate. However, in practice, it is difficult to collect an infinite number of data samples, which makes the error rate infinitely reduced, and it is difficult to achieve ideal classification results. And blindly compressing the redundant areas of training data will lead to overfitting, which is also one of the representative shortcomings of neural networks in practice.

Comparatively speaking, a support vector machine has better classification ability when it cannot obtain a considerable number of data sets. Support vector machine is good at solving the classification problem of small samples and nonlinear data, and the global optimal solution obtained by optimization method can prevent the problem of over-learning. Therefore, this paper chooses the SVM algorithm for the automatic classification of infrared images of power equipment.

SVM is different from ANN in that it can learn from small samples and has the advantage of high learning efficiency. SVM is widely used in regression and classification research directions and has a high accuracy for nontraining samples.

As shown in Figure 4, it is a schematic diagram of SVM classification [20, 21] under the condition that two-dimensional data are linearly separable. It can be seen from the diagram that if two types of samples are completely divided by a linear function, they are linearly separable samples; otherwise, they are nonlinear separable samples.

Real and empty points represent two kinds of samples in a two-dimensional plane, in which the positive example label is l and the negative example label is -1, and the samples are linearly separable. L is the final linear classifier; and $L1$ and

$L2$ are the straight lines where the support vectors are located, parallel to L; and their intervals are classification intervals. To get the optimal classification line is to get a classification line that can divide the data into two categories with the largest interval.

The above-mentioned SVM theory is only suitable for binary classification, but the actual situation is often like the power equipment classification studied in this paper. In order to extend SVM to solve multiclass classification problems, the "one-to-many" method of indirect method, "one-to-one" method, and multiclass support vector machine of direct method are generally used.

4. Thermal Fault Diagnosis of Power Equipment Based on the Infrared Image

The operation of substation equipment is easily affected by the external environment, such as artificial construction, temperature, humidity, and electric field interference. In the long run, some problems such as poor contact of power equipment will inevitably occur, which will lead to local heating of equipment and cause thermal failure. If it is not found in time and hidden dangers are not eliminated in advance, it is easy to cause major accidents such as power failure and even explosion. Infrared diagnosis is used to detect the infrared radiation signal of power equipment. After the temperature distribution map is obtained, it is preliminarily judged whether the equipment has fault signs through analysis. On the basis of image segmentation and

recognition, this chapter designs an infrared diagnosis method of power equipment based on the combination of relative temperature difference method and temperature data and realizes the basic thermal fault diagnosis function of the infrared image of power equipment.

4.1. Infrared Thermal Fault Diagnosis Principle.

The wavelength of the infrared spectrum is between 0.75 and 1,000 μm, which is outside the visible red light region, so it is also called the infrared ray. In the natural environment, because the molecules and atoms in any object move irregularly, they will continuously radiate energy outward, so infrared rays exist most widely compared with other electromagnetic radiation.

There are too many external interference factors and complicated conditions when studying actual objects, so people set up a physical model with maximum radiation power, that is, blackbody, whose incident radiation absorptivity is a fixed value 1 under all conditions. Although the blackbody hypothesized here is only an ideal model for studying infrared radiation, it is used as a standard to measure the radiation amount, and its thermal radiation law shows that the thermal radiation has a relationship with wavelength and temperature, so it is widely used in the precision calibration of infrared equipment and other fields.

There is a close relationship between the energy of an object and the motion of molecules and atoms inside it. The more violent the irregular motion inside the object, the greater the energy contained in the object and the higher its own temperature. On the contrary, for stationary particles, their temperature is absolute zero, and there is no radiant energy that will be emitted.

4.2. Infrared Thermal Fault Diagnosis Means.

Infrared detection technology, which is a noncontact detection technology, will inevitably bring the defect of low-temperature measurement accuracy in fault detection of substation equipment, which will directly affect the detected temperature value. In order to improve the accuracy of infrared diagnosis of power equipment, researchers and professional inspectors put forward three applicable diagnosis methods after a large number of measurement practices:

(1) Surface temperature judgment method

After measuring the surface temperature of power equipment with a thermal imager, the overheated part of the equipment is judged according to the relevant data standard of Common Technical Requirements of High-Voltage Switchgear and Control Equipment Standard, combined with the current environmental temperature condition and equipment load degree. This judgment method is simple and effective, but because of the influence of external factors on the accuracy of temperature value, there is a situation that the fault of small temperature rise cannot be accurately judged.

(2) Relative temperature difference judgment method

The relative temperature difference judgment method measures the temperature difference between two corresponding temperature measuring points of two devices that meet the same or similar basic states such as category and temperature and calculates the temperature rise at the higher point. Finally, the relative temperature difference σ_t defined by the relative temperature difference method is calculated by the ratio of the two, and its expression is shown in the following equation:

$$\sigma_t = \frac{\tau_1 - \tau_2}{\tau_1} \times 100\% = \frac{(T_1 - T_0) - (T_2 - T_0)}{T_1 - T_0} \times 100\%$$

$$= \frac{T_1 - T_2}{T_1 - T_0} \times 100\%,$$

(17)

where σ_t is the relative temperature difference of the measured point; τ_1 and T_1 are the temperature rise (unit: K) and temperature (unit: K) of the starting point, respectively, while τ_2 and T_2 are the temperature rise and temperature of the normal measured point respectively; and T_0 is the ambient temperature.

When the ambient temperature of the equipment is low or the measured temperature is within the normal value specified in GB736-90, it cannot be directly determined that the equipment does not have an abnormal heating problem, but it can be determined in the actual infrared diagnosis that the probability of thermal failure increases with the increase of ambient temperature. By calculating the relative temperature difference of equipment, this method is generally applicable to diagnose the fault of equipment caused by current or voltage.

(3) Similar comparison method

similar comparison method, as its name implies, is a method to compare and judge the temperature difference of similar equipment with the same working state and external environment including temperature and noise, which is used for fault diagnosis of voltage-induced equipment.

4.3. Diagnosis and Result Analysis of Substation Equipment.

In this paper, the relative temperature difference is calculated by the temperature data measured by the infrared thermal imager, and then whether there is a fault in suspected power equipment is diagnosed, and the severity of equipment fault is graded. This chapter uses the relative temperature difference method of power equipment combined with the fault diagnosis standard of common power equipment and carries out thermal fault diagnosis and fault classification experiments for three common substation equipment.

4.3.1. Selection of Thermal Fault Diagnosis Methods.

Through the definition of the relative temperature difference method in the previous section, it can be determined that

relative to other thermal fault diagnosis methods, such as surface temperature judgment method and similar comparison method, the relative temperature difference method can accurately judge whether there is a thermal fault in the measured power equipment. When a load of power equipment increases or a short circuit occurs, the relative temperature difference method can judge whether the power equipment is faulty or not by measuring the temperature of the heating part of the equipment and calculating the change of temperature rise, so it can judge the severity of the electrical equipment fault, so it is widely used in the thermal fault diagnosis of power equipment. The relative temperature difference is judged as shown in Table 1.

Attention should be paid to the use of the relative temperature difference judgment method in actual operation:

(1) When calculating the relative temperature difference, the three values measured, that is, the normal position, the suspected fault position, and the external environment, must be the values measured by the same instrument at the same time or at short intervals. It should be noted that the ambient temperature must be the value obtained by measuring the equipment in nonoperation by the thermal imager, and the atmospheric temperature cannot be simply measured by other methods; otherwise, the error of ambient temperature will lead to an inaccurate relative temperature difference value, which will affect the diagnosis and fault of suspected faulty equipment.

(2) Although the relative temperature difference is theoretically independent of the environment and load conditions, it is not applicable to all situations. In order to reduce the influence of solar radiation on the test results, the test value of the equipment should be carried out in the infrared environment with insufficient light as much as possible.

(3) For the fault of voltage-induced equipment, although theoretically, there is a similar relationship between the relative temperature difference and equipment loss. However, in general, the relative temperature difference method does not diagnose such faults.

4.3.2. Determination of Thermal Fault Level.

In this paper, the relative temperature difference method is used to judge the fault level of the infrared image of the power equipment to be measured, and the relative temperature difference σ_t of the equipment is calculated according to the expression. According to the σ_t of different power equipment under different thermal fault levels in Table 1, the threshold value of the three common power equipment required in this paper is set.

Firstly, the relative temperature difference σ_1 of high-pressure casing is determined. According to the criteria for judging the thermal defects of high-pressure casing in Table 1, σ_1 is set: if $\sigma_1 < 0.20$, the high-pressure casing is diagnosed to be in normal operation; If $0.2 \leq \sigma_1 < 0.8$, the

general fault is diagnosed; if $0.8 \leq \sigma_1 < 0.95$, serious faults will be diagnosed; and if $\sigma_1 \geq 95$, the emergency fault is diagnosed, and the diagnosis process is shown in Figure 5.

Secondly, the isolation relative temperature difference σ_2 is determined, which is similar to that of a high-voltage sleeve. According to the thermal defect judgment standard of isolating switch in Table 1, σ_2 is set: the diagnostic isolating switch is diagnosed as normal when $\sigma_2 < 0.35$ is different from σ_1; when $0.35 < \sigma_2 < 0.8$, it is diagnosed as a general fault; when $0.8 < \sigma_2 < 0.9$, it is diagnosed as a serious fault; and when $\sigma_2 \geq 0.95$, it is diagnosed as a critical fault, and its diagnosis flow is shown in Figure 6.

In the same step as the above two types of equipment, the relative temperature difference σ_3 of the clamp is finally determined. σ_3 is set according to the judgment standard of thermal defect of the clamp in Table 1. The diagnostic criteria of relative temperature difference σ_3 of clamp are the same as σ_2, and the diagnostic flow is shown in Figure 7.

4.3.3. Thermal Fault Diagnosis and Result Analysis.

As an important component of the substation system, high-voltage bushing plays an irreplaceable role and function, which ensures that the leads keep insulated to the ground by leading the internal leads of the transformer to the external insulating bushing of the oil tank. An infrared thermal imager can be used for daily inspection and maintenance of casing and prevent the occurrence of faults and defects.

The high-voltage bushing adjusts the lead direction and protects the lead to insulate it from the ground. When the internal overheating is serious, it will lead to the failure of the circuit connection and the breakdown accident of the dielectric, which is an important hidden danger that seriously affects the normal operation and power supply of the substation. In order to realize the thermal fault diagnosis of the high-voltage casing, according to the above processing steps, firstly, the infrared images of high-voltage casing are segmented and classified by SVM, and then the thermal fault diagnosis method combining relative temperature difference method and temperature data is used to diagnose and classify the thermal fault of the high-voltage casing.

This experiment runs under MATLAB 2020b platform, and its diagnosis results are shown in Figure 8. In the left infrared image, the stigma at the top of the casing is regarded as an abnormal hot spot. The diagnosis results show that the power equipment in the image is a high-voltage casing, and its relative temperature difference is 72.6%, which belongs to general defects. It can be concluded that the classification is correct, the fault grade obtained by the relative temperature difference method is accurate, and the operation countermeasures are given.

Isolating switch, as an interrupt device in substation, mainly completes the conversion between circuit devices and is a switch breaker without a special arc-extinguishing device. Because of its lack of arc extinguishing ability, the isolating switch is generally limited to operating without load. However, if an oil circuit breaker is installed in the circuit, it can also be operated with a load under certain technical conditions. Due to a series of outage maintenance

TABLE 1: Determination of relative temperature difference of current heating equipment.

Type of equipment	Relative temperature difference fault range		
	General fault type	Major thermal failure	Emergency thermal failure
SF6 circuit breaker	≥20%	≥80%	≥95%
Vacuum circuit breaker	≥20%	≥80%	≥95%
Oil-filled casing	≥20%	≥80%	≥95%
High-voltage switchgear	≥35%	≥80%	95%
Isolating switch	≥35%	≥80%	≥95%
Other conductive equipment	≥35%	≥80%	≥95%
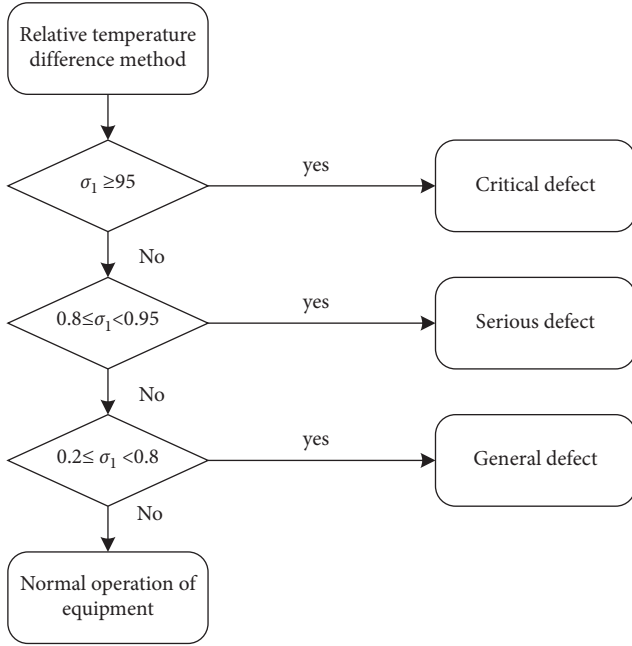
FIGURE 5: Diagnostic flow of high-pressure casing.
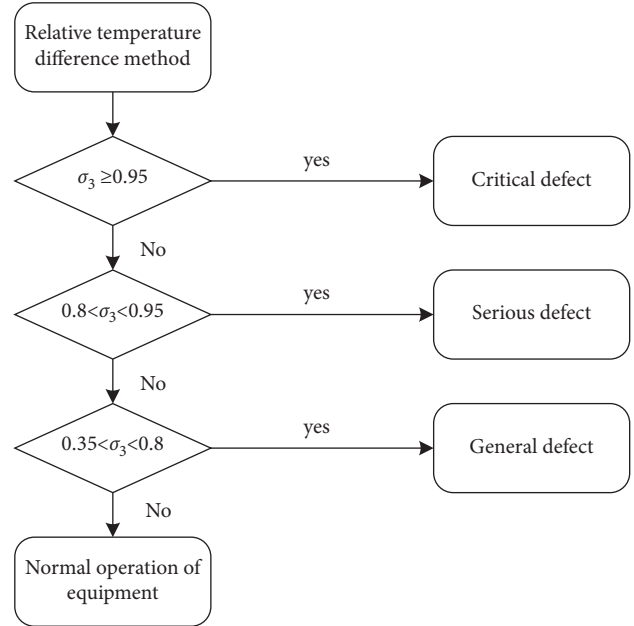
FIGURE 7: Diagnostic flow of device clamp.
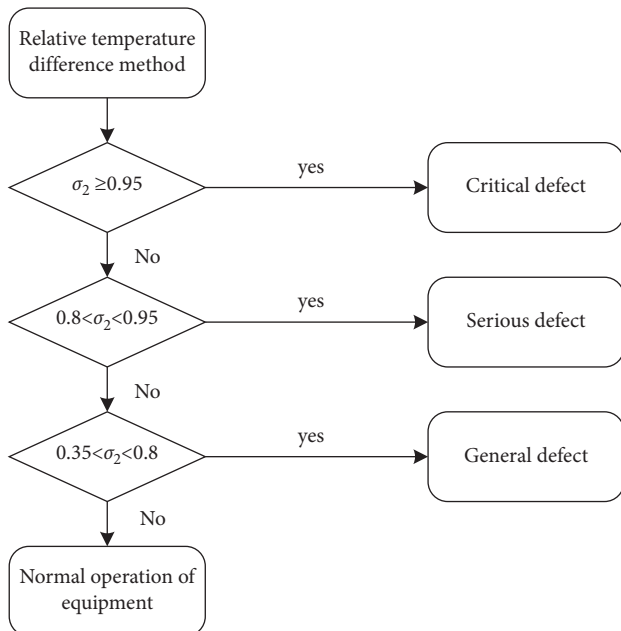
FIGURE 6: Diagnostic flow of disconnector.

work in the substation, the disconnector equipment operates frequently, and its conductive part is exposed to air, which leads to poor contact and overheating of equipment, so the probability of thermal failure is higher than the other two types.

This study is also carried out under MATLAB 2020b platform, and its diagnosis results are shown in Figure 9, in which the contact heads are divided into hot spots. The diagnosis results show that the power equipment in the image is an isolating switch, and its relative temperature difference is 80.2%, which is a serious defect. It can be concluded that the classification results are correct, the fault grade is accurate, and the relative treatment opinions are given.

As a common connecting component in the substation transmission line, the thermal fault of wire clamp generally has the following reasons: improper operation of hydraulic wire clamp during hydraulic installation leads to a thermal fault; bolt clamp in daily operation or maintenance of bolt fastening is not in place; or vibration phenomenon in operation leads to poor contact of clamp caused by thermal failure.

The diagnosis results of the disconnector under MATLAB 2020b platform in this study are shown in Figure 10.

Maximum temperature	63.9°C	Normal temperature of equipment	32.0°C
Type of equipment	High pressure casing		
Atmospheric temperature	20.0°C	Relative temperature difference	72.6%
Diagnostic results	General defect	Coping treatment	Pay close attention
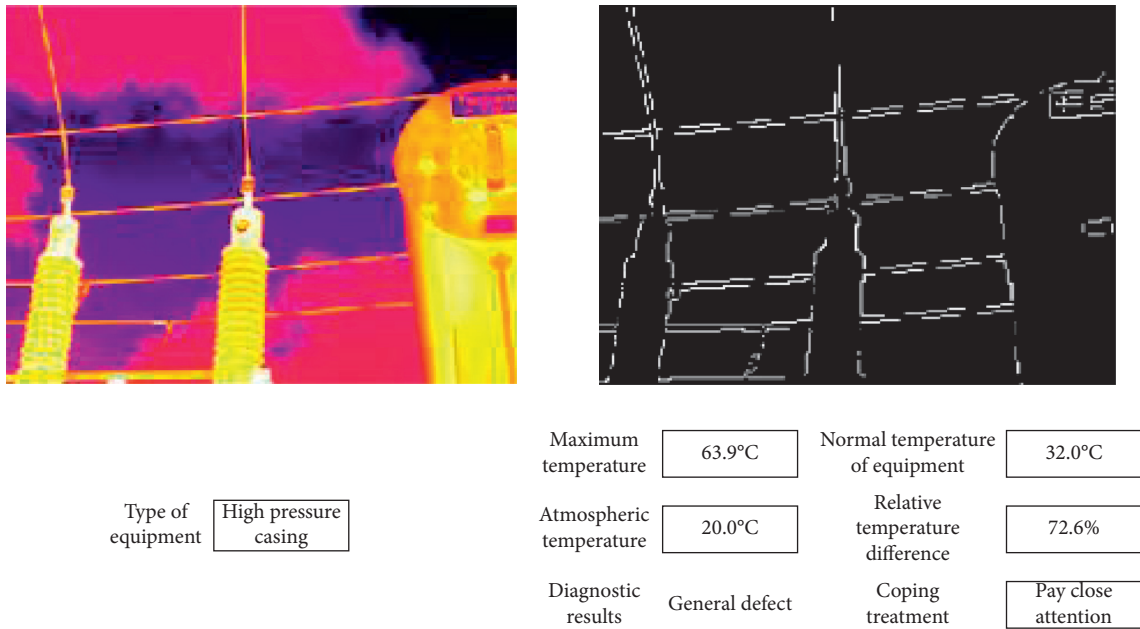
Figure 8: Thermal fault diagnosis results of high-voltage casing.

Maximum temperature	101°C	Normal temperature of equipment	36.0°C
Type of equipment	Isolating switch		
Atmospheric temperature	20.0°C	Relative temperature difference	80.2%
Atmospheric temperature	Serious defect	Coping treatment	Emergency treatment
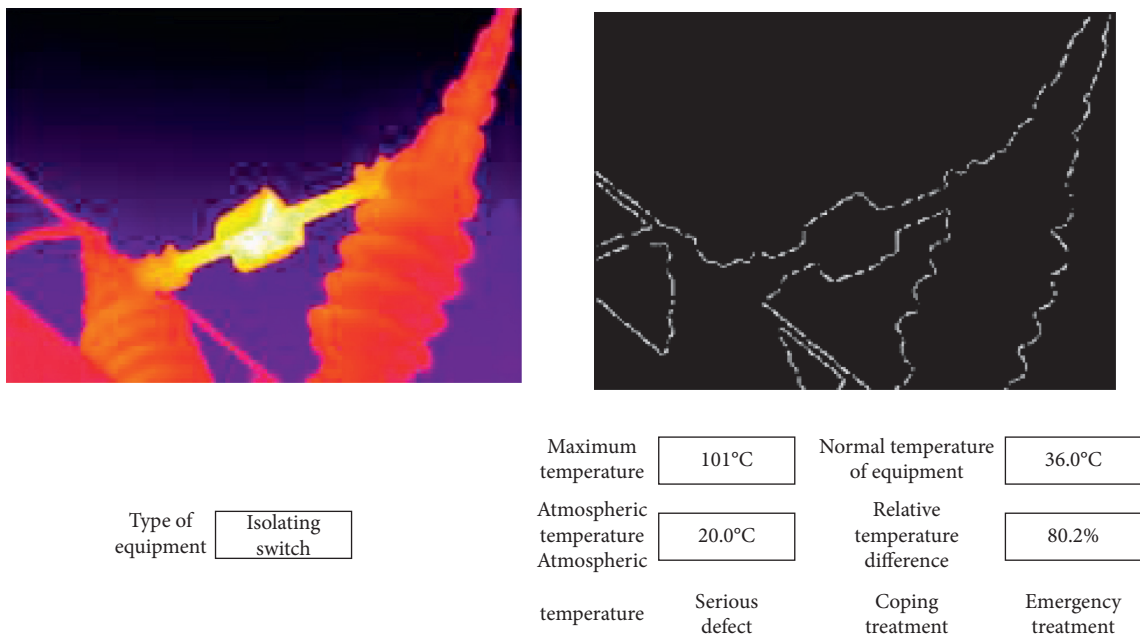
Figure 9: Disconnector thermal fault diagnosis result.

It can be seen from the above figure that the joint part of the wire clamp is obviously heated in the infrared image, and the diagnosis result shows that the power equipment in the image is an equipment wire clamp, and its relative temperature difference is 88.3%, which is a serious defect. It can be concluded that the classification result is correct, and the accurate fault level is also given, and there are relative treatment opinions.

4.3.4. Recognition Accuracy of Classification Experiment.
In this experiment, the infrared image classification recognizer based on HOG-SVM proposed in this paper is used to classify and identify substation equipment. The classification experiment flow is shown in Figure 11.

One hundred and fifty infrared images of high-voltage bushing, 100 isolating switches, and 100 wire clamps in substation equipment are selected. Among them, 100 infrared images of high-voltage bushing, 50 infrared images of isolating switches, and 50 infrared images of wire clamps are taken as training samples of HOG-SVM. The remaining 50 infrared images of high-voltage bushing, 50 isolating switches, and 50 wire clamps are used as the images to be classified by the HOG-SVM substation equipment classifier.

The classification and recognition experiment of substation equipment in this paper is implemented under the
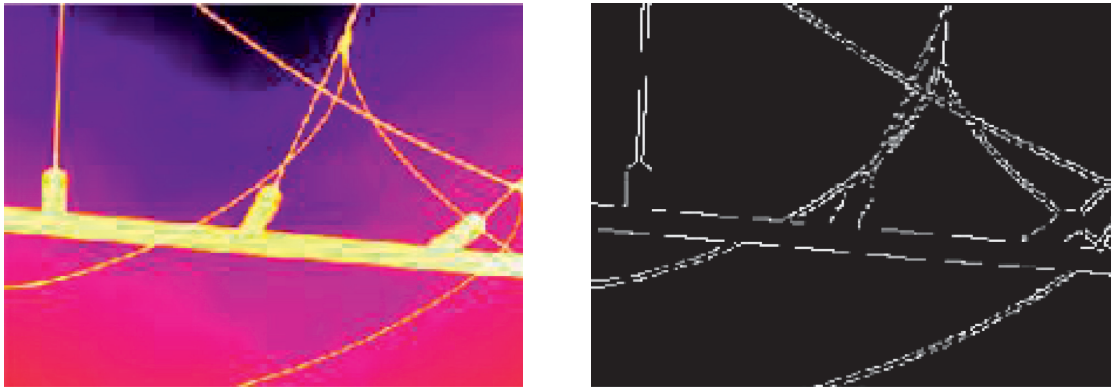
Type of equipment	Equipment clamp	Maximum temperature	71.5°C	Normal temperature of equipment	26.0°C
		Atmospheric temperature	20.0°C	Relative temperature difference	88.3%
		Diagnostic results	Serious defect	Coping treatment	Emergency treatment

FIGURE 10: Thermal fault diagnosis results of equipment clamp.
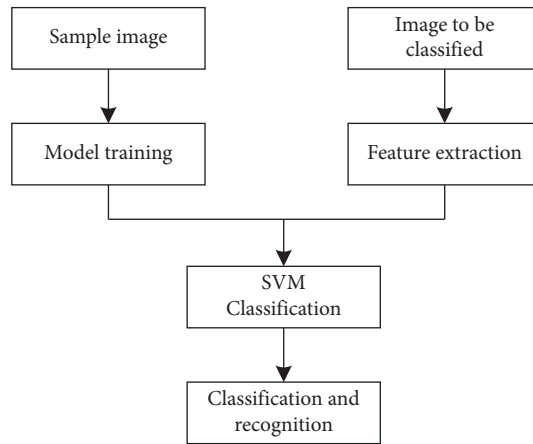
FIGURE 11: SVM infrared image classification process.

LIBSVM toolbox of MATLAB 2016b platform. The specific classification steps are as follows:

(1) Firstly, the infrared image preprocessing and image segmentation methods given in Sections 3 and 4 are used for image denoising, image enhancement, and image segmentation of all selected experimental samples.

(2) The HOG feature extraction and SVM multi-classification training are carried out on 100 infrared images of high-voltage sleeve, 50 isolating switches, and 50 wire clamps.

(3) At last, the classification method of HOG-SVM is used to classify and identify 50 images of high-voltage bushing, isolating switch, and clamp equipment in turn.

The classification results of each device are shown in Table 2.

It can be seen from Table 2 and Figure 12 that the HOG-SVM classification method is used to classify 50 infrared images of high-pressure casing, and the classification accuracy rate is 96%; the classification accuracy of isolating switches is 90%; and the classification accuracy of clamps is 100%. Through the comprehensive analysis of three common substation equipment, the total number of correct classification can be calculated to be 143, and the comprehensive classification accuracy of the three kinds of equipment reaches 95.3%. It can be seen that the infrared image classification method of substation equipment based on HOG-SVM proposed in this study has a good classification effect.

TABLE 2: Substation equipment classification results based on HOG-SVM.

	High-pressure casing	Isolating switch	Wire clamp
Total number of samples	150	100	100
Number of training samples	100	50	50
Number of samples to be classified	50	50	50
Correct classification number	48	45	50
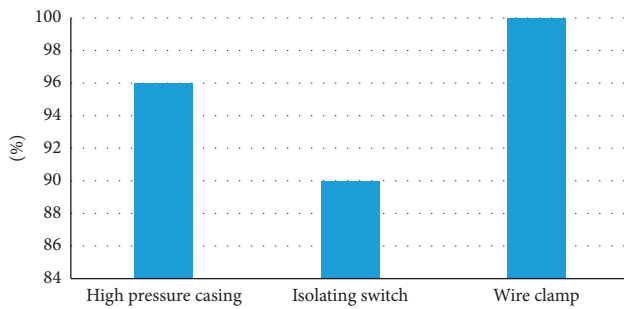
FIGURE 12: Classification accuracy.

5. Conclusion

After segmenting and classifying the infrared images of the equipment, this paper designs a fault diagnosis system for substation equipment based on infrared thermal image technology and realizes the accurate identification and diagnosis of thermal faults of three common power equipment, namely, high-voltage bushing, isolating switch, and equipment clamp. Based on the analysis of the features of the infrared image of power equipment in feature extraction, this paper designs a classifier for infrared power equipment image by using SVM and realizes the recognition of three kinds of power equipment, and the recognition accuracy reaches the predetermined expectation. In addition, on the basis of summarizing the types of electrical equipment faults and their causes and characteristics, in view of how to judge the thermal failure of equipment from the abnormal temperature rise of the thermal image, through the temperature difference discriminant method combined with infrared image temperature data on three types of common equipment in substations, the accurate diagnosis results and corresponding treatment suggestions are obtained, which verifies that the thermal fault diagnosis method of substation equipment designed in this paper is of practical significance.

With the rapid evolution of related infrared technology, it is believed that it will continue to exert its energy in the detection of power equipment. In the future, infrared diagnosis technology will be highly intelligent and integrated into the smart grid vigorously developed by State Grid. Infrared detection devices will be used to automatically collect infrared images, automatically process image information to diagnose faulty equipment, and upload them to the control terminal for the overall deployment of staff to carry out efficient detection and maintenance operations. In this paper, the classification and thermal fault diagnosis of substation equipment based on infrared images are basically realized. However, due to the limited time, there are still some shortcomings. For example, the number of infrared power equipment data sets collected is insufficient, which leads to the inability to classify equipment images more accurately. This is the place that needs to be improved in the future.

Data Availability

The experimental data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest regarding this work.

References

[1] H. Yanjun, C. Honggang, Z. Taoyun et al., "Design of transformer secondary voltage drop on-line monitoring system based on Internet of things," *IOP Conference Series: Materials Science and Engineering*, vol. 677, no. 5, Article ID 052028, 2019.

[2] C. Li, Z. Yu, and M. Zhuo, "Research on fault detection method of infrared thermal imaging for power equipment based on deep learning," *IOP Conference Series: Earth and Environmental Science*, vol. 714, no. 4, Article ID 042045, 2021.

[3] X. Gao, H. Liu, L. Du, and T. Wu, "Study on the online status monitoring technology of intelligent substation virtual circuit based on comprehensive evaluation and recognition method," *Dianli Xitong Baohu yu Kongzhi/Power System Protection and Control*, vol. 47, no. 3, pp. 182–188, 2019.

[4] S. Fan, T. Li, Y. Liu, Y. Gong, and K. Yu, "Infrared image-based detection method of electrical equipment overheating area in substation," *E3S Web of Conferences*, vol. 185, no. 10, Article ID 01034, 2020.

[5] X. Zhao, Z. Peng, and S. Zhao, "Substation electric power equipment detection based on patrol robots," *Artificial Life and Robotics*, vol. 25, no. 385–386, 2020.

[6] Q. Yu, H. Liu, Q. He, G. Zhang, Z. Zhang, and X. Zhang, "Research on pressure-based detection technology for partial overheat insulation defect of oil-less power equipment," *IOP Conference Series: Earth and Environmental Science*, vol. 632, Article ID 042009, 2021.

[7] A. Kadechkar, J. R. Riba, M. Moreno-Eguilaz, and J. Sanllehí, "Real-time wireless, contactless, and coreless monitoring of the current distribution in substation conductors for fault diagnosis," *IEEE Sensors Journal*, vol. 5, pp. 1693–1700, 2018.

[8] F. Bahri and H. Hasini, "Flow and heat dissipation analysis in transformer substation with minimal ventilation using CFD," *International Journal of Engineering & Technology*, vol. 7, no. 4, pp. 327–332, 2018.

[9] P. Shang, H. Dong, X. Li, and W. Ren, "Research on fault diagnosis method of 750kV substation based on Bayesian network and fault recording information fusion," *Journal of*

Physics: Conference Series, vol. 1550, no. 5, p. 5, Article ID 052020, 2020.

[10] S. Biricik and H. Komurcugil, "Three-level hysteresis current control strategy for three-phase four-switch shunt active filters," *IET Power Electronics*, vol. 9, no. 8, pp. 1732–1740, 2016.

[11] Z. Lu, T. Wang, J. Zhang, T. Jin, and S. Lu, "Research and application of substation site maintenance assistant safety management system based on computer control system," *Journal of Physics: Conference Series*, vol. 1574, Article ID 012112, 2020.

[12] A. Nasiri, A. Taheri-Garavand, M. Omid, and G. M. Carlomagno, "Intelligent fault diagnosis of cooling radiator based on deep learning analysis of infrared thermal images," *Applied Thermal Engineering*, vol. 163, Article ID 114410, 2019.

[13] F. Hu, H. Chen, and X. Wang, "An intuitionistic kernel-based fuzzy C-means clustering algorithm with local information for power equipment image segmentation," *IEEE Access*, vol. 8, pp. 4500–4514, 2020.

[14] M. A. Basaran, B. Simonetti, and A. A. Basaran, "Quantification of qualitative assessments using computing with words: in framework of fuzzy set theory," *Soft Computing*, vol. 24, no. 18, pp. 13565–13577, 2020.

[15] J. Fang, X. Song, N. Yao, and M. Shi, "Application of FCM Algorithm combined with articial neural network in TBM operation data," *Computer Modeling in Engineering and Sciences*, vol. 126, no. 1, pp. 397–417, 2021.

[16] M. Chen, C. Tang, M. Xu, and Z. Lei, "Binarization of optical fringe patterns with intensity inhomogeneities based on modified FCM algorithm," *Optics and Lasers in Engineering*, vol. 123, pp. 14–19, 2019.

[17] X. Zhao, H. Wei, H. Wang, T. Zhu, and K. Zhang, "3D-CNN-based feature extraction of ground-based cloud images for direct normal irradiance prediction," *Solar Energy*, vol. 181, pp. 510–518, 2019.

[18] L. Wang, J. Gui, Z. M. Lu, and C. Liu, "Fast pedestrian detection and tracking based ON vibe combined hog-svm scheme," *International Journal of Innovative Computing Information and Control*, vol. 15, no. 6, pp. 2305–2320, 2019.

[19] X. Zhang, T. Xue, and H. Eugene Stanley, "Comparison of econometric models and artificial neural networks algorithms for the prediction of baltic dry index," *IEEE Access*, vol. 7, pp. 1647–1657, 2019.

[20] G. Chen and S. Li, "Research on location fusion of spatial geological disaster based on fuzzy SVM," *Computer Communications*, vol. 153, pp. 538–544, 2020.

[21] G. Chen, X. Xie, and S. Li, "Research on complex classification algorithm of breast cancer chip based on SVM-RFE gene feature screening," *Complexity*, vol. 2020, Article ID 1342874, 12 pages, 2020.

Research Article

Application of QR Code Online Testing Technology in Nursing Teaching in Colleges and Universities

Hai-yan Zheng [1] **and Xing-cheng Ran**[2,3]

[1]*Nursing School, Hexi University, Zhangye, Gansu Province, 734000, China*
[2]*The School of Information Science and Engineering, Lanzhou University, Lanzhou, 730000, Gansu Province, China*
[3]*Information Construction Management Service Center, Hexi University, Zhangye, 734000, Gansu Province, China*

Correspondence should be addressed to Hai-yan Zheng; lzu_zhenghy@lzu.edu.cn

In view of the current situation that the nursing teaching content is lacking pertinence and cannot effectively cultivate students' critical thinking, this paper puts forward the application analysis of two-dimensional code online testing technology in nursing teaching in colleges and universities. Through the analysis of two-dimensional code recognition technology architecture and common application fields, an intelligent nursing teaching platform for two-dimensional code online testing technology was designed. Systematically, we summarize the advantages of intelligent nursing teaching platform, use flash and 3D technology to complete the virtualization of nursing teaching scene, and use XML technology to update and save teaching resources; we further build an intelligent nursing teaching platform by using the basic plate, nursing station plate, ward plate, dispensing room plate, and case and operation review plate. The experimental results show that under the two-dimensional code online test technology, the nursing teaching method in colleges and universities has strong teaching resource processing efficiency, improves the operation level of students' nursing technology, and effectively cultivates the academic thinking of nursing students, which is of great significance to the progress of nursing teaching.

1. Introduction

Nursing teaching is the extension and expansion of higher education. It is an important period for nursing interns to organically combine theoretical knowledge and practical skills into clinical practice, cultivate their corresponding professional competence, and make a smooth transition from school stage to clinical stage [1, 2]. With the advent of the era of information globalization and the rapid development of new media forms that provide information and entertainment to users through terminals such as computers, mobile phones, and digital televisions, the pace of learning and life of modern people has become more rapid and convenient. With the wide application of smart phones, college students put forward new requirements for nursing teaching in colleges and universities [3, 4]. In the evaluation system of college teachers, it is required to increase the

evaluation of teaching technology design and development ability in order to improve the teacher teaching evaluation system. It can be seen that college teachers need to use the new media platform to improve the quality of teaching work, whether it is the requirements of student groups or the education and teaching system. In recent years, educational experts, teachers, and students in educational practice are trying to explore new educational ideas and methods based on network. Teachers' new media ability not only includes the operation and use of various hardware equipment but also should flexibly use hardware to design and develop new teaching methods and display and transmit the teaching content to school students or social learners through various electronic media. The interaction in the classroom is not only the interaction of language, gestures, and looks between teachers and students, but also the interaction between teachers and media technology, students and media

technology. Therefore, the combination of traditional teaching mode and modern mobile information technology has become an inevitable trend.

Traditional nursing teaching uses teachers to teach at fixed places and at fixed times, which can complete the specified teaching tasks. However, due to the large number of students and the large classroom area, the traditional desk teaching has the disadvantages of poor teaching interaction, and it is not easy to eliminate skipping classes. The teaching method is boring [5–7]; therefore, two-dimensional code online testing technology is applied to guide nursing teaching in colleges and universities. Through the analysis of the basic application principle of two-dimensional code recognition technology, the focus of nursing teaching in colleges and universities is determined, the application database of multiplatform information nursing teaching is constructed, and the application principle of multiplatform information teaching in nursing teaching is determined [8], according to which the multiplatform information teaching is carried out. Relying on the intelligent nursing teaching platform for a systematic evaluation of each student, according to different assessment, results will be divided into different types of students, according to different types of targeted interest training and targeted interest guidance, to build a concise internal medicine nursing system, so that college students have a comprehensive understanding of internal medicine nursing. Breaking the traditional teaching method of fixed time and fixed place, relying on two-dimensional code online testing technology, training different types of students, assigning different homework at the same time, relying on the multiplatform information supervision system, supervising students in class, eliminating truancy, not finishing homework on time, and other phenomena are considered. We construct student-student learning group and student-teacher learning group to discuss nursing knowledge, increase communication with internal medicine nursing staff, and explain nursing according to the actual situation. We also complete the application analysis of two-dimensional code online testing technology in nursing teaching in colleges and universities.

2. Basic Application Principle of Two-Dimensional Code Recognition Technology

2.1. Architecture Analysis of QR Code Recognition Technology. In the process of practical application of QR code recognition technology [9, 10], data information can be recorded according to a certain plane distribution law through specific geometric figures, which belongs to a kind of bar code. The recognition process mainly realizes the automatic recognition of symbols by using the corresponding binary rules. Before the generation of two-dimensional code identification technology [11], the widely used one-dimensional code must be matched with the corresponding computer database in order to read the data information in the one-dimensional code and can only identify the product identification, while the two-dimensional code identification technology can realize the detailed description of the product. Figure 1 is a schematic diagram of QR code composition.
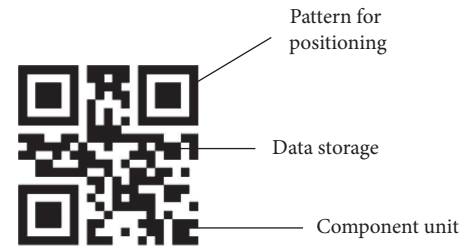
Figure 1: Schematic diagram of QR code composition.

At present, common QR code formats include PDF486, MaxiCode, Han Xin code, Aztec code, and QR code. Among them, QR code matrix is the most commonly used format in the industry, and the QR code in this format is called fast response code. As can be seen from Figure 1, the QR code has a square structure as a whole, and the patterns can be roughly divided into three types: positioning graphics, data storage graphics, and various organization units. The figure for positioning is the "back" character pattern in Figure 1, which is mainly used to realize the positioning in the recognition process. In the process of scanning the QR code, the user does not need to align the graphics and can effectively extract the relevant information in the QR code at any scanning angle. In QR code recognition technology, the storage form of data information is usually distributed according to matrix. In addition to supporting corresponding data and text information, it also supports the storage and display of various audio, video, and image information.

2.2. Common Application Fields of QR Code Recognition Technology. At present, QR code recognition technology has been widely used. According to different business forms of application, it can be divided into read type and main read type [12]. When the two-dimensional code recognition technology is applied in the read business field, the business information can generate the corresponding two-dimensional code graphics through different encryption and compilation methods. Send the graphics to the user's mobile data terminal, and the user can scan the QR code graphics with professional identification software. After obtaining the identification authority, the user can extract the corresponding information, which is often used as the transaction voucher for various applications [13, 14]. When applying QR code recognition technology in the main reading business field, users can install the corresponding QR code production client, scan the QR code on read-only media and other similar media, and obtain all information stored in the QR code graphics after decoding. At the same time, some QR codes also contain corresponding applications. Through scanning, you can directly open the corresponding programs to realize the functions of fast browsing of web pages and fast query of graphics and text.

In nursing teaching, the teacher assessment system requires to increase the assessment of teaching technology design and development ability and improve the teacher teaching evaluation system. Whether for students or the

requirements of education and teaching system, college teachers need to use two-dimensional code recognition technology to improve teaching quality. Experts in the field of education, educational practitioners, teachers, and students are trying new ideas and methods of network education. The teaching content is presented to school students or social learners through two-dimensional code recognition technology, so that it can be displayed and disseminated.

3. Construction of Intelligent Nursing Teaching Platform

In order to help students effectively realize the role transformation from nursing students to interns and then nurses and follow the trend of the times, an intelligent nursing teaching platform for QR code online testing technology is designed. Its multiplatform information-based teaching interest exploration system is shown in Figure 2.

As can be seen from Figure 2, multiplatform information-based teaching interest mining starts from students' consciousness, analyzes the behaviors that can stimulate teaching interest, classifies students into five categories, designs personalized teaching platform and personalized teaching model, and puts forward the design of personalized learning platform on this basis, so as to complete the multiplatform information-based teaching interest mining system, The specific design contents are as follows. As a new storage medium, the intelligent nursing teaching platform based on two-dimensional code online testing technology carries a large amount of information, small volume, low cost, and simple production. In addition, it is convenient for users to store two-dimensional code into mobile phones and use fragment time to learn and consolidate at any time. College students in the new era like innovation and are no longer confined to the traditional methods of teaching nursing in colleges and universities. The application of QR code in nursing teaching activities in colleges and universities meets the development requirements of the times and is worthy of clinical application and promotion.

3.1. Platform Design Concept.
According to the different modes of users entering virtual reality (VR) and the differences of immersion level, combined with the research results of researchers at home and abroad in the field of VR and the real situation of its application in education, VR is divided into four categories: desktop VR, immersive VR, augmented reality VR, and distributed VR.

The platform should focus on simulating the real operating environment and process, not simply simulating the visual experience. Using the intelligent nursing teaching platform [15, 16], teachers can more thoroughly teach the operation process under the real nursing state and actively update teaching resources such as drugs and cases. Students can review and evaluate their learning results by themselves.

3.2. Development Environment.
The key function of the platform is to standardize the nursing process and enhance students' ability to consult and analyze real cases. The
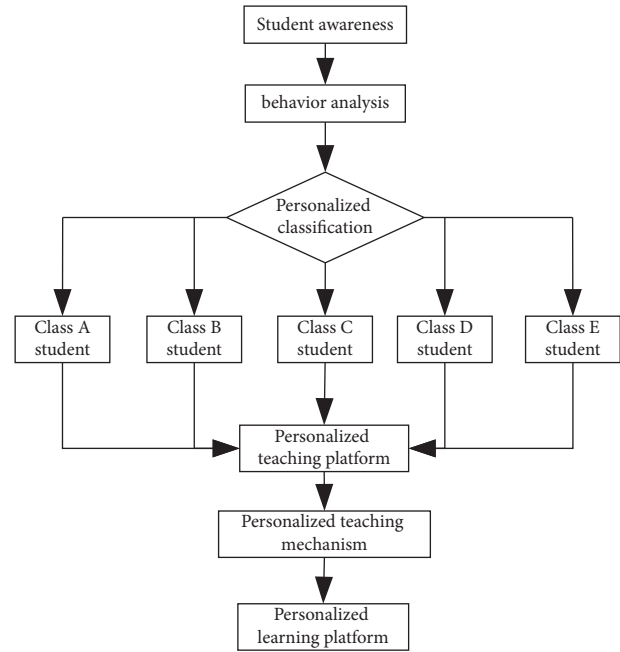
FIGURE 2: Interest exploration system of multiplatform information teaching.

platform is developed by using the current relatively complete flash integrated extensible markup language (XML) technology. Flash is a kind of animation software with rich animation content, short time, and strong interaction. It is widely used in web animation design. However, the application platform developed by flash also has some defects: it cannot be directly connected with the database, resulting in poor applicability of the platform application in data update. XML technology can take structural analysis of documents and information for transmitting and saving data. XML technology can operate and be applied independently of the database for lightweight data storage. Coupling flash and XML technology can make the nursing platform under QR code online testing technology have a better feeling of use.

3.3. Platform Teaching Resource Compression Algorithm.
By compressing the data of teaching resources, it is convenient to complete the timely interaction between teachers and students and the efficient learning of students. Compression rate and compression deviation are the fundamental elements to weigh the compression of nursing teaching resources [17–19]. The resource compression ratio is the ratio of the number of compressed data to the number before compression, and the result is

$$\eta = \frac{N_b - N_e}{N_b} \times 100\%. \tag{1}$$

Set a total of N_b nodes in curve F and record them as

$$F = \{f_1, f_2, \ldots, f_{Nb}\} = \{(x_1, y_1), (x_2, y_2), \ldots, (x_{Nb}, y_{Nb})\}. \tag{2}$$

The fixed resource compression ratio $(0 \leq \eta \leq 1)$ is equal to the number of nodes N_e and $N_e = N_b \times (1 - \eta)$ of the

curve after compression, keeping N_e always an integer. Compress the curve F into a curve $F\prime$ with N_e nodes and minimize the compression deviation E.

Thus, the compressed curve F' is expressed as

$$F' = [f'_1, f'_2, \ldots, f'_{Ne}]. \tag{3}$$

It can be seen that F' is a subset of F, and any point in curve F' belongs to curve F. In the process of curve compression, the initial point and end point of the curve shall be saved to obtain

$$f'_1 = f_1, \quad f'_{Ne} = f_{Nb}. \tag{4}$$

The total displacement compression $E_\sum$ is regarded as the evaluation criterion of compression deviation. If the connecting straight line T between point f_i and point f_j in the initial curve F is $y = ax + b$, the distance from the point $f_k(x_k, y_k)$ between the point f_i and the point f_j in the initial curve F to the straight line T is d. If some curves in curve F are compressed into subsegments of curve F', the compression deviation generated is

$$E_\sum (f'_s, f'_{s+1}) = E_\sum (f_i, f_j) = \sum_{k=i+1}^{j=1} \frac{(y_k - ax_k - b)^2}{1 + a^2}. \tag{5}$$

Calculate the curve F' after the compression of nursing teaching resources [20, 21], that is, each node of N_e in curve F', and minimize the value of compression deviation function:

$$E_\sum (F\prime) = \min \sum_{s=1}^{Ne-1} E_\sum (f'_s, f'_{s+1}). \tag{6}$$

The fundamental theorem of dynamic programming algorithm is to resolve the problem to be solved into multiple subproblems, calculate the subproblems in advance, and then obtain the calculation results of the original problem from the solutions of many subproblems. For the optimization calculation of the deviation function of the compression of nursing teaching resources in the platform [22], first set a two-dimensional state space Ω without continuous relationship, and establish the correlation between the number of nodes N_b of curve F before compression and the number of subsegments of curve $F\prime$ after compression. If the number of $F\prime$ subsegments of the compressed curve is H, the expression of state space Ω is

$$\Omega = \{(n_b, h) | n_b = 1, 2, \ldots, N_b; h = 0, 1, \ldots, H\}. \tag{7}$$

The random point (n_b, h) of the state space Ω represents the compression optimization of the curve $\{f_1, f_2, \ldots, f_{n_b}\}$ with n_b nodes. The compressed curve contains h subsegments; that is, $n_e = h + 1$ nodes. Therefore, the final solution state of the nursing resource compression problem is (N_b, H), and the compressed curve $F\prime$ is a line from state $(0, 1)$ to state (N_b, H), which is recorded as P. In the state space Ω, the cost function $D(n_b, h)$ of the state (n_b, h) is also set, which is the compression deviation of the curve $\{f_1, f_2, \ldots, f_{nb}\}$ with n_b nodes that becomes h subsegments after compression.

The process of calculating the minimum deviation is to find the best path from state $(0, 1)$ to (N_b, H). Simplify the state space Ω and obtain the following boundary function constraints:

$$L(h) = \frac{h + 1}{N_b},$$
$$R(h) = \frac{1}{N_b - H + h},$$
$$B(n_b) = \frac{1}{n_b - N_b + H}, \tag{8}$$
$$T(n_b) = \frac{n_b - 1}{H - 1}.$$

The solution process of resource compression optimization problem is described by the following recursive analytical formula:

$$D(n_b, h) = \min_{L(h-1) \leq j < n_b} \left\{ D(j_h - 1) + E_\sum (f_j, f_{n_b}) \right\},$$
$$A(n_b, h) = \arg\min_{L(h-1) \leq j < n_b} \left\{ D(j_h - 1) + E_\sum (f_j, f_{n_b}) \right\}. \tag{9}$$

In order to improve the compression speed of nursing teaching resources [23–25], an improved method of adaptively modifying the search range is designed. A reference compression path is derived by Douglas–Peucker method and expressed as formula (10), and then the next iteration calculation is carried out by virtue of the reference path P, and the last calculation result is used as the reference of the next iteration.

$$P = \{p(0), p(1), \ldots, p(H)\}. \tag{10}$$

Construct the lowest deviation search space $\Omega\prime$ and enhance the search rate. Since the reference path has been obtained in the previous step, the path can be used to create a search strip to reduce the search range. The corner shape of each point in the path is different; that is, the compression degree is not equal. The search strip width W will be adjusted adaptively according to the change of compression degree. The boundary function of the improved search space $\Omega\prime$ is described as

$$L(h) = \frac{\max\{h + 1, h + (p(h) - p(h-1)) \times H/N_b\}}{h + 1},$$
$$R(h) = \frac{\min\{N_b, h + 2 \times (p(h) - p(h-1)) \times H/N_b\}}{N_b, H = h}. \tag{11}$$

The key of the improved method lies in the compression deviation and searching for the minimum compression deviation. For each node, the deviation from other nodes shall be calculated, and the complexity of its calculation time shall be set to $O(WNb^2/H)$. There are WH state space points in the process of resource compression. The time complexity of minimum deviation search is set to $O(WNb/H)$.

Therefore, it is obtained that the time complexity of nursing teaching resource compression under QR code online testing technology is $O(Nb^2W^2/H)$.

Through the above process, the compression of nursing teaching resources can be completed in the shortest time [26], which not only ensures the stability and fluency of the platform, but also solves the practical problem of small memory of the mobile terminal.

3.4. Platform Design Process. Based on the real investigation and research and the guidance of relevant experts, the platform planning and design process is shown in Figure 3. After logging into the platform, the user selects the nursing time, browses the patient's condition information and relevant nursing planning, and matches the patient with reasonable medication time according to the nursing planning and cases and patient's physical condition [27, 28]. The design process refers to the real nursing process and records all the operation behaviors of users in the process as a powerful voucher for follow-up investigation.

3.5. Platform Structure Implementation. Using browser/server mode structure and modular design strategy, the global platform functions are divided into basic plate, nursing station plate, ward plate, dispensing room plate, case and operation review plate, as shown in Figure 4.

In the whole platform, the basic plate is the most critical. Its design principle is based on the complete investigation of the real nursing environment and process [29, 30]. It uses system analysis to extract public attributes for several application goals and program goals and envisages the scalability of the platform in the future. The basic section includes the environmental variables required to initialize the platform, adjust user operation records, and extract basic data such as department and examination indicators according to the real nursing situation of the hospital [31, 32].

The nursing station plate, ward plate, dispensing room plate, case plate, and other plates present independent characteristics, which can provide students with corresponding nursing data under simulated ward detailed environmental conditions [33–35]. Such sections use the information stored in a large number of XML files. Dispensing room and ward are the core communication plates, including communication between patients and nursing students and drug matching [36]. The platform will record the user's operations in all sections as the basis for the final assessment.

The operation review section compares the existing data such as cases and medical orders according to the operation records of the platform users, comprehensively judges the user's global simulation process, finds out the wrong behavior and missing nursing process, and gives corresponding reminder information, so as to complete the application analysis of QR code online testing technology in nursing teaching in colleges and universities.
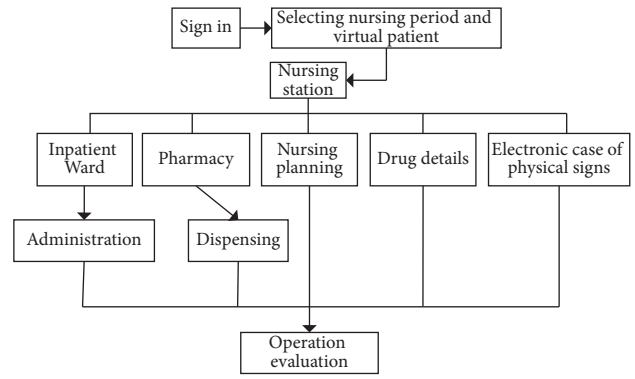
FIGURE 3: Platform planning and design process.
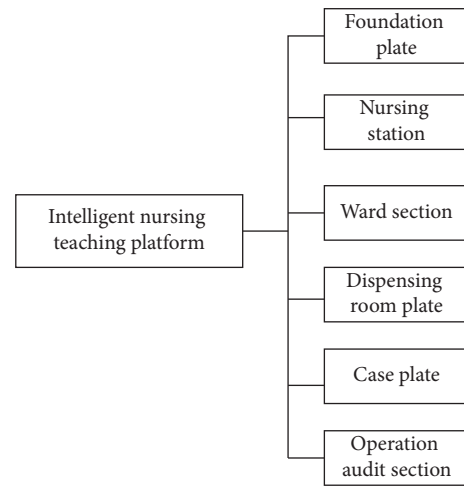
FIGURE 4: Schematic diagram of platform structure.

4. Experimental Analysis

In order to ensure the effectiveness of the two-dimensional code online testing technology proposed in this paper in college nursing teaching, simulation experiments were carried out. In the process of experiment, students studying in the course of internal medicine nursing in colleges and universities were taken as experimental objects, and numerical quantification was carried out for students studying in the course of internal medicine nursing in colleges and universities, which was conducive to the rapid progress of simulation experiment. The different learning abilities and the difficulty of learning knowledge points of the experimental objects are simulated and quantified. According to the quantitative results, the simulation experiments of educational interest cultivation, educational diversity, and educational interaction are carried out. Through the experimental demonstration of interest cultivation, educational diversity and interaction, the effectiveness of the application analysis of the two-dimensional code online test technology proposed in this paper in nursing teaching in colleges and universities is verified. In order to ensure the preciseness of the experiment, the traditional teaching method is used as the comparison object, the traditional

teaching method is quantified, the results of two simulation experiments are compared, and the experimental data are presented in the same data chart.

4.1. Data Preparation. In order to ensure the accuracy of the simulation experiment process and set the experimental parameters of the test, in this simulation experiment, students of internal medicine nursing courses in colleges and universities were used as experimental objects, using two different teaching methods to carry out the simulation experiment of cultivating educational interest and educational diversity and analyzing the simulation experiment results. Because the analysis results and analysis methods obtained by different methods are different, it is necessary to ensure the consistency of experimental environment parameters in the experimental process. The experimental data setting results in this paper are shown in Table 1.

4.2. Simulation Experiment of Cultivating Educational Interest. The data of 6000 students were quantified to construct different learning ability parameter values, with the unit of D. From 0 to 1.0 d, 0.1d to 0.2 d are class A students, 0.3 d to 0.4 d are class B students, 0.5 d to 0.6 d are class C students, 0.7 d to 0.8 d are class D students, and 0.9 d to 1.0 d are class E students. According to different types of students and different learning abilities, the simulation experiment of cultivating educational interest is carried out in three environments: simple, moderate, and difficult. In the process of experiment, we should pay attention to the quantification of students and the quantitative expression of the nature of simulated experiment. Because interest is a subjective parameter, human consciousness is heavy, and different quantitative methods lead to different conclusions. Therefore, Jeffrey Hall's interest analysis model is used in the simulation experiment of educational interest cultivation, and the interest simulation constant is used as the parameter to quantify interest, with % as the parameter unit. The analysis principle is to construct different internal medicine nursing teaching options and use students with different learning abilities to click, judge, and click the modules they are interested in according to the degree of difficulty. If the number of clicks is high, it is considered that there is strong interest. On the contrary, the interest in medical nursing is relatively flat. Using the same personnel and the same quantization mechanism, build the same learning ability as the abscissa and the interest simulation constant as the ordinate. The experiment is simple, moderate, and difficult. The experimental results are recorded in the simulation experiment result curve of educational interest cultivation. During the experiment, because the analysis results cannot be directly compared by using two different education and teaching methods, the third-party analysis and recording software is used to record and analyze the experimental process and results, and the results are displayed in the curve of the experimental comparison results. In the simulation experiment result curve, the analysis function is used to eliminate the uncertainty caused by the factors of simulation laboratory personnel operation and simulation computer

equipment. The simulation experiment of cultivating educational interest is carried out only for different experimental objects using different education and teaching methods. The comparison curve of the simulation experiment results of educational interest cultivation is obtained, and the comparison curve of the experimental results is shown in Figure 5.

By analyzing the comparison curve of the simulation experiment results of educational interest cultivation in Figure 5, it is concluded that the teaching application method proposed in this paper is more obvious than the traditional teaching method, and the improvement of students' interest is more prominent in different stages. In the simple experimental environment, with the traditional teaching method, the learning ability increases from 0 to 0.5 d, and the interest simulation constant increases by 10% for each 0.1d increase in learning ability, of which the 0 to 0.5 d range is class A weariness, class B passive, and class C mechanical students. However, in general, class A, class B, and class C students have low ability to cultivate interest in traditional teaching methods. The average interest simulation constant of the three types of students is 55%. Cultivating interest is not ideal. When the learning ability exceeds 0.6 d, the interest simulation constant is higher than 80% on average, indicating that the simulation experiment is in a simple state, which has little impact on the interest of class D enterprising students and class E autonomous students, and can obtain high interest through traditional education methods. It is concluded that the traditional teaching method is not obvious enough to cultivate interest in students with low learning ability. This paper puts forward the teaching application method. In a simple simulation environment, the overall change is moderate and close to a straight line, indicating that the change of learning ability has no obvious change on interest behavior. Students with poor learning ability also have high learning interest. When the learning ability is 0.1d, the interest simulation constant of the teaching application proposed in this paper is 80%, which is close to the learning ability of 0.6 d in the traditional teaching method. This paper puts forward the teaching application. When the learning ability exceeds 0.6 d, the interest simulation constant is greater than 86%. Compared with the traditional teaching method, the interest simulation constant of the same type of students is increased by at least 6%. At the same time, this paper puts forward teaching application methods. Different types of students have similar interests, the span of interest constant is small, and the overall operation is good. However, traditional teaching methods are easy to cause interest polarization, resulting in academic performance polarization. Class E and class D have better academic performance, and class C, class B, and class A have poor academic performance; it shows that the teaching method proposed in this paper plays a positive role in promoting students' interest in a simple simulation environment.

When the simulated experimental environment is moderate and difficult, using traditional teaching methods, students with different learning abilities have different specific manifestations of interest. Under the moderate

TABLE 1: Experimental parameter setting.

Project	Execution range/parameters	Remarks
Number of students	6000	Male : Female = 1 : 1
Analog quantization parameters	0.1–1.0	Normal distribution
Student level	Colleges and universities	Students majoring in internal medicine nursing
Test difficulty	0.4,0.7,0.9	Simple, moderate, and difficult
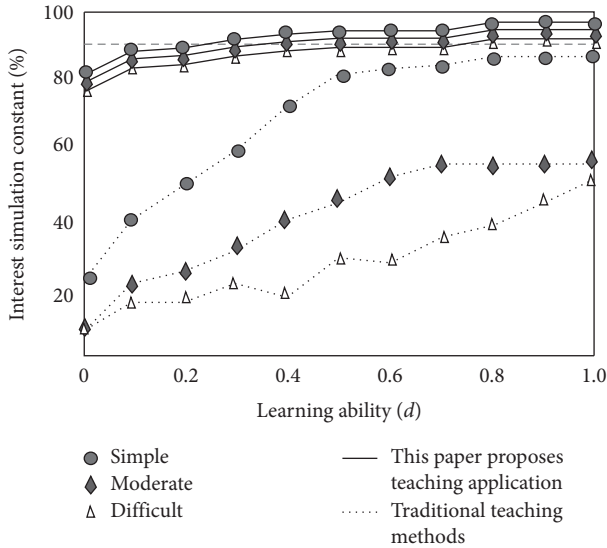
FIGURE 5: Comparison curve of simulation experiment results of educational interest cultivation.

experimental simulated environment, the curve trend is basically similar to that under the simple environment, but the interest simulation constant is lower than that under the simple environment. Under the moderate difficulty, the learning ability is before 0.7 d, rise in restricted line. When the learning ability is 0.7 d, the interest simulation parameter is 55%. In the range of 0 to 0.7 d, the learning ability is improved by 0.1d and the interest simulation constant is increased by 5%. When the learning ability is greater than 0.7 d, the interest simulation constant presents a stable state of 55%. Similarly, it has a greater impact on class A, class B, and class C students and less impact on class D and class E students, but the interest simulation constants of class D and class E students are not good enough at this time. In the difficult experimental environment, the traditional teaching methods improve linearly as a whole and have different effects on different types of students. When the learning ability is 0.4 d, the interest simulation constant decreases briefly. The highest interest simulation constant is 50%. In the range of learning ability from 0 to 0.6 d, the interest simulation constant increases in stages, and two growth platforms appear. Class A, class B, and class C students with learning ability ranging from 0 to 0.6 d have low interest. The learning ability increases linearly from 0.6 d to 1.0 d. For every 0.1d increase in learning ability, the interest simulation constant increases by about 5%, and the overall performance is poor. In this paper, the simulation constant of interest in teaching application is proposed, which is basically consistent with the curve in simple case. The average interest simulation constant in difficult environment is 2.5%

different from that in moderate environment, and the average interest simulation constant in moderate environment is 2.0% different from that in simple environment. The overall curve is stable. The minimum interest simulation constant in complex environment is 75% and the maximum interest simulation constant is 89%. In moderate environment, the lowest environmental simulation constant is 78%, and the highest interest simulation constant is 91%.

The weighted calculation by the third-party analysis software shows that, under the simple environment, the interest simulation constant of the traditional teaching method is 63%; under the moderate environment, the interest simulation constant of the traditional teaching method is 41%; and under the difficult environment, the interest simulation constant of the traditional teaching method is 35% and the average interest simulation constant is 46.3%. Then it is considered that the interest simulation constant of traditional teaching methods is 46.3%. In a simple environment, this paper proposes that the interest simulation constant of applied teaching methods is 96%. In a moderate environment, this paper proposes that the interest simulation constant of applied teaching methods is 93%. In a difficult environment, this paper proposes that the interest simulation constant of applied teaching methods is 90%, and the arithmetic weighted analysis takes the average interest simulation constant as 93%. Then it is considered that the interest simulation constant of the applied teaching method proposed in this paper is 93%. It is concluded that the application of internal medicine nursing teaching in colleges and universities is done using more traditional teaching methods, which improves students' interest in internal medicine nursing courses by 46.7%, and promotes students' continuous learning.

4.3. Educational Diversity Simulation Experiment. Similarly, the data of 6000 students are quantified to construct different learning ability parameter values, with the unit of d. From 0 to 1.0 d, similarly, 0.1d to 0.2 d are class A students, 0.3 d to 0.4 d are class B students, 0.5 d to 0.6 d are class C students, 0.7 d to 0.8 d are class D students, and 0.9 d to 1.0 d are class E students. According to different types of students and different learning abilities, educational diversity simulation experiments are carried out in three environments: simple, moderate, and difficult. The teaching diversity index is used as the experimental measurement standard. Its teaching diversity index includes the richness of teaching resources, the ways students obtain knowledge points, and teachers' teaching methods. The richness of the three teaching resources, the ways students obtain knowledge points, and teachers' teaching methods are quantified to judge the size of teaching diversity. It is the teaching

diversity index, in %. At the same time, due to the different abilities of students, the acceptable diversity index is also different, so the comparison curve of educational diversity simulation experiment results is obtained. At the same time, the same personnel and the same quantitative mechanism are used to build the same learning ability as the abscissa. The experiment is simple, moderate, and difficult. The experimental results are recorded in the comparison curve of educational diversity simulation experiment results. The comparison curve of educational diversity simulation experiment results is shown in Figure 6. During the experiment, the same as above, the third-party analysis and recording software is used to record and analyze the experimental process and results, and the results are displayed in the comparison result curve of this experiment. The analysis function is used to eliminate the uncertainty caused by the operation of simulation laboratory personnel and simulation computer equipment. The educational diversity simulation experiment is carried out only for different educational and teaching methods and different experimental objects.

By analyzing the comparison curve of the simulation experiment results of educational diversity in Figure 6, it is concluded that the teaching application method proposed in this paper has significantly improved the educational diversity as a whole and made outstanding progress in different stages compared with the traditional teaching method. Under the simple experimental environment, students with different learning abilities received different diversity indexes and showed volatility. When the learning ability increases from 0 to 0.2 d, the diversity index increases linearly, which means that the teaching diversity of class A students increases briefly under the traditional teaching methods, and the highest teaching diversity index is 42%. When the learning ability is from 0.2 d to 0.3 d, the teaching diversity index tends to decline. When the learning ability is from 0.3 d to 0.5 d, the teaching diversity index increases rapidly, and its teaching diversity index is 42%. At the same time, it decreases rapidly and improves rapidly. Under the traditional teaching methods, the diversity of students with different learning abilities is low, up to 42%. When the experimental environment is moderate, the change of teaching diversity index curve is similar to that of simple environment, both of which have uncertainty and volatility. The highest teaching diversity index is no more than 38%, which appears when the learning ability is 0.5 d and 0.7 d, respectively, representing class C and class D students with moderate learning ability. Students with poor learning ability and high learning ability perform mediocrely in the teaching diversity index under the traditional teaching methods. When the simulation experiment environment is difficult, the teaching diversity curve of traditional teaching methods moves down as a whole, with the highest teaching diversity index of 28% and the lowest teaching diversity index of 12%. Under the three simulation experiment environments, the separation of the three curves is obvious and the overall span is large, indicating that the teaching diversity of traditional teaching methods is poor. For example, when the learning ability is 0.2 d, the educational diversity index is
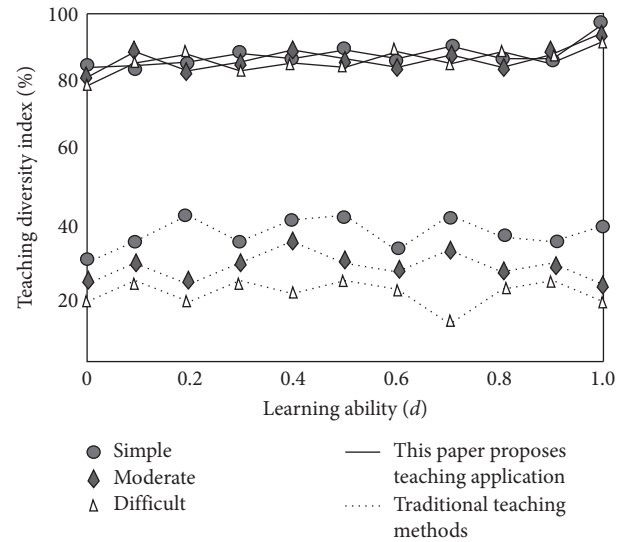
FIGURE 6: Comparison curve of simulation experiment results of educational diversity.

42% in simple simulation environment, 25% in moderate simulation environment, and 20% in difficult simulation environment. When the learning ability is 0.7 d, the educational diversity index is 42% in simple simulation environment, 27% in moderate simulation environment, and 12% in difficult simulation environment. It shows that traditional educational methods are not suitable for diversity teaching in difficult environment.

In this paper, the teaching application method is proposed. Under the condition of simple simulation experiment, the change range of teaching diversity index curve is unknown and tends to a straight line. The lowest teaching diversity index appears when the learning ability is 0.6 d, and the teaching diversity index is 85%; the highest teaching diversity index appears when the learning ability is 1.0 d, and the teaching diversity index is 95%; and the average change of teaching diversity index is 85% to 90%. In the moderate simulation environment, this paper puts forward the teaching application method, teaching diversity index, and the curve change which is basically the same as that in the simple simulation environment. The lowest teaching diversity index appears when the learning ability is 0.2 d, 0.6 d, and 0.8 d, and the teaching diversity index is 82%; the highest teaching diversity index appears when the learning ability is 1.0 d, and the teaching diversity index is 92%; and the average change of teaching diversity index is 82% to 90%. The lowest point of teaching diversity index appears in multiple positions, which shows that the teaching application method proposed in this paper changes gently in a moderate environment. In the difficult simulation environment, this paper puts forward the teaching application method and the teaching diversity index, which is basically similar to the change curve under the simple and moderate simulation experimental environment. The lowest teaching diversity index appears when the learning ability is 0.3 d, and the teaching diversity index is 81%; the highest teaching diversity index appears when the learning ability is 1.0 d, and the teaching diversity index is 90%, and the average change

of teaching diversity index is between 80% and 90%. Under different simulated experimental environments, the three curves are intertwined. The overall change of the curve is stable, which shows that the teaching application method proposed in this paper has a good embodiment in teaching diversity and can provide diverse teaching modes for students with different learning abilities, improving their learning ability and academic performance.

The weighted calculation by the third-party analysis software shows that in a simple environment, the teaching diversity index is 36%; in a moderate environment, the teaching diversity index is 29%; in a difficult environment, the teaching diversity index is 22%; and the arithmetic weighted analysis takes the average teaching diversity index as 29%; then, it is considered that the teaching diversity index of traditional teaching methods is 29%. In a simple environment, this paper proposes that the teaching diversity index of applied teaching methods is 84%. In a moderate environment, this paper proposes that the teaching diversity index of applied teaching methods is 83%. In a difficult environment, this paper proposes that the teaching diversity index of applied teaching methods is 84%, and the average teaching diversity index is 83.67%. Then it is considered that the teaching diversity index of applied teaching methods proposed in this paper is 83.67%. It is concluded that the application of internal medicine nursing teaching in colleges and universities is 54.67% higher than the traditional teaching methods.

5. Conclusion

This paper puts forward the application analysis of two-dimensional code online testing technology in nursing teaching in colleges and universities. Through the difficulties of nursing teaching in higher vocational colleges and the construction of multiplatform information-based teaching system, relying on multiplatform information-based teaching system, we can promote students' interest in internal medicine nursing. Relying on different types of evaluation of students, carry out diversified teaching and targeted assignment. At the same time, use the multiplatform information-based teaching supervision system to prevent students from skipping classes and not completing their homework on time. Establish different types of learning groups, and construct the communication mechanism with in-service internal medicine nurses, so as to improve the teaching level of internal medicine nursing in higher vocational colleges and realize the research of this paper. The experimental data show that the method designed in this paper is highly effective.

There are still improvements in the actual operation process, which are discussed and analyzed as follows:

(1) The quality of nursing teaching video in QR code is difficult to guarantee. The video recording must be carried out in the daily working environment to make students feel more situational. The operators must move smoothly and be standardized; the commentators must enunciate clearly, use standard terms, and focus on interpretation of key contents, and each video recording shall be simplified as much as possible and controlled within about 5 min. The produced video must be sent to the teaching department or nursing department of the hospital for review before use.

(2) Teachers tend to weaken their teaching dominance over students. The application of QR code online testing technology is a teaching mode dominated by students' autonomous learning and teachers. Teachers play an important role in supervising and guiding students' learning process, which can be further studied in the future.

Data Availability

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Conflicts of Interest

The authors declared that they have no conflicts of interest regarding this work.

Acknowledgments

References

[1] B. L. Reedy, P. R. Philips, and D. J. Newell, "Nurses and nursing in primary medical care in England," *British Medical Journal*, vol. 2, no. 6047, pp. 1304–1306, 2019.

[2] K. T. Unroe, E. Mary, and C. John, "The IOM report on dying in America: a call to action for nursing homes," *Journal of the American Medical Directors Association*, vol. 16, no. 2, pp. 90–92, 2015.

[3] N. Listed, "Medicare program; schedule of limits for skilled nursing facility inpatient routine service costs--HCFA. Final

notice with comment period," *Federal Register*, vol. 56, no. 62, Article ID 46177, 2019.

[4] A. A. Mwakanyamale, A. M. A. Mukaja, M. Ndomondo, A M A. Mukaja, and J P. Zenas, "Nursing practice on post-operative wound care in surgical wards at muhimbili national hospital, dar-es-salaam, Tanzania," *Open Journal of Nursing*, vol. 9, no. 8, pp. 870–890, 2019.

[5] S. Perkisas, C. Amd, M. Vandewoude, and V. Verhoeven, "Prevalence of sarcopenia and 9-year mortality in nursing home residents," *Aging Clinical and Experimental Research*, vol. 31, no. 7, pp. 951–959, 2019.

[6] N. Mcclure and C. Brame, "Asthma medications: a think-pair-share-square activity for nursing students," *Nurse Educator*, vol. 45, no. 1, p. 16, 2020.

[7] B. Burford, P. Greig, M. Kelleher et al., "Effects of a single interprofessional simulation session on medical and nursing students' attitudes toward interprofessional learning and professional identity: a questionnaire study," *BMC Medical Education*, vol. 20, no. 1, pp. 1–11, 2020.

[8] F. T. Li and S. Seetharaman, "Preventing the spread of COVID-19 to nursing homes: experience from a Singapore geriatric centre," *Journal of the American Geriatrics Society*, vol. 68, no. 5, pp. 942–950, 2020.

[9] M. Zhou, J. Hu, and Q. Wang, "Design of semi-open intelligent access control system based on QR code," *Microcontrollers & Embedded Systems*, vol. 20, no. 10, pp. 50-51+55, 2020.

[10] L. Cong, W. He, and N. A. University, "Design of transfer management system based on QR code," *Microcontrollers & Embedded Systems*, vol. 20, no. 7, pp. 43–46, 2020.

[11] K. Latha and T. Sheela, "Security scheme in cloud system using hyper elliptic curve and SIFT technique with QR code," *Wireless Personal Communications*, vol. 111, no. 2, pp. 1327–1340, 2020.

[12] J. Qian, B. Xing, B. Zhang, and H. Yang, "Optimizing QR code readability for curved agro-food packages using response surface methodology to improve mobile phone-based traceability," *Food Packaging and Shelf Life*, vol. 28, no. 1, Article ID 100638, 2021.

[13] S. Zhang, J. Liao, S. Wu, J. Zhong, and X. Xue, "A traceability public service cloud platform incorporating IDcode system and colorful QR code technology for important product," *Mathematical Problems in Engineering*, vol. 2021, Article ID 5535535, 15 pages, 2021.

[14] D.-C. Wu and Y.-M. Wu, "Covert communication via the QR code image by a data hiding technique based on module shape adjustments," *IEEE Open Journal of the Computer Society*, vol. 1, no. 4, pp. 12–34, 2020.

[15] C. M. Santostefano, E. M. White, R. A. Feifer, and V. Mor, "Accuracy of ICD-10 codes for identifying skilled nursing facility residents with lab-confirmed COVID-19," *Journal of the American Geriatrics Society*, vol. 10, no. 8, pp. 45–53, 2021.

[16] C. A. Hossler, "Competency-based nursing: reducing cost while maintaining or improving quality," *Journal of Competency-Based Education*, vol. 6, no. 1, pp. 1–15, 2021.

[17] V. M. Boscart, S. Sidani, J. Ploeg et al., "Neighbourhood Team Development to promote resident centred approaches in nursing homes: a protocol for a multi component intervention," *BMC Health Services Research*, vol. 19, no. 1, pp. 1–7, 2019.

[18] M. Musabwasoni, M. Kerr, Y. Babenko-Mould, M. Nzayirambaho, and A. Ngabonzima, "Assessing the impact of mentorship on nurses' and midwives' knowledge and self-efficacy in managing postpartum hemorrhage," *International Journal of Nursing Education Scholarship*, vol. 17, no. 1, pp. 98–106, 2020.

[19] S. Zhang, Y. Bai, and L. Liu, "Value evaluation of application of mu class teaching in clinical nursing teaching," *The Science Education Article Collects*, vol. 481, no. 1, pp. 100–102, 2020.

[20] R. . Zou, "Effect analysis of multimedia on improving interest of students in obstetrics and Gynecology nursing teaching," *Smart Healthcare*, vol. 6, no. 3, pp. 27-28, 2020.

[21] Y. Zhang, D. Zou, Y. Zhang, and X. Wang, "Accurate application of "Hand as Foot" in nursing teaching," *Asian Journal of Surgery*, vol. 44, no. 9, pp. 1232-1233, 2021.

[22] D. R. Hanna, "Proposing standards for teaching authentic nursing knowledge," *Advances in Nursing Science*, vol. 43, no. 1, pp. 42–49, 2020.

[23] L. Dumas and F. d. Montigny, "Quality perinatal nursing education through coteaching," *The Journal of Perinatal Education*, vol. 8, no. 4, pp. 27–35, 2020.

[24] H. Ma, "For the love of nursing. and teaching!" *Journal of Christian Nursing*, vol. 37, no. 4, pp. 251–260, 2020.

[25] L. J. Labrague, D. M. McEnroe-Petitte, M. S. D'Souza, K. S. Hammad, and J. N. A. Hayudini, "Nursing faculty teaching characteristics as perceived by nursing students: an integrative review," *Scandinavian Journal of Caring Sciences*, vol. 34, no. 1, pp. 23–33, 2020.

[26] F. F. Lv, H. He, and H. M. Bi, "Application of "Internet+" question-based teaching mode in theteaching of characteristic nursing of Dai medicine," *TMR integrated care*, vol. 004, no. 2, pp. 45–49, 2020.

[27] X. Gao, "Application of inquiry learning in nursing teaching of cardiac surgery operating room," *Health Horizon*, no. 11, pp. 45–49, 2019.

[28] K. A. Kalb and S. O'Conner-Von, "Holistic nursing education: teaching in a holistic way," *Nursing Education Perspectives*, vol. 40, no. 3, pp. 162–164, 2019.

[29] K. F. Phillips, L. Mathew, N. Aktan, and P. Sandanapitchai, "The effectiveness of shared clinical teaching in nursing," *International Journal of Nursing Science*, vol. 10, no. 4, pp. 211–215, 2019.

[30] D. Lillekroken, "Teaching basic nursing care: nurse preceptors' perceptions about changing the teaching context from the clinical setting to a school simulation lab," *Nephron Clinical Practice*, vol. 16, no. 1, pp. 60–135, 2019.

[31] K. Wang and F. Qian, "Clinical application analysis of nursing teaching ward round in ICU teaching," *China Continuing Medical Education*, vol. 12, no. 9, pp. 54-55, 2020.

[32] J. Pellet, M. Weiss, J. Rapin, C. Jaques, and C. Mabire, "Nursing discharge teaching for hospitalized older people: a rapid realist review," *Journal of Advanced Nursing*, vol. 76, no. 11, pp. 2885–2896, 2020.

[33] G. Basit and F. Korkmaz, "The effect of web-based nursing process teaching on senior nursing students' care planning skills#," *International Journal of Nursing Knowledge*, vol. 32, no. 2, pp. 1–16, 2020.

[34] D. Mcenroe-Petitte and C. Farris, "Using gaming as an active teaching strategy in nursing education," *Teaching and Learning in Nursing*, vol. 15, no. 1, pp. 61–65, 2020.

[35] Y. Zolkefli, "The use of debate in teaching nursing ethics," *Teaching and Learning in Nursing*, vol. 4, no. 10, pp. 365–367, 2021.

[36] E. J. B. Thrower, R. Fay, L. Cole et al., "A systematic process for evaluating teaching methods in nursing education," *Nurse Educator*, vol. 45, no. 5, pp. 257–260, 2020.

Research Article

Research on Parallel Support Vector Machine Based on Spark Big Data Platform

Yao Huimin ⓘ

Management School, University of Bath, Bath, UK

Correspondence should be addressed to Yao Huimin; huimin.yao@bath.edu

With the development of cloud computing and distributed cluster technology, the concept of big data has been expanded and extended in terms of capacity and value, and machine learning technology has also received unprecedented attention in recent years. Traditional machine learning algorithms cannot solve the problem of effective parallelization, so a parallelization support vector machine based on Spark big data platform is proposed. Firstly, the big data platform is designed with Lambda architecture, which is divided into three layers: Batch Layer, Serving Layer, and Speed Layer. Secondly, in order to improve the training efficiency of support vector machines on large-scale data, when merging two support vector machines, the "special points" other than support vectors are considered, that is, the points where the nonsupport vectors in one subset violate the training results of the other subset, and a cross-validation merging algorithm is proposed. Then, a parallelized support vector machine based on cross-validation is proposed, and the parallelization process of the support vector machine is realized on the Spark platform. Finally, experiments on different datasets verify the effectiveness and stability of the proposed method. Experimental results show that the proposed parallelized support vector machine has outstanding performance in speed-up ratio, training time, and prediction accuracy.

1. Introduction

As the mainstream part of today's media industry, images and videos are rich in information and easy to understand, which makes them an indispensable part of life. Computer vision analysis is also the key development direction of the Internet communication industry at present. For example, character recognition has great application value in many scenes, such as vehicle license plate detection, image-text conversion, image content translation, and image search. However, because the precision of text recognition technology is not ideal, its application scenarios are relatively simple, such as content search in images [1–6].

2. Literature Review

With the increasing amount of data, how to efficiently store, organize, and analyze these massive data has become a hot issue for both academia and industry. Distributed computing, which can distribute computing tasks on a large number of computers connected together through the network and cooperate with each other to complete computing tasks, is expected to become an effective means to solve this problem [1–7].

Faced with the pressure brought by the storage and calculation of big data, Google proposed and designed its own distributed file storage system GFS (Google FileSystem) [8, 9] in 2003, which used a large number of cheap commercial computers as distributed clusters to store and manage large-scale data. Later, it published MapReduce, a distributed parallel computing technology, and Bigtable, a distributed storage framework for structured data, which were applied to fast parallel processing of large-scale data. Then Apache implemented the open-source distributed file system HDFS (Hadoop Distributed File System) and the distributed computing engine Apache Hadoop MapReduce as the storage and computing solutions of big data, respectively. With the contribution of the open-source community, many projects around it are constantly emerging, such as Hive, a data warehouse tool originated from Facebook, HBase, a column-based distributed open-source database, and of course, big data projects such as Pig, a data

stream-based processing framework, and ZooKeeper, a distributed coordination framework [10–13]. In 2009, MateiZaharia founded Spark's big data processing and computing framework based on memory computing at the University of California, Berkeley. Compared with the previous MapReduce framework, the intermediate results of computation are kept in memory and the optimization of execution plan based on DAG (Directed Acyclic Graph) makes Spark process data in parallel faster than MapReduce. It also provides support for Hadoop components and Hive and supports SQL-like large-scale structured data calculation and query [14, 15].

The real significance of big data lies not in having a huge amount of data but in mining valuable information for people and how to use it. Therefore, data mining and artificial intelligence based on big data are the ultimate destination of big data. Machine learning provides a variety of learning algorithms and models for data mining and artificial intelligence and has been applied in the fields of urban management, finance, entertainment, security, medical care, and so on, which has produced great influence and important value on people's lives. As an important supervised learning algorithm, support vector machine (SVM) has a complete mathematical theory and has been widely used in the fields of text recognition and speech recognition. However, because its ultimate learning problem is to solve a convex quadratic programming problem, the corresponding time complexity and space complexity are relatively high. With the continuous growth of training data, the space occupation and training time of stand-alone training will increase dramatically, which is not suitable for model training on big data. This also directly leads to the fact that SVM is usually used to solve small sample problems, which limits its application and analysis on big data.

How to use the distributed parallel technology of big data to improve the training efficiency of SVM on big data has become a very meaningful research direction. Spark is a mature distributed parallel computing framework for big data at present, and it is of practical significance to realize parallel SVM based on Spark to solve the application in large-scale data. Therefore, the purpose of this study is to accelerate the training process of SVM on large-scale data while maintaining certain model accuracy.

When faced with a large training set, the whole training process of SVM needs a lot of memory. Jindal et al. [16] proposed that, by eliminating nonsupport vectors step by step, the storage space required in the training process was reduced, and each time a part of samples were selected as the training set, the support vectors obtained by training were combined with the samples that most seriously violated the training results among the remaining samples as a new training set. Aslahi-Shahri et al. [17] put forward the idea based on decomposition, which decomposes a relatively large quadratic programming problem into several relatively small subproblems, solves only one subproblem at a time, and iterates until the global optimal solution is obtained, thus obtaining the training model of SVM. Chen et al. [18] proposed a fast implementation algorithm based on this strategy-a series of minimum optimization

algorithms, which only considered the optimization problem of two variables at a time until all the variables met the requirements. This algorithm has also become a widely used SVM tool LIBSVM. Xu et al. [19] put forward an incremental algorithm similar to the block algorithm, which takes the training scale tolerated by the single training algorithm as an increment and combines it with the support vector of the previous sample for training until all the training samples are processed.

3. Literature Review

Although the above algorithms have different effects on speeding up SVM training and reducing memory usage, these strategies still have their limitations when the scale of training data reaches a certain level. Therefore, how to solve SVM in parallel has become a hot research direction. Cao et al. [20] proposed a parallel SVM algorithm based on the distributed memory system. Das et al. [21] proposed Cascade SVM based on cascade and feedback architecture. In the initial stage, the support vector machine randomly divides the whole training set into even subsets. Experiments show that since each layer of SVM training can be carried out in parallel and a large number of nonsupport vectors can be filtered in the initial stage, it can effectively reduce the training time on large-scale data. On the basis of group training, Singh and Jaiswal [22] proposed a parallel SVM algorithm based on Hadoop. In the prediction stage, the distance between the point to be predicted and the center of each subset is calculated, and the nearest subset model is used to predict it. Although the training time has been shortened to the extreme, the generalization ability of the algorithm remains to be discussed.

Aiming at the problem that the prediction accuracy of traditional Cascade SVM is lower than that of single machine training, the merging algorithm based on support vector is improved, and the process of realizing parallelized support vector machine on Spark platform [23] is studied. Firstly, based on the HDFS distributed file system and Spark distributed computing engine, a three-tier architecture of the machine learning platform is constructed. Then, aiming at the problem that only a single support vector is considered when merging two support vector machines, when merging two support vector machines, "special points" and support vectors are taken as the input of the next layer, and a parallelized support vector machine based on cross-validation is proposed. Finally, by deploying HDFS and Spark clusters, the traditional cascade support vector machine and the proposed support vector machine are compared in real environment in terms of acceleration ratio, training time, and prediction accuracy.

4. Overall Architecture of Machine Learning Platform

4.1. Spark Architecture Ideas. Spark operation mode abstracts the memory through the technology of Resilient Distributed Datasets (RDD) [24], realizes the data exchange

of the whole memory, and greatly improves the speed of frequent iteration and other operations. The schematic diagram of the Spark operation process is shown in Figure 1.

The most important reason why the Spark platform can realize iteration in memory is that RDD is adopted, which is also the greatest advantage of the Spark platform. In this paper, sample points are stored in RDD of computing nodes in the Spark platform, which greatly reduces the time of reading and writing interaction with hard disk. Spark, as a common engine for large-scale data processing, uses master-slave node management mode to complete data processing tasks together, but in terms of functional structure, master-slave nodes have the same computing capability, and its main structure is shown in Figure 2.

In addition to this master-slave node cooperative working mode, the Spark platform has another advantage that most data operations are completed in RDD of platform node memory, which greatly improves data access efficiency.

4.2. Overall Platform Architecture. Modeling in big data scenarios usually has two basic requirements: one is modeling with real-time data flow, and the other is modeling with mass data analysis. At present, these two requirements have their own solutions. However, in common applications, these two requirements are generated at the same time. For example, a website of a recommendation system requires not only mining large-capacity historical data but also real-time modeling for quick feedback of users' clicks and purchases. In order to meet this challenge, many projects adopt the form of hybrid architecture, which is called Lambda architecture [25]. Lambda architecture provides a series of clear architectural design principles for mixed and diverse data scenes (which need batch processing and real-time or streaming data processing), which are mainly divided into three layers, Batch Layer, Serving Layer, and Speed Layer, as shown in Figure 3.

In the design of distributed Spark recommendation scheme, HDFS distributed file system based on Parquet data storage is used for data storage [26], and Spark SQL is used for database table query. The recommendation scheme analyzes the user's preferences by studying the user's dynamic and static data (including friend information, historical search information, interests and hobbies information, and registration information) and completes personalized recommendations. The core module of the overall platform architecture is the recommendation engine module, which adopts the recommendation scheme based on the parallel SVM algorithm. The overall architecture of the Spark-based platform is shown in Figure 4.

For example, for the recommendation task of e-commerce websites, the system needs to analyze the historical preferences of current users on the one hand and generate real-time recommendations based on the feedback of browsing and clicking behaviors of current users on the other hand. First, at a fixed time, the batch processing layer will analyze all the historical data collected at present and obtain the user and project models of the recommendation system by matrix decomposition of the large-scale user

feedback matrix of the current historical records. The decomposition target is calculated as follows [27]:

$$R = U \times V^T, \tag{1}$$

where U is the user interest matrix and V is the project theme matrix.

The decomposed user interest matrix and project theme matrix are stored in the database and then inform the service layer to update the user interest index and theme feature index. When users browse the website and generate feedback such as browsing products, clicking links, and searching for products, the system will quickly collect these data and store them in the batch processing layer and the speed layer, respectively. At this time, the batch processing layer is still in the data accumulation stage, while the speed layer can quickly update the decomposition model online in memory, give a new list of product recommendations, and return to the website recommendation page.

5. Parallel Support Vector Machine Based on Spark

The traditional Cascade SVM based on decomposition provides a new idea for the parallel solution of support vector machine, which improves the training efficiency on large-scale datasets and achieves certain model accuracy. However, compared with stand-alone training, there is still a certain loss. Without changing the overall architecture of Cascade SVM, this paper studies the impact of merging algorithm on the accuracy of the final model and proposes a parallelized support vector machine model based on cross-validation.

5.1. Merge Algorithm of Cross-Validation. The basic idea of the traditional Cascade SVM is to decompose the global support vector machine solving process into several sub-problems of support vector machine, then merge the models on these subproblems in pairs by using a tree-like form, and finally get a global solution. The merging algorithm of Cascade SVM only considers the support vectors on two subsets, but this local support vector cannot completely contain two subsets as global support vectors on a training set, so in every small merging process, the loss will be gradually enlarged with the increase of iteration layers, resulting in the ultimate loss of model accuracy, which is reflected in the number of support vectors of the final model and the prediction accuracy of the test set.

Cross-validation [28] is often used to evaluate the generalization ability and reliability of a model or algorithm. In the related fields of machine learning, its basic idea is to group the original training data, one of which is used as a training set and the other as a verification set or a test set, train the corresponding model by using the training set, and verify the model by using the test set as an index to evaluate the learning algorithm. In the learning process of the support vector machine, the prediction accuracy obtained by cross-validation is usually used to measure the selection of parameters in the learning process. Here, we do not care about
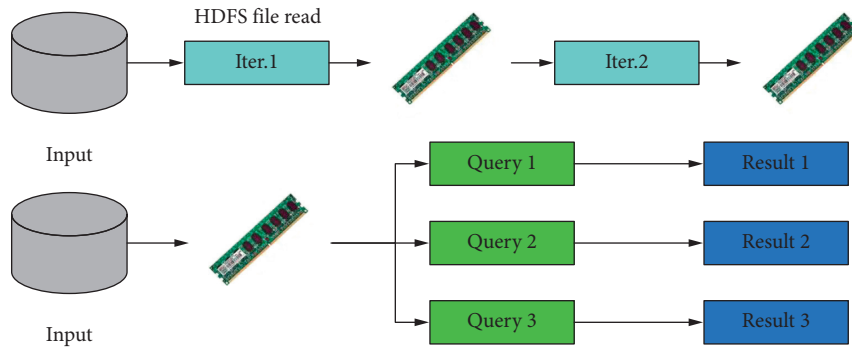
FIGURE 1: Operation process of Spark.
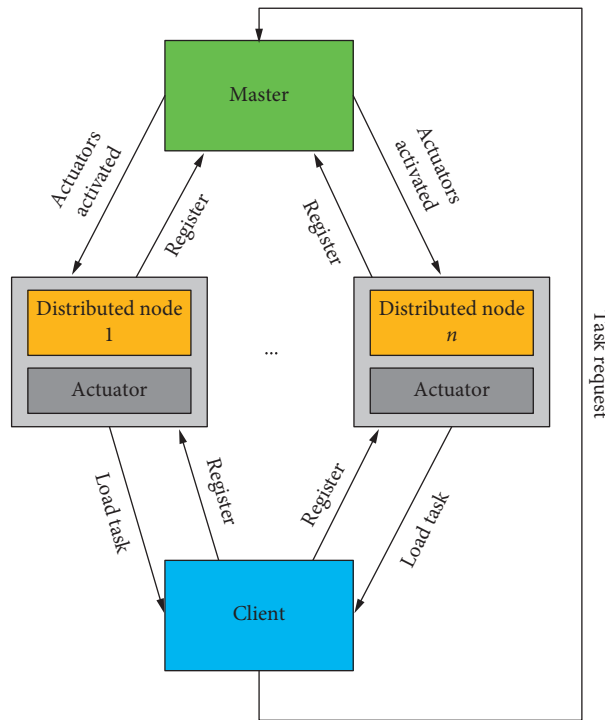
FIGURE 2: Spark structure.
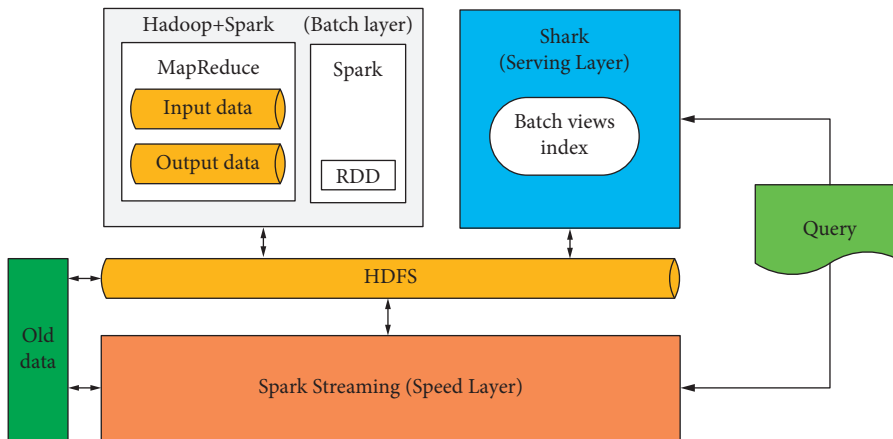
FIGURE 3: Lambda architecture.
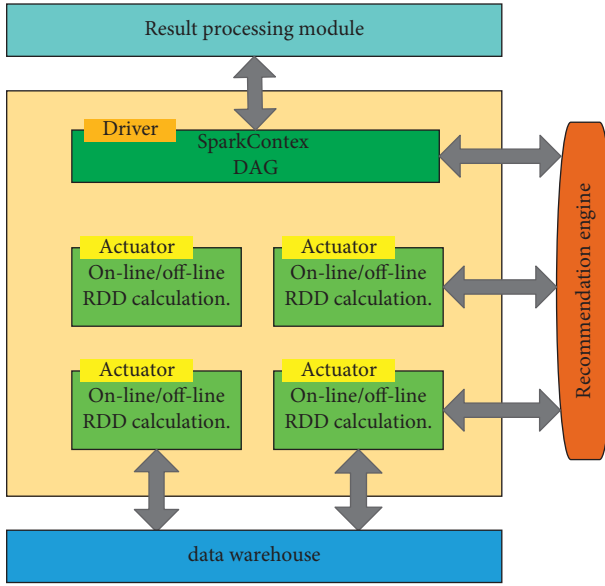
FIGURE 4: Overall platform architecture design based on Spark.

the prediction accuracy of the model on the training set to the test set but care about those "special points" in the test set that violate the training results on the training set.

Considering the merging of two support vector machines, TD_1 and TD_2, respectively, represent two datasets, SVM_1 and SVM_2, respectively represent the support vector machines trained on the two datasets, and $(w^*_{TD_1}, b^*_{TD_1})$ and $(w^*_{TD_2}, b^*_{TD_2})$ represent their respective classification hyperplanes, respectively. From the point of view of SVM_2, although SVM_2 will not misclassify x_1 and x_2, these points are located between the boundary planes of SVM_2, which may become a new subset, which means that adding these points will produce a new support vector machine model. From the point of view of geometric space, the distance between these points and the classification hyperplane of SVM_2 is small. Because of the relationship between function interval and geometric interval, adjusting parameters of the hyperplane in equal proportion will not affect the size of the geometric interval. In the process of establishing the interval maximization problem, the functional interval of the interval boundary hyperplane with respect to the classification hyperplane is defined as 1. Therefore, the points in TD_1 located between the boundary of TD_2 interval satisfy formula (2) and formula (3).

$$0 \leq \frac{y\left(w^*_{TD_2} \cdot x + b^*_{TD_2}\right)}{\left\| w^*_{TD_2}\right\|} \leq \frac{1}{\left\| w^*_{TD_2}\right\|}, \ (x, y) \in TD_1. \tag{2}$$

$$0 \leq y\left(w^*_{TD_2} \cdot x + b^*_{TD_2}\right) \leq 1, \ (x, y) \in TD_1. \tag{3}$$

Therefore, the merging algorithm of cross-validation is to take each training subset as a training set and a test set, respectively. When merging two support vector machines, we should not only consider the support vectors on the two subsets but also consider these "special points." In addition

to considering each support vector machine, the new merging algorithm also needs to consider the KKT conditions (Karush–Kuhn–Tucker conditions) of nonsupport vectors in a subset. Therefore, the merged support vector machine is a better model on two subtraining sets, and each merging is a local optimum, and this local optimum will also get a global optimum in layer-by-layer iteration.

5.2. Implementation Flow Based on Spark. The process of the proposed parallel SVM is similar to that of the traditional Cascade SVM, except that the merging algorithm based on cross-validation is adopted when merging the two SVMs. The specific implementation process is shown in Figure 5.

Initially, the dataset on HDFS is randomly divided with restrictions to ensure that the ratio of positive and negative samples in each divided subset is equal so as to avoid the reduction of the global support vector in the first layer caused by extreme cases. In the same way, all the training subsets are encapsulated in an RDD, and each training subset corresponds to a partition by setting partitions. Then, the data corresponding to each partition is trained in parallel by foreachPartition operation. After the training process of all subsets is completed, the support vector (SV) and nonsupport vector (NoSV) obtained by each training are copied to HDFS to be used as the data input of the next layer, or the two support vector machines are prepared for cross-verifying and merging data.

When the two support vector machines are merged, not only is the support vector merged as the input of the next layer, but also the cross-validation is completed through the parallel prediction process. The points where the nonsupport vector in each subset violates the model in another subset are screened out, and these points are combined with the support vector as the input of the next layer, corresponding to the Across Validation and Merge stage in Figure 5. Similarly, all training sets in the next layer can be trained in parallel by setting partitions. In this way, after many iterations, the cross-validation parallelized SVM trained model is obtained.

The prediction process in the whole training process is not to predict the labels of the test set samples but to filter out the "special points" that the nonsupport vectors in one training set violate the model of another training set; that is, the points that satisfy formula (2) or formula (3) need to be filtered out. Compared with the training process, the prediction process of support vector machine is relatively simple, so formula (4) needs to be obtained, and the form of kernel function is formula (5).

$$F(x) = w^* x + b^*. \tag{4}$$

$$F(x) = \sum_{i=1}^{N_S} \alpha_i y_i K\left(x_i, x\right) + b^*, \tag{5}$$

where x denotes a point to be predicted, (w^*, b^*) denotes an SVM model on a certain training set, N_S denotes a support vector on the model, $K(x_i, x)$ denotes the kernel function
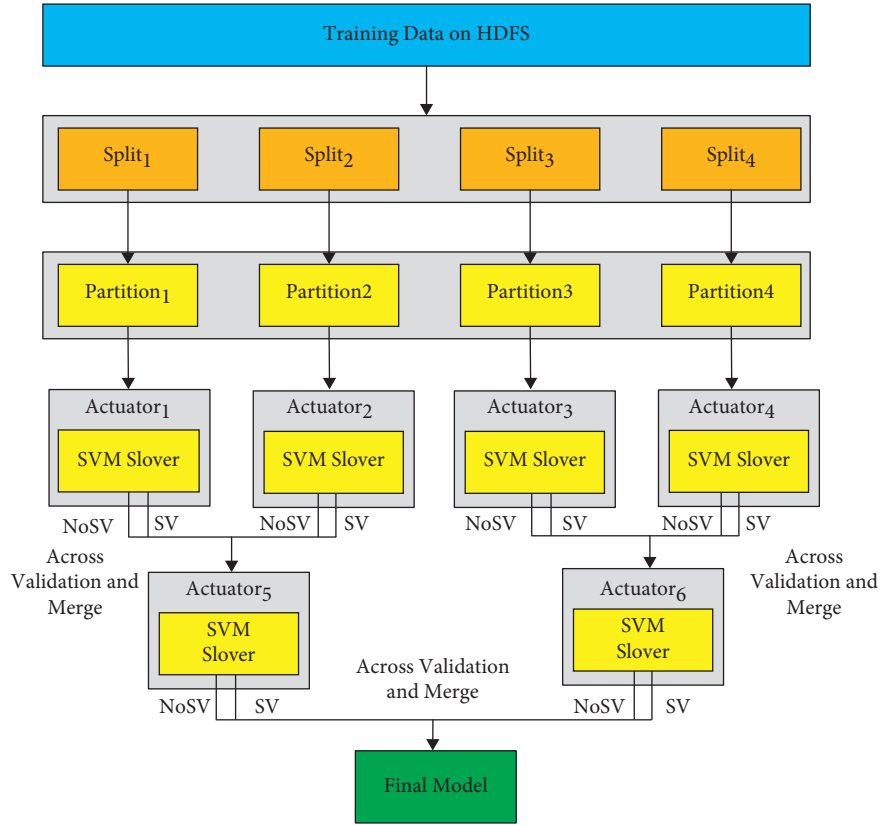
Figure 5: Specific implementation flow based on Spark.

value of the point to be predicted and the support vector, and α_i denotes the Lagrange coefficient corresponding to the support vector.

6. Experimental Results and Analysis

6.1. Experimental Environment and Dataset. There are 6 machines in the cluster environment, and the hardware configuration parameters are shown in Table 1. Spark version 0.9.0 is installed on all machines, the Hadoop version is 1.0.1, and the JDK environment is OpenJDK 1.7.0 64-bit version. Hadoop and Spark environment configuration parameters are shown in Table 2.

In the configuration of Hadoop and Spark, the experiment mainly focuses on memory usage. In Hadoop, the maximum number of map tasks that can run simultaneously is 16, the maximum number of reduced tasks is 2, and each task can occupy up to 4 GB of memory. In Spark, the computing memory of each node is 20G, which is mainly the space occupied by computing data and RDD linear dependency storage. The experimental dataset is the Movie-Lens dataset provided by the School of Computer Science and Engineering, University of Minnesota (https://grouplens.org/datasets/movielens/), which contains information of 6,000 users and 4,000 movies and is the most commonly used test dataset for the recommendation system.

6.2. Speed-Up Ratio. In the aspect of parallelization programming, an important performance index is the speed-up ratio, which is described by the following formula:

$$S = \frac{T_a}{T_s}, \tag{6}$$

where T_a is the running time required by the serial program and T_s is the running time after parallelization.

It is difficult to obtain a perfect linear acceleration ratio for parallelized programs. However, with the increase of the problem scale, the proportion of nonparallelizable parts in the program will gradually decrease. Therefore, when the total parallel execution time of the program is assumed to be constant, the parallelized program can still obtain a good linear acceleration ratio. The speed-up ratio of Hadoop and Spark with different numbers of nodes is shown in Figure 6.

It can be seen from Figure 6 that the Spark-based parallelized support vector machine has good horizontal scalability. With the increase of the number of computing nodes, more computing resources are put into the task, and the running time of the computing task can show an obvious downward trend. The result of the speed-up ratio reflects the efficiency improvement of parallelization. Therefore, Spark-based parallelization SVM can greatly improve the efficiency of the program running.

TABLE 1: Hardware configuration.

Number	Node name	CPU	Internal storage capacity (GB)	Hard disc capacity
1	Master	i7-3820 8-Core	64	1 TB
2	Slave01	E31230 8-Core	32	500 GB
3	Slave02	E31230 8-Core	32	500 GB
4	Slave03	E31230 8-Core	32	500 GB
5	Slave04	i5-2300 4-Core	32	500 GB
6	Slave05	i5-2300 4-Core	32	500 GB

TABLE 2: Configuration of test cluster.

Number	Node name	Hadoop's configuration	Spark's configuration
1	Master		
2	Slave01		
3	Slave02	dfs.replication = 3; map.tasks.maximum = 16; reduce.tasks.maximum = 2; child.java.	SPARK_MEM = 20g
4	Slave03	opts = -Xmx4096 M	
5	Slave04		
6	Slave05		
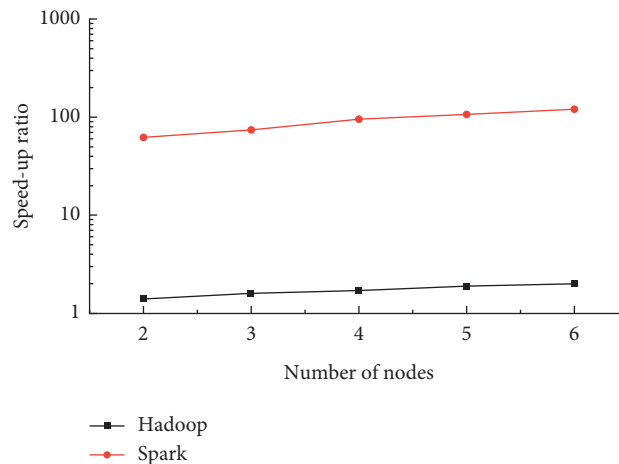
FIGURE 6: Experimental comparison of speed-up ratio.

6.3. Prediction Accuracy and Training Time. In the experiment, the training time and prediction accuracy of LIBSVM, Cascade SVM, and our proposed parallel SVM are compared. The MovieLens dataset is randomly divided into 7 subdatasets. The prediction accuracy ratio of the three algorithms on each dataset is shown in Figure 7.

For 7 subdatasets, the prediction accuracy of cross-validation parallelized SVM is higher than that of Cascade SVM, and it is very close to the prediction accuracy of single machine training. The prediction accuracy of cross-validation parallelized SVM is almost the same as that of single machine training, and it has been improved compared with layered support vector machine. The biggest reason is that the merging algorithm based on cross-validation adds more training data, and these data are very likely to be global

support vectors. Comparison of training time on 7 datasets is shown in Table 3.

It can be seen from Table 3 that when the scale of the training set is small, the training time of Cascade SVM and proposed SVM is longer than that of stand-alone SVM because the data scale is small and stand-alone SVM can train to get the final model in a very short time, while Cascade SVM and proposed SVM take longer training time than stand-alone training because each layer needs to communicate with the next layer. However, with the increase of data scale, the advantages of hierarchical divide-and-conquer are manifested. The training time of Cascade SVM and the proposed SVM is shorter than that of a single machine, and compared with Cascade SVM, the training time of the proposed SVM will be reduced.
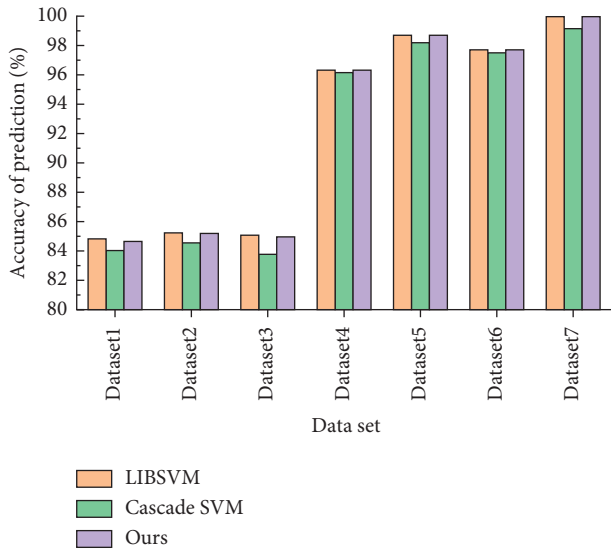
Figure 7: Prediction accuracy.

Table 3: Training time.

Number	Dataset	LIBSVM (s)	Cascade SVM (s)	Ours (s)
1	Dataset1	31.30	46.16	44.57
2	Dataset2	42.03	49.41	47.03
3	Dataset3	94.23	63.47	59.06
4	Dataset4	25.10	18.21	16.42
5	Dataset5	1042.73	541.43	446.83
6	Dataset6	143.44	30.25	24.73
7	Dataset7	30144.50	2503.07	1601.30

7. Conclusions

This paper proposes cross-validation parallelized SVM based on the Spark big data platform. Firstly, based on the HDFS distributed file system and Spark distributed computing engine, a three-tier architecture of machine learning platform is constructed. Then, the merging algorithm based on the support vector is improved, "special points" and support vector are taken as the input of the next layer, and a parallel SVM based on cross-validation is proposed. Experimental results show that the prediction accuracy of the proposed parallelized SVM is higher than that of Cascade SVM, it is very close to the prediction accuracy of single machine training, and the training time is further reduced. Further research will be carried out to solve the problem of excessive differences in subdatasets caused by randomly dividing datasets.

Data Availability

The experimental data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] L. Xu, C. Jiang, J. Wang, J. Yuan, and Y. Ren, "Information security in big data: privacy and data mining," *IEEE Access*, vol. 2, no. 2, pp. 1149–1176, 2017.

[2] E. Baccarelli, N. Cordeschi, A. Mei, M. Panella, M. Shojafar, and J. Stefa, "Energy-efficient dynamic traffic offloading and reconfiguration of networked data centers for big data stream mobile computing: review, challenges, and a case study," *Computers & Chemical Engineering*, vol. 91, no. 2, pp. 182–194, 2016.

[3] A. Gani, A. Siddiqa, S. Shamshirband, and F. Hanum, "A survey on indexing techniques for big data: taxonomy and performance evaluation," *Knowledge and Information Systems*, vol. 46, no. 2, pp. 241–284, 2016.

[4] S. Wang, J. Wan, D. Zhang, D. Li, and C. Zhang, "Towards smart factory for industry 4.0: a self-organized multi-agent system with big data based feedback and coordination," *Computer Networks*, vol. 101, no. 6, pp. 158–168, 2016.

[5] X. Wang, "Application of network protocol improvement and image content search in mathematical calculus 3D modeling video analysis," *AEJ - Alexandria Engineering Journal*, vol. 60, no. 5, pp. 4473–4482, 2021.

[6] S. Fong, R. Wong, and A. Vasilakos, "Accelerated PSO swarm search feature selection for data stream mining big data," *IEEE Transactions on Services Computing*, vol. 11, pp. 33–45, 2016.

[7] A. Barbu, Y. She, L. Ding, and G. Gramajo, "Feature selection with annealing for computer vision and big data learning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 2, pp. 272–286, 2017.

[8] M. Giannakis and M. Louis, "A multi-agent based system with big data processing for enhanced supply chain agility," *Journal of Enterprise Information Management*, vol. 29, no. 5, pp. 108–121, 2016.

[9] M. Agathe, B. Paulo, and S. George, "Learning analytics: from big data to meaningful data," *Journal of Learning Analytics*, vol. 2, no. 3, pp. 4–8, 2016.

[10] H. Xing, A. Qian, R. C. Qiu, W. Huang, L. Piao, and H. Liu, "A big data architecture design for smart grids based on random matrix theory," *IEEE Transactions on Smart Grid*, vol. 8, no. 2, pp. 674–686, 2017.

[11] R. F. Babiceanu and R. Seker, "Big Data and virtualization for manufacturing cyber-physical systems: a survey of the current status and future outlook," *Computers in Industry*, vol. 81, pp. 128–137, 2016.

[12] L. Zhou, S. Pan, J. Wang, and A. V. Vasilakos, "Machine learning on big data: opportunities and challenges," *Neurocomputing*, vol. 237, no. 5, pp. 350–361, 2017.

[13] R. Lokers, R. Knapen, S. Janssen, Y. van Randen, and J. Jansen, "Analysis of Big Data technologies for use in agro-environmental science," *Environmental Modelling & Software*, vol. 84, pp. 494–504, 2016.

[14] M. Bilal, L. O. Oyedele, J. Qadir et al., "Big Data in the construction industry: a review of present status, opportunities, and future trends," *Advanced Engineering Informatics*, vol. 30, no. 3, pp. 500–521, 2016.

[15] H. Wang, Z. Xu, H. Fujita, and S. Liu, "Towards felicitous decision making: an overview on challenges and trends of Big Data," *Information Sciences*, vol. 367-368, pp. 747–765, 2016.

[16] A. Jindal, A. Dua, K. Kaur, M. Singh, N. Kumar, and S. Mishra, "Decision tree and SVM-based data analytics for theft detection in smart grid," *IEEE Transactions on Industrial Informatics*, vol. 12, no. 3, pp. 1005–1016, 2016.

[17] B. M. Aslahi-Shahri, R. Rahmani, M. Chizari et al., "A hybrid method consisting of GA and SVM for intrusion detection

system," *Neural Computing & Applications*, vol. 27, no. 6, pp. 1–8, 2016.

[18] W. Chen, H. R. Pourghasemi, A. Kornejady, and N. Zhang, "Landslide spatial modeling: i," *Geoderma*, vol. 305, pp. 314–327, 2017.

[19] P. Xu, F. Davoine, H. Zha, and T. Denœux, "Evidential calibration of binary SVM classifiers," *International Journal of Approximate Reasoning*, vol. 72, no. may, pp. 55–70, 2016.

[20] L. J. Cao, S. S. Keerthi, C.-J. Ong et al., "Parallel sequential minimal optimization for the training of support vector machines," *IEEE Transactions on Neural Networks*, vol. 17, no. 4, pp. 1039–1049, 2006.

[21] P. Das and I. Banerjee, "An hybrid detection system of control chart patterns using cascaded SVM and neural network-based detector," *Neural Computing & Applications*, vol. 20, no. 2, pp. 287–296, 2011.

[22] S. P. Singh and U. C. Jaiswal, "Classification of audio signals using SVM-WOA in Hadoop map-reduce framework[J]," *SN Applied Sciences*, vol. 2, no. 12, pp. 1–22, 2020.

[23] X. Meng, J. Bradley, B. Yavuz et al., "MLlib: machine learning in Apache Spark," *Journal of Machine Learning Research*, vol. 17, no. 1, pp. 1235–1241, 2015.

[24] J. L. Reyes-Ortiz, L. Oneto, and D. Anguita, "Big data analytics in the cloud: Spark on Hadoop vs MPI/OpenMP on b," *Procedia Computer Science*, vol. 53, no. 1, pp. 121–130, 2015.

[25] M. Zaharia, R. S. Xin, P. Wendell et al., "Apache Spark," *Communications of the ACM*, vol. 59, no. 11, pp. 56–65, 2016.

[26] S. Gopalani, R. Arora, and S. Gopalani, "Comparing Apache Spark and map reduce with performance analysis using K-means," *International Journal of Computer Application*, vol. 113, no. 1, pp. 8–11, 2015.

[27] R. Shyam, S. Bharathi Ganesh, S. Kumar, P. Poornachandran, and K. P. Soman, "Apache spark a big data analytics platform for smart grid," *Procedia Technology*, vol. 21, pp. 171–178, 2015.

[28] A. P. Kyprioti, J. Zhang, and A. A. Taflanidis, "Adaptive design of experiments for global Kriging metamodeling through cross-validation information[J]," *Structural and Multidisciplinary Optimization*, vol. 62, no. 3, pp. 11–19, 2020.

Research Article

Research on the Application of Generative Adversarial Networks in the Generation of Stock Market Forecast Trend Images

Daiyou Xiao (iD)

School of Finance, Central University of Finance and Economics, Beijing 100081, China

Correspondence should be addressed to Daiyou Xiao; 2019110026@email.cufe.edu.cn

Investors make capital investment by buying stocks and expect to get a certain income from the stock market. When buying stocks, they need to draw up investment plans based on various information such as stock market historical transaction data and related news data of listed companies and collect and analyze these data. The data are relatively cumbersome and require a lot of time and effort. If you only rely on subjective analysis, the reference factors are often not comprehensive enough. At the same time, Internet social media, such as the speech in stock forums, also affect the judgment and behavior of investors, and investor sentiment will have a positive or negative effect on the stock market. This has an impact on the trend of stock prices. Therefore, this article proposes a stock market prediction model that uses data preprocessing technology based on past stock market transaction data to establish a stock market prediction model, and secondly, an image description generation model based on a generative confrontation network is designed. The model includes a generator and a discriminator. A time-varying preattention mechanism is proposed in the generator. This mechanism allows each image feature to pay attention to the image features of other stock markets to predict stock market trends so that the decoder can better understand the relational information in the image. The discriminator is based on the recurrent neural network and considers the degree of matching between the input sentence and the 4 reference sentences and the image features. Experiments show that the accuracy of the model is higher than that of the stock pretrend forecast model based on historical data, which proves the effectiveness of the data used in this paper in the stock price trend forecast.

1. Introduction

Generative Adversarial Network (GAN) is a new neural network architecture born in recent years [1]. A traditional neural network is usually composed of an input layer, several hidden layers, and an output layer. In the context of the generation task, the output layer directly outputs the generated results. For the image description generation task of stock market prediction trend, the goal of generation is a highly flexible language description. Compared with the maximum likelihood estimation, the structure of the generation confrontation network can provide more flexible and diverse results [2]. In addition to the generation confrontation, the network can overcome the exposure bias problem caused by the maximum likelihood estimation in the sequence generation task. Therefore, the stock market trend prediction image description model based on the generation of the confrontation network has a good research value.

2. Related Work

Most of the current research papers on text-generated images are based on Generative Adversarial Networks. In 2016, Reed [3, 4] and others first proposed the problem of generating images based on GAN text descriptions. The article proposed to extract effective semantic information from text descriptions, and the method of letting the computer recognize, by generating an adversarial neural network, produces an image that is more consistent with the text content. Subsequently, Reed [5] and others proposed a Generative Adversarial What-Where Network (GAWWN). The article proposed a Generative Adversarial What-Where Network

model that takes text information, position, and posture as conditions. This method enriches the details of image generation but increases the difficulty of the training process. Zhang [6] and others proposed the Stacked Generative Adversarial Network Architecture (StackGAN). The article proposed the use of two Generative Adversarial Networks to achieve text-to-image generation because when the generator is just a simple upsampling method, it cannot improve the quality of the generated samples, so StackGAN's task of generating text images is divided into two stages. The first stage generates images with lower accuracy. This stage is mainly used to generate basic information such as the outline, background, and color of the image. In the second stage, the output samples of the first stage are used as input to generate high-quality high-definition images, and the detailed information lost in the generated samples is complemented. Zhang [7–9] and others proposed the StackGAN++ model on the basis of previous research. In the paper, the Generative Adversarial Network was designed to be similar to a tree structure, and multiple generative neural networks and adversarial neural networks were trained in parallel, which effectively reduced the model's training time.

This paper designs a stock market trend prediction stock market trend image description generation model based on a generative confrontation network. The model includes a generator and a discriminator. The generator proposes a time-varying preattention mechanism, which allows each stock market trend to predict the stock market trend. Image features pay attention to other image features to introduce the relationship between different image regions so that the decoder can better understand the relationship information in the stock market trend prediction stock market trend image. The discriminator takes the recurrent neural network as the main body and considers the input sentence and the degree of matching between the reference sentences and the image features of the stock market trend.

3. Related Theoretical Methods for Generating Confrontation Networks

Generative Adversarial Network is a neural network that uses unsupervised learning methods. It does not require complex processing of input data and is very suitable for processing data with large amounts of data and incomplete information. The basic structure of the generative confrontation network includes a generator and a discriminator. The generator is generally composed of a multilayer neural network, which reconstructs the data obtained by random sampling in a certain way to be consistent with the sample dimension. When the characteristics of the generated stock market trend prediction image need to be restricted, a control vector is added to the data and input to the judgment. In the device, finally update the network with the judgment result, as shown in Figure 1 [10–12].

GAWWN uses a pair of generators and discriminators to build a model, introduces bounding box conditions and key point information to control the pose and position of the subject in the composite image, and generates a 128×128

resolution image. Figure 2 shows a typical GAN framework in the field of image generation.

4. Generating the Model of Stock Market Forecast Trend Image Based on the Generative Confrontation Network

This section discusses the specific structure of the stock market prediction trend image generation model based on the generative confrontation network and the training method of the network. Figure 3 shows the overall architecture of the model. This section is divided into three parts: generating network, discriminating network, and training method.

4.1. Generating Network

4.1.1. Encoder. In the model, the encoder uses a deep residual neural network (Resnet). The residual neural network is an improvement of the traditional neural network. After the traditional neural network reaches a certain depth, it is difficult to increase the depth to obtain performance improvement. The residual neural network greatly deepens the depth of the network by increasing the residual connection. For the encoder, in terms of tasks, the residual network can also better extract feature vectors from the data. The principle and structure of the residual network are explained as follows [13–15].

The starting point of the residual network is to solve the problem that the single-layer neural network is difficult to fit the unit mapping. Assuming that the input of the network is X and the fitting target is $H(X)$, experiments show that, under the condition of $H(X) = X$, fitting is very difficult, so the residual network introduces a new fitting target $F(X) = H(X) - X$. At this time, the fitting goal of $H(X) = X$ is equivalent to that of $F(X) = 0$. This fitting is easy to achieve, so the fitting goal can be obtained as $F(X) = H(X) + X$, as shown in Figure 4(a).

In the figure, the first linear transformation and activation function is equivalent to a layer of the neural network, and the function of the subsequent linear transformation is to transform the dimension to be the same as the input. The main contribution of the residual network is to increase a path from input to output. The learning of the original objective function is transformed into the learning of the residual. The addition of residuals allows the neural network to train normally after reaching a depth of tens of layers and achieve better performance than networks with fewer layers. In practical applications, in order to further deepen the network depth, Resnet also uses a structure called a bottleneck building block [16], which is shown in Figure 4(b).

Each box in the figure represents a convolution operation. The input of the block is a piece of data with 256 channels, after a convolution operation with a convolution window of 1×1 and a convolution kernel of 64, the function of this step is to reduce the dimensionality of the data, and the data continues to pass through a convolution window to 3×3. The convolution kernel is a convolution operation of
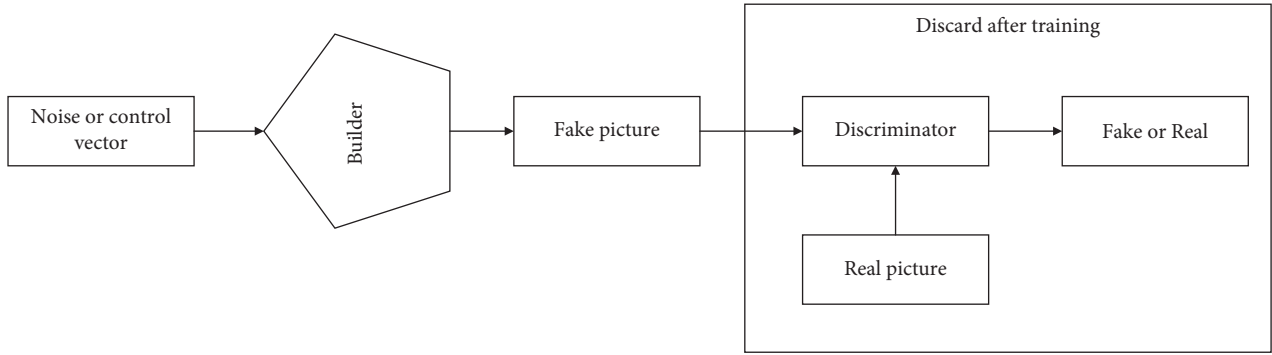
FIGURE 1: The structure diagram of the generated confrontation network.
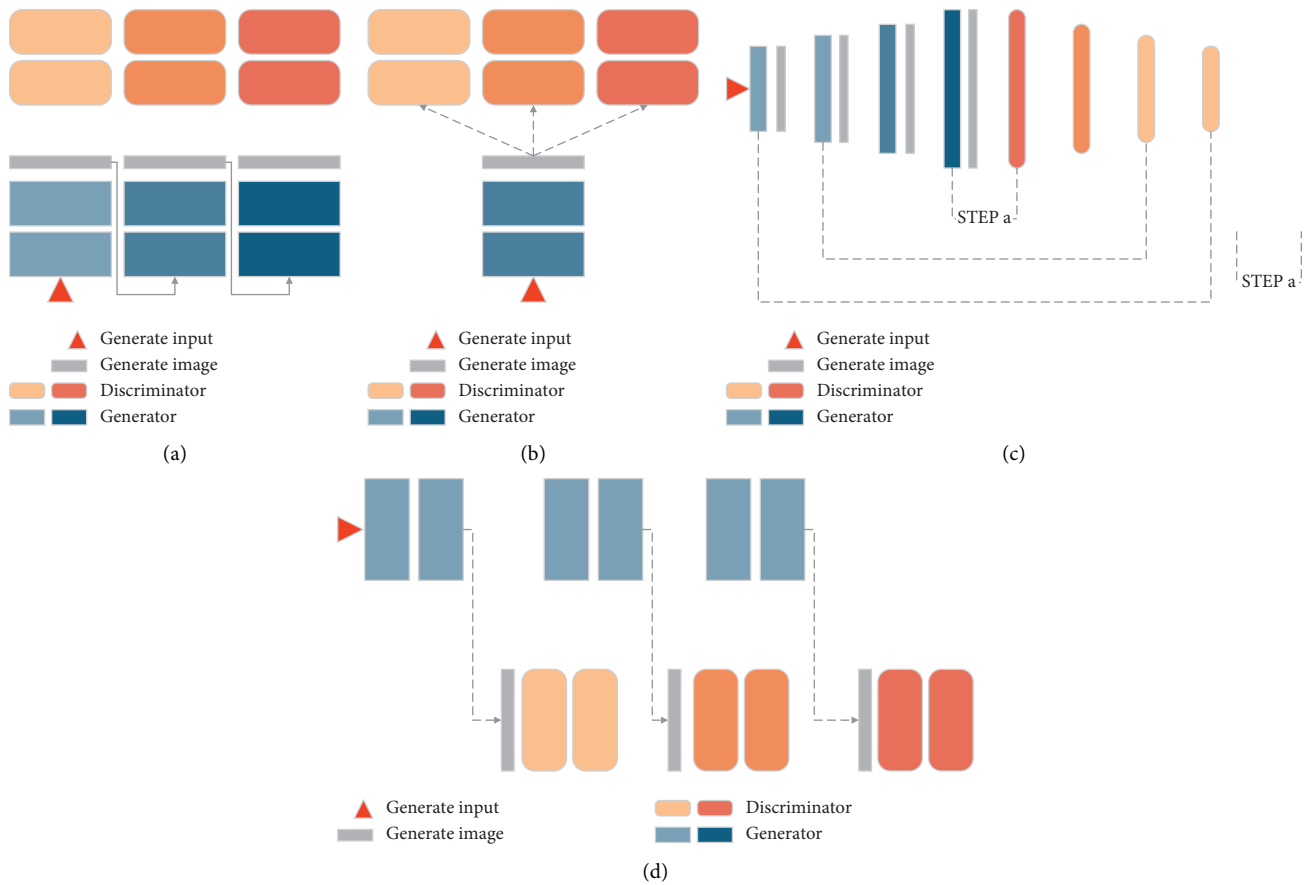
FIGURE 2: A typical GAN framework in the field of image generation. (a) Stacked. (b) Multidiscriminator. (c) Progressive. (d) Nested.

64, and finally, after a convolution operation with a convolution window of 1×1 and a convolution kernel of 256, the data dimension is restored to 256 channels, and the data after the restored dimension is the same as the original data. Add the restored dimension data to the original data, that is, convert the original pixel feature data to more specific and physical features and then solve the classification This block structure is the basic structure in Resnet, and extremely deep networks are stacked by this block. The experiment in this paper uses the 101-layer Resnet, denoted as Resnet-101, and its complete structure is shown in Table 1.

First, input the historical raw stock market data to do preliminary convolution and pooling. After the data passes through multiple bottleneck building blocks, each matrix in the table corresponds to the above basic block structure. Except for the number and step size of the convolution kernel, the others are consistent with the above example. After the matrix, $\times N$ represents stacking N of the same structure. The output after all block operations is $7 \times 7 \times 2048$, as 49 2048-dimensional stock market trend image spatial feature vectors, and each feature vector is linearly transformed into D-dimensional feature vectors used by the model [17–19].
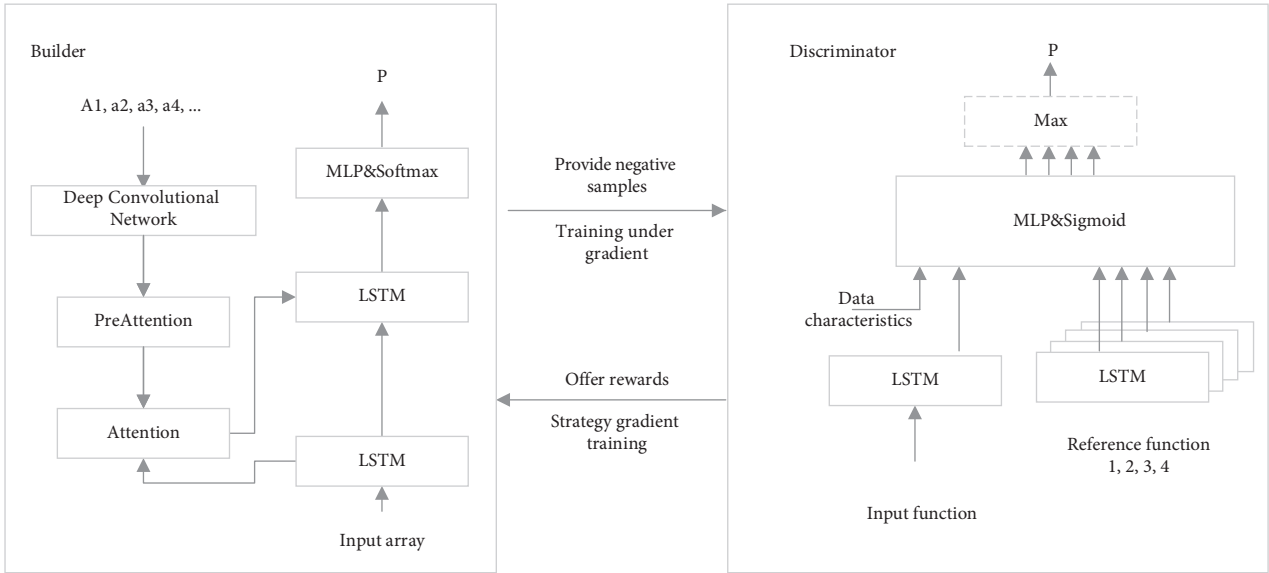
FIGURE 3: The overall architecture of the stock market prediction trend image generation model based on the generative confrontation network.
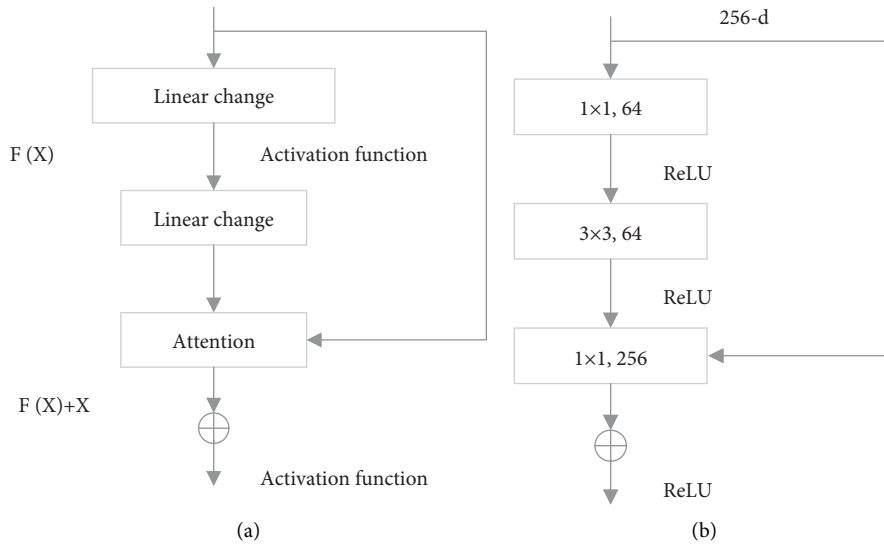
FIGURE 4: Schematic diagram of residual neural network. (a) Schematic diagram of residual learning. (b) Bottleneck building block structure.

4.1.2. Time-Varying Preattention Mechanism. The attention mechanisms used in previous research work, such as commonly used soft attention and top-down attention, directly use the data spatial features extracted from the deep convolutional network, and the final attention vector is expressed in the form of the weighted sum of all feature vectors. However, some studies have pointed out that although convolutional neural networks and fully connected networks have achieved excellent results in many tasks of predicting stock market trend images, even simple reasoning about relationships in stock market trend images is difficult for them. Relational reasoning ability is very important in the image description task of predicting stock market trends. A simple image description of predicting stock market trends may only contain entities and attributes, but a slightly more complicated description often involves the

association between entities or attributes. The lack of relational reasoning capabilities of convolutional neural networks means that the encoder only extracts the local spatial information of the data but does not include the correlation between these pieces of information. In this case, if the attention mechanism is only a simple image feature weighted sum, it is difficult to fully express the relationship in the image, and considering that the image information predicting the stock market trend that the decoder can access only comes from the attention mechanism, the decoder may not be able to accurately understand the relationship in the image predicting the stock market trend.

There are rich relationships in predicting stock market trend images. The relationship between any two data trends is intricate. Such a large number of relationships lead to a large number of relationships in the aggregate feature vector,

Table 1: Complete structure of Resnet-101.

Operate	Output dimension
Resize and crop	$224 \times 224 \times 3$
7×7 convolution, convolution kernel 64, step size 2	$112 \times 112 \times 64$
3×3 max pooling, step size 2	$56 \times 56 \times 64$
$\begin{bmatrix} 1 \times 1 & 64 \\ 3 \times 3 & 64 \\ 1 \times 1 & 256 \end{bmatrix} \times 3$	$56 \times 56 \times 256$
$\begin{bmatrix} 1 \times 1 & 128 \\ 3 \times 3 & 128 \\ 1 \times 1 & 512 \end{bmatrix} \times 4$	$28 \times 28 \times 512$
$\begin{bmatrix} 1 \times 1 & 256 \\ 3 \times 3 & 256 \\ 1 \times 1 & 1024 \end{bmatrix} \times 23$	$14 \times 14 \times 1024$
$\begin{bmatrix} 1 \times 1 & 512 \\ 3 \times 3 & 512 \\ 1 \times 1 & 2048 \end{bmatrix} \times 3$	$7 \times 7 \times 2048$

Average pooling, 1000-dimensional fully connected network, softmax 1000.

which makes it difficult for the decoder to determine if the required relationship is not limited. What relationship is important, so it is necessary to limit the relationship considered in the preattention mechanism. What kind of relationship is required cannot be determined in advance. It should be determined by the sentence that you want to describe or has been described. That is, a semantic context is needed to help the preattention mechanism find the required relationship. Here, the semantic context vector is represented as h_{t-1}^2, and its source will be introduced in the next section. This context vector changes with time because every moment a new word is generated, which affects the content to be described next and the corresponding preattention mechanism aggregate feature vector [20, 21].

It is also time-varying, so it is called time-dependent preattention (TDPA). After introducing the semantic context h_{i-1}^2, the formula is as follows:

$$
\begin{aligned}
a_{ti} &= w_p^T \tanh \left(W_{vp} + \left(W_{sp} v_i + W_{hp} h_{i-1}^2 \right) 1^T \right), \\
\alpha_{ti} &= \text{softmax} \left(a_{ti} \right), \\
v_{ti}^{'} &= \sum_{j=1}^{L} \alpha_{tij} v_j,
\end{aligned}
\tag{1}
$$

where $1 \in R^L$ is an L-dimensional all-one vector and $W_{vp} \in R^{L \times D}$, $W_{sp} \in R^{L \times D}$, $w_{hp} \in R^{L \times D}$, and $w_p \in R^L$ are parameters to be learned.

$v_{ti}^{'}$ contains the relationship information between v_j and other feature vectors. Connect it with the image feature vector v_i and describe the features of this area at the current moment through a fully connected network:

$$
v_{ti} = \text{MLP} \left(\left[v_{ti}^{'}; v_i \right] \right).
\tag{2}
$$

Figure 5 shows the above process.

4.1.3. Decoder.
Inspired by some previous research work using hierarchical LSTM (hierarchical LSTM), this paper designs a hierarchical LSTM combined with a time-varying
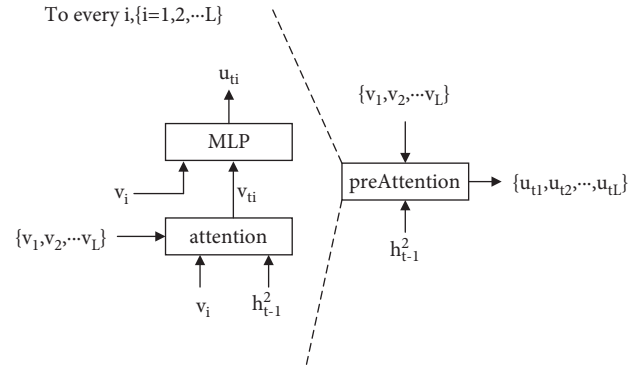
Figure 5: The structure of the time-varying preattention mechanism.

preattention mechanism as a decoder. The advantage of hierarchical LSTM is that it has multiple LSTM units, and the special LSTM units can be used to encode the information required by the attention mechanism, which helps to better play the role of the time-varying preattention mechanism. In the overall framework, the decoder contains the following parts: a preattention module, which is responsible for generating aggregate feature vectors; an attention module, which helps LSTM pay attention to the aggregate feature vectors; 2 LSTM units, one of which is used to encode the attention module , the required information, one for text generation; MLP and softmax layers Figure 6 shows the complete decoder framework.

The following describes the specific work of each module one by one.

The bottom LSTM: the function of this LSTM unit is to encode the information required by the attention mechanism from the memory of the network, the current input, and the global image features, so its input is the hidden state of the top LSTM at the previous moment h_{t-1}^2, the currently input word vector x_t, and the average image feature $\bar{v} = 1/L \Sigma_{i=1}^{L} v_i$, these vectors are connected and input into the
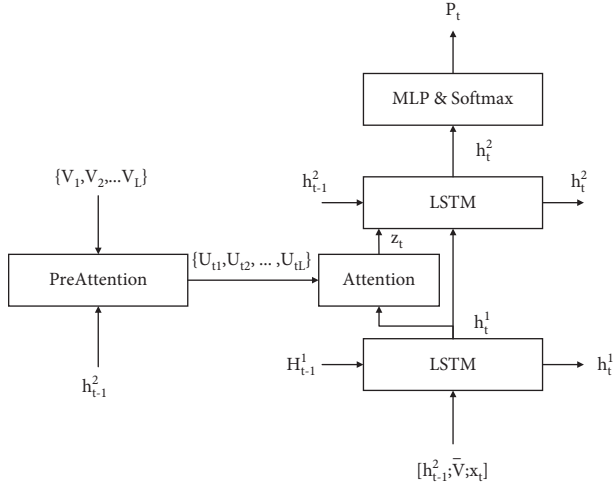
FIGURE 6: Decoding architecture.

bottom LSTM, and the bottom LSTM combines its own memory to get the output.

Preattention layer: this layer includes the use of pre-attention mechanism to generate aggregated and predicted stock market trend image features and the use of the underlying LSTM output information to guide the generation of attention to aggregate predicted stock market trend image features. The hidden state h_{t-1}^2 of the top LSTM at the previous moment encodes all the semantic information generated before, which is used in preattention, that is, in the previous section:

$$U_t = \text{preattention}\left(V, h_{t-1}^2\right). \tag{3}$$

The output U_t contains the relational information and the image features of aggregated and predicted stock market trends. Then another attention module will notice the local special image feature U_t according to the hidden state h_t^1 of the underlying LSTM, that is, do the following calculation:

$$
\begin{aligned}
a_i &= w_a^T \tanh\left(W_{ua} U_t + \left(W_{ha} h_t^2\right) 1^T\right), \\
\alpha_t &= \text{softmax}\left(a_t\right), \\
z_t &= \sum_{i=1}^{L} \alpha_{ti} U_{ti}.
\end{aligned}
\tag{4}
$$

Among them, $1 \in R^L$ is an L-dimensional all-one vector, $W_{ua} \in R^{L \times D}$, $W_{ha} \in R^{L \times M}$, and $w_a \in R^L$ are the parameters to be learned, and z_t is on U_t attention vector.

The top LSTM: the function of this LSTM unit is to integrate the attention vector z_t and the output information h_t^1 of the bottom LSTM and combine its own memory to generate the information of the next word:

$$h_t^2 = \text{LSTM}\left(\left[h_t^1; z_t\right], h_{t-1}^2\right), \tag{5}$$

where $h_t^2 \in R^M$ represents the hidden state of the top LSTM at time t.

MLP and softmax layers: the hidden state h_t^2 of the top LSTM becomes a probability distribution through a single-layer fully connected network and softmax function.

$$P_t = \text{softmax}\left(\text{MLP}\left(h_t^2\right)\right). \tag{6}$$

Interpret this probability distribution as the distribution on the dictionary, and then you can get the generated words at the current moment by sampling.

4.2. Discriminating the Network. The discriminator is used to distinguish whether a sentence is a real data description or generated by a model. Therefore, the input of the discriminator is a sentence and a set of data, and the output is the probability P that the sentence is a true description of the predicted data image. Each group of data in the training data has 5 description sentences, 4 of which are randomly selected as reference sentences and 1 is used as a positive sample to train the discriminator.

Figure 7 shows the architecture of the discriminator. The input sentence (denoted as in) will first be mapped to the word vector sequence for neural network processing and then input into an LSTM word by word, and the hidden state of the LSTM at the last moment will be taken out as the sentence encoding vector of the input sentence, denoted as

$$h_{\text{in}} = \text{LSTM}\left(s_{\text{in}}\right). \tag{7}$$

For the 4 reference sentences (denoted as $s_{\text{ref}}^i = 1, 2, 3, 4$), the same LSTM module is used to do the same processing, and they are, respectively, turned into corresponding sentence encoding vectors:

$$h_{\text{ref}}^i = \text{LSTM}\left(s_{\text{ref}}^i\right). \tag{8}$$

Each reference sentence encoding vector is connected to the input sentence encoding vector and the data global feature vector and then input to a fully connected network. The data global feature vector is the average of all the feature vectors. This fully connected network contains 3 layers, and each layer is used. In addition to the residual connection, the activation function of the last layer is the sigmoid function, which is used to map the input to a value between 0 and 1. It is interpreted as the probability of the input sentence and the image under the auxiliary reference of the reference sentence:

$$P^i = \sigma\left(\text{MLP}\left[\overline{v}; h_{\text{in}}; h_{\text{ref}}^i\right]\right). \tag{9}$$

Then take the maximum of the 4 probabilities as the probability that the input sentence is the true description of the image, namely,

$$P = \max_i P^i. \tag{10}$$

4.3. Training Method. This section discusses how to train the generative network and the discriminant network because when training the generative network, you will encounter a discrete problem, which makes the gradient unable to be
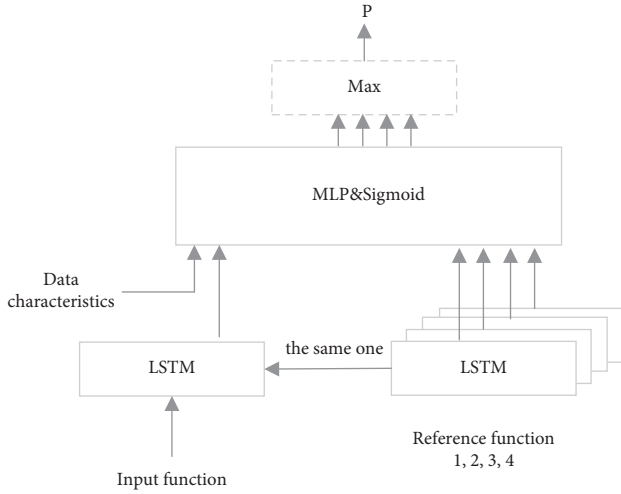
Figure 7: Discriminator architecture.

backpropagated. This section introduces a reinforcement learning framework and uses the policy gradient method in reinforcement learning to solve this problem. Discuss the training objectives and training algorithms of the generated network and the discriminant network through the reinforcement learning side.

4.3.1. Reinforcement Learning. Reinforcement learning is a decision-making model that describes how to seek the largest long-term reward in the interaction between the subject and the environment. There are several key concepts in reinforcement learning. The agent in reinforcement learning has the ability to act, where the action is the interaction with the environment, and the action will get new observations from the environment and change the subject's state (state, a specific state corresponds to a specific reward (reward) obtained from the environment). Use the symbols a, o, s, and r to represent actions, observations, states, and rewards; then, a series of behavior history can be expressed as

$$o_1, r_1, a_1, \cdots, a_{t-1}, o_t, r_t. \tag{11}$$

The subscript represents time. This sequence represents the observation O_1, the reward r_1 is obtained, and the action a_1 is performed, the observation is changed to O_2, the reward r_2 is received, and the action a_2 is performed, and the observation is changed to o_3, the process is repeated until the action at -1 is observed and the observation is changed to o_t, and reward r_t is obtained. State s is a summary of the entire behavior history, namely,

$$s_t = f(o_1, r_1, a_1, \cdots, a_{t-1}, o_t, r_t). \tag{12}$$

In many reinforcement learning tasks, it is assumed that the environment is fully observable (full observability); that is, the subject can observe the complete current environmental state, and the state can be determined by current observations, namely,

$$s_t = f(o_t). \tag{13}$$

In reinforcement learning, the main body contains one or more parts of policy, value function, and model.

Strategy refers to the method of determining action from the state, which is divided into deterministic strategy and random strategy. The mathematical representation of a deterministic strategy is

$$a = \pi(s). \tag{14}$$

The mathematical representation of randomness strategy is

$$\pi(a|s) = P(a_t = a|s_t = s). \tag{15}$$

Here, π represents the strategy.

The value function is used to estimate the value of a specific action in a certain state. Unlike rewards, rewards are obtained with the state and are short-term benefits at the current moment, while the value function focuses on the fact that an action will bring how many long-term rewards. The value function can be defined as

$$Q^\pi(s, a) = E(r_{t+1} + \gamma r_{t+2} + \gamma^2 r_{t+3} + \cdots | s, a). \tag{16}$$

Among them, γ is a conversion factor between 0 and 1 because future rewards are uncertain, and discounts need to be discounted until now. The longer the reward, the greater the uncertainty and the greater the discount.

The model is the prediction of the state and reward of the environment at the next moment under the current state and the selected action. The mathematical description is

$$P^a_{ss'} = P(s_{t+1} = s'|s_t = s, a_t = a), \tag{17}$$

$$R^a_s = E(r_{t+1}|s_t = s, a_t = a). \tag{18}$$

Formula (17) expresses the probability of transition to state s' after taking action a in state s, and formula (18) expresses the expected value of environmental rewards after taking action a in state s.

Subjects can be roughly divided into three categories according to their composition: (1) policy based, such as policy gradient and actor-critic algorithm; (2) value based, such as Q-learning; (3) model based.

4.3.2. Train Discriminant Grid and Generating Grid. Use θ and w for all the parameters contained in the generator and discriminator, respectively. To represent, the generator is represented by the symbol G_θ, and the discriminator is represented by the symbol D_w. Then, discuss how to train these two networks.

In the traditional training method, the goal of the training generator G_θ is to maximize the likelihood of generating the true description sentence $y = \{y_1, y_2, \ldots, y_T\}$ corresponding to the predicted data image in a given set of data. The objective function at this time is

$$L_{XE}(\theta) = -\frac{1}{N} \sum_y \sum_{i=1}^{T} \log P(y_i|y_{1:\, i-1}, I, \theta). \tag{19}$$

Likelihood here is equivalent to an artificially set model evaluation index, but as described in Section 1, this index has obvious flaws. The design of the generative confrontation network introduces the idea of confrontation in the neural network by introducing a discriminator D_w. To determine whether a sample is real or generated by the generator, the discriminator D_w in fact plays a role in evaluating the quality of the samples generated by the generator. It can be seen as an evaluation method for generators, scoring the samples generated by the generator to help improve the generator. This evaluation method is not obtained by setting rules but has the ability to self-learn and progress. A simple strategy is used when training the discriminator: the real sample is good, give 1 point; the generated sample is bad, give 0 points; this paper uses the least square loss function to train the discriminator:

$$\min_{w} \frac{1}{2}E_{y \sim P_{\text{real}}}\left[\left(D_w(y, I) - 1\right)^2\right] + \frac{1}{2}E_{y \sim P_{\text{fake}}}\left[\left(D_w(y, I)\right)^2\right], \tag{20}$$

where P_{real} is the distribution of the real language description corresponding to image I and P_{fake} is the distribution of the language description of the generator corresponding to the image I. In the image description task, the work of the discriminator includes two aspects. On the one hand, it judges whether the generated sentence meets the natural language specification, and on the other hand, it judges whether the generated sentence semantically matches the image. Therefore, in the above formula, introduce additional items to require the discriminator to learn the matching relationship between sentence semantics and images. The final optimization goal is

$$\min_{w} \frac{1}{2}E_{y \sim P_{\text{real}}}\left[\left(D_w(y, I) - 1\right)^2\right] + \frac{\alpha}{2}E_{y \sim P_{\text{fake}}}\left[\left(D_w(y, I)\right)^2\right]$$
$$+ \frac{\beta}{2}E_{y \sim P_{\text{unm}}}\left[\left(D_w(y, I)\right)^2\right]. \tag{21}$$

Among them, P_{unm} represents the distribution of image I that does not match but comes from the real language description, and α and β are hyperparameters. This objective function can be optimized by the backpropagation algorithm.

For the generator, the goal of optimization is to make the samples generated by themselves get high scores under the evaluation index of the discriminator. According to the original design of the adversarial network, the goal of optimization should be

$$\min_{\theta} \frac{1}{2}E_I\left[\left(D_w(G_\theta(I), I) - 1\right)^2\right]. \tag{22}$$

However, this optimization goal cannot be trained by backpropagation. This is because the generator $G_\theta(I)$ (approximately needs to sample the output distribution when generating words). This process is not differentiable, causing the gradient signal of the loss function to fail to generate. The output of the generator continues to propagate back, so the

parameters of the generator cannot be updated through backpropagation. In order to solve this problem, this paper introduces a reinforcement learning method in the training.

First, explain the data description model as a reinforcement learning problem. Because every word generated is known, the problem of data description is a completely observable problem, so no distinction is made between observation and state. The main body in reinforcement learning is the generator in the confrontation network. The action a_t is the word y_t generated by the generator at each moment, and the state s_t is the generated word sequence $y_{1: t-1}$. The intermediate state is not obtained. Only when the complete sentence is generated, the reward is obtained, and the reward for the complete sentence is the score given by the discriminator.

Then determine the strategy, value estimation function, and model included in the subject. The meaning of the strategy is how to generate the distribution of the next word according to the state (the generated word sequence), that is, the output distribution of the generator model $G_\theta(y_{t+1}|y_{1: t}, I)$ approximately. The value estimation function is represented by $Q^{G_\theta}(s_t, y_{t+1}, I)$ approximately, which means that starting from the state s_t, select the action y_{t+1}, and follow the strategy G_θ. Generate words, the long-term cumulative rewards are obtained, and the calculation of this reward depends on the data I. In the model, under the current state $y_{1: t-1}$ and the selected action y_t, the state and reward at the next moment are determined. In this way, the original Generative Adversarial Network structure is interpreted as a reinforcement learning model. Then, set the generator's optimization goal to start from the initial state, 0, and follow the generator's strategy to generate the sentence with the largest expected cumulative reward, namely,

$$L(\theta) = \sum_{y_1} G_\theta(y_1|s_0, I)Q^{G_e}(s_0, y_1, I). \tag{23}$$

In order to train the generator, it is necessary to obtain the gradient of the optimization target to the network parameters and use the policy gradient to solve this problem.

Introduce the symbol $V^{G_e}s_t, I$ to represent the expected cumulative reward of the state s_t under the strategy G_e, namely,

$$V^{G_e}s_t, I = \sum_{y_{t+1}} G_\theta(y_{t+1}|s_t, I)Q^{G_e}(s_t, y_{t+1}, I) = Q^{G_e}(s_{t-1}, y_t, I). \tag{24}$$

Derivation of the objective function is as follows:

$$\nabla_\theta L(\theta) = \nabla_\theta\left[\sum_{y_1} G_\theta(y_1|s_0, I)Q^{G_e}(s_0, y_1, I)\right]$$
$$= E_{y_{1: t-1} \sim G_\theta} \sum_{y_t} \nabla_\theta G_\theta(y_t|s_{t-1}, I)Q^{G_e}(s_{t-1}, y_t, I). \tag{25}$$

Formula (25) gives an estimate of the target gradient, where the expectation can be approximated by sampling, and the gradient descent method can be used to train the

generator after the gradient is obtained. G_θ in the formula only defines complete sentences. For incomplete sentences, let it share the same value of G_θ with complete sentences. At this time,

$$\nabla_\theta L(\theta) = \frac{1}{T} \sum_{t=1}^{T} E_{y_t \sim G_\theta(y_t|s_{t-1},I)}$$

$$[D_w(y_{1:T}, I) \nabla_\theta \log G_\theta(y_t|s_{t-1},I)]. \tag{26}$$

In order to reduce the variance of the estimated gradient, consider not adding the reference term b at the end of the gradient estimation, as long as b has nothing to do with $W:T$; then,

$$\nabla_\theta L(\theta) = \frac{1}{T} \sum_{t=1}^{T} E_{y_t \sim G_\theta(y_t|s_{t-1},I)}$$

$$[D_w(y_{1:T}, I) - b] \nabla_\theta \log G_\theta(y_t|s_{t-1},I). \tag{27}$$

For the selection of b, this paper uses the self-critical benchmark $D_w(\hat{y}_{1:T},I)$ proposed in," where $\hat{y}_{1:T}$ is the sentence obtained by the current model using the inference method. The inference method can choose greedy decoding, beam search, and so on. The greedy decoding method is selected in the paper. From this formula, if the quality of the sampled sentence is better than the quality of the greedy decoding during the training process, then the probability of generating such a sentence is increased. If the sentence is sampled, the quality is worse than greedy decoding, thus reducing the probability of generating such a sentence. If the generator is well trained at this time, the quality of the sampled sentence and the sentence of greedy decoding should be better. If the generator is poorly trained at this time, then the quality of the sampled sentence and the sentence of greedy decoding should be relatively poor, so the self-criticism benchmark can always be maintained at the level near the sampled sentence, so the variance of the estimated gradient can be effectively reduced.

Now it is possible to train the entire adversarial network, but after starting the adversarial training, the generator and the discriminator need to be pretrained because if the pretraining is not done, the quality of the samples generated by the generator is very poor, and the discriminator is easy to distinguish. Given a very low score, the generator can only learn which samples are bad but cannot learn good samples, so the maximum likelihood estimation is used to pretrain the generator and the discriminator, respectively. After the pretraining is completed, the generator and the discriminator are trained alternately until the two converge. Algorithm 1 shows the complete training process.

5. Model Application and Result Analysis

5.1. Experimental Environment and Data Set Research Object.
Randomly select a number of stocks in the A-share market to carry out the price prediction experiment, verify the basic data generation through the database, and finally select all the daily data of the stock market since 2020, including the opening price, closing price, highest price, and lowest price. The selected indicators mainly include opening price, closing price, highest price, lowest price, trading volume, total trading value, etc. Research methods: according to the key indicators of investors' attention, the 14-day RSI, 10-day CR, 10-day BR, WR difference from the previous day, 10-day BIAS, MACD value, 9-day KDJ value, ROC difference, price change, and price, economic indicators such as the range of change and the range of change in trading volume are used as data characteristics inspection criteria to simulate the psychological state of investors. At the same time, the selection of technical indicators is constantly evaluated and adjusted to ensure the validity of data features. Using the LSTM-GAN network, adjust the network structure through continuous experimentation (number of hidden layers, number of hidden layer neurons, weight initialization method of each layer, excitation function used by each network layer, dropout function setting, model training iteration number, iteration Batch_size, and other parameters) to predict the increase in the closing price of the stock market. Today's closing price increase = (today's closing price-yesterday's closing price)/ yesterday's closing price. How many days' data are used as a window to predict the next day's stock price increase data requires careful consideration. After a large number of experimental tests, the final model uses the data of the previous 14 days to predict the closing price increase on the 15th day, and the positive or negative increase value information represents the rise or fall of the closing price of the stock market. Through this method, the stock price rise and fall predictions are made, and at the same time, a rise value is obtained as a reference. n represents the number of data records in the training set.

5.2. Using the Model to Conduct Experiments on Important A-Share Indexes.
Among the A shares, 30 stock markets were randomly selected. Each stock market selected about 700 data, which is the historical stock market data since 2020. Among them, 80% of the data are used as training data to train the model, and the remaining 20% are used as test data to evaluate the model. The BP neural network is used to predict 30 stocks.

The forecast accuracy rates of each stock market are 51.33%, 52.94%, 54.19%, 58.21%, 51.20%, 52.94%, 56.39%, 54.55%, 55.56%, 51.47%, 55.15%, 53.91%, 50.86%, 52.10%, 51.89%, 53.23%, 53.79, 54.41%, 55.56%, 54.81%, 53.68%, 56.81%, 51.13%, 57.01%, 52.59%, 55.45%, 51.47%, 52.94%, 51.47%, and 52.94%, the highest of which is 58.21% and the minimum is 50.8%. The average forecast accuracy rate of all stock markets is 53.67%, as shown in Figure 8.

The same stock market data directly use the LSTM-GAN model; after the tuning process, the average accuracy rate obtained is 54.03%, as shown in Figure 9. The accuracy of the stock price fluctuations predicted by the two is roughly the same, with a slight difference. Comparison of results obtained by LSTM reveals that the BP network has a weak advantage, which also verifies the potential of LSTM in stock market forecasting.

Input: generator G_e, discriminator D_w. Training data <stock market, real trend> collection;
Output: Trained generator G_e, discriminator D_w
(1) Randomly initialize generator G_e, discriminator D_w
(2) Use the objective function announcement (19) to pretrain the generator G_e;
(3) Use the generator G_e to generate the stock market trend prediction graph, and use three kinds of data <stock market, real trend>, <stock market, generated trend>, and <stock market, unmatched real trend> to train the discriminator D_w according to the objective function (21)
(4) **repeat**
(5) for g-turn times do
(6) Use generator G_e to generate sentence $y_{1:\, r}$
(7) Estimate the gradient according to formulas 3–45 and update the parameters of the generator G_e
(8) **end for**
(9) for g-turn times do
(10) Use the generator G_e to generate the image description, use three kinds of data <stock market, real trend>, <stock market, generated trend>, and <stock market, unmatched real trend> to train the discriminator D_w according to the objective function formula (25)
(11) **end for**
(12) until generator G_e, discriminator D_w converges

ALGORITHM 1: A network training algorithm for stock prediction trend image generation based on a generative confrontation network.
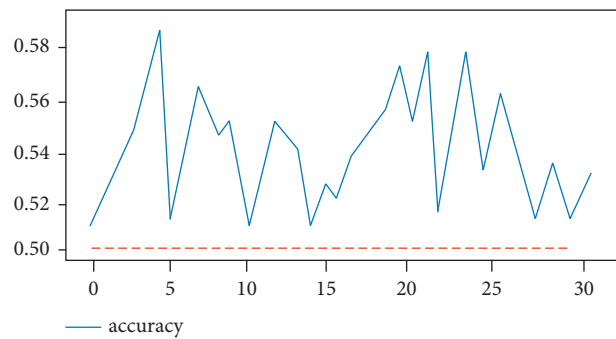
FIGURE 8: BP network closing price forecast results.
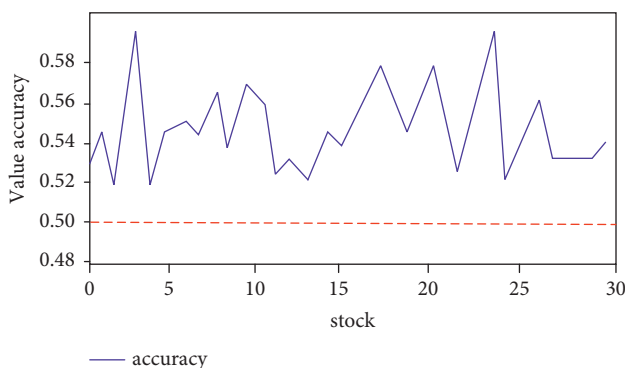
FIGURE 9: The forecast result of the closing price after BP is changed to LSTM-GAN.

First, use the data of the Shanghai Composite Index to train the model. The parameter tuning of the model is all carried out around the Shanghai Composite Index. When a relatively satisfactory result is achieved, the model will be used to verify other large-cap indexes, such as Shenzhen Component Index, Shanghai and Shenzhen 300, and Small and Medium-Sized Index. Because these indexes are not easy to be manipulated by "bookmakers," and their stock prices and transaction data can more reflect investors' judgments on the market, these large-cap indexes are more suitable for evaluating forecasting models. Specific experimental results are shown in the following.

5.2.1. Shanghai Composite Index. The data set uses the historical daily data of the Shenzhen Component Index from April 1, 2009, to February 27, 2020, 1894 strip. Considering that historical stock market data before 2009 did not have much reference value, especially the economic crisis in 2008, data collection began in 2009. 75% of the data is used as training data, and the remaining 25% is used for verification testing.

Daily line data include a total of 1921 data: training set, 1447; test set, 474.

After 20 iterations, the predicted result is shown in Figure 10:

5.3. Experimental Results. The experimental results prove that the designed LSTM-GAN regression prediction model is effective. And the accuracy of the model after tuning for a
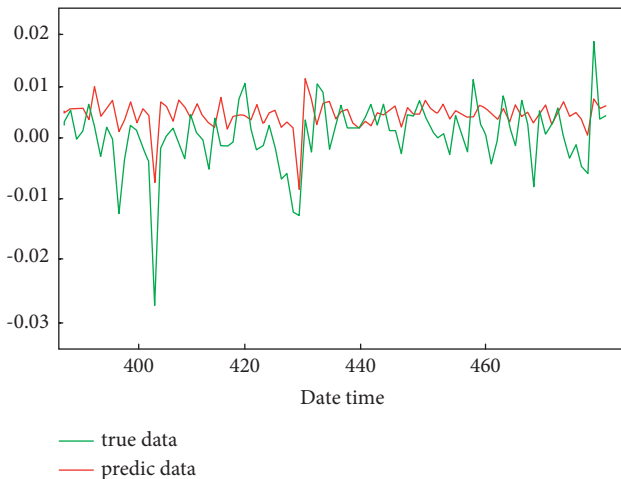
FIGURE 10: Details of the forecast results of the closing price increase of the Shenzhen Component Index. The accuracy rate of forecasting changes: 61.603%; increase forecast error value: 0.0120.

single stock market reaches 61.603%. Moreover, without modifying the model parameters, it is applied to other large-cap indexes, and its average accuracy rate exceeds 60%, which fully verifies the effectiveness and universality of the designed model for the A-share large-cap index. The potential of LSTM in stock market forecasting is worthy of in-depth study and exploration. At the same time, another innovative point of the experiment is that the regression prediction value is the predicted value of the closing price increase, and the average predicted increase error value is also obtained during the process of training the model, which can be provided to investors for reference.

Through experiments, it can also be shown that although China's stock market is a policy market, its rise and fall are more affected by policies, and stock price data will have a certain "singularity," but the market index is relatively stable and quite measurable.

6. Conclusion

This paper systematically expounds on the theoretical basis of Generative Adversarial Networks in deep learning and applies these two models to the stock market to predict the ups and downs of A-share stocks in the next day. The main research results of this paper are as follows.

First, from the perspective of deep learning, while introducing neural networks, this paper also clarifies how LSTM-GAN neural networks solve the shortcomings of traditional BP neural networks one by one.

Second, apply deep learning technology to the stock price prediction. Take historical transaction data as input data and use the LSTM-GAN model to predict the rise and fall (up, adjustment, and fall) of stocks. The prediction accuracy of the model reached 60%.

Compared with the traditional neural network, the average error of this new network method is smaller, and in most cases, the relative error is better than the traditional neural network. However, there are still deficiencies in the

experiment process that need to be improved. Although the prediction model based on the generative confrontation network introduces a preattention mechanism, it solves the problem of difficulty in convergence and difficulty in training and avoids the problem of nonconvergence and collapse of the algorithm in the budget process. However, this model is compared with other generative models. The GAN model does not need to be modeled in advance. In the case of more pixels, the model is too free to cause uncontrollable problems.

Data Availability

The dataset can be accessed upon request to the author.

Conflicts of Interest

The author declares no conflicts of interest.

References

[1] I. Goodfellow, J. Pouget-Abadie, and M. Mirza, "Generative adversarial nets," in *Proceedings of the Advances in Neural Information Processing Systems*, pp. 2672–2680, Cambridge MIT Press, Cambridge, MA, USA, June 2014.

[2] S. Talwar, M. Viberg, and A. Paulraj, "Blind separation of syn‐chronous co-channel digital signals using an antenna array (part I. algorithms)," *IEEE Transactions on Signal Processing*, vol. 44, no. 5, pp. 1184–1197, 1995.

[3] Y. P. Wei, L. K. Liu, and H. Guo, "Approach to blind estimation of pseudo random code sequence for DSSS signaling multipath," *Computer Engineering and Applications*, vol. 49, no. 12, pp. 195–199, 2013.

[4] S. Reed, Z. Akata, and X. Yan, "Generative adversarial text to image synthesis," in *Proceedings of the International Conference on Machine Learning*, pp. 1060–1069, ACM, New York, NY, USA, May 2016.

[5] S. E. Reed, Z. Akata, and S. Mohan, "Learning what and where to draw," *Advances in Neural Information Processing Systems*, vol. 12, pp. 217–225, 2016.

[6] X. Zhang, J. Zhao, and Y. Lecun, "Character-level convolutional networks for text classification," *Advances in Neural Information Processing Systems*, vol. 1, pp. 649–657, 2015.

[7] H. Zhang, T. Xu, and H. Li, "StackGAN++:Realistic image synthesis with stacked generative adversarial networks," in *Proceedings of the IEEE Transactions on Pattern Analysis and Machine Intelligence*, p. 1, IEEE, Venice, Italy, October 2017.

[8] Z. Xu, J. Du, and J. J. Wang, "Satellite image prediction relying on GAN and LSTM neural networks," in *Proceedings of the IEEE International Conference on Communications*, Shanghai, China, May 2019.

[9] A. Joulin, E. Grave, and P. Bojanowski, "Bag of tricks for efficient text classification," in *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics*, ACL, Valencia, Spain, September 2016.

[10] S. Reed, Z. Akata, and H. Lee, "Learning deep representations of finegrained visual descript-tions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 49–58, IEEE, Las Vegas, NV, USA, June 2016.

[11] E. Denton, S. Chintala, and A. Szlam, "Deep generative image models using a Laplacian pyramid of adversarial networks," in *Proceedings of the 28th International Conference on Neural Information Processing Systems (NIPS'15)*, pp. 1486–1494, ACMDL, Montréal, Canada, December 2015.

[12] T. Karras, T. Aila, and S. Laine, "Progressive growing of GANs for improved quality, stability, and variation," in *Proceedings of the International Conference on Learning Representations*, ICLR, Vancouver, Canada, May 2018.

[13] C. Ledig, L. Theis, and F. Huszár, "Photorealistic single image superresolution using a generative adversarial network," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4681–4690, IEEE, Honolulu, HI, USA, July 2017.

[14] K. He, X. Zhang, and S. Ren, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778, IEEE, Las Vegas, NV, USA, June 2016.

[15] A. Santoro, D. Raposo, and D. G. Barrett, "A simple neural network module for relational reasoning," in *Proceedings of the Conference and Workshop on Neural Information Processing Systems*, NIPS, Denver, CO, USA, June 2017.

[16] J. Song, L. Gao, Z. Guo, W. Liu, D. Zhang, and H. T. Shen, "Hierarchical LSTM with adjusted temporal attention for video captioning," in *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence*, August 2017.

[17] X. Mao, Q. Li, and H. Xie, "Least squares generative adversarial networks," in *Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 2813–2821, IEEE, Venice, Italy, October 2017.

[18] B. Dai, D. Lin, and R. Urtasun, "Towards diverse and natural image descriptions via a conditional GAN," in *Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV)*, IEEE, Venice, Italy, October 2017.

[19] L. Yu, W. Zhang, and J. Wang, "SeqGAN: sequence generative adversarial nets withPolicy gradient," in *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence (AAAI 2017)*, Cornell University, Ithaca, NY, USA, August 2017.

[20] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*, MIT Press Cambridge, London, England, 1998.

[21] S. J. Rennie, E. Marcheret, and Y. Mroueh, "Self-critical sequence training for image captioning," in *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Venice, Italy, July 2017.

Research Article

Intelligent Classification Method of Low Occupancy Big Data Based on Grid Index

Haiyan Zhao [ID][1] **and Shuangxi Li**[2]

[1]*School of Information Science and Engineering, Tianjin Tianshi College, Tianjin 301700, China*
[2]*School of Computer Science and Engineering, Tianjin University of Technology, Tianjin 300384, China*

Correspondence should be addressed to Haiyan Zhao; glister@email.tjut.edu.cn

In order to enhance the load balance in the big data storage process and improve the storage efficiency, an intelligent classification method of low occupancy big data based on grid index is studied. A low occupancy big data classification platform was built, the infrastructure layer was designed using grid technology, grid basic services were provided through grid system management nodes and grid public service nodes, and grid application services were provided using local resource servers and enterprise grid application services. Based on each server node in the infrastructure layer, the basic management layer provides load forecasting, image backup, and other functional services. The application interface layer includes the interfaces required for the connection between the platform and each server node, and the advanced access layer provides the human-computer interaction interface for the operation of the platform. Finally, based on the obtained main structure, the depth confidence network is constructed by stacking several RBM layers, the new samples are expanded by adding adjacent values to obtain the mean value, and the depth confidence network is used to classify them. The experimental results show that the load of different virtual machines in the low occupancy big data storage process is less than 40%, and the load of each virtual machine is basically the same, indicating that this method can enhance the load balance in the data storage process and improve the storage efficiency.

1. Introduction

In the process of operation and development, enterprise networks will accumulate a large number of low occupancy big data [1, 2]. Effective analysis of this low occupancy big data can obtain implicit data and knowledge, resulting in data value added and providing and attaching a variety of services [3, 4], which shows that low occupancy big data is very key. These low occupancy big data are stored in the form of message file or database and rise exponentially [5], so high-quality storage methods are required. Intelligent classification technology is a very important big data management technology. At present, it has been successfully used in various fields, but the problems of high occupancy rate and low classification efficiency of traditional big data classification are still difficult to solve. In view of the above phenomena, scholars at home and abroad have put forward the following solutions.

In [6] with the rapid development of sensing and digital technology, network physical system is regarded as the most feasible platform to improve architectural design and management. It investigates the possibility of integrating energy management system with network physical system to form energy network physical system to promote building energy management. However, due to the dynamics of building occupants, minimizing energy consumption while realizing the architectural function of energy network physical system is a challenge. Because occupant behavior is the main source of uncertainty in energy management, ignoring it will usually lead to energy waste caused by overheating and undercooling, as well as discomfort caused by insufficient heat and ventilation services. In order to alleviate this uncertainty, an energy network physical system related to occupancy is proposed, which combines occupancy detection based on WiFi probe. The framework uses integrated classification algorithm to extract three forms of

occupancy information. It creates a data interface connecting the energy management system and the network physical system and realizes automatic occupancy detection and interpretation by assembling multiple weak classifiers for WiFi signals. A validation experiment was conducted in a large office to check the performance of the proposed network physical system occupying linked energy. Experimental and simulation results show that the proposed model can save about 26.4% of refrigeration and ventilation energy consumption with appropriate classifiers and occupancy data types. However, this method does not conduct in-depth research on data classification methods. In [7] in complex image environment, text detection and location from natural images embedded with text is still a challenging problem. Foreground object segmentation and classification are common methods for this task. Therefore, component level target classification in a clutter environment is an important subproblem. Proper extraction of foreground objects can achieve effective classification, so as to improve the performance of text detection. A feature vector based on equidistant pixel area distribution is proposed for text/nontext classification. The generated feature descriptor is script invariant and is very effective in the actual scene. Five different pattern classifiers are used to evaluate the proposed feature set on our data set. The experimental results show that the classification accuracy of the feature set is more than 86% regardless of the script. However, this method only classifies image data and has high limitations. Reference [8] uses three kinds of random forest classifiers to solve the problem of anomaly classification of sensor data. Consider the sensor deployment scenario specifically, where the sensor fields of view may overlap. According to the sensor type, we design signal features in time, frequency, and space-time domain. The results show that the proposed random forest classifier has higher true positive rate and lower false positive rate than unsupervised k-means method and random forest classifier with single signal energy feature. However, the classification effect of this method for low occupancy data is still unknown.

Grid technology can transform different computer resources distributed in a wide range of space into a computing power and data processing power characterized by universality, standardization, accuracy, and economy, so as to realize wide area resource sharing [9]. In essence, grid technology can be understood as maximizing the use of existing resources (software and hardware) in the network, meeting the storage, sharing and calculation of data and resources, improving the problem of resource island [10], and realizing the value-added network resources. Grid technology is widely used in data storage fields such as e-commerce, academic research, and so on. Based on this, this paper studies the low occupancy big data classification method based on grid technology and uses grid technology to integrate low occupancy big data and realize high-quality storage of low occupancy big data.

2. Low Occupancy Big Data Processing and Classification

2.1. Big Data Preprocessing Stage. Multisource network data is massive and high-dimensional. Cross-source classification algorithm feature extraction of multisource network data needs dimension reduction function. For multisource heterogeneous data, the traditional dimensionality reduction method can not determine the dimensionality of all kinds of data. The amount of data contained in multisource heterogeneous data is high, and its workload is also high. Feature extraction methods with high computational efficiency are needed.

Incremental orthogonal component analysis (IOCA) is selected to reduce the dimension and extract features of multisource network data, so as to improve the time complexity of massive multisource network data classification. IOCA method does not need to set a fixed target dimension, and the target dimension can be adjusted according to the changes of input data in the learning process [11]. Using this method to preprocess massive data will form better orthogonal components and avoid data redundancy and good compression dimension.

IOCA method can use the prefetched multisource network data to obtain the orthogonal component space $\{d_1, d_2, \cdots, d_n\}$ that can automatically determine the dimension, so as to realize the rapid dimensionality reduction of multisource network data. The dimensionality reduction of multisource network data needs to be realized through the following two steps:

(1) Let the existing new data be represented by φ_{k+1} and the learned orthogonal component space be represented by $D = \{d_1, d_2, \cdots, d_k\}$, and calculate the new potential orthogonal component d_{k+1} possibly generated by φ_{k+1} and the linear independence between them.

(2) Set the adaptive threshold, and use the set adaptive threshold to judge whether d_{k+1} can be used as a new orthogonal component added to D.

The specific calculation process of extracting data features by IOCA method is as follows:

(1) Initialize the orthogonal component space $D = \varnothing$ with the initial dimension $k = \dim(D) = 0$.

(2) Use x_j to represent the new input data and meet $j > k$.

(3) Use $r_{i,k+1}$ to represent the eigenvector and calculate $r_{i,k+1} = x_j^T d_i$.

(4) Calculate $d_{k+1}' = x_j - \sum_{i=1}^{k} r_{i,k+1} d_i$.

(5) Calculate $r_{k+1,k+1} = \|d_{k+1}'\|_2$.

(6) Calculate $d_{k+1} = d_{k+1}'/r_{k+1,k+1}$.

(7) g represents the original data dimension. When $r_{k+1,k+1}/\|x_j\| \geq \dim(D)/g$, it means that d_{k+1} belongs to the new orthogonal component. At this time, d_{k+1} is added to D.

(8) Let $k = k + 1$ repeat the above process until all data preprocessing is completed.

The features of multisource network data are extracted by using the corresponding orthogonal space to obtain multiple groups of feature vectors with lower dimensions

[12]. The feature vectors with lower dimensions are sent to the multicore learning support vector machine classifier to realize the cross-source classification of multisource network data.

2.2. Intelligent Classification Algorithm Selection.

Support vector machine is a machine learning method based on structural risk minimization principle and statistical learning theory [13], and support vector machine has good generalization performance. Support vector machine realizes multisource network data classification by searching the optimal hyperplane [14]. The solution formula of the optimal hyperplane obtained through the quadratic optimization problem is as follows:

$$\begin{cases} \min \dfrac{1}{2}\|w\|^2 + C\displaystyle\sum_{i=1}^{N}\xi_i \\ \\ s.t.\, y_i\left[(w,x)+b\right] \geq 1-\xi_i \end{cases} \tag{1}$$

In formula (1), $\|w\|^2$ and ξ_i represent the regularization term and the relaxation variable measuring the training error of the sample, respectively. The smaller the value of ξ_i, the lower the training error. At this time, the support vector machine has higher classification accuracy; C represents the penalty coefficient of the balance adjustment parameter between the training error and the regularization term. The smaller the value, the lower the degree of penalty error classification; $(w,x)+b = 0$ represents the hyperplane to be solved, and w and b represent the normal vector and offset of the hyperplane, respectively.

2.2.1. Multicore Learning Method.

Multisource heterogeneous data can replace a single kernel with multiple kernels to improve the interpretability of decision function and the classification performance of support vector machine classifier [15, 16]. The convex combination formula based on kernel function $K(x, x')$ is as follows:

$$K(x, x') = \sum_{m=1}^{M} s_m K_m(x, x'). \tag{2}$$

In formula (2), $s_m \geq 0$, $\sum_{m=1}^{M} s_m = 1$. M and K_m represent the total number of cores and the number of positive definite cores M in the same input space χ, respectively. The classical kernel with different parameters is represented by each basic kernel K_m.

Through the above process, the weight s_m is selected to replace the data representation of the kernel.

Based on the gradient descent of the target value of support vector machine, the gradient descent of support vector machine solver is used to determine the combination of different problem kernels, that is, multicore learning method. The multicore learning is realized by clarifying the coefficient $\{s_m\}$ of the learning process of the decision function. The multicore learning formula is as follows:

$$\begin{cases} \displaystyle\min_{\{f_m\},b,\xi,s} \dfrac{1}{2}\sum_{m=1}^{M}\dfrac{1}{s_m}\|f_m\|_{H_m}^2 + C\sum_{i=1}^{l}\xi_i \\ \\ s.t.\, y_i\displaystyle\sum_{m=1}^{M} f_m(x_i) + y_i b \geq 1-\xi_i, \forall i \\ \\ \xi_i \geq 0, \forall i \\ \\ \displaystyle\sum_{m=1}^{M} s_m = 1, s_m \geq 0, \forall m \end{cases} \tag{3}$$

In formula (3), $C > 0$ and $\xi_i \geq 0$ represent the penalty coefficient and relaxation variable, respectively, and l represents the number of samples. When $s_m = 0$, the square norm $\|f_m\|_{H_m}$ of f_m in each s_m control objective function needs to be 0 so that the objective value is limited.

Let the Lagrange function exist as follows:

$$L = \dfrac{1}{2}\sum_{m=1}^{M}\|f_m\|_{H_m}^2 + C\sum_{i=1}^{l}\xi_i + \sum_{i=1}^{l}\alpha_i$$

$$\left(1 - \xi_i - y_i\sum_{m=1}^{M} f_m(x_i) - y_i b\right) \tag{4}$$

$$-\sum_{i=1}^{l}v_i\xi_i + \lambda\left(\sum_{m=1}^{M} s_m - 1\right) - \sum_{m=1}^{M}\eta_m s_m.$$

In formula (4), both α_i and v_i represent Lagrange multipliers related to support vector machine problems. Both λ and η_m represent Lagrange multipliers related to constraints on s_m.

Relative to the original variable, set the Lagrange function gradient to 0, substitute the set optimization conditions into the Lagrange function, and obtain the dual problem formula as follows:

$$\begin{cases} \displaystyle\max_{\alpha} \quad -\dfrac{1}{2}\sum_{i,j=1}^{l}\alpha_i\alpha_j y_i y_j\sum_{m=1}^{M} s_m K_m(x_i, x_j) + \sum_{i=1}^{l}\alpha_i \\ \\ s.t. \displaystyle\sum_{i=1}^{l}\alpha_i y_j = 0 \\ \\ C \geq \alpha_i \geq 0, \forall i \end{cases} \tag{5}$$

The formula (5) is transformed into the dual formula of standard support vector machine by using kernel combination $K(x_i, x_j) = \sum_{m=1} s_m K_m(x_i, x_j)$. $J(s)$ is used to represent the optimal objective function of cross-source classification of multisource network data. Multicore learning has strong duality, so $J(s)$ can be used as the

objective function of the dual problem of formula (5) at the same time. Select the descending direction S of the gradient descent method and update s through the obtained $J(s)$ gradient. The updating process is $s \longleftarrow s + \gamma S$, where γ represents the step size. It is necessary to search the maximum allowable step according to the descent direction S to judge whether there is a decrease in the target value. When the target value decreases, s needs to be updated. Repeat the above process until the target value does not decrease.

2.2.2. Support Vector Machine Classification of Multicore Learning. Support vector machine needs to use the combination of multiple binary classifiers to solve the multiclass classification problem. Suppose that the category in the massive multisource network data is q, and the number of binary classification tasks is $q(q-1)/2$ through pairing the categories with the number of q. Send the multisource network data to the trained support vector machine classifier, and finally obtain the classification results with the number of $q(q-1)/2$.

Use $J(s)$ to represent the target value of cross-source classification of multisource network data, search for the kernel combination that can jointly optimize all decision functions as an even number [17], and obtain the objective function of optimizing cross-source classification of multisource network data according to kernel weight $\{s_m\}$ as follows:

$$J(s) = \sum_{q \in Q} J_q(s). \tag{6}$$

In formula (6), Q and $J_q(s)$, respectively, represent all binary classifier sets to be considered and the target value of support vector machine for binary classification problems related to binary classifier. The Lagrange multipliers of each binary classification problem are obtained according to the gradient descent method, and the kernel combination of all binary classification problems is obtained, that is, the sum of maximized intervals, so as to realize the cross-source classification of multisource network data.

3. Build a Big Data Classification Platform with Low Occupancy

The low occupancy big data classification platform based on grid technology is composed of four parts: advanced access layer, application interface layer, basic management layer, and infrastructure layer, as shown in Figure 1.

Using the open grid services architecture (OGSA) and GT3 toolkit, build the infrastructure layer of the low occupancy big data private cloud platform on the basis of ensuring the low occupancy big data system structure. Develop the basic management layer, application interface layer, and advanced access layer according to the actual application requirements of low occupancy big data storage.

The infrastructure layer is the foundation of the physical and storage devices in the low occupancy big data system classification platform, which mainly includes four parts: grid system management node, grid public service node, local resource

server, and enterprise grid application service [7, 18]. Each part of the equipment is scattered, and some types of server equipment have a large number and scattered locations. Therefore, WAN connection is adopted between different server devices [19].

The main function of the basic management layer is to use integrated or distributed control to manage the collaborative operation of all storage devices in the low occupancy big data private cloud, complete the optimal utilization of resources, and build a work contradiction mechanism for problems such as workflow contradiction. At the same time, this layer is also responsible for data backup, data encryption, and data disaster recovery in the process of low occupancy big data classification.

The application interface layer includes all interfaces used by the low occupancy big data classification platform to connect with other equipment and servers and provides interfaces and corresponding services to power management institutions at all levels according to the actual application requirements and corresponding levels of different power enterprises. The user can successfully log into the platform by inputting the corresponding account password through the public interface of the low occupancy big data classification platform and collect the corresponding data resources according to the account level authority.

The main function of the advanced access layer is to provide the low occupancy big data classification platform with the interface required for operation. On the basis of ensuring the primary application of low occupancy big data classification, the advanced application is developed according to the actual application requirements of power enterprises to realize the interpersonal interaction function.

3.1. Hardware Design

3.1.1. Infrastructure Layer. Grid technology has a wide range of applications and can be effectively deployed on LAN, Wan, and Internet [20]. In order to meet the internal low occupancy big data sharing of power enterprises, a low occupancy big data network including shared areas is established by using grid technology. The application of grid technology can ensure the safe access and sharing of different low occupancy big data resources in the grid system.

The high-speed network inside the power enterprise is used to connect different application system servers, database servers, storage backup servers, and other servers. Set up the grid operating system, set up professional grid scheduling servers, registration servers, server pools, and other devices, and integrate these servers and devices into a grid system to build a low occupancy big data grid. Using the open grid service architecture and GT3 toolkit, build the infrastructure layer of the low occupancy big data private cloud platform on the basis of ensuring the low occupancy big data system structure. The infrastructure layer grid is shown in Figure 2.

The infrastructure layer designed by using grid technology includes grid basic services provided by grid system management node and grid public service node, grid
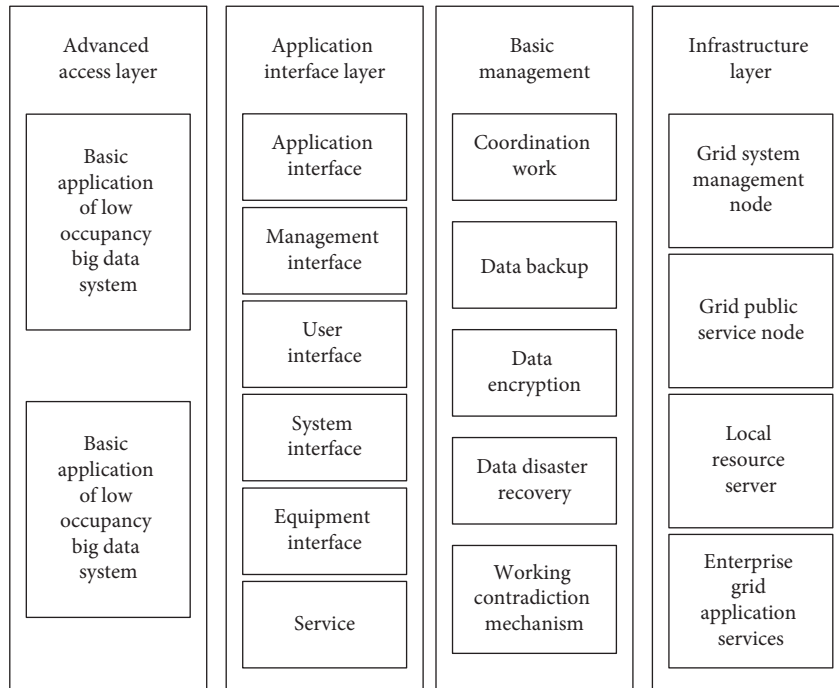
FIGURE 1: Overall structure of low occupancy big data classification platform.
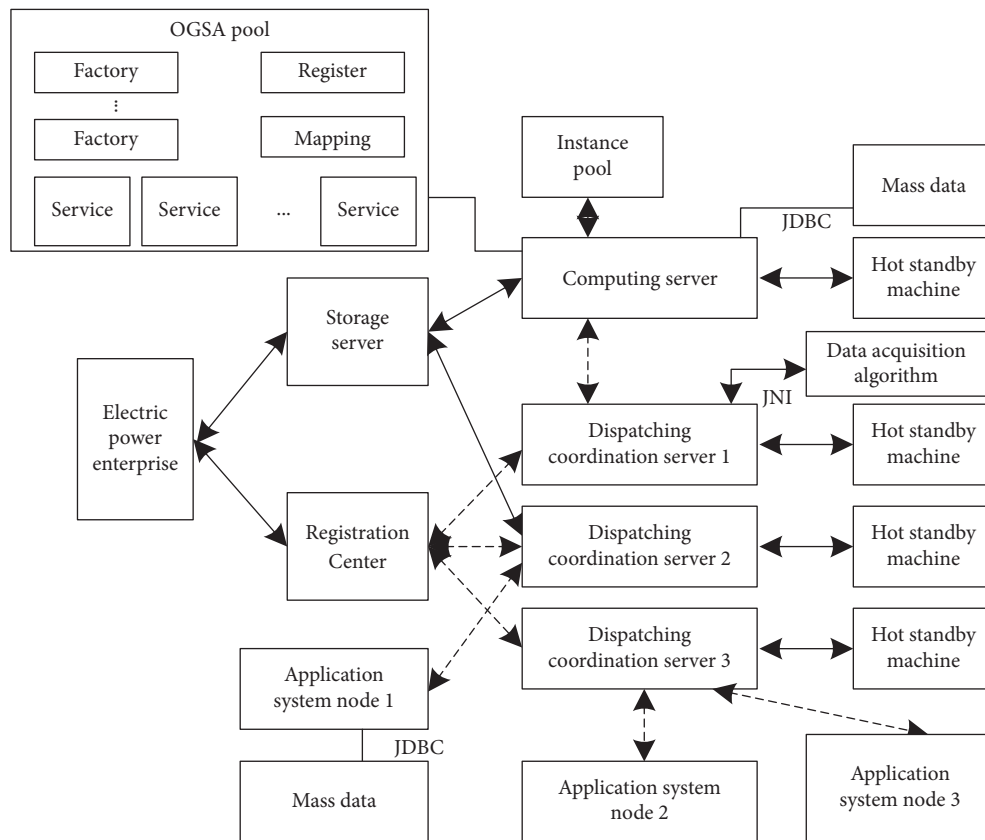
FIGURE 2: Infrastructure layer construction based on grid technology.

application services provided by local resource server, and enterprise grid application services.

Dispatching coordination server, registration server, and other servers and equipment that realize basic grid technology form a grid system management node, and different servers are equipped with hot standby machines to ensure their uninterrupted operation [21, 22]. The main function of each server in the grid system management node is to provide basic grid

services and ensure the effective work of the grid system. Dispatching coordination server is the main component of grid technology and is responsible for controlling and coordinating all grid operations and services.

Storage server, computing server, and instance pool form a grid common service node [23]. The main function of the storage server is to provide data storage services in the grid, which has the characteristics of high capacity and reliability. The computing server usually adopts high-performance computer, which has high-precision computing performance. The instance pool contains different services in the grid, such as web services (the basis for building grid services) and public services.

As the infrastructure of grid technology, local resource server is mainly composed of heterogeneous systems such as application system and automatic office system [24].

Enterprise grid application service is an advanced service that defines and implements different grid applications according to the needs of power enterprises based on the foundation and public services provided by public service nodes. Based on the concept of grid, different computing resources and storage resources in low occupancy big data grid can be realized through grid services.

3.1.2. Basic Management.
The basic management layer includes various functional modules applied in the process of low occupancy big data classification, such as load prediction module, image backup module, data security module, physical host module, etc. The structure hierarchy of each functional module is shown in Figure 3.

3.1.3. Load Forecasting Layer.
Based on the application requirements of low occupancy big data classification platform, it is necessary not only to provide the historical load change trend of virtual machine (storage server) to power enterprises, but also to display the load prediction results of virtual machine to power enterprises, so as to meet the purpose of diversified application of low occupancy big data resources [25, 26]. Using load forecasting, you can apply for an appropriate amount of virtual machines to balance the load before the virtual machine load reaches the upper limit [27], so as to alleviate the access pressure. When the virtual machine load forecast is lower than a certain load value within a certain time, the virtual machine can be recovered to improve the utilization rate of the virtual machine and meet the requirements of environmental protection.

Load forecasting is based on the load monitoring results. After preliminary processing, the load monitoring results are calculated by BP neural network [28], and the prediction results are obtained. The prediction results are stored in the database and displayed through the interface in the advanced access layer. The load forecasting process is shown in Figure 4.

3.1.4. Image Backup Module.
The main function of the image backup module is to backup each virtual machine in the low occupancy big data classification platform within a fixed time under the control of the timer, so as to restore the virtual machine in case the platform cannot run due to an accident. The image backup process is shown in Figure 5.

In the process of image backup, it is necessary to judge whether the virtual machine has image backup. If so, delete the old image backup and build a new image backup to ensure the uniqueness of image backup.

3.2. Software Optimization.
Based on the above hardware platform, the low occupancy data after part 2 training and classification are input into the deep confidence network to optimize the software part of the low occupancy big data classification platform.

3.2.1. RBM Training.
In the process of classifying big data with low occupancy in complex scenes by using depth confidence network, the depth confidence network is obtained by stacking several RBM (restricted Boltzmann machine) layers [29], and each RBM layer is trained separately by contrast divergence method.

RBM includes visible layer and hidden layer, which are represented by v and h, respectively. The connection weight only exists in v and h nodes, and there is no connection weight between nodes of each layer. Under the condition that v is known, all h nodes have conditional independence. When v is input, h can be obtained based on conditional probability, and v can also be obtained according to h. Under the condition of optimizing the internal parameters of RBM, if the v' obtained by h is the same as the initial v, it indicates that the obtained h is another description form of v.

As a standard energy model, the energy function of RBM can be described by

$$E(v, h) = -\sum_{i=1}^{m}\sum_{j=1}^{n} w_{ij} v_i h_j - \sum_{i=1}^{m} a_i v_i - \sum_{i=1}^{m} b_j h_j. \qquad (7)$$

In formula (7), w_{ij}, a_i, and b_j represent the weight between E and R and the offset between them, respectively. Based on the joint configuration energy function shown in formula (7), the joint probability [30, 31] of v and h can be obtained. The formula is described as follows:

$$P(v, h) = \frac{e^{-E(v,h)}}{\sum_{v,h} e^{-E(v,h)}}. \qquad (8)$$

Since there is no connection between nodes in the layer in RBM, the conditional probability can be obtained according to the joint probability. The formula is described as follows:

$$P(v) = \frac{\sum_h e^{-E(v,h)}}{\sum_{v,h} e^{-E(v,h)}},$$

$$P(h) = \frac{\sum_v e^{-E(v,h)}}{\sum_{v,h} e^{-E(v,h)}},$$

$$P(v|h) = \frac{e^{-E(v,h)}}{\sum_v e^{-E(v,h)}}, \qquad (9)$$

$$P(h|v) = \frac{e^{-E(v,h)}}{\sum_h e^{-E(v,h)}}.$$
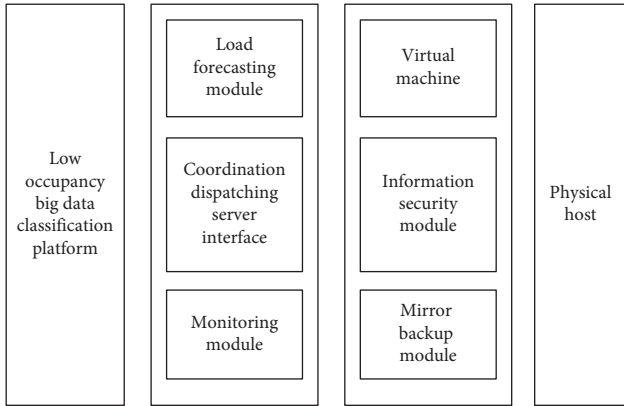
Figure 3: Structural hierarchy diagram of each functional module.
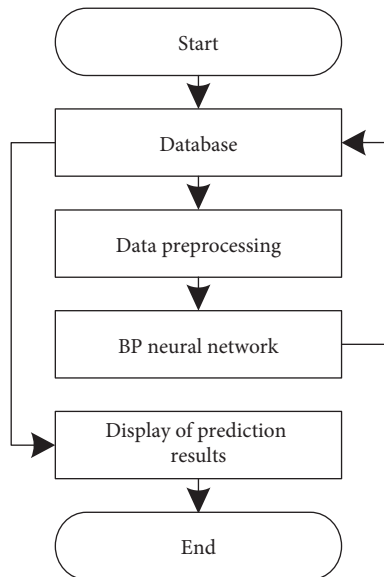
Figure 4: Load forecasting process.
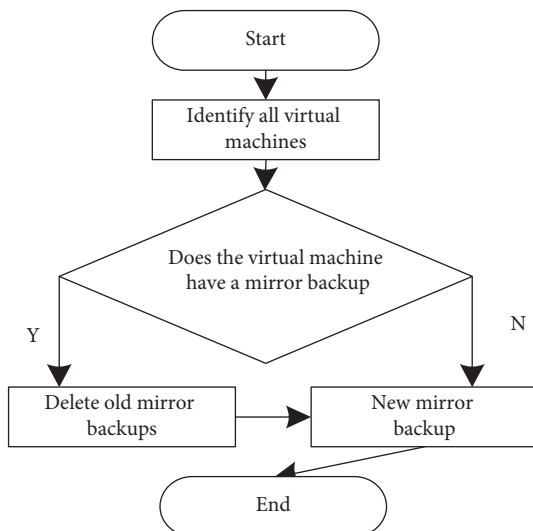
Figure 5: Image backup process.

After obtaining the conditional probability of RBM network, in order to make the Gibbs distribution described infinitely approximate to the fitting input data [32], it is necessary to learn (w, a, b) and other parameters. Generally, the parameter solution is obtained by solving the upper likelihood limit of the input sample.

In view of the long training time, the contrast divergence method can be used to improve the training efficiency. In this method, the gradient of the number likelihood function is solved by two approximations: (1) the average summation in the gradient calculation process is approximately replaced by the samples obtained in the conditional distribution; (2) Gibbs sampling was performed only once. The algorithm based on contrast divergence can obtain the optimization results of (w, a, b) and other parameters at a faster speed, so as to realize RBM network training.

3.2.2. Classification Process Based on Deep Confidence Network. The main feature of the deep confidence network is that it requires a large number of training samples. Therefore, when using the deep confidence network to divide the categories of low occupancy big data, it is necessary to reconstruct the data, expand the samples, and reduce the dimension of the samples [33].

When expanding samples, new samples can be obtained by adding adjacent to each other to find the mean. This method not only improves the number of training samples, but also considers the spatial correlation of samples. The specific process is described as follows:

Calculate the adjacent sum of each and its surrounding four directions, and divide the sum by 3 to obtain 4 times the number of new samples in the original low occupancy big data.

The same category of big data with low occupancy rate is relatively unified. Therefore, optimizing the input data based on spatial combination can greatly improve the classification accuracy of deep confidence network. If the input is simply added to the surrounding neighbors through training, the dimension and redundancy will be improved.

In order to reduce the dimension on the basis of spatial combination, principal component analysis or self-encoder is used to reduce the dimension of object metadata. The former method mainly reduces the dimension of linear data and the latter method mainly reduces the dimension of unstructured data. Compared with the two dimensionality reduction methods, the self-encoder can maximize the protection of the characteristics of low occupancy big data metadata, while the principal component analysis method can describe the data through fewer dimensions [34, 35].

The image after expanding the sample data and dimensionality reduction is used as the input of the first RBM visible layer of the depth confidence network, and several training depth confidence networks are randomly determined.

RBM layer training is to obtain parameter values through iteration, so as to describe low occupancy big data in other ways. The activation probability of the j-th node of the hidden layer in RBM can be obtained through formulas (8)–(10). The formula is described as follows:

$$P(h_j = 1|v) = \frac{1}{1 + \exp\left(-\sum_i w_{ij} v_i - b_j\right)}. \quad (10)$$

Meanwhile, under the condition that h is known, the activation probability of the i-th node can be described by

$$P(v_i = 1|h) = \frac{1}{1 + \exp\left(-\sum_j w_{ij} h_j - a_i\right)}. \quad (11)$$

The training is carried out according to the contrast divergence method, and the (w, a, b) isoparametric value in the RBM layer can be obtained after several iterations. Under this condition, h_1 is another way to describe big data with low occupancy.

Taking h_1 as the RBM input of the next layer and implementing iterative training in the same way, $h_2, h_3, \cdots$ will be obtained, and the corresponding (w, a, b) isoparametric values in each layer can be obtained. The pretraining process of the deep confidence network model can be realized through the layer-by-layer training process.

In order to obtain higher classification accuracy, a BP layer optimization parameter is introduced after the last hidden layer. If d and y^l are used to represent the expected classification result and the final hidden layer output result, respectively, the residual of the two is determined by

$$\delta^{(l)} = -\left(d - y^l\right) y^l \left(1 - y^l\right), \quad (12)$$

$\delta^{(l)}$ is transmitted from back to front. In each layer of $l = n^l - 1, n^l - 2, \cdots, \delta^{(l)}$ is determined as

$$\delta^{(l)} = \left(w^{(l)}\right)^T \delta^{(l+1)} y^l \left(1 - y^l\right). \quad (13)$$

Determine the partial derivatives of w and b according to $\delta^{(l)}$ of each layer:

$$\begin{aligned} \nabla w^{(l)} J &= y^l \delta^{(l+1)}, \\ \nabla b^{(l)} J &= \delta^{(l+1)}. \end{aligned} \quad (14)$$

After obtaining the partial derivative, the weights of w and b can be updated. After several iterations, the depth confidence network model optimization after pretraining can be realized, and the low occupancy big data classification method based on grid technology can be realized.

4. Experimental Analysis

In order to verify the effectiveness of the low occupancy big data classification method based on grid technology, taking the low occupancy data of an enterprise network in China as the research object, the low occupancy big data classification test was carried out using this method. The test results are as follows.

4.1. Experimental Environment and Parameter Setting.
The experimental environment is the hardware environment of Intel Celeron Turing 1 GHz CPU and 384 mb SD memory

TABLE 1: Parameter setting.

Parameter name	Numerical value
Threshold	—
Initial data	One
Node communication radius	11 m
Number of iterations	19 times
Test duration	$50 \, \mu$m
Transmission speed	15 Gb/s
Data overhead size	7984 B

and the software environment of Matlab6.1. The simulation system includes data interference module, resource scheduling module, and task generation module. Based on the composition of the above parts, ms-coco data set is used as the data acquisition source, and 1000 task data are arbitrarily taken in the data set. At the same time, the task data is proposed to be stored in the simulation system, and the data bytes are controlled between [256~568 kb]. The experiment termination condition was set to 300 times. The simulation algorithm parameters in this paper are shown in Table 1.

4.2. Virtual Machine Load Prediction Results.
Using this method, according to the load of the research object in the past 30 days, the load trend in the next 5 days is predicted and compared with the current load. The results are shown in Figure 6.

According to the analysis of Figure 6, the load prediction results of this method in the storage platform during the implementation of low occupancy big data classification are basically consistent with the actual load, which shows that this method can accurately predict the load in the low occupancy big data storage process, make use of load balancing control, and meet the actual application requirements of the research object.

4.3. Load Balancing Test.
The experiment verifies the load balancing performance of this method in the process of low occupancy big data classification with two indicators of load balancing and response time. The results are shown in Figure 7.

By analyzing Figure 7(a), it can be seen that the load of different virtual machines is less than 40% and the load of each virtual machine is basically the same in the process of low occupancy big data classification of the research object by using the method in this paper. By analyzing Figure 7(b), it can be seen that the response time of different virtual machines in the process of big data classification with low occupancy is controlled between 0.5 s and 0.7 s. This shows that the load balancing degree is high in the process of low occupancy big data classification, and the classification efficiency can be significantly improved through load balancing.

4.4. Storage Synchronization Test.
Storage synchronization is one of the key indexes to evaluate the performance of storage methods. The storage performance of this method is verified

(a)

(b)
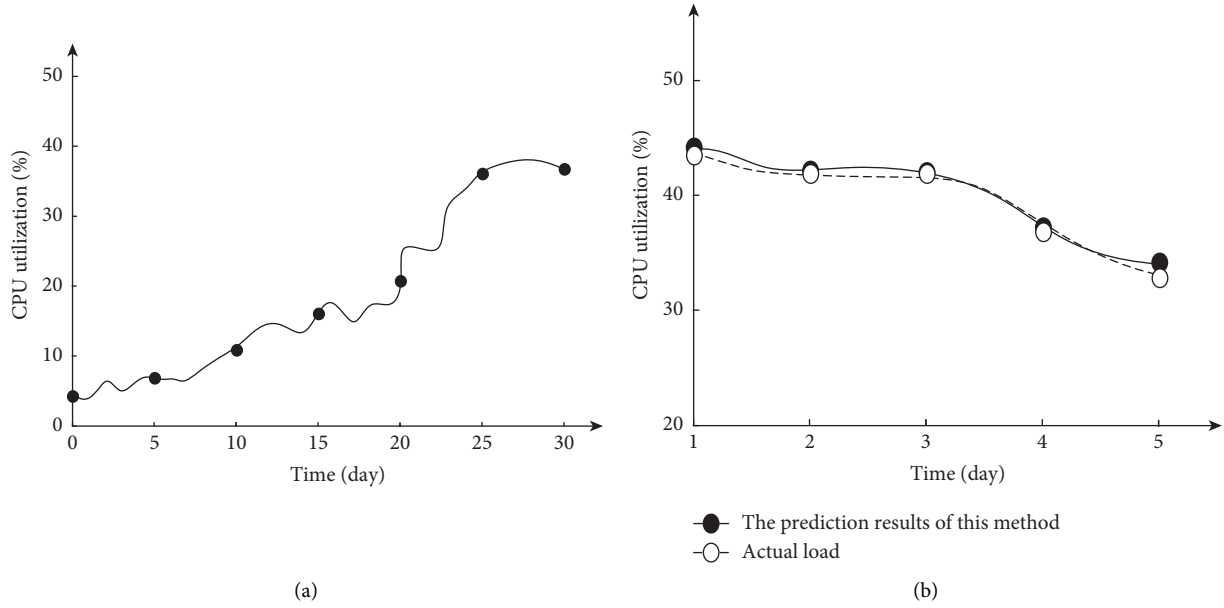
FIGURE 6: Load prediction results of this method. (a) Load in the last 30 days. (b) Forecast results for the next five days.

(a)

(b)
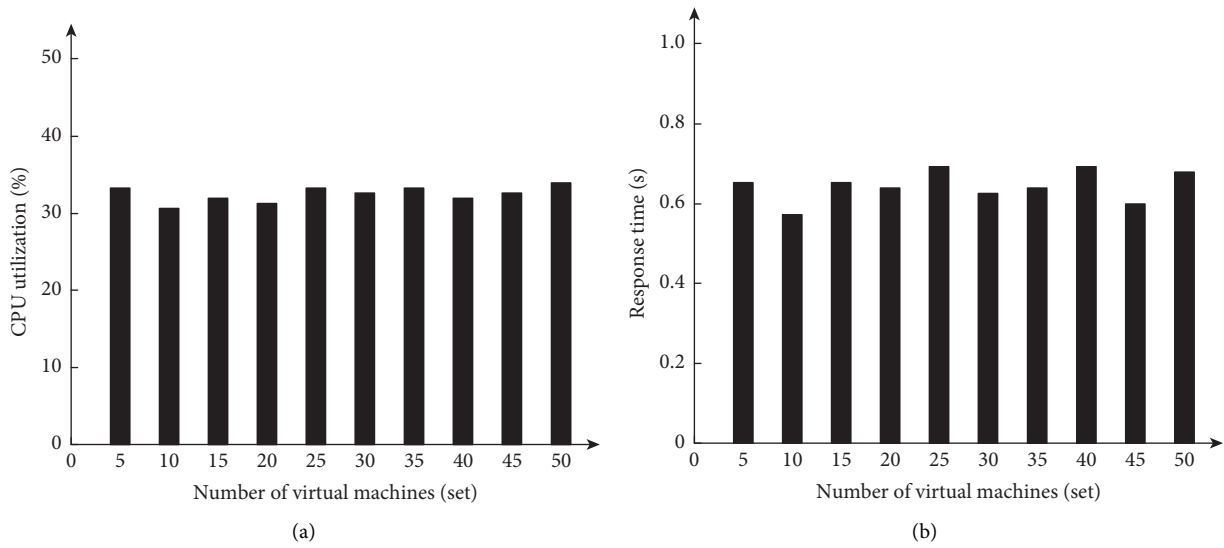
FIGURE 7: Load balancing test results. (a) Load balancing. (b) Response time.

by taking the storage synchronization as the index, and the results are shown in Table 2.

The calculation formula of bit error rate W is as follows:

$$W = \frac{M}{M_z} \times 100\%. \tag{15}$$

In the above formula, M and M_z, respectively, represent the error code and the total number of codes input in the process of power application data input.

Analysis of Table 2 shows that when the low occupancy big data input frequency gradually increases, the synchronization bit error rate of this method under the condition of different number of virtual machines shows an upward trend with the increase of data input frequency. When the number

of virtual machines is 10, the synchronization error rate of this method basically maintains a linear upward trend in the low occupancy big data input frequency. When the input data frequency reaches 100 Hz, the synchronization error rate of this method is 0.013%. When the number of virtual machines rises to 30, the synchronization error rate of the method in this paper is basically the same as that under the condition of 10 virtual machines. When the number of virtual machines increases to 50, the synchronization error rate of this method increases obviously, and when the data input frequency is less than 70 Hz, the synchronization error rate fluctuation of this method is small. When the data input frequency reaches more than 80 Hz, the synchronization error rate of this method increases rapidly. When the input

Table 2: Storage synchronization test results.

Low occupancy big data input frequency/Hz	Synchronous bit error rate/%		
	10 virtual machines	30 virtual machines	50 virtual machines
10	0.004	0.004	0.013
20	0.006	0.005	0.013
30	0.007	0.005	0.01
40	0.006	0.007	0.012
50	0.008	0.008	0.014
60	0.008	0.009	0.013
70	0.008	0.01	0.014
80	0.01	0.12	0.063
90	0.012	0.012	0.078
100	0.013	0.014	0.089

Table 3: Comparison results of storage capacity.

Storage time (s)	Storage capacity (W)	
	Before using this method	After adopting the method in this paper
10	372	456
20	349	461
30	361	459
40	377	452
50	360	448
60	355	451
70	373	450
80	348	446
90	380	453
100	370	460
110	352	452
120	389	457

data frequency reaches 100 Hz, the synchronization error rate of this method is 0.089%. This shows that the number of virtual machines and low occupancy big data input frequency have a significant impact on the synchronization bit error rate of this method. When the number of virtual machines is 30, it can ensure that the synchronization bit error rate is low and meets the application needs of power enterprises.

4.5. Storage Comparison. Storage capacity is one of the main indicators to verify the data storage method. Under the same network environment and hardware facility environment, compare the storage capacity of the research object in the low occupancy big data storage process before and after using this method. The results are shown in Table 3.

According to the analysis in Table 3, from the perspective of average storage capacity, the average storage capacity of the research object data before using this method is 365.50 W/s, and after using this method, the average storage capacity increases by 88.25 W/s. From the perspective of storage volume fluctuation, the data storage volume in the low occupancy big data storage process has obvious fluctuation before using this method. After using this method, the data storage volume in the low occupancy big data storage process is relatively stable. The experimental results

show that this method can effectively improve the storage capacity of large data with low occupancy and optimize the storage process.

5. Conclusion

High-precision classification of low occupancy data helps to achieve efficient network transmission and improve data utilization. Therefore, aiming at the problems of low efficiency and large space of intelligent classification of low occupancy big data, this paper proposes an intelligent classification method of low occupancy big data based on grid index. The main contents and process of this paper are as follows:

(1) Feature selection of big data is done by constructing intelligent classification model.

(2) Build a low occupancy big data intelligent classification platform based on grid index technology.

(3) The constructed depth confidence network model is trained, and the selected features are intelligently classified.

(4) Simulation results show that this method has the advantages of high classification efficiency and low occupancy in the intelligent classification of low occupancy big data based on grid index.

Data Availability

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

References

[1] X. Q. Yang and Y. Zhu, "Research on-line monitoring method of noise pollution in robot master-slave control system," *Computer Simulation*, vol. 38, no. 6, pp. 301–305, 2021.

[2] S. Salimi and A. Hammad, "Sensitivity analysis of probabilistic occupancy prediction model using big data," *Building and Environment*, vol. 172, Article ID 106729, 2020.

[3] S. A. Petersen, Z. Pourzolfaghar, I. Alloush, D. Ahlers, and M. Helfert, "Value-added services, virtual enterprises and data spaces inspired enterprise architecture for smart cities," in *Proceedings of the Collaborative Networks and Digital Transformation-20th IFIP WG 5.5 Working Conference on Virtual Enterprises, PRO-VE 2019*, Turin, Italy, September 2019.

[4] A. S. Cabezuelo, "Using open data repositories and geo-location to create value-added services for tourism," in *Proceedings of the 2020 24th International Conference Information Visualisation (IV)*, Melbourne, Australia, September 2020.

[5] V. Yesin, M. Karpinski, M. Yesina, V. Vilihura, and K. Warwas, "Hiding the source code of stored database programs," *Information*, vol. 11, no. 12, p. 576, 2020.

[6] W. Wang, T. Hong, N. Li, R. Q. Wang, and J. Chen, "Linking energy-cyber-physical systems with occupancy prediction and interpretation through wifi probe-based ensemble classification," *Applied Energy*, vol. 236, no. FEB.15, pp. 55–69, 2019.

[7] T. Khan and A. F. Mollah, "Text non-text classification based on area occupancy of equidistant pixels," *Procedia Computer Science*, vol. 167, 2020.

[8] G. Violatto and A. Pandharipande, "Anomaly classification in people counting and occupancy sensor systems," *IEEE Sensors Journal*, vol. 20, p. 1, 2020.

[9] G. Ravikumar, M. Govindarasu, S. Misra, S. Brahma, and R. Tourani, "Anomaly detection and mitigation for wide-area damping control using machine learning," *IEEE Transactions on Smart Grid*, vol. 18, no. 5, p. 1, 2020.

[10] M. Mohamma Dd Oust, A. Harounabadi, and M. Neizari, "Evaluation of grid computing environment using topsis," *The International arab Journal of Information Technology*, vol. 16, no. 3, pp. 323–331, 2019.

[11] D. Toshniwal and M. Dipasquale, "Counting the dimension of splines of mixed smoothness: a general recipe, and its application to planar meshes of arbitrary topologies," *Advances in Computational Mathematics*, vol. 47, no. 1, 2021.

[12] S. Zhu, D. Wang, L. Liu, W. Yan, and D. Guo, "Inferring region significance by using multi-source spatial data," *Neural Computing & Applications*, vol. 32, no. 22, 2020.

[13] Z. Rustam, J. Pandelaki, D. A. Utami, R. Hidayat, and A. A. Ramli, "Comparison support vector machine and fuzzy possibilistic c-means based on the kernel for knee osteoarthritis data classification," *International Journal of Advanced Science, Engineering and Information Technology*, vol. 9, no. 6, pp. 2142–2146, 2019.

[14] A. A. Ribeiro and M. Sachine, "On the optimal separating hyperplane for arbitrary sets: a generalization of the svm formulation and a convex hull approach," *Optimization*, vol. 8, no. 10, pp. 1–14, 2020.

[15] N. Ali, D. Neagu, and P. Trundle, "Evaluation of k-nearest neighbour classifier performance for heterogeneous data sets," *SN Applied Sciences*, vol. 1, no. 12, pp. 1–15, 2019.

[16] R. P. Monteiro, M. Cerrada, D. R. Cabrera, R. V. Sánchez, and C. J. A. Bastos-Filho, "Using a support vector machine based decision stage to improve the fault diagnosis on gearboxes," *Computational Intelligence and Neuroscience*, vol. 2019, Article ID 1383752, 13 pages, 2019.

[17] M. T. Sattari, H. Feizi, M. . Sevba olak et al., "Estimation of sodium adsorption ratio in a river with kernel-based and decision-tree models," *Environmental Monitoring and Assessment*, vol. 192, no. 9, p. 575, 2020.

[18] E. Hugues-Salas, F. Ntavou, D. Gkounis, G. T. Kanellos, R. Nejabati, and D. Simeonidou, "Monitoring and physical-layer attack mitigation in sdn-controlled quantum key distribution networks," *Journal of Optical Communications and Networking*, vol. 11, no. 2, p. A209, 2019.

[19] W. Huang, J. Fang, S. Wan, C. Xie, and X. He, "Design and evaluation of a risk-aware failure identification scheme for improved ras in erasure-coded data centers," *IEEE Transactions on Parallel and Distributed Systems*, vol. 32, 2020.

[20] S. S. Hameed, W. H. Hassan, L. Abdul Latiff, and F. Ghabban, "A systematic review of security and privacy issues in the internet of medical things; the role of machine learning approaches," *PeerJ Computer Science*, vol. 7, no. 4, p. e414, 2021.

[21] N. Sridhar and M. Kowsalya, "Enhancement of power management in micro grid system using adaptive alo technique," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, no. 2, pp. 2163–2182, 2021.

[22] J. Kiljander, D. Gabrijelcic, O. Werner-Kytola, A. Krpic, and M. Taumberger, "Residential flexibility management: a case study in distribution networks," *IEEE Access*, vol. 7, 2019.

[23] O. Khutsoane, B. Isong, N. Gasela, and A. M. Abu-Mahfouz, "Watergrid-sense: a lora-based sensor node for industrial iot applications," *IEEE Sensors Journal*, vol. 20, 2019.

[24] I. González, A. José Calderón, and J. María Portalo, "Innovative multi-layered architecture for heterogeneous automation and monitoring systems: application case of a photovoltaic smart microgrid," *Sustainability*, vol. 13, no. 4, p. 2234, 2021.

[25] Y. Kim, U. Shahab, and S. H. Bae, "Local augment: utilizing local bias property of convolutional neural networks for data augmentation," *IEEE Access*, vol. 9, 2021.

[26] G. S. Krishna and N. Prakash, "Deep learning for efficient and multi-labelled classification of synthetic aperture radar images," *Evolving Systems*, vol. 12, no. 1, pp. 1–14, 2021.

[27] W. Guo, W. Ge, X. Lu, and H. Li, "Short-term load forecasting of virtual machines based on improved neural network," *IEEE Access*, vol. 7, 2019.

[28] Q. Liu, S. Liu, G. Wang, and S. Xia, "Social relationship prediction across networks using tri-training bp neural networks," *Neurocomputing*, vol. 401, pp. 377–391, 2020.

[29] G. Wang, J. Qiao, J. Bi, Q. S. Jia, and M. C. Zhou, "An adaptive deep belief network with sparse restricted Boltzmann machines," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, pp. 1–12, 2019.

[30] M. Kakimoto, Y. Endoh, H. Shin, R. Ikeda, and H. Kusaka, "Probabilistic solar irradiance forecasting by conditioning joint probability method and its application to electric power trading," *IEEE Transactions on Sustainable Energy*, vol. 10, no. 2, pp. 983–993, 2019.

[31] K. Qiu and Q. Zhang, "The m-ducg methodology to calculate the joint probability distribution of directed cycle graph with local data and domain causal knowledge," *IEEE Access*, vol. 9, 2021.

[32] J. Cabessa and O. Finkel, "Computational capabilities of analog and evolving neural networks over infinite input streams," *Journal of Computer and System Sciences*, vol. 101, no. MAY, pp. 86–99, 2019.

[33] K. R. M. Fernando and C. P. Tsokos, "Dynamically weighted balanced loss: class imbalanced learning and confidence calibration of deep neural networks," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 14, no. 1, pp. 1–12, 2021.

[34] A. K. Seghouane, N. Shokouhi, and I. Koch, "Sparse principal component analysis with preserved sparsity pattern," *IEEE Transactions on Image Processing*, vol. 28, 2019.

[35] N. Gilbert, R. E. Mewis, and O. B. Sutcliffe, "Classification of fentanyl analogues through principal component analysis (PCA) and hierarchical clustering of GC-MS data," *Forensic Chemistry*, vol. 21, Article ID 100287, 2020.

Research Article

Spatial Transformer Network-Based Automatic Modulation Recognition of Blind Signals

Yuxin Huang (ID)

School of Electronic and Electrical Engineering, Minnan University of Science and Technology, Quanzhou 362700, Fujian, China

Correspondence should be addressed to Yuxin Huang; yiasom@163.com

Modulation recognition of communication signals plays an important role in both civil and military uses. Neural network-based modulation recognition methods can extract high-level abstract features which can be adopted for classification of modulation types. Compared with traditional recognition methods based on manually defined features, they have the advantage of higher recognition rate. However, in actual modulation recognition scenarios, due to inaccurate estimation of receiving parameters and other reasons, the input signal samples for modulation recognition may have large phase, frequency offsets, and time scale changes. Existing deep learning-based modulation recognition methods have not considered the influences brought by the above issues, thus resulting in a decreased recognition rate. A modulation recognition method based on the spatial transformation network is proposed in this paper. In the proposed network, some prior models for synchronization in communication are introduced, and the priori models are realized through the spatial transformation subnetwork, so as to reduce the influence of phase, frequency offsets, and time scale differences. Experiments on simulated datasets prove that compared with the traditional CNN, ResNet, and the CLDNN, the recognition rate of the proposed method has increased by 8.0%, 5.8%, and 4.6%, respectively, when the signal-to-noise ratio is greater than 0. Moreover, the proposed network is also easier to train. The training time required for convergence has reduced by 4.5% and 80.7% compared to the ResNet and CLDNN, respectively.

1. Introduction

Modulation recognition of communication signals plays an important role in both civil and military applications. In civil use, modulation recognition technology is the basis for both the communication parties to automatically adjust the modulation type according to the current channel conditions or transmission quality, i.e., adaptive communication. In military applications, especially when receiving signals blindly, it is often impossible to know the relevant information of the received signal in advance, especially the information of modulation type. However, for further processing, the modulation type should be known in advance in many military applications.

For current modulation recognition technologies, there are mainly two types of methods: traditional methods and deep learning-based methods. The two methods are introduced as follows. In traditional methods, the features of the

signal are manually defined, such as spectral characteristics, instantaneous feature statistics, high-order moments, high-order cumulants, and so on [1–3]. Then, classification models can be established according to classic classifiers, such as decision tree (DT), support vector machine (SVM), and so on. The advantage for the type of methods is that the manually defined features can have better theoretical support. Because they have clear physical meanings, the synchronization parameters of the signal can be analyzed and extracted in the process of modulation recognition. The shortcomings for this type of methods are mainly as follows: (1) there is a lack of generalization ability, which affects the extraction of features under different channel conditions, resulting in a decrease in accuracy; (2) when there are many modulation types for recognition, the methods will also lead to a decline in the recognition rate due to limited number of manual features. For methods based on deep learning, the features for modulation recognition are automatically

extracted through training sample adopting deep neural networks, which can effectively avoid the shortcomings of traditional methods and achieve a higher recognition rate. However, deep learning-based methods also have shortcomings, such as the lack of interpretability of features and the inability to estimate signal parameters, such as symbol rate, in the process of modulation recognition.

Artificial intelligence has been successfully applied in the field of image and natural language processing (NLP). As the modulation recognition problem can be transferred to an image recognition problem, deep learning-based modulation recognition has also become a research hot spot. The following publications have summarized the application of deep learning in modulation recognition. References [4–6] directly adopt baseband samples for modulation recognition, which assumes that the input contains the same number of symbols. The effects of different neural network structures on modulation recognition rates are compared. The authors in [7, 8] adopt the constellation diagram for modulation recognition after preprocessing. The pre-processing includes procedures such as sampling time synchronization and symbol rate synchronization. Note that there are blind estimation processes in the preprocessing, including symbol rate estimation, frequency offset estimation, and so on. The author in [2] directly assumes that under the condition of cooperative communication, the phase jitter has been eliminated, and the symbol synchronization has been completed. The convolutional neural network is then adopted for modulation recognition.

From the aforementioned modulation recognition methods, the influence of different symbol rates, frequencies, and phase offsets is eliminated through the receiving synchronization process under both cooperative communication condition and blind receiving condition. However, in real application, due to the inaccurate estimation of blind receiving parameters, the input signal samples for modulation recognition still have large phase and frequency offsets and different time scales. Existing deep learning modulation recognition methods have not taken the mentioned situations into consideration, which can lead to a decrease in the modulation accuracy. A novel modulation recognition method based on the spatial transformation network is proposed in this paper. In the network, prior models for synchronization in communication are introduced, and the priori models are realized through the spatial transformation subnetwork, which can reduce the phase and frequency offsets. The influence of different time scales or the number of symbols on modulation recognition can also be reduced. Through the simulation dataset generated adopting gnuradio [9, 10], the experiments are carried out. The difference between the proposed method and the spatial transformer-based method in [11] is threefold. (1) The structure of the parameter regression module is different. The paper has adopted both time and frequency-domain samples as input, which have better ability to extract features from both domains. (2) The paper has given more details about the spatial transformer-based model. (3) For the model, the training process has added the supervision of symbol rate according to the baseline symbol rate model. Therefore, the

proposed method has the ability for symbol rate estimation. Overall, the paper can be regarded as an improvement of [11], which has better parameter regression capability and ability for symbol rate estimation. The results show that in the presence of different symbol rates and different frequency offsets and phase offsets, the proposed method has a recognition rate of 8.3%, 4.9%, and 5.2% higher compared with the traditional CNN, ResNet, and CLDNN, but the training convergence time has reduced by 3.5%, 27%, and 85%, respectively.

2. Methods

The overall structure of the proposed method is shown in Figure 1. In the proposed network, the spatial transformation subnetwork is inserted into the traditional convolutional neural network, where other parts are similar to the traditional convolutional neural network. The structure of the spatial transformation subnetwork is also shown in Figure 1, which is mainly composed of three substructures: the parameter regression estimation module, the time compensation module, and the phase frequency offset compensation module. Among them, the parameter regression estimation module is composed of a few convolutional layers, the input of which is the feature extracted by the previous layer. The output of the last convolutional layer can output some parameter estimations. In our implementation, these parameters include time scaling parameters, frequency offset, and phase offset parameters. The dimension of the output parameters is 5, where the time compensation-related dimension is 3, and the phase frequency offset compensation-related dimension is 2. The design of output parameters is related to the subsequent parameter-based transformation model, which will be discussed in detail in the following section. The time compensation module adopts the time-related parameters to perform the corresponding transformations on the input samples, thereby compensating for problems introduced by different number of symbols in the signal sample. For the frequency and phase offsets, they are compensated according to the phase and frequency offsets estimations obtained by the parameter regression module. The compensated samples are then adopted to identify the modulation type. The following is a detailed discussion of the proposed spatial transformation-based method.

2.1. The Parameter Regression Module. The function of the parameter regression estimation module is to estimate the parameters for subsequent transformation. The estimated parameters can transform the samples accordingly to compensate for the time offset, scale changes, frequency offset, and phase offset of the input signal samples. In modulation recognition, the mentioned parameters have an impact on the accuracy of modulation recognition. The cost function of the recognition network also has a correlation with the mentioned parameters. Therefore, the parameter can be estimated through network training. In addition, in order to enable more direct extract of frequency-based
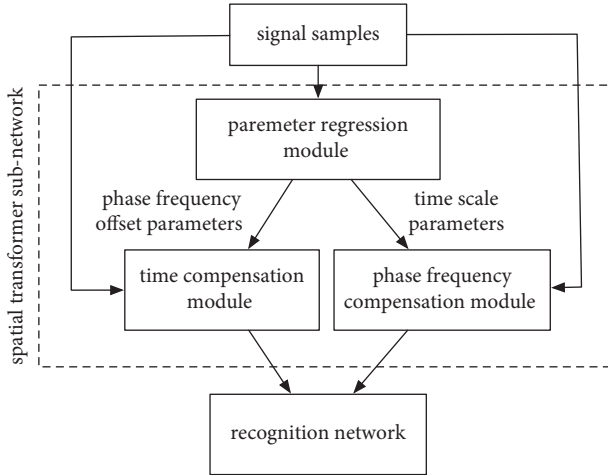
Figure 1: The overall structure of the method.

features, the frequency spectrum of the signal is also adopted as input to the network. This can better guide the network to learn the frequency-domain-based features, thereby avoiding the time-frequency domain conversion learning in the neural network. The structure of the parameter regression estimation module is shown in Figure 2. The time-domain and frequency-domain-based signals go through two feature extraction networks, respectively. Then, the obtained feature vectors are joined to form a larger feature vector. Another feature extraction network is added with input of the vector to obtain the estimation of the 5-dimension transformation parameters. Features A, B, and C are made of two convolutional layers and an all connected layer. In the model, there are in total 9 layers in depth. However, there are parallel layers in the model.

2.2. The Time Compensation Module.
In real scenario modulation recognition applications, the signal samples may have different time scales and frequency offsets. For the explanation convenience, it is assumed that the signal samples have been already converted to baseband and the signal-to-noise ratio is high. For neural network input, although the length of the signal samples or the number of sampling points are the same, the following situations may exist: (1) due to different SPS (samples per symbol), there are different numbers of symbols in the signal sample with the same number of sampling point; for example, if one signal sample has SPS twice as the other, the number of symbols is also twice; (2) the number of symbols in the signal samples is the same, but the position of the signal starting point is different. The difference can be regarded as the offset in time. The factors of time offset and different numbers of symbols in the signal samples may cause the decrease of the modulation recognition rate. Therefore, after the time offset and time scale transformation parameters are obtained through parameter regression, the corresponding model is adopted to transform the signal samples. The transformation can compensate the time scale and offset, thereby reducing the effects in recognition rate. The transformation herein is
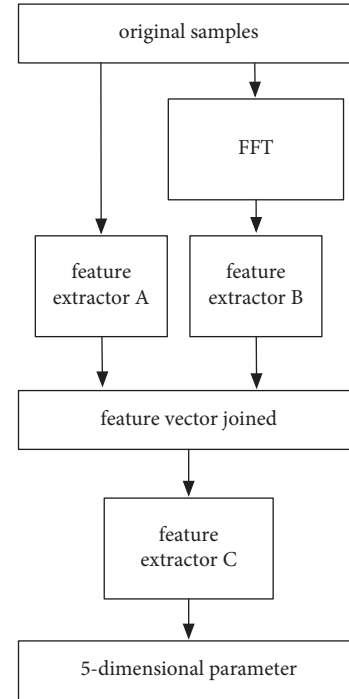
Figure 2: The structure of the parameter regression module.

based on the two-dimension affine transformation in the field of image processing [11, 12]:

$$\begin{bmatrix} x^s \\ y^s \end{bmatrix} = \begin{bmatrix} \theta_1 & \theta_2 & \theta_3 \\ \theta_4 & \theta_5 & \theta_6 \end{bmatrix} \begin{bmatrix} x^t \\ y^t \\ 1 \end{bmatrix}, \tag{1}$$

where x^s, y^s denote the pixel position of the original image, x^t, y^t denote the positions of the transformed image, and $\theta_1 \sim \theta_6$ are the transformation parameters. In the aforementioned transformation, translation, rotation, and scaling are all included. In our application, since there is only translation and scaling in time, the transformation model can be simplified as follows:

$$\begin{bmatrix} x^s \\ y^s \end{bmatrix} = \begin{bmatrix} \theta_1 & 0 & \theta_2 \\ 0 & \theta_3 & 0 \end{bmatrix} \begin{bmatrix} x^t \\ y^t \\ 1 \end{bmatrix}. \tag{2}$$

After simplification, only 3 parameters are adopted to for translation (representing the offset in time) and scaling (representing the change in scale due to the difference in symbol rate). Note that x^s, y^s denote the position of the input signal feature and x^t, y^t denote the position of the transformed signal feature. After the transformation, the corresponding original coordinate position may be a decimal number and can exceed the range of feature dimension. Therefore, in practical applications, it is also necessary to interpolate for the value of the feature in the following equation, and the value of at position l after interpolation can be written as

$$V_l^t = \sum_i^I \sum_j^J V_{ij}^s k\left(x_l^s - i; \Phi_x\right) k\left(y_l^s - j; \Phi_y\right), \tag{3}$$

where V_l^t represents the value at the corresponding position l after interpolation, I and J denote the dimensions of the features, V_{ij}^s denotes the value of the input feature at the position of ij, $k(.)$ represents the metric between two variables defined by the kernel function, Φ_x and Φ_y represent the corresponding kernel functions, and x_l^s and y_l^s denote the corresponding position of the input after the simplified affine transformation at position l. Note that here the position may be decimal. Generally speaking, if bilinear transformation is selected as the corresponding kernel function, the above equation can be written as

$$V_l^t = \sum_i^I \sum_j^J V_{ij}^s \max\left(0, \left|x_l^s - i\right|\right)\max\left(0, \left|y_l^s - j\right|\right). \tag{4}$$

The above equation can be regarded as the weighted average of the values near the position before the simplified affine transformation [13, 14]. Overall, the corresponding position can be obtained through affine transformation, and the value of the corresponding position can be obtained through bilinear interpolation. According to the mentioned processes, the feature output after time translation and scaling can be obtained. Figure 3 shows the processing flow of the time compensation module. It can be seen that the position of the output domain x^t, y^t is subjected to affine transformation according to the estimated parameters to obtain the position in the corresponding input domain x^s, y^s firstly. Then, the values at different positions in the output domain can be obtained through interpolation. Since the calculation of the bilinear transformation is differentiable, the time compensation module can be trained through the network.

For the training process, the supervision of symbol rate is added for enhancing parameter regression. As a matter of fact, the parameters of θ_1 and θ_3 represent the time scaling, which is symbol rate related. If the symbol rate is equal to the baseline symbol rate, then the parameter should be

$$\begin{aligned}\theta_1 &= 1, \\ \theta_3 &= 1.\end{aligned} \tag{5}$$

During the training, we added the following loss to the overall loss function:

$$\left|\theta_1 - \widehat{\theta}\right|^2 + \left|\theta_3 - \widehat{\theta}\right|^2, \tag{6}$$

where θ_1 and θ_3 represent the estimated parameters and $\widehat{\theta}$ represents the actual relative symbol rate. Note that for each input sample, the value of $\widehat{\theta}$ is different. Then, in the testing process, after the parameters θ_1 and θ_3 are acquired. The actual symbol rate can be estimated as

$$\frac{(\theta_1 + \theta_3) * \theta_{\text{base}}}{2}, \tag{7}$$

where θ_{base} denotes the baseline symbol rate signal.

2.3. The Phase Offset Compensation Module.
The phase and frequency offset compensation is more intuitive, and it is processed directly according to the following equation:
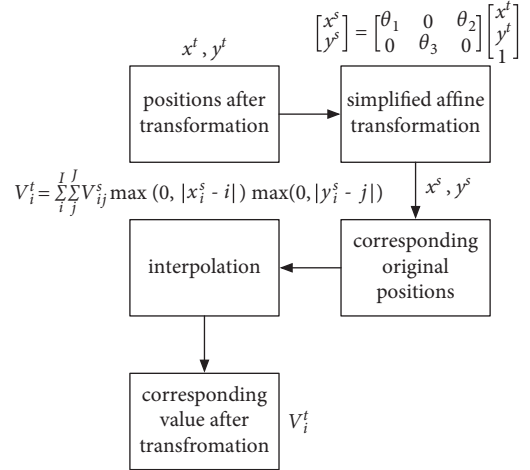
FIGURE 3: The processing flow of the time compensation module.

$$x_{\text{out}} = x_{\text{in}} * \exp\left((n\theta_4 + \theta_5)j\right). \tag{8}$$

Assuming that the processing of the above formula is complex value based, where x_{in} is the input and x_{out} is the output, the parameter θ_4 represents the frequency offset estimation, the parameter θ_5 represents the phase offset estimation, and n represents the time. In real implementation, since the input data are IQ time domain based, which can be regarded as the real and imaginary parts of complex values, the actual transformation can be written as

$$x_{\text{out}} = \left(x_{\text{in}}^I + jx_{\text{in}}^Q\right) * \left(\cos\left(n\theta_4 + \theta_5\right)\right) + j\sin\left(n\theta_4 + \theta_5\right). \tag{9}$$

The corresponding real part output is

$$x_{\text{out}}^I = x_{\text{in}}^I * \cos\left(n\theta_4 + \theta_5\right) - x_{\text{in}}^Q * \sin\left(n\theta_4 + \theta_5\right). \tag{10}$$

The corresponding imaginary part output is

$$x_{\text{out}}^Q = x_{\text{in}}^Q * \cos\left(n\theta_4 + \theta_5\right) + x_{\text{in}}^I * \sin\left(n\theta_4 + \theta_5\right). \tag{11}$$

As the transformation is also differentiable, the network can be trained. For the frequency compensation module, the number of n can be estimated. Adopting the estimation of n, the relative SPS can be estimated according to the original samples.

3. Experiments

In this paper, the spatial transformation network is adopted for modulation recognition, which takes into account different time scales, frequencies, and phase offsets in the model. The following describes the experimental process and results from the aspects of experimental dataset generation and method comparisons.

3.1. Dataset Generation.
In order to verify the effectiveness of the proposed method, the open-source software radio platform gnuradio [15] is adopted for generating the dataset. The generated dataset contains 11 different modulation

types, including digital modulation types for BPSK, QPSK, 8PSK, PAM4, QAM16, QAM64, and CPFSK and analog modulation types for GFSK, AM-DSB, AM-SSB, and FM. When generating the dataset, the methods in [9] are adopted for reference. The source for the dataset includes real text and audio sources. For generation of modulated signals, including BPSK, QPSK, 8PSK, PAM4, QAM16, and QAM64, a root raised cosine filter is adopted to shape the transmitted signal to obtain a baseband modulated signal. For the GFSK signal, Gaussian filter is adopted for shaping, and the analog frequency modulation signal is adopted to obtain the corresponding two frequency peaks. After the modulated signals are generated, they are truncated in time to generate signal samples, in order to obtain samples of the same length. The dimension of the sample is $2*128$, where 2 denotes the IQ channels and 128 denotes number of sampling points in time. In the process of dataset generation, different from the dataset in [9], the different frequency offset and time scale changes are added. For frequency offsets, the related frequency offset parameter is set in the channel function dynamic_channel_model in gnuradio. Figure 4 shows the normalized frequency comparison of the same 8PSK signal with and without frequency offset. It can be seen that the frequency offset of the baseband signal can be generated by the frequency offset parameter in the channel model. In practical applications, such frequency offsets are prevalent due to inaccurate signal detection. In our experiment, we have adopted the open-source RadioML dataset for modulation recognition. In the dataset, we have additional added different frequency offsets and time scales adopting the gnuradio software. In the software, there are multirate signal processing modules for decimation and interpolation.

The scale changes are generated by changing the SPS accordingly. In Figure 5, the signal sample for the 8PSK modulation type is shown, and the corresponding dimension of the sample is 2 * 128. In the figure, the upper one has set SPS to 4, and the bottom one has set the SPS to 6. It can be seen that the two signal samples contain different numbers of symbols, which are 6 and 4, respectively. This figure can intuitively show the problem of different time scales under the same modulation mode. The problems of different SPS are common in actual situations due to different signal bandwidths.

After the dataset is generated according to the above method, each signal sample is normalized according to the sample energy. Then, the dataset is split randomly into 50% for training and 50% for testing.

3.2. Method Comparisons. In order to fully illustrate the effectiveness of the proposed method, the recognition rate of the proposed method is compared with that of several classical neural network-based methods. The recognition rate of different methods under different signal-to-noise ratio conditions is shown in Figure 6. From the statistics in Table 1, it can be seen that when the signal-to-noise ratio is greater than 0, the recognition rate of the proposed method in this paper is 8.0%, 5.8%, and 4.6% higher than that of the traditional CNN, ResNet, and CLDNN, respectively [16].
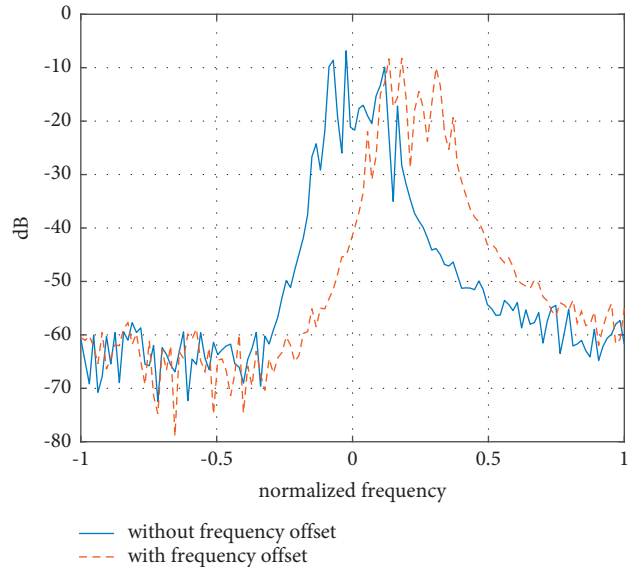
FIGURE 4: The comparison of the spectra with and without frequency offset for the 8PSK signal (noting that here the FFT is with 128 samples).
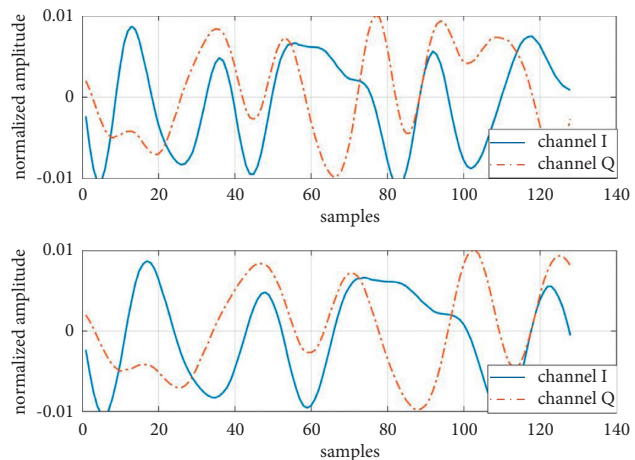
FIGURE 5: 8PSK modulated signal samples of different time scales (the SPS are 4 in the upper figure and 6 in the bottom figure).

Table 2 lists the comparisons of the total number of parameters and the training convergence time between the proposed method in this paper and the traditional CNN, ResNet, and CLDNN. It can be seen that, compared with the classic CNN, the proposed method has increased the number of network parameters by about 300% due to the addition of the parameter regression module, the time compensation module, and the phase frequency offset compensation module. However, as the proposed network structure is designed with a priori time scale change and frequency phase offset model, it is easier to train. The training time required for convergence is reduced by 4.5% and 80.7% compared to the ResNet and CLDNN, which has fully demonstrated the effectiveness of the proposed method.
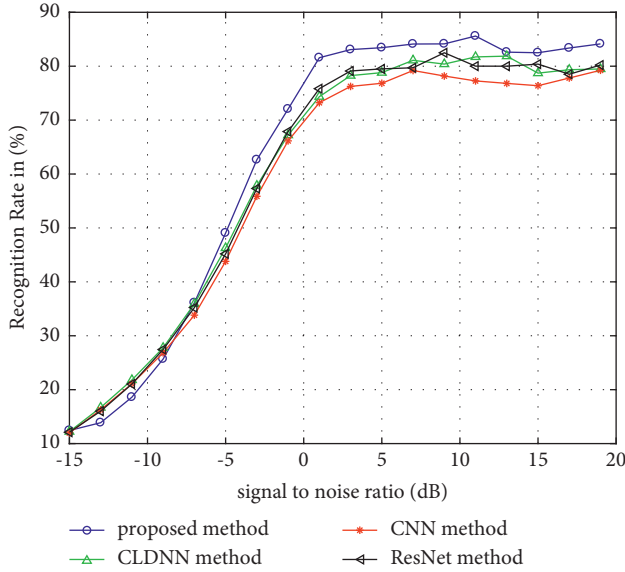
FIGURE 6: Comparison of correct recognition rates of different methods under different signal-to-noise ratio conditions.

TABLE 1: Comparisons of average recognition rate of different methods when SNR>0.

Method	Recognition rate (%)	Increase (%)
Proposed	84.5	—
CLDNN	80.8	4.6
CNN	78.2	8.0
ResNet	79.9	5.8

TABLE 2: Comparisons of the number of parameters and training time for different methods.

Method	Parameter number	Training time (s)
Proposed	4.9×10^6	319
CLDNN	5.8×10^5	1654
CNN	1.6×10^6	297
ResNet	1.8×10^6	334

3.3. SPS Estimation. As mentioned, from the estimation of n in the phase and frequency offset compensation module, the SPS of the signal sample can be estimated. The SPS estimation accuracy of the proposed method can reach 98.8%. This is another advantage over other deep neural network-based methods, which are not able to extract knowledge on SPS on the signal samples. Noting that as the estimated n can be a decimal number, the original estimated SPS can also be decimal. For the accuracy statistics, the estimated SPS are rounded. For the training process, the supervision of symbol rate is added for enhancing parameter regression. For the experiment, we have chosen the SPS 4 as the baseline symbol rate model. For both training and testing, the symbol rate ranges from 2 to 8. Then, ground truth parameter of θ_1 and θ_3 should be in the range of 0.5 to 2 accordingly.

4. Conclusions

Blind signal modulation recognition has great application potential in both civil and military uses. For real scenario modulation recognition applications, signals of the same modulation type may have encountered the effects of different time scales and frequency offsets. A modulation recognition method based on spatial transformation network is proposed in this paper. Compared with the classic CNN recognition network, a parameter regression estimation module, a time compensation module, and phase frequency offset compensation module are added. Among them, the parameter regression module can estimate the time scale transformation parameters (3 dimensions) and the frequency and phase offset parameters (2 dimensions). The time compensation module and the phase frequency offset compensation module can perform the corresponding compensational transformations on the original signal samples according to the estimated parameters. Through the open-source software radio gnuradio, the experimental dataset is generated. The dataset includes signal samples of 11 modulation types with different signal-to-noise ratios, different SPS, and different frequency offsets. Experiments adopting the generated dataset prove that compared with the traditional CNN, ResNet, and CLDNN, the recognition rate of the proposed method has increased by 8.0%, 5.8%, and 4.6%, respectively, when the signal-to-noise ratio is greater than 0. Moreover, the proposed network in this paper is easier to train, and the training time required for convergence has reduced by 4.5% and 80% compared with the ResNet and CLDNN, respectively.

Data Availability

The RadioML dataset is publicly available.

Conflicts of Interest

The author declares that there are no conflicts of interest.

Acknowledgments

References

[1] Z. Zhu and A. K. Nandi, *Automatic Modulation Classification: Principles, Algorithms and Applications*, Wiley Publishing, Hoboken, New Jersey, 2015.

[2] Z. K. Lei, *Research on Deep Learning Based Modulation Recognition Technologies*, University of Electronic Science and Technology of China, China, Sichuan, 2019.

[3] Y. Lin, Y Tu, Z Dou, and Z Wu, "The application of deep learning in communication signal modulation recognition," in *Proceedings of the 2017 IEEE/CIC International Conference on Communications in China (ICCC)*, October 2017.

[4] T. J. O'Shea, J. Corgan, and T. C. Clancy, "Convolutional radio modulation recognition networks," Edited by C. Jayne and L. Iliadis, Eds., in *Proceedings of the Engineering Applications of Neural Networks. EANN 2016*, vol. 629Cham, September 2016.

[5] N. E. West and T. O'Shea, "Deep architectures for modulation recognition," in *Proceedings of the 2017 IEEE International Symposium on Dynamic Spectrum Access Networks (DySPAN)*, pp. 1–6, Baltimore, MD, USA, March 2017.

[6] T. J. O'Shea, N. West, M. Vondal, and T. C. Clancy, "Semi-supervised radio signal identification," in *Proceedings of the 2017 19th International Conference on Advanced Communication Technology (ICACT)*, pp. 33–38, PyeongChang, Korea (South), February 2017.

[7] C. Yang, Z. He, Y. Peng, Y. Wang, and J. Yang, "Deep learning aided method for automatic modulation recognition," *IEEE Access*, vol. 7, Article ID 109063, 2019.

[8] K. Karra, K. Scott, and J. Petersen, "Modulation recognition using hierarchical deep neural networks," in *Proceedings of the 2017 IEEE International Symposium on Dynamic Spectrum Access Networks (DySPAN)*, March 2017.

[9] T J. O'Shea and N. West, "Radio machine learning dataset generation with GNU radio," *Proceedings of the GNU Radio Conference, [S.l.*vol. 1, no. 1, pp. 1–5, 2016.

[10] R. Gandhiraj and K. P. Soman, "Modern analog and digital communication systems development using GNU Radio with USRP," *Telecommunication Systems*, vol. 56, no. 3, pp. 367–381, 2014.

[11] T. J. O'Shea, L. Pemula, D. Batra, and T. C. Clancy, "Radio transformer networks: attention models for learning to synchronize in wireless systems," in *Proceedings of the 2016 50th Asilomar Conference on Signals, Systems and Computers*, pp. 662–666, Pacific Grove, CA, November 2016.

[12] E. W. Weisstein, "Affine transformation," 2004, https://mathworld. wolfram. com.

[13] M. Jaderberg, K. Simonyan, A. Zisserman, and K. Kavukcuoglu, "Spatial transformer networks," in *Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 2 (NIPS'15)*, MIT Press, Cambridge, MA, USA, December 2015, https://arxiv.org/abs/1506.02025.

[14] L. Tan, Z. Li, and Q. Yu, "Deep Face Attributes Recognition Using Spatial Transformer Network," in *Proceedings of the 2016 IEEE International Conference on Information and Automation (ICIA)*, pp. 1928–1932, Ningbo, August 2016.

[15] *Gnuradio Software. Website*, https://www.gnuradio.org.

[16] T. Zhang, C. Shuai, and Y. Zhou, "Deep learning for robust automatic modulation recognition method for IoT applications," *IEEE Access*, vol. 8, Article ID 117689, 2020.

Retraction

Retracted: Mind Map Construction for English Grammar Teaching Based on Knowledge Map

Scientific Programming

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] Q. Fu and S. Kuang, "Mind Map Construction for English Grammar Teaching Based on Knowledge Map," *Scientific Programming*, vol. 2021, Article ID 4921825, 10 pages, 2021.

Research Article

Mind Map Construction for English Grammar Teaching Based on Knowledge Map

Qijun Fu and Shouzhong Kuang [ID]

General Education School, Hunan University of Information and Technology, Changsha 410151, Hunan, China

Correspondence should be addressed to Shouzhong Kuang; kuangshouzhong@nudt.edu.cn

The analysis of the frontier issues of the English language teaching method in China is of great guidance for English language teaching. Based on the ontology model of English teaching domain, the knowledge map of English teaching in colleges and universities is constructed by fusing heterogeneous English subject data from multiple sources. Firstly, we obtain domain knowledge from relevant websites and existing documents through web crawlers and other techniques and clean the data based on BERT model; then, we use Word2Vec to judge the similarity between the research directions of characters and solve the entity alignment problem; based on the scientific knowledge map theory, we count the frequency of keywords in each year and analyze them to describe the association and union between keywords. It can explain the current situation and trend, rise and fall, disciplinary growth points, and breakthroughs of ELT. Through keyword analysis, the hot issues mainly revolve around ELT, English teaching, college English, grammar-translation method, curriculum reform, and so forth, to realize the quick query and resource statistics of ELT basic data, in order to promote the subsequent English discipline assessment work to be completed more efficiently.

1. Introduction

Since the birth of English in the 17th century, discussions, reforms, and researches on teaching methods have been in full swing [1]. At the same time, English teaching theory has been greatly enriched. Since the reform and opening up, the situation of English teaching in China has undergone great changes [2]. The traditional grammar-translation method has been broken, and some new teaching methods from abroad have been introduced and brought in, which has injected vitality into English teaching. English teaching workers have been actively involved in the reform, research, and practice of English teaching methods, and the face of English teaching has become a new one. However, many English teachers also reflect that while they are encouraged by the good situation of English teaching since the reform and opening up, they are always "circling" around the foreign teaching methods, resulting in, to some extent, some misunderstandings in the research and practice of English

teaching methods. In order to make the reform of English teaching to get out of the misunderstanding and go further and really receive effective results, it is necessary to make a serious analysis and comparison of English teaching methods, so as to provide scientific guidance for English teaching [3]. Knowledge mapping is an emerging research field developed on the basis of citation analysis theory and information visualization technology. It displays the development process and structural relationship of scientific knowledge in a visual way by visualizing the complex scientific knowledge field through data mining, information processing, knowledge measurement, and graphical drawing, reveals scientific knowledge and its activity law, and shows the knowledge structure relationship and evolution law. Through scientific knowledge mapping, the authors, keywords, abstracts, and references contained in the literature are analyzed in two-dimensional horizontal and vertical time, and the research development path of the field and its frontier hot issues are visually analyzed [4].

In recent years, the problem of information visualization has received more and more attention [5]. People want to analyze a large amount of data at a deeper level and have no way to do so, so people want to analyze it at a deeper level in order to make better use of it. Knowledge mapping can graphically display the overall image, affinity, and evolution law of frontier fields that are difficult to obtain by personal experience alone and has become an important tool for grasping the development dynamics of disciplines, the direction of disciplinary research, and assisting scientific and technological decision-making. The study in [6] introduced CiteSpace to China, which has rapidly created a boom in related research. CiteSpace is a Java application for identifying and visualizing new trends and developments in scientific research in the scientific literature and has become an influential information visualization software application in the field of information analysis.

Competition among universities is mainly based on discipline competition, and the strength of a discipline can represent the level of the institution to some extent. The assessment of disciplines can help in effectively and comprehensively understanding the current status of discipline construction, and, through the correct assessment of disciplines, problems in construction can be identified, so as to further clarify the direction of the discipline and achieve better development [7]. Since the results of discipline construction involve many aspects, storing and displaying information about the discipline in the form of scattered documents and web resources cannot show the correlation between all data comprehensively, and it is difficult to dig out the information statistics and potential relationships, which is not conducive to the subsequent evaluation work.

As a new and efficient knowledge organization method in the era of big data, knowledge map can fuse and correlate heterogeneous data from multiple sources based on graphs [8]. In this paper, we apply the knowledge map technology to the field of English teaching in colleges and universities; firstly, we obtain the domain knowledge related to English teaching from the resource-rich data sources such as knowledge networks, university official websites, and discipline assessment documents through the method of web crawlers and rule mapping. With the possible problem of impurity data, a fine-tuned BERT (Bidirectional Encoder Representations from Transformers) model is used to classify the data. By searching the keywords of ELT and the frequency of occurrence by year and analyzing them, the association and combination between the keywords are described, which can explain the current status and trends, the rise and fall of ELT research, the growth points, and breakthroughs of the discipline. Visualization mapping can clearly show the changes of ELT method in a certain time period and the process of its change.

2. Related Knowledge and Research Ideas

The concept of knowledge map was first introduced by Google in 2012, with the aim of improving the quality of search engine results and enhancing the user search experience. According to the different coverage, the knowledge map can be divided into general knowledge map and domain knowledge map. Among them, the general knowledge map has a wider coverage, covering a lot of common knowledge in the real world, and some well-known large-scale general knowledge maps are DBpedia, Wikidata, Freebase, and so forth [9, 10]. These knowledge maps are very large in scale, but the quality of extracted knowledge is not strict, and the structure of the knowledge of each domain is simple, so they do not perform well when applied to specific domains. Domain knowledge maps are built for specific domains and have very strict requirements on the accuracy and depth of knowledge in the domain and can provide good support for the upper layer applications in the target domain. Knowledge maps have been used in medical, e-commerce, and legal fields, such as chatting bots based on knowledge maps for users to learn about healthcare and drugs [11] and designing inference rules to provide reference for sentencing in similar cases based on the constructed knowledge maps of legal documents of theft cases [12].

The knowledge map model is based on the graph structure $G=(V.E)$ in graph theory, where V is the set of vertices and E is the set of edges. The knowledge map can be perceived as a factual knowledge, which can be represented as a triple (H, r, t), where h is the head entity, t is the tail entity, and r is the relationship between the two entities. In constructing the knowledge map, there are two main ways to construct the knowledge map, top-down and bottom-up. The top-down approach refers to extracting the relevant ontology and pattern information directly from the high-quality dataset, while the bottom-up approach refers to extracting the resource patterns from the collected large amount of data and then selecting the ones with high confidence as the basis for the subsequent knowledge map construction [13]. For some more mature domains with complete knowledge systems, the top-down approach is usually adopted; that is, the schema ontology is defined first, and then knowledge is extracted using supervised, semi-supervised, and unsupervised methods, and finally the domain knowledge map is improved by combining knowledge fusion and knowledge inference mechanisms.

The general construction process of knowledge map is as follows: firstly, we determine the knowledge representation model; then we select different technical means to acquire knowledge according to different sources of data and import it into the knowledge map database; then we make comprehensive use of knowledge fusion, knowledge inference, and knowledge mining technologies to improve the scale and quality of the constructed knowledge map; finally, we design effective knowledge access and presentation channels according to different requirements of target scenarios, as shown in Figure 1, such as human-computer interaction and Q&A, graph visualization and analysis, and similar recommendation.

3. ELT Ontology Construction

Ontology defines the class set, relationship set, attribute set, and so forth of the knowledge map, which mainly emphasizes the relationship between concepts and is the
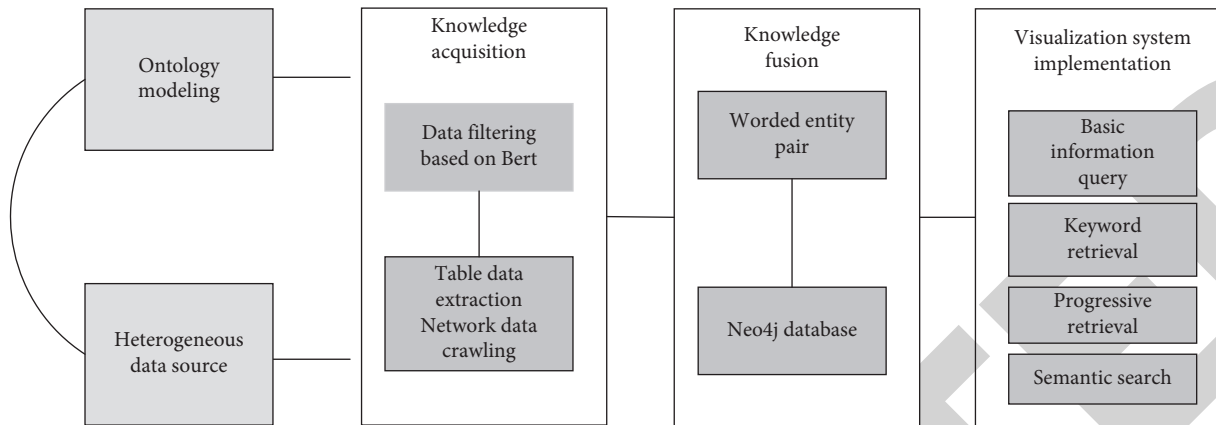
FIGURE 1: Overall build process.

management of the schema layer of the knowledge map. By constructing an ontology model, entities, relationships, and entity attributes can be constrained and standardized as a guide for subsequent knowledge extraction and organization [14]. In this paper, we use the ELT Computer Science and Technology 4th round discipline assessment brief as the main knowledge source, combine with specific ELT domain related websites, use OWL language as the ontology description language, and use Protégé ontology development tool to complete the construction of ELT ontology in universities.

The concepts included in the ELT ontology and the structure of their relationships are shown in Figure 2 using the OntoGraf tool in Protégé. In this ontology model, there are 10 categories: teachers, alumni, students, foreign students, institutions, national projects, provincial projects, journal papers, conference papers, and patents, and the subcategories are related to each other through various relationships. The ontology represents the relationships between concepts as semantic relationships, which are also called object attributes in Protégé, including generic semantic relationships and custom semantic relationships [15]. The ontology constructed in this paper contains a variety of custom semantic relations, and the related concepts and their detailed descriptions are shown in Table 1.

4. ELT Knowledge Mapping Construction

4.1. Knowledge Acquisition. In the process of knowledge map construction, data is a very important underlying support, and only by obtaining a large amount of data in the research domain can we build a good quality knowledge map. Generally, the knowledge sources used to build the knowledge map can be structured data, semistructured data, unstructured data, IoT sensors, and artificial crowdsourcing [16]. It is found that the data in the field of English teaching in universities are mainly distributed in electronic documents and various websites, such as subject evaluation documents, university official websites, and national knowledge infrastructures, which cover different types of subject area data, including teachers' information, papers, patents, and research projects, respectively. Therefore, this paper mainly obtains domain knowledge from the sources shown in Table 2.

For the structured data stored in the form of table documents, such as the English teaching assessment profile, a mapping-based information extraction method can be used; that is, first establish a one-to-one mapping between the table header fields to be extracted and the data attributes in the subject ontology constructed above, and then use the vocabulary defined in the ontology to describe the extracted structured information, thus preventing the occurrence of synonymy between attribute names; complete the extraction of data from the target table cell.

In order to crawl the data stored in the Internet web pages, the content organization of different web pages varies greatly, so it is necessary to develop specific crawling methods according to different target websites. Requests downloads a web page through an initial URL, parses the content of the tags contained in it with a web page parsing library, and obtains a new URL to crawl in turn [7]. Selenium, on the other hand, runs directly in the browser by simulating user actions, such as clicking buttons and typing text, to achieve the correct jumping between web pages [17]. The different implementation principles also determine the advantages and disadvantages of each type of crawler and their respective scenarios: Requests is fast, but the crawl is interrupted when the URL of the jumped page is not available, so it is suitable when the target URL is available; when the target URL is not available directly, Selenium can be used for page jumping, but it has the disadvantage that it needs to wait for the browser to open and load. When the target URL is not available, Selenium can be used for page jumping.

In this paper, we propose a web crawler algorithm that can flexibly invoke the above two tools according to the different forms of web page organization to obtain the target data while improving the crawling efficiency as much as possible. The specific crawler workflow is shown in Figure 3.

The algorithm needs to determine the organization of the jump URLs on the web page after the crawling starts. For example, for the official website of a university, the faculty list page usually contains the URL of the faculty details page, so we can crawl it by the following steps: (1) Starting from
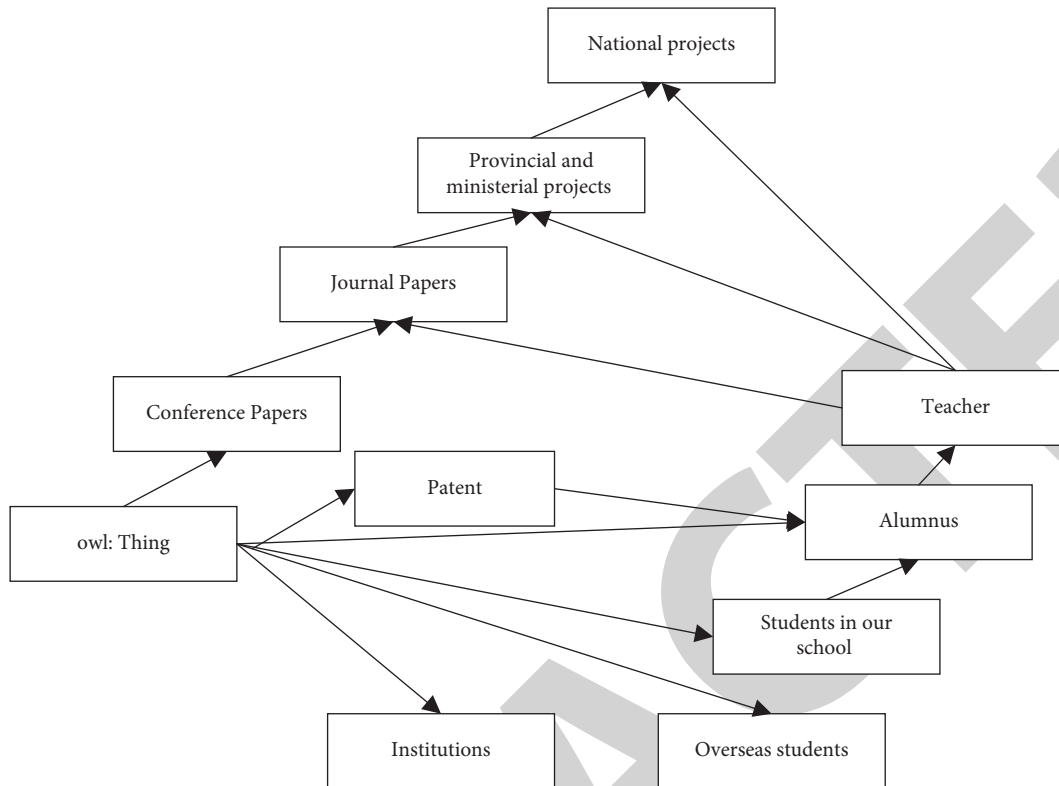
Figure 2: Ontology model of subject knowledge mapping.

Table 1: Table of custom semantic relations.

Association concept	Semantic relation	Associated concept	Semantic relation description
Teacher	Teach in	Institutions	Teach at
Teacher	Graduated from	Institutions	Graduated from
Alumnus	Graduated from	Institutions	Graduated from
Students in our school	Study in	Institutions	Study in
Overseas students	Study in	Institutions	Study in
Teacher	Author in	Journal papers	Publish one's thesis
Teacher	Author in	Conference papers	Publish one's thesis
Teacher	Lead	National projects	Host the project
Teacher	Lead	Provincial and ministerial projects	Host the project
Teacher	Participate in	National projects	Participation project
Teacher	Participate in	Provincial and ministerial projects	Participation project
Teacher	Inventor in	Patent	Invention patent

Table 2: Statistics of various data volumes before and after cleaning.

Data category	Before cleaning	After cleaning
Paper	3297	3208
Patent	1189	1150
Teacher	87	89
Institutions	35	36

the URL of the faculty list page, we obtain the content of the page through Requests library. (2) According to the defined page extraction rules, we extract the URL of the faculty details page and put it into the URL queue to be crawled. If the URL is incomplete, the missing fields are added according to the URL construction of similar pages. (3) Download the details page according to the queue of URLs to be crawled, extract the target data from it, and save it to the data storage file. (4) The whole process is executed cyclically until all URLs in the queue are crawled [18]. For some websites that cannot get the URL of the jump page directly, such as China Knowledge Network, the Selenium tool can be chosen to crawl, and the process is as follows: (1) Configure the URL address and related parameters, and call Selenium's web driver to open the browser page. (2) Wait for the page to finish loading, locate the search box and button
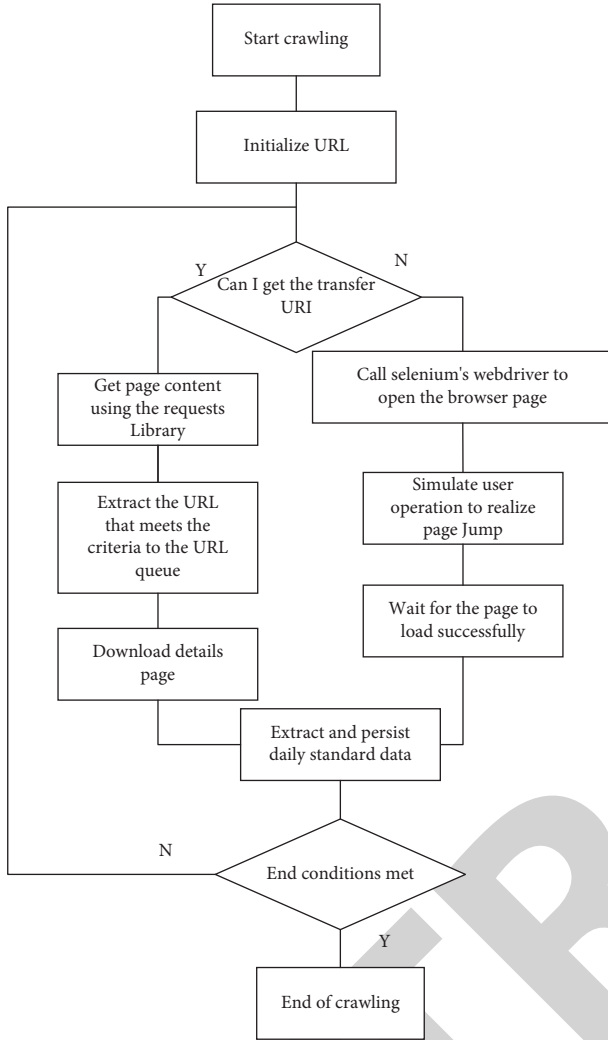
Figure 3: Crawler workflow.

elements, and, after completing the input of search conditions, simulate. (3) After the page is loaded successfully, use XPath to extract the target data and perform the data persistence operation. (4) Repeat the previously mentioned process until the number of crawlers is satisfied or all pages are crawled.

4.2. Knowledge Integration. When fusing knowledge from different sources, the problem of instance heterogeneity may arise; that is, entities with the same name may point to different objects, while entities with different names may point to the same object. Therefore, if so, construct corresponding alignment relationships between the entities to complete the knowledge fusion. In the process of constructing the knowledge map, the ambiguity of characters may arise in the process of collecting data related to English teaching in universities from data sources such as Knowledge Network and SooPAT. For example, the research results published by university teachers at different time points, such as papers and invention patents, are determined as different person

entities due to job transfer, or the teachers with the same name in the same university are incorrectly pointed to the same entity, resulting in incorrect statistics of research results. Therefore, in order to build an accurate knowledge map of English teaching in universities, we need to design a suitable entity alignment algorithm to solve the above problems.

The algorithm firstly extracts the renamed characters from multiple data sources to obtain the set of entities to be aligned; then, the basic information of the characters, including gender, ethnicity, date of birth, and other attributes that cannot be easily changed, is used for preliminary screening; finally, based on the set of keywords in the published papers or patent applications of the characters, the corresponding word vectors are obtained using Word2Vec and the cosine similarity between the word vectors is calculated [19]. If the similarity exceeds a defined threshold, it can be considered that both of them have the same research direction and refer to the same entity.

In order to determine the similarity threshold, the following experiment is designed to investigate. Firstly, we select some university teachers' papers as the original data, and each teacher randomly selects 3 papers' keywords to form his or her research keyword set, and suppose that the length of a teacher's research keyword set is m; then the set can be expressed as

$$\{K_{s1}, K_{s2}, \ldots, K_{sm}\}. \tag{1}$$

The remaining papers of the faculty member are then compared with this set, and the set of keywords for comparison is assumed to be n if the remaining papers contain n keywords:

$$\{K_{t1}, K_{t2}, \ldots, K_{tm}\}. \tag{2}$$

After that, the Word2Vec model is used to obtain the word vector of the keyword set, and the word vector of the keyword set in the research direction is represented as

$$\{K_{s1}, K_{s2}, \ldots, K_{sm}\}. \tag{3}$$

The word vector of the set of contrasted keywords is represented as

$$\{K_{t1}, K_{t2}, \ldots, K_{tn}\}. \tag{4}$$

Finally, the mean value of the cosine of the word vector between the two sets of keywords is calculated as the similarity between the paper and the corresponding faculty research direction:

$$\text{similarity} = \frac{\sum_{j=1}^{n} \sum_{k=1}^{m} \cos\left(V_{sk}, V_{tj}\right)}{n \times m}. \tag{5}$$

The cosine function $\cos(\bullet)$ between two word vectors is defined as

$$\cos\left(V_s, V_t\right) = \frac{V_s \cdot V_t}{\|V_s\| \times \|V_t\|} = \frac{\sum_{i=1}^{L} V_s^i V_t^i}{\sqrt{\sum_{i=1}^{L} \left(V_s^i\right)^2} \sqrt{\sum_{i=1}^{L} \left(V_t^i\right)^2}}, \tag{6}$$

where L is the dimension of the word vector obtained by Word2Vec and V_i is the i-th component of the word vector.

In this paper, a total of 2400 sets of test data were randomly selected, and the final distribution of the keyword similarity values is shown in Figure 4. From Figure 4, we can see that the keyword similarity of papers with the same research direction is above 0.5, so the similarity threshold is set to 0.5 in the entity alignment algorithm.

In order to verify the feasibility of the algorithm, several teachers with the same name but different research directions were selected and their published papers were crawled from the Internet, and the set of keywords of papers of the same teacher was taken as the positive data and the set of keywords of papers of different teachers was taken as the negative data. Then 200, 400, 600, and 800 pieces of data were randomly selected and the accuracy was analyzed and calculated with the results of manual annotation [20]. The experimental results are shown in Table 3. The accuracy rates of the four random tests are above 90%, which indicates that the Word2Vec-based character entity alignment method identifies less erroneous data and can be used in the knowledge fusion scenario of university subject areas.

4.3. Knowledge Storage. After cleaning and aligning the data, its content and format have met the requirements of subject knowledge map construction, and the next step is to import the data into the underlying database. Neo4j is a high-performance nonrelational graph database that stores data on a very large network, which is very suitable for storing knowledge maps based on graph structures [21]. In this paper, we import various types of data into Neo4j in the form of nodes and edges by using the operations provided by the Py2Neo third-party library supported by Python and can perform corresponding operations such as adding, deleting, and checking.

The data scale statistics of the finalized subject knowledge map are shown in Table 4. The various types of knowledge in the map for a large and complex multi-relationship network are helpful for the subsequent implementation of various functions and performance optimization.

5. Visualization System Implementation

In this paper, we develop a visualization system for teaching English in higher education based on the above knowledge mapping, which is implemented in a B/S (Browser/Server) front-end and back-end model and built with Python's Flask framework. In the front end, ECharts tool is used to visualize the data [22], and the subject domain knowledge is visualized in various forms such as text and force-oriented diagrams.

5.1. System Functions. The functions of this visualization system mainly include basic information query, keyword search, and progressive search and semantic search, which can complete the search and display of knowledge from
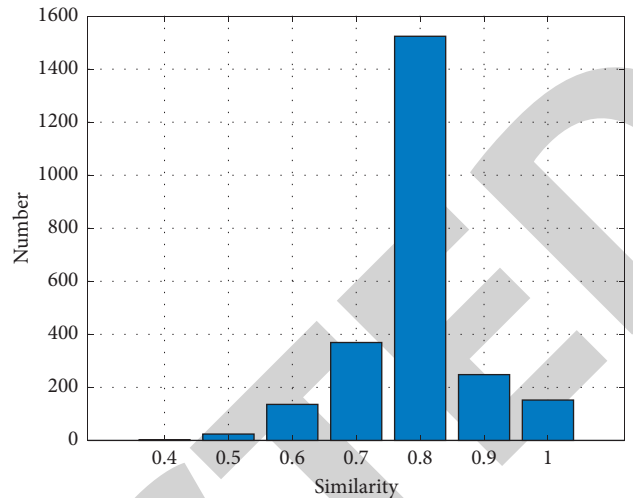
Figure 4: Keyword similarity distribution.

Table 3: Character entity alignment test results.

Number of data pieces	Number of correct classifications	Accuracy (%)
200	189	94.50
400	367	91.75
600	560	93.33
800	741	92.63

Table 4: ELT knowledge mapping data statistics.

Element	Type	Total quantity
Entity	10	3504
Relationship	24	4668
Attribute	39	22584

multiple dimensions such as entities, attributes, and relationships.

The purpose of the basic information query function is to count all entities and relationships related to the queried entity and then represent the entity relationships in the form of a force-oriented graph through a graphical interface. At the same time, a recommendation algorithm conforming to the storage structure of the graph database is used to select some similar entities with the highest similarity to the queried entity as recommendations that may be of interest to the user.

This function consists of two main data processing modules: direct query and similar recommendation. In the direct query module, the corresponding matching paths are constructed based on the user input [5], and then all the related entities and their relationships are found from the Neo4j graph database by Cypher statements. In the similar recommendation module [23], we first construct the multihop matching path "(qe:Stype)-[r1]-(e)-[r2]-(me:Stype)," where qe refers to the queried entity, me refers to the matched entity, Stype indicates that they are of the same data type, and r1, r2, and e represent the relationships and entities that do not make specific requirements. After that, all the matched entities and their

corresponding paths are counted and ranked in descending order, and the top-k entities are selected as similar recommendations (the value of k in this paper is 3; i.e., at most 3 similar entities are recommended). Finally, the node and link types and label values are determined from the obtained data attribute values and passed to the drawing function of ECharts to complete the drawing and display of the graph.

Figure 5 shows the results of the information search by entering "English teaching," which includes a description of the nodes of the various entities directly related to this entity and the relationships between them and also suggests the most relevant entities of the same type for the user: "English test," "English teachers," and "New Oriental." The force-oriented diagram supports zooming in and out and moving the graph, and when clicking on the category tabs at the top of the interface, all entity nodes in that category can be hidden or reproduced, making it easy for users to observe and count.

The keyword search function displays all entity nodes related to the input keywords and supports the task of multikeyword search. The system first uses the HIT LTP language processing tool to annotate the user input keywords with lexical types, including people, time, and nouns, and then constructs the corresponding regular expressions based on the lexical distribution to find the eligible entities from the knowledge map [24].

For example, when multiple keywords are entered as "neural network," "recognition," and "2019," the LTP lexical annotation module labels them as "n," "v," and "nt," and the corresponding regular expressions are "(? =. * [god][warp][net][network]). *," "(? =. * [knowledge][other]). *," and "(? =. * [2][0] [1], [8]). *." These regular expressions are then used as attribute fields to form a Cypher statement to retrieve the entities that satisfy the conditions, and the final result is shown in Figure 6.

Semantic search can be used to match the user's real intention by mining the semantic meaning behind the input question. In the process of semantic search, the input questions and defined question templates are first divided by using the LTP word division tool, which usually may overdivide the entities and concepts in the proprietary domain [25] and cause troubles in the subsequent search. Then, each question template and the input question are combined to perform one-hot coding to obtain the word vector representation of this template and the input question, and the highest similarity is selected as the type of the input question by calculating the cosine similarity between the two word vectors. Finally, the relevant data based on the template and the question keywords are returned to the front-end interface for integration and display.

Figure 7 shows the search results of the question "the coauthored journal papers of Ms. Zheng Qiumei and Ms. Huang Tingpei," and the interface shows the coauthored journal papers and the collaboration between the two teachers through graphical drawings, which achieves the goal of semantic search [26, 27].
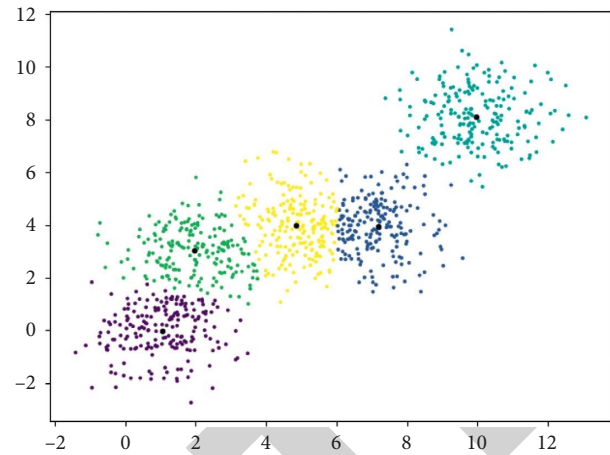
Figure 5: Results for "English Language Teaching."

5.2. System Performance Evaluation. In order to verify whether the performance of the system can meet the requirements of users, a dozen of English teaching staffs were invited to test the system after it was built. The testers participating in the testing process were divided into two groups according to the different testing methods: one group adopted the α-testing method; that is, the participants were given certain instructions and instructions on how to use the system, mainly to verify the reliability of the visualization system; the other group adopted the β-testing method, in which the participants were allowed to explore the functions of the system on their own without any guidance and help, mainly to verify the robustness and ease of use of the system. The overall feedback from the α-testing group was that the system had a wide range of data coverage and was user-friendly and reliable, while the overall feedback from the β-testing group was that the system was easy to use and operate, and no anomalies were generated. As shown in Figure 8, the average satisfaction rate of all test participants was 91.67%.

Chinese foreign language teaching has made remarkable achievements, but the traditional English major training mode aims to create tool-oriented talents, and such English talents can no longer meet the demand for English talents in today's social development. Therefore, the curriculum of English majors must be reformed, and the new training objectives and training mode should be based on the cultivation of complex English talents with innovative quality. A foreign language university in northeast China has reformed the CBI curriculum for college English majors. Content-dependent teaching reform is guided by the content-dependent teaching theory, and a curriculum system that integrates content-dependent courses and skill-based course in the basic stage of English majors has been built. After the curriculum reform, students' competencies were significantly improved, and experimental studies showed that content-dependent teaching achieved better results in terms of language knowledge teaching, language skill development, and subject knowledge transfer compared to language skill-oriented teaching and better achieved the overall teaching objectives specified in the national syllabus [1].
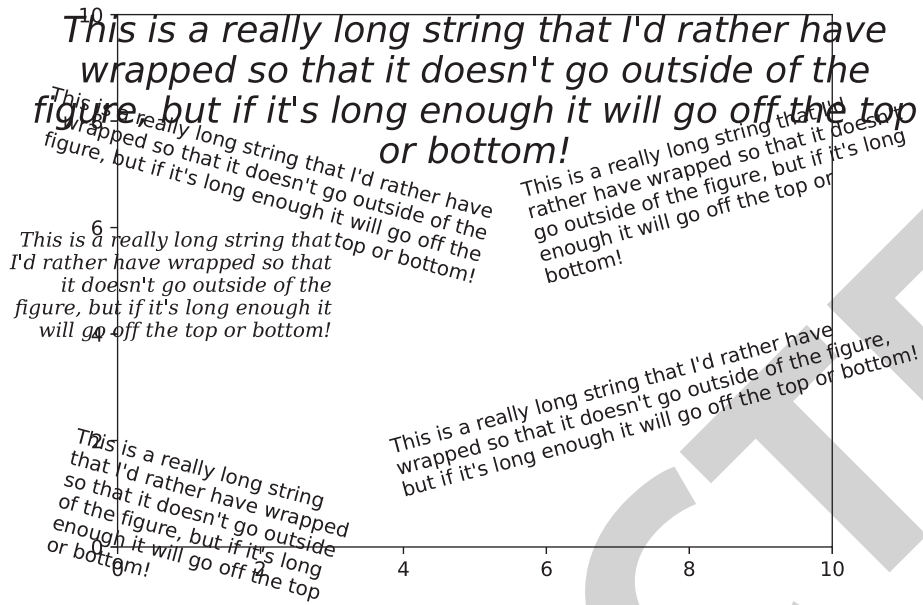
Figure 6: Multikey word search results.
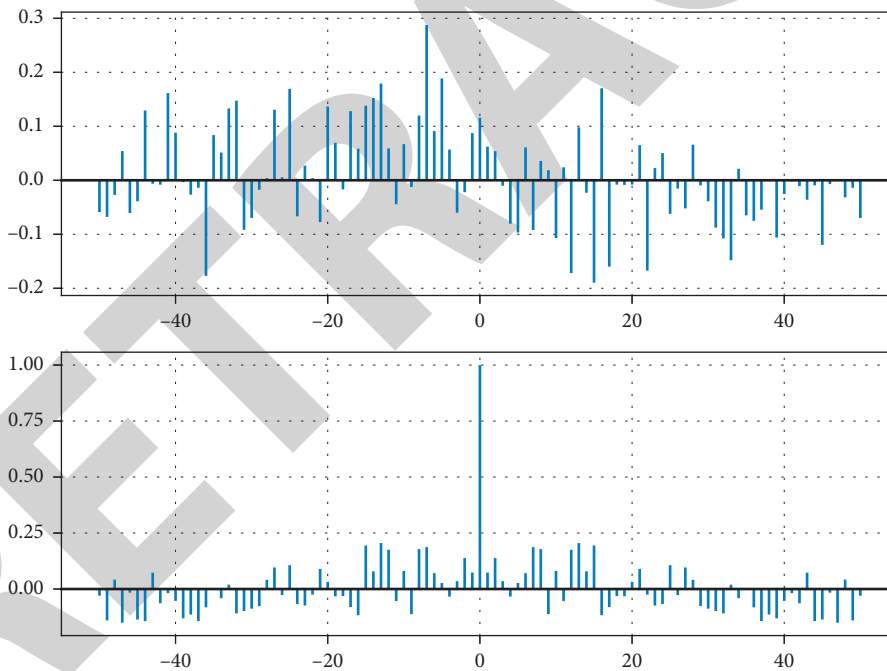
Figure 7: Semantic search results.
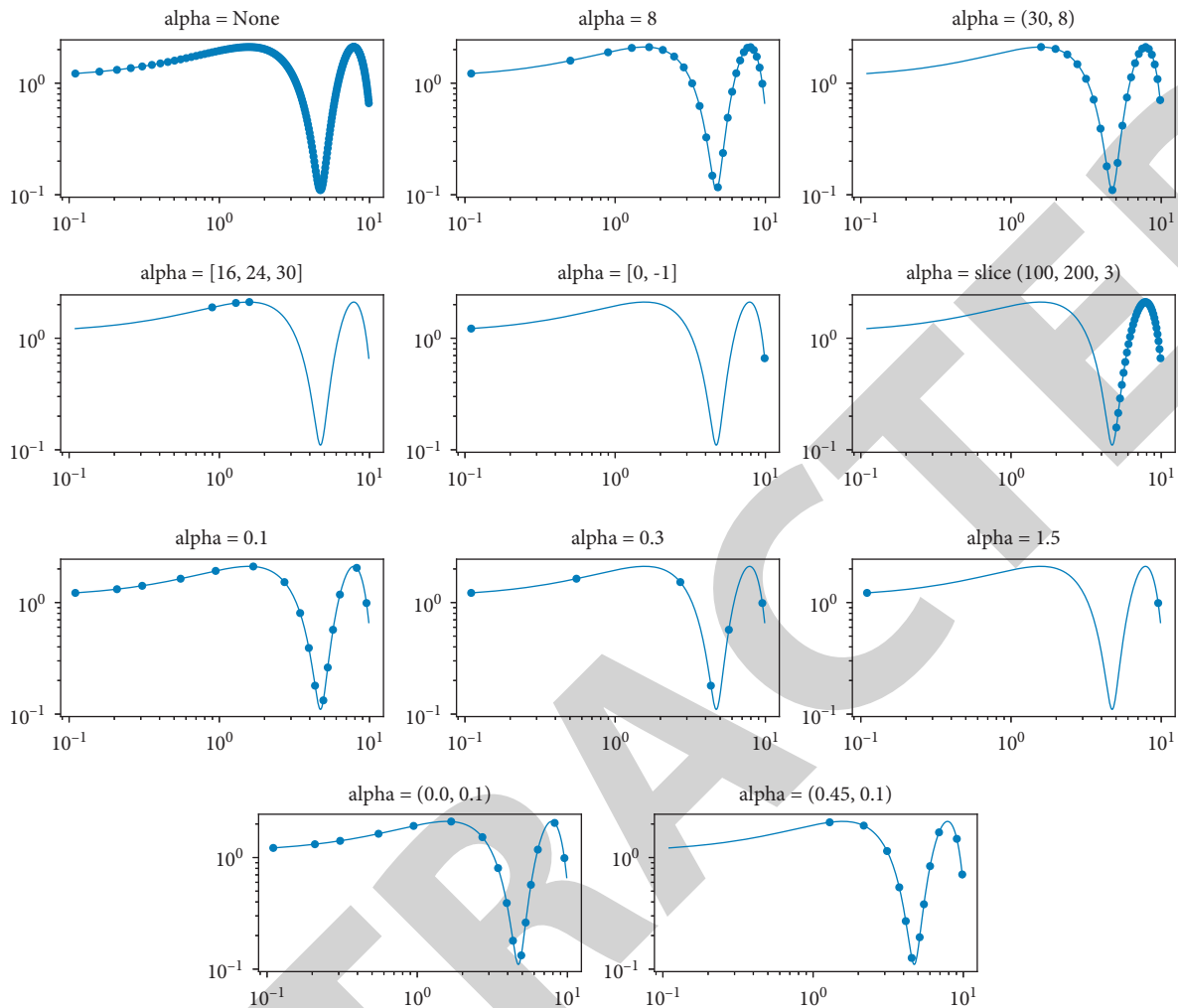
FIGURE 8: System performance with different alpha settings.

6. Conclusions

In this paper, we present a complete domain knowledge map construction scheme in the field of English teaching in colleges and universities and demonstrate the usability of the scheme through experimental results. For multisource heterogeneous domain data, we design a data acquisition method based on a combination of rule-based mapping and improved web crawlers and then use a BERT classification model after fine-tuning to clean and filter the data. In the fusion of knowledge from different sources, a Word2Vec-based entity alignment method is proposed to effectively solve the data conflict problem in the fusion process. Finally, knowledge is imported into Neo4j graph database for storage, and the implementation of English teaching visualization system is completed based on this knowledge map, which provides convenient and fast resource query and relationship display services for future discipline assessment work. Since the data sources of ELT include some unstructured data, the knowledge extraction method of unstructured text will be improved in the follow-up work to make the constructed subject knowledge map more comprehensive.

Data Availability

The datasets used in this paper are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest regarding this work.

References

[1] S. Zhou and T. Zhang, "Research on the construction of flipped classroom model for English teaching based on SPOC," *Revista de la Facultad de Ingenieria*, vol. 32, no. 14, pp. 267–273, 2017.

[2] Carmen Pérez-Llantada Auría, "Social pragmatics in technical writing: a corpus-based analysis of thematic articles," *Iberica*, vol. 5, pp. 19–34, 2003.

[3] D. Disney, "'Is this how it's supposed to work?': poetry as a radical technology in L2 Creative Writing classrooms," *New Writing*, vol. 9, no. 1, pp. 4–16, 2012.

[4] S. Moisik, "Frequency, phonology, the brain, and second language acquisition: laying the ground work for a hybrid

Research Article

Research on Methods of English Text Detection and Recognition Based on Neural Network Detection Model

Chunlan Li [ID]

Geely University of China, Chengdu, Sichuan 610095, China

Correspondence should be addressed to Chunlan Li; lichunlan@bgu.edu.cn

With the rapid development of computer science, a large number of images and an explosive amount of information make it difficult to filter and effectively extract information. This article focuses on the inability of effective detection and recognition of English text content to conduct research, which is useful for improving the application of intelligent analysis significance. This paper studies how to improve the neural network model to improve the efficiency of image text detection and recognition under complex background. The main research work is as follows: (1) An improved CTPN multidirectional text detection algorithm is proposed, and the algorithm is applied to the multidirectional text detection and recognition system. It uses the multiangle rotation of the image to be detected, then fuses the candidate text boxes detected by the CTPN network, and uses the fusion strategy to find the best area of the text. This algorithm solves the problem that the CTPN network can only detect the text in the approximate horizontal direction. (2) An improved CRNN text recognition algorithm is proposed. The algorithm is based on CRNN and combines traditional text features and depth features at the same time, making it possible to recognize occluded text. The algorithm was tested on the IC13 and SVT data sets. Compared with the CRNN algorithm, the recognition accuracy has been improved, and the detection and recognition accuracy has increased by 0.065. This paper verifies the effectiveness of the improved algorithm model on multiple data sets, which can effectively detect various English texts, and greatly improves the detection and recognition performance of the original algorithm.

1. Introduction

With the rise and popularity of the Internet of Things, there will be a huge amount of data every day, these data with the development and change of society. However, with the rapid growth of data volume, there will be a large number of data that is difficult to understand and difficult to manage, including pictures and video image data accounting for a considerable proportion. The image not only contains the shape, color, and other underlying information, but also contains the text and other high-level semantic information, which plays an indispensable role in the analysis and utilization of the image. In order to solve the problem of multidirectional text detection, the SegLink [1] algorithm cuts the text into smaller text blocks that are easier to detect and then connects the small text blocks into complete text areas. TextBoxes [2] algorithm, with SSD as the basic framework, adjusts the text area candidate box's length and width ratio and convolution core into rectangles, proposing an end-to-end text detector, so that it is more suitable for detecting slender lines of text. Mask TextSpotter [3–5] algorithm, in order to solve the problem of text that can detect any shape, combines FPN network, Fast RCNN network, and RPN network and introduces the idea of segmentation to propose an end-to-end text detection and recognition algorithm. The neural network model under deep learning can automatically extract image features, refine the feature matrix through special convolution and pooling operations, and automatically optimize network parameters. At the same time, with the help of high-performance computing platforms and large-scale data sets, methods based on deep learning have made great breakthroughs in the field of computer vision in recent years and are currently the main technical direction for studying text recognition problems from all walks of life. Aiming at the unique structure of Chinese characters, this paper uses convolutional neural

network to achieve effective overcoming of interference with natural scenes and to detect and recognize scene text.

2. Improved CTPN English Text Detection Algorithm

English text detection refers to extracting the text area in the image. There are many ways to extract, including rectangular box extraction, polygon extraction, and pixel-level extraction. This paper draws on the multiscale network structure of inception and designs an improved English text detection algorithm based on CTPN. The algorithm uses a multiscale convolution structure to extract English text features and uses adaptive text lines to improve the CTPN algorithm. Use the improved convolutional neural network based on inception to extract the features of the input image, use the RPN to obtain the feature sequence, and input the feature sequence to the bidirectional long and short period memory network to achieve feature fusion, and then input each feature after fusion into two to predict the position and confidence of the text in the parallel fully connected network, and finally stitch the obtained text area to get the final text area. It can better adapt to the detection of English text and improve the detection efficiency of English text [4] as shown in Figure 1.

2.1. CTPN Algorithm. The CTPN algorithm is a text proposal network that combines a convolutional neural network and a recurrent neural network. This algorithm introduces the recurrent neural network to the task of English text detection for the first time. It can directly locate the English text sequence in the convolutional layer to a certain extent. The above solves the limitations of traditional character detection methods [5–8]. The main structure of CTPN is shown in Figure 2:

CTPN converts the problem of English text detection into a series of fine-grained texts and proposes a new anchor box regression mechanism based on the RPN network. First, the English text area is subdivided into each small area, and then the area of each area is predicted. The vertical position of the English text and the confidence of the text finally obtain a priori information of the position of the text area with high precision. The algorithm uses a recurrent neural network to connect the convolutional feature maps. This seamless connection allows the network to obtain the context information of the English text line, so that it can detect more challenging text lines. The algorithm can process multiscale English text in a single process, avoiding subsequent filtering and refinement operations [6, 9, 10].

The advantage of the CTPN algorithm: The English text box that needs to be detected is divided into a series of small text boxes with a fixed width, making the detection horizontal.

2.2. English Text Detection Model

2.2.1. Input Image Preprocessing. The input picture size and the number of channels are not the same. In order to make it conform to the structure of the network, the input picture needs to be converted into a single-channel image and its size is scaled to 32×100.

2.2.2. Sequence Feature Extraction Based on CNN. Sequence recognition models can be divided into explicit segmentation models and implicit segmentation models, as shown in Figure 3. The main challenge of traditional text recognition algorithms based on explicit segmentation models is the correct segmentation at the pixel or character level. The quality of the segmentation effect directly affects the subsequent recognition effect. However, implicit segmentation only needs to perform simple segmentation on the sample, and there is no requirement for the quality and accuracy of character-level segmentation. A word is a continuous string. In many natural scenes, there may be interfering factors such as adhesion or uneven lighting between characters, which makes it impossible for us to correctly segment each character. Therefore, we adopt the method of implicit segmentation and use CNN to slide the window on the image to extract the feature sequence [11].

Since the length of the words existing in nature generally does not exceed 26, a total of 26 subwindows are extracted as input. The window size of CNN is 32×25, and the sliding step size is 3. The sequence feature extracted by CNN is expressed as $x = \{x_1, x_2 \ldots x_T\}$, where $T = 26$.

The CNN model of the sequence feature extraction part is based on the structure of the VGG-Very Deep network, but the final full link layer is removed, as shown in Table 1. CNN extracts sequence features from the left-to-right sliding window. The feature at each moment is the union of all the feature maps corresponding to the window position, and it is pulled into a column vector as the final input feature with a dimension of 512.

2.2.3. Processing of Context Information Based on Two-Way LSTM. CNN is based on the interdependence of the sequence features extracted by hidden segmentation and contains rich context information, which will greatly improve the recognition effect of fuzzy, nonuniform illumination and occluded words [11]. At the same time, RNN has a powerful ability for sequence learning. For the problems of RNN, we use LSTM here to replace it.

LSTM is directional; it only uses past contextual information, but for word pictures, the contextual information before and after it is meaningful for recognition. Therefore, similar to [46], we merge two LSTMs, one of which is used for forward propagation and the other is used for backward propagation to realize a two-way LSTM. The abstraction level of the deep network structure is higher than that of the shallow network, and it has achieved remarkable results in the task of speech recognition. Therefore, this method also uses two-way LSTM to process the sequence. Each layer of LSTM has 512 memory cell modules, which correspond to the dimensions of the features extracted by CNN. The input layer has a total of 512 neurons, which are fully connected to their hidden layer. After the hidden layer, they are fully connected to the output layer. Then softmax is used to classify them. There
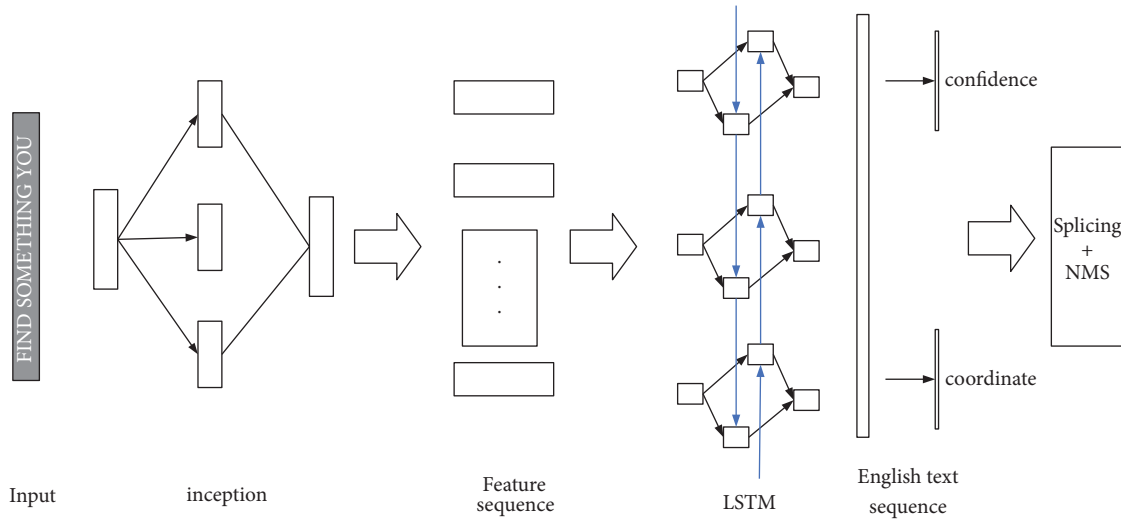
Figure 1: Improved CTPN flowchart.
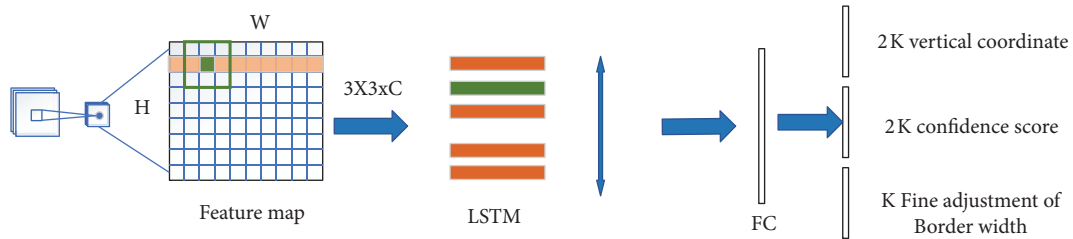
Figure 2: CTPN algorithm structure.

Figure 3: The difference between explicit segmentation and implicit segmentation.

Table 1: The structure and parameters of convolutional neural network.

Type	Number of channels	Nuclear size	Step size	Padding size
Convolutional layer 1	64	3×3	1	1
Maximum pooling layer 1		2×2	2	
Convolutional layer 2	128	3×3	1	1
Maximum pooling layer 2		2×2	2	
Convolutional layer 3	256	3×3	1	1
Convolutional layer 4	256	3×3	1	1
Maximum pooling layer 3		1×2	2	
Convolutional layer 5	512	3×3	1	1
Convolutional layer 6	512	3×3	1	1
Maximum pooling layer 4		1×2	2	
Convolutional layer 7	512	2×2	1	

are 36 categories (including ten numbers from 0 to 9 and twenty-six English letters from a to z that are not case sensitive). The result of the bidirectional LSTM prediction output is $y = \{y_1, y_2 \cdots y_T\}$, and its length is the same as the length of the input sequence feature.

2.2.4. CTC-Based Transcription. CTC is specifically designed for sequence labeling tasks, especially those input sequences that are difficult to segment into specific targets. In our recognition network, the CTC layer is directly connected to the output of LSTM, and its effect is similar to

the output layer of LSTM. It not only eliminates the need for presegmentation of the input image [12], but also allows us to achieve end-to-end training by minimizing the loss function.

2.3. Overall Framework Description and Algorithm Evaluation

2.3.1. Overall Framework Description. Assuming that N pictures are entered, then we have the following.

(1) First, scale N pictures to the same size, and normalize the pixels to the range of $[−0.5, 0.5]$ to ensure that gradient descent can be used to accelerate training.

(2) Input the normalized picture into the improved inception network, extract the feature map with multiple receptive fields, and the size of the obtained feature map is $N \times C \times H \times W$.

(3) Then use a 3×1 convolution window to slide on the feature map. Each sliding window generates k anchor boxes on the original image and obtains a feature vector of size $N \times C$ to represent the features of the original image area.

(4) The feature vector of each row is used as a set of feature sequences and input into the two-way long- and short-term memory network for encoding. The encoded features incorporate contextual information.

(5) Input the encoded features into two parallel fully connected networks, and predict 2k confidence scores and 2k vertical regression box positions.

(6) In the training phase, in order to facilitate the calculation of IOU, the original English text line splicing method of CTPN is still used. During the test, in order to obtain a more accurate proposal area for the slanted English text, the improved text line splicing method proposed in this paper is used.

(7) Use the improved NMS algorithm to suppress the regression box to obtain the optimal English text prediction box, and use the perspective transformation to correct the English text box as the input of the English text recognition network. The entire network is composed of multiple parts cascaded, but through ingenious construction, the various modules are combined into a whole, so that the network can be mapped from end to end. In the process of training, the network extracts features from the original picture without complicated preprocessing.

2.3.2. Algorithm Evaluation. The performance evaluation of the proposed algorithm is carried out by using the competition evaluation criteria corresponding to the IC15 database, which is measured by dividing the area of the overlapping part of the inspection result rectangle and the ground-truth rectangle by the area of the union part. The evaluation indicators used are recall rate, training speed, and prediction speed.

3. Improved CRNN English Text Recognition Algorithm

For scanned English text, traditional optical character recognition and other algorithms have become mature, but natural scene text is still very challenging due to the interference of complex backgrounds, fonts, lighting, textures, and viewing angles. After text detection, a series of long texts are obtained. Recognize these long texts. This chapter uses the CRNN algorithm to unify the feature extraction, sequence modeling, and transcription of the long image into a framework to realize the end-to-end recognition algorithm of the image. The CRNN algorithm has the characteristics of end-to-end training, recognition of any length, undefined vocabulary recognition, lightweight structure, and strong generalization ability [13].

The network architecture of CRNN consists of three parts in total, including a convolutional network used to extract features, a recurrent network used to identify sequence information, and a CTC layer used to transcribe and map the final label.

The processing process is as follows: the input English text image is passed through the convolutional layer to obtain a series of feature maps, these feature maps are compressed into features acceptable to the recurrent network, these features are predicted by frame, and then they are decoded and transcribed by CTC.

A sequence of labels is in the same format. The whole can be jointly trained through a loss function. The network architecture of the CRNN algorithm is shown in Figure 4. The algorithm is as follows.

In this chapter, on the CRNN network infrastructure, the role of increasing the antineural network layer is to artificially increase the learning difficulty of the algorithm, so as to achieve the function of the algorithm to recognize the occluded text, improve the recognition accuracy of the algorithm, and increase the robustness of the algorithm [14]. At the same time, the transcription layer adopts the combination of CTC and attention mechanism to improve the accuracy of transcoding. The structure diagram of the improved CRNN text recognition algorithm is shown in Figure 5.

3.1. Feature Extraction Layer. Based on the VGG16 improved convolutional neural network to construct the convolutional layer, use the convolutional neural network to extract. Enter the text characteristics of the English text image. When inputting the network, the height of the English text image needs to be scaled to a fixed value, such as 32 pixels, and the width is proportional. Then, the feature map extracted by the convolutional network is expanded from left to right in columns, converted into a feature sequence, and input to the antineural network layer [15].

3.2. Against the Neural Network Layer. Use the VGG16 network pair to counter the network to share features, but not to share parameters. The feature map generated by the VGG16 network is used as input, and the mask is generated through the mask network. The mask is used to
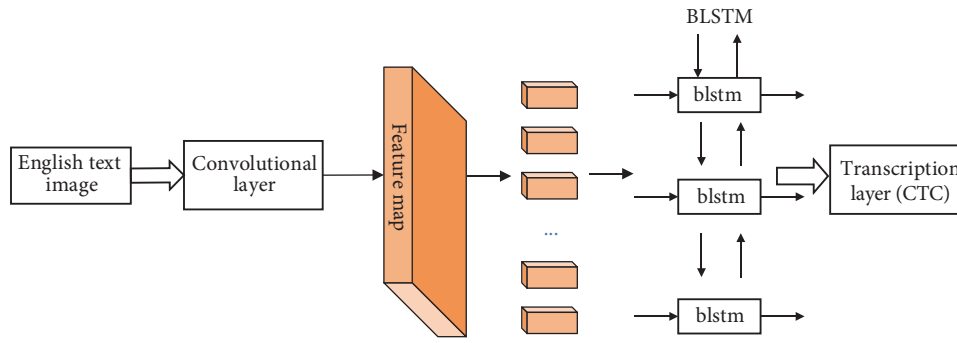
FIGURE 4: The network architecture of the CRNN algorithm.
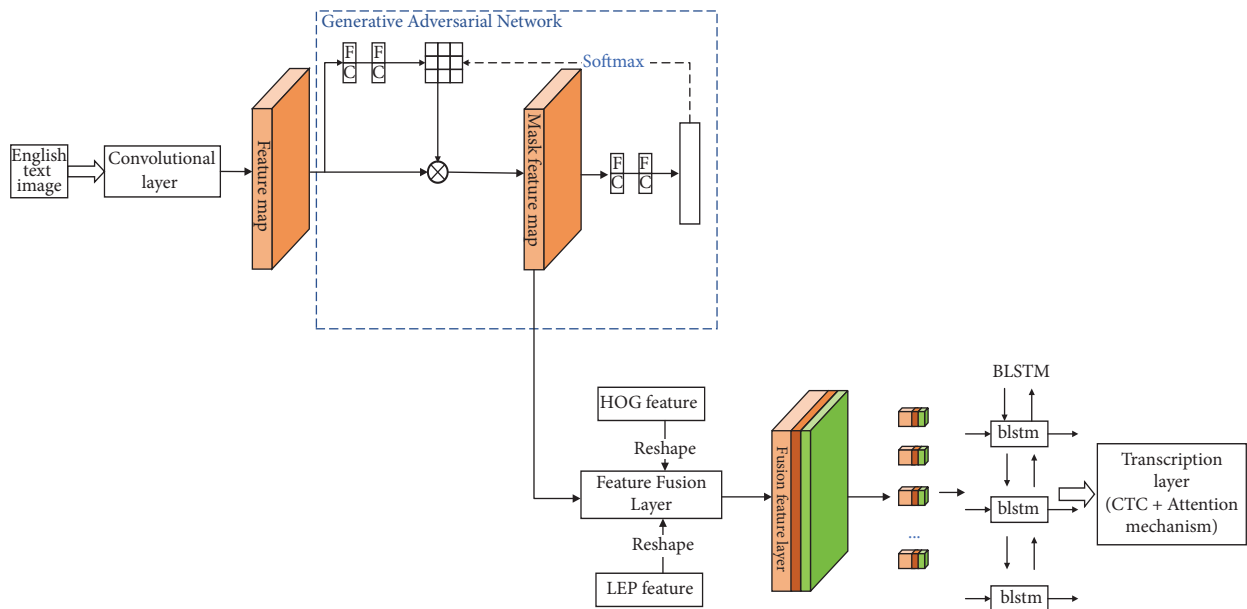
FIGURE 5: Improved CRNN text recognition algorithm.

determine which parts of the feature map should be subtracted to generate a masked feature map, and then send it to the classifier for judgment. The mask will automatically make corresponding adjustments according to the loss function. First, train the iterative CRNN network about 10,000 times. First, get a model that can basically be recognized; second, train the adversarial neural network separately to predict specific occluded parts. It first divides the feature map into 9 grids. In order to generate training information against the neural network, these 9 grids are occluded in sequence, and the grid with the largest classification loss after occlusion is the grid that is most worthy of occlusion. The training loss function of the antineural network is to classify and judge these 9 grids, and whether each grid is the most worthy of occlusion, so the output is a graph composed of classification probabilities. When using the output results, take the 1/2 pixels with the highest probability of being classified as "most worthy of occlusion" and randomly select 1/3 of these pixels for occlusion, and the remaining 2/3 without occlusion, to increase a certain random factor. Finally, the adversarial neural network and CRNN are jointly trained,

and inspired by reinforcement learning, they focus on training the binarization mask that significantly reduces the classification effect.

3.3. Feature Fusion Layer.
Extract the HOG feature and LBP feature of the text line image, and fuse them with the mask feature extracted in the previous step to generate a fusion feature map.

3.4. Cyclic Network Layer.
Input the feature matrix into the two-way LSTM network to extract the sequence features of the text.

3.5. Transcription Layer.
The essence of text recognition is a translation process, which converts an image sequence into a text sequence. The conversion system consists of two parts. The feature extraction of the text image is completed before and the features are encoded. Then the last step is to generate a feature decoder to convert features into text. Therefore, the conditional probability can be defined:

$$\log p(y|x) = \sum_{j=1}^{m} \log p(y_j|y_{<j}, s). \tag{1}$$

This article uses the structure of encoding and decoding to realize the recognition of the text. Specifically, each character decoded can be converted into a probability:

$$p(y_j|y_{<j}, s) = \text{soft max}(g(h_j)). \tag{2}$$

$g(h_j)$ is the conversion function, h_j is the hidden unit of the RNN, and the calculation method is

$$h_j = f(h_{j-1}, s). \tag{3}$$

The function $f(h_{j-1}, s)$ uses the output of the hidden layer at the previous moment to calculate the current hidden state.

The model based on attention mechanism decoding directly predicts the output of the entire sequence without conditional independence, sexual hypothesis. Using the attention mechanism to decode without building a character-level language model at the output end, compared with CTC, the calculation speed and the recognition rate have been significantly improved. The model can be defined by the following recursive equation, using the previously calculated label sequence. Predict the probability distribution of the label at the current moment: the specific process is shown in equations (4)–(6):

$$P(y|x) = \prod_{u} P(y_u|x, y_{1:u-1}), \tag{4}$$

$$h = \text{Encoder}(x), \tag{5}$$

$$y_u = \text{Attention Decoder}(h, y_{1:u-1}). \tag{6}$$

In practice, CTC decoding only considers the feature vector at the current moment, which is very restrictive, but there is no contextual information obtained. Supplement: therefore, the sequence obtained is not necessarily globally optimal. The attention mechanism scans all encoding vectors and calculates the value of each feature as shown in Figure 6. Weights, using global information for decoding, with lack of constraints, lead to excessive freedom in the decoding stage and slow network convergence. This paper uses the multitask loss function method to jointly train CTC and attention; using CTC as the attention a priori constraint [16], the loss function is as follows:

$$L_{\text{MTL}} = \lambda L_{\text{CTC}} + (1 - \lambda)L_{\text{Attention}}. \tag{7}$$

Practice has proved that the grid convergence speed of the joint training of CTC and attention mechanism is faster, and the accuracy of decoding is greatly improved.

4. Algorithm Simulation Test and Result Analysis

In order to evaluate the improved CRNN English text detection and recognition algorithm, experiments were conducted on the standard data set of English text detection and recognition.
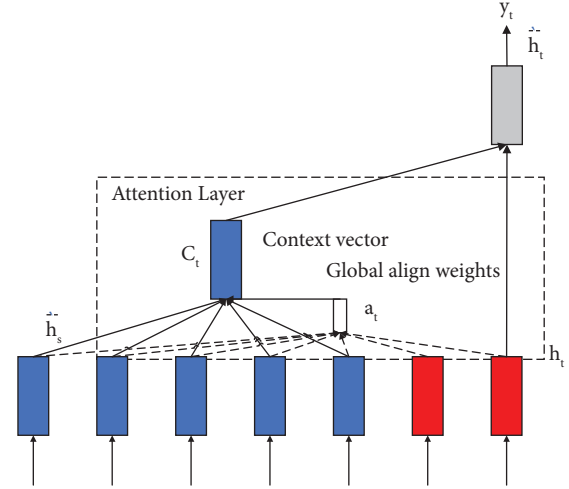
FIGURE 6: Attention mechanism.

TABLE 2: Mjsynth data set.

Category	Quantity
Total image	8919273
Number of images in the training set	7224612
Number of images in validation set	802734
Number of images in the test set	891927
LABEL type	88172

4.1. Experimental Design

4.1.1. Training Set.
The training set used by the algorithm in this chapter is the Mjsynth data set. This data set is a synthetic data set published by Jaderberg et al. At present, the number of English text images used in this data set has reached more than 9 million, which provides sufficient training samples for users to train. The data set contains nearly 90,000 labels. The algorithm in this chapter was trained on the Mjsynth data set and tested on the standard English text detection and recognition benchmark data set. The specific conditions of the Mjsynth training set are shown in Table 2.

It should be pointed out that when training the model, first iterate the CRNN network on the Mjsynth data set 10,000 times, then train the improved CRNN text detection and recognition network, and finally train the entire model.

4.1.2. Test Set.
The algorithm in this chapter selects the benchmark data set ICDAR 2013 (IC13) and Street View Text (SVT) as the test set. ICDAR2013 (IC13) was proposed for scene text detection in the ICDAR 2013 Robust Reading Competition. It contains high-resolution images, 229 training images, and 233 test images, including English text. The comment is a rectangular box of words. It contains 1015 actual cropped word images. The SVT test data set consists of 249 Street View images collected from Google Street View. 647 word images were cut out from them. Each word image has a 50-word dictionary. This test set mainly tests English text [17–21].

TABLE 3: Time table consumed by different models.

Model structure	Recall rate (%)	Training speed (hours)	Forecast speed (seconds)
CTPN_convID	90.86	2.9	1–1.6
CTPN_LSTM	92.22	4.6	1.3–2.5
EAST	89.73	4	1.2–2

Chapter I. The Shadow of Change

"Harvest is ended and summer is gone," quoted Anne Shirley, gazing across the shorn fields dreamily. She and Diana Barry had been picking apples in the Green Gables orchard, but were now resting from their labors in a sunny corner, where airy fleets of thistledown drifted by on the wings of a wind that was still summer-sweet with the incense of ferns in the Haunted Wood.

But everything in the landscape around them spoke of autumn. The sea was roaring hollowly in the distance, the fields were bare and sere, scarfed with golden rod, the brook valley below Green Gables overflowed with asters of ethereal purple, and the Lake of Shining Waters was blue—blue—blue; not the changeful blue of spring, nor the pale azure of summer, but a clear, steadfast, serene blue, as if the water were past all moods and tenses of emotion and had settled down to a tranquility unbroken by fickle dreams.

"It has been a nice summer," said Diana, twisting the new ring on her left hand with a smile. "And Miss Lavendar's wedding seemed to come as a sort of crown to it. I suppose Mr. and Mrs. Irving are on the Pacific coast now."

"It seems to me they have been gone long enough to go around the world," sighed Anne.

"I can't believe it is only a week since they were married. Everything has changed. Miss Lavendar and Mr. and Mrs. Allan gone—how lonely the manse looks with the shutters all closed! I went past it last night, and it made me feel as if everybody in it had died."

I. An Irate Neighbor

A tall, slim girl, "half-past sixteen," with serious gray eyes and hair which her friends called auburn, had sat down on the broad red sandstone doorstep of a Prince Edward Island farmhouse one ripe afternoon in August, firmly resolved to construe so many lines of Virgil.

But an August afternoon, with blue hazes scarfing the harvest slopes, little winds whispering elfishly in the poplars, and a dancing slendor of red poppies outflaming against the dark coppice of young firs in a corner of the cherry orchard, was fitter for dreams than dead languages. The Virgil soon slipped unheeded to the ground, and Anne, her chin propped on her clasped hands, and her eyes on the splendid mass of fluffy clouds that were heaping up just over Mr. J. A. Harrison's house like a great white mountain, was far away in a delicious world where a certain schoolteacher was doing a wonderful work, shaping the destinies of future statesmen, and inspiring youthful minds and hearts with high and lofty ambitions.

To be sure, if you came down to harsh facts . . . which, it must be confessed, Anne seldom did until she had to . . . it did not seem likely that there was much promising material for celebrities in Avonlea school; but you could never tell what might happen if a teacher used her influence for good. Anne had certain rose-tinted ideals of what a teacher might accomplish if she only went the right way about it; and she was in the midst of a delightful scene, forty years hence, with a famous personage . . . just exactly what he was to be famous for was left in convenient haziness, but Anne thought it would be rather nice to have him a college president or a Canadian premier . . . bowing low over her wrinkled hand and assuring her that it was she who had first kindled his ambition, and that all his success in life was due to the lessons she had instilled so long ago in Avonlea school. This pleasant vision was shattered by a most

(a)   (b)

FIGURE 7: Original image of text detection. (a) Test sample 1. (b) Test sample 2.

4.2. Result Analysis

4.2.1. Analysis of English Text Detection and Recognition Results.
The optimization function uses the Adam optimizer, which can directly optimize the algorithm for end-to-end training, and each labeled anchor point needs to be calculated in advance before entering the model. Using momentum of 0.9 and weight decay of 0.0005, batch data is 1024, period is 50000, and each period is 50 steps. In the first 16000 iterations, the learning rate is set to 0.0001. On the core of Conv1D, after several experiments, 5 was finally adopted as the parameter. In the ICDAR2013 data set, the Conv1D kernel uses 7 as a parameter.

Because the feature extractor uses the VGG16 model, the transfer learning method is introduced here. Load the training weights of ImageNet directly into the VGG16 model. At the same time, some public text data sets are used to pretrain the model. In this way, when training self-labeled data, the overall loss has a starting point of about 1.3. The front-end network adopts VGG, and two cases where the loop module adopts LSTM and Conv1D are, respectively, tested. Also use 10,000 pictures as an experiment and train for 50,000 cycles. Training loss is recorded every 10 cycles.

In order to verify the performance of the model on the public data set, several algorithms were verified on the ICDAR2019 data set. The results are shown in Table 3.

It can be seen from 4.2 that, in the process of comparing the three models, no matter which data set, the accuracy rate and recall rate in the evaluation index are relatively close, which is acceptable in engineering. The training speed is a qualitative leap in the Conv1D model. This speed-up can speed up model iteration and accelerate the prediction process. The final effect is shown in Figures 7 and 8.

4.2.2. Analysis of English Text Recognition Results.
The English text recognition experiment data uses 1 million English pictures, and the optimization function uses the SGD optimizer. Each batch of data uses 2048 pictures. Each cycle has 500 steps. The learning rate is divided into three stages: 0.004, 0.0004, and 0.00004. Both GRU and Conv1D use 40 cycles, and LSTM uses 50 cycles. First, experiments were conducted on the feature extraction network, using a shallow structure similar to VGG pairs and experimenting with different architectures of recurrent networks, using LSTM and GRU, respectively. The experimental process loss is shown in Figures 9(a) and 9(b). The coordinate is the function loss, and the abscissa is the period. Conv1D loss and acc are shown in Figures 10(a) and 10(b). In Figure 10(b), the ordinate is the accuracy of the verification set, and the abscissa is the period.

In the experiment, it was found that the accuracy rate is basically the same, but the training speed is very different. Here, we use 1 million pieces of data with a width of 286 pixels and a height of 32 pixels, and the training time is 95% accurate. The training time of each cycle is shown in Table 4.

(a)

(b)

Figure 8: Text detection effect diagram. (a) Identification sample 1. (b) Identification sample 2.

(a)

(b)

Figure 9: LSTM and GRU training curve. (a) LSTM loss curve. (b) GRU loss curve.

(a)

(b)

Figure 10: Conv1D training curve. (a) Conv1D loss curve. (b) Conv1D acc curve.

TABLE 4: Cycle structure diagram.

Model structure	Accuracy (%)	Life cycle	Training time per cycle (minutes)
LSTM	95	48	32
GRU	95	32	17
convID	95	36	5

TABLE 5: Accuracy test table.

Model structure	Accuracy (%)	Life cycle	Training time per cycle (minutes)
LSTM	98.38	59	32
GRU	98.38	41	17
convID	98.36	37	5
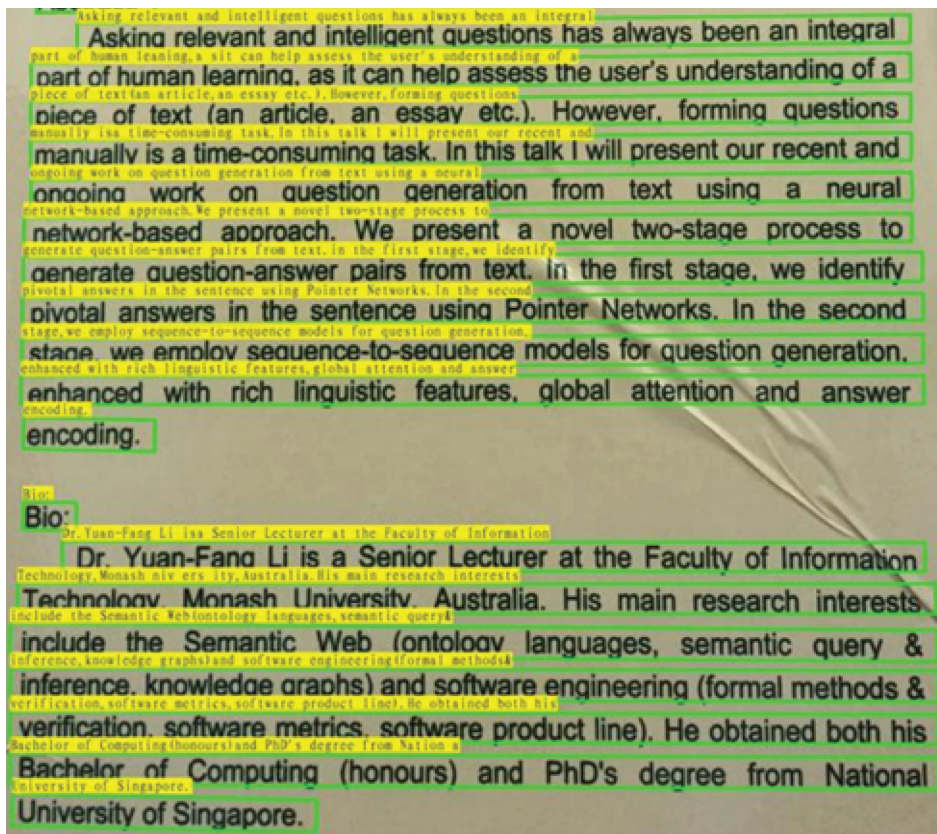
FIGURE 11: English text recognition graphics.

From Table 4, it can be found that the training duration of a single cycle of Conv1D is the smallest, which is 6.4 times higher than that of LSTM, while the cycle is reduced by one-third. Compared to the fastest GRU structure in the RNN structure, it is also 3.2 times faster. And from the loss curves in Figures 9(a) and 9(b) and 10(a), it can be seen that when the same loss function is used, the loss images are similar. It can be shown that the model can accomplish the task well under the current data set. The highest accuracy rate of the final experiment is shown in Table 5.

The final result is shown in Figure 11.

5. Conclusion

This paper studies how to improve the neural network model to improve the efficiency of text detection and recognition of English text images. The main research work is as follows.

In the aspect of English text detection, an improved CTPN multidirectional English text detection algorithm is proposed to solve the problem that the CTPN network can only detect the text in the approximate horizontal direction. The algorithm consists of a preprocessing model, a CTPN network English text positioning model, and a text box fusion model. Using the proposed text box fusion algorithm,

multiple candidate boxes obtained by CTPN are fused to obtain the best text box. Experiments show that, compared with the CTPN algorithm, the accuracy of this algorithm increases by 0.07, the recall rate increases by 0.217, and the comprehensive index increases by 0.16, which proves that the algorithm can detect multidirectional English text and has a higher detection accuracy.

In terms of natural scene text recognition, an improved CRNN English text recognition algorithm is proposed. The algorithm is based on the CRNN model. On the basis of the model, the adversarial network and fusion features are added to enable the recognition of occluded English text. The algorithm was tested on the IC13 and SVT data sets. Compared with the original CRNN model, this algorithm has a higher recognition accuracy. Especially on the SVT data set, the accuracy is increased by 0.065 without the constraint dictionary. It is proved that the algorithm can effectively recognize the occluded English text.

This paper uses the transfer training CRNN model with better English recognition effect to realize the variable length recognition of Chinese and English texts. And the recognition accuracy rate of 97.85% was obtained on the verification set. Since the recognition result of the entire English text line is a string with no spaces, this paper uses Viterbi algorithm to segment the English text string in order to present a better readable effect. This paper compares and analyzes the final detection and recognition results with the results of CTPN and DenseNet models and has obtained significant advantages in accuracy and test speed.

For the text detection algorithm proposed in this article, it is found that there is a high time complexity and space complexity, which will occupy a large amount of computing resources during detection. When running on embedded hardware such as tree mold, it can not meet the requirements at all, so the model needs to be further compressed in the future.

Data Availability

The dataset can be accessed upon request to the author.

Conflicts of Interest

The author declares that there are no conflicts of interest.

References

[1] M. H. Liao, B. G. Shi, X. Bai, X. G. Wang, and W. Y. Liu, "Textboxes: a fast text detector with a single deep neural network," in *Proceedings of the 31st AAAI Conference on Artificial Intelligence*, pp. 4161–4167, AAAI, San Francisco, CA, USA, March 2017.

[2] P. Lyu, M. Liao, C. Yao, W. Wu, and X. Bai, "Mask Text-Spotter: an end-to-end trainable neural network for spotting text with arbitrary shapes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, 2019.

[3] T. Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and H. Belongie, "Feature pyramid networks for object detection," 2016, https://arxiv.org/abs/1612.03144.

[4] X. Zhao, W. Li, Y. Zhang, T. A. Gulliver, S. Chang, and Z. Feng, "A faster RCNN-based pedestrian detection system," in *Proceedings of the 2016 IEEE 84th Vehicular Technology Conference (VTC-Fall)*, IEEE, Montreal, Canada, 18-September 2016.

[5] J. Ma, W. Shao, H. Ye et al., "Arbitrary-oriented scene text detection via rotation proposals," *IEEE Transactions on Multimedia*, vol. 20, pp. 3111–3122, 2017.

[6] Z. Tian, W. Huang, T. He, P. He, and Y. Qiao, "Detecting text in natural image with connectionist text proposal network," in *Proceedings of the 14th European Conference on Computer Vision*, pp. 56–72, Springer, Cham, Switzerland, October 2016.

[7] Y. L. Liu and L. W. Jin, "Deep matching prior network: toward tighter multi-oriented text detection," in *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3454–3461, IEEE, Honolulu, HI, USA, July 2017.

[8] B. G. Shi, X. Bai, and S. Belongie, "Detecting oriented text in natural images by linking segments," in *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3482–3490, IEEE, Honolulu, HI, USA, July 2017.

[9] A. Bissacco, M. Cummins, Y. Netzer, and H. N. Photoocr, "Reading text in uncontrolled conditions," in *2013 Proceedings of the 2013 IEEE International Conference on Computer Vision*, pp. 785–792, Sydney, Australia, December 2013.

[10] M. Busta, L. Neumann, and J. Matas, "Deep textspotter: an end-toendtrainable scene text localization and recognition framework," in *Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 2223–2231, Venice, Italy, October 2017.

[11] Z. Cheng, F. Bai, Y. Xu, G. Zheng, S. Pu, and S. Zhou, "Focusing attention: towards accurate text recognition in natural images," in *Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 5086–5094, Venice, Italy, October 2017.

[12] Z. Cheng, Y. Xu, F. Bai, Y. Niu, S. Pu, and S. Zhou, "Aon: towards arbitrarily-oriented text recognition," in *Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5571–5579, Salt Lake City, UT, USA, June 2018.

[13] C. K. Chng and C. S. Chan, "Total-text: a comprehensive dataset for scene text detection and recognition," in *Proceedings of the 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, pp. 935–942, Kyoto, Japan, November 2017.

[14] J. Dai, K. He, Y. Li, S. Ren, and J. Sun, "Instance-sensitive fully convolutional networks," in *Proceedings of the Computer Vision - ECCV 2016*, pp. 534–549, Amsterdam, The Netherlands, October 2016.

[15] Y. Dai, Z. Huang, Y. Gao et al., "Fused text segmentation networks for multi-oriented scene text detection," in *Proceedings of the 2018 24th International Conference on Pattern Recognition (ICPR)*, pp. 3604–3609, Beijing, China, August 2018.

[16] J. Donahue, L. Anne Hendricks, S. Guadarrama et al., "Long-term recurrent convolutional networks for visual recognition and description," in *Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2625–2634, Boston, MA, USA, June, 2015.

[17] J. Gehring, M. Auli, D. Grangier, D. Yarats, and Y. N. Dauphin, "Convolutional sequence to sequence learning," in *Proceedings of the ICML*, pp. 1243–1252, Sydney, Australia, August 2017.

[18] R. B. Girshick, "Fast R-CNN," in *Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV)*, pp. 1440–1448, Santiago, Chile, December 2015.

[19] R. B. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 580–587, Columbus, OH, USA, June 2014.

[20] L. G´omez and D. Karatzas, "Textproposals: a text-specific selective search algorithm for word spotting in the wild," *Pattern Recognition*, vol. 70, pp. 60–74, 2017.

[21] R. Gomez, B. Shi, L. Gomez et al., "Icdar2017 robust reading challenge on coco-text,"vol. volume 1, pp. 1435–1443, in *Proceedings of the 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, vol. volume 1, IEEE, Kyoto, Japan, November 2017.

Research Article

Research on Deep Learning-Based Financial Risk Prediction

Boning Huang [1] **and Junkang Wei**[2]

[1]*Shenzhen University Webank Institute of Fintech, Shenzhen University, Shenzhen, Guangdong 518052, China*
[2]*School of Pharmaceutical Sciences, Sun Yat-sen University, Guangzhou, Guangdong 510630, China*

Correspondence should be addressed to Boning Huang; 2068511002@email.szu.edu.cn

Financial text-based risk prediction is an important subset for financial analysis. Through automatic analysis of public financial comments, fundamentals on current financial expectations can be evaluated. A deep learning method for financial risk prediction based on sentiment classification is proposed in this paper. The proposed method consists of two steps. Firstly, the abstract of the financial message is extracted according to the seq2seq model. During the extraction process, the seq2seq model can cope with the situation of different input message lengths. After the abstraction, invalid information in the financial messages can be effectively filtered, thus accelerating the subsequent sentiment classification step. The sentiment classification step is performed through the GRU model according to the abstracted texts. The proposed method has the following advantages: (1) it can handle financial messages of different lengths; (2) it can filter out the invalid information of financial messages; (3) because the extracted abstract is more refined, it can speed up the subsequent sentiment classification step; and (4) it has better sentiment classification accuracy. The proposed method in this paper is then verified through financial message dataset from the financial social network StockTwits. By comparing the classification performances, it can be seen that compared with the classical SVM and LSTM methods, the proposed method in this paper can improve the accuracy of sentiment classification by 5.57% and 2.58%, respectively.

1. Introduction

Despite having new tools and algorithms, financial market analysis is still a complex subject. The main purpose for financial market analysis is to assist investors in making decisions by analyzing the price fluctuations of the financial products. Financial risk analysis is a subtopic for financial market analysis, which can be adopted to predict the future price trends of financial products, and it is a hot topic of current research. Traditional financial analysis methods can be divided into two categories: fundamental analysis [1–3] and technical analysis [4–6]. Among them, fundamental analysis is to predict the future by studying the basic attributes of the company, which is suitable for relatively long-term forecasts. Technical analysis does not explicitly consider the company's internal and external characteristics but directly predicts the future through price fluctuations. Technical analysts believe that price fluctuations include all fundamental factors. Technical analysis models price fluctuations into time series and transforms them into pattern

recognition problems. With the in-depth application of machine learning in financial analysis in recent years, an effective financial analysis method has become more and more popular; that is, by mining important features from financial messages, such as financial news [7, 8], financial comments [9, 10], and social networks [11, 12], the sentiment tendency for a large number of users or authors can be evaluated, thereby capable of predicting financial prices. The mentioned method not only has the factors in fundamental analysis but also has the advantages of automatic technical analysis, which has been a research hotspot in recent years.

With the accumulation of data samples in the financial text field, financial message sentiment analysis based on deep learning becomes possible. Through the deep neural networks, the public's sentiment tendency for individual stocks or the overall economic situation can be obtained. There are usually two categories of the sentiment tendency: bullish and bearish. The authors in [13] conducted a research on the correlation between financial sentiment tendency and financial conditions in reality. The conclusion points out that

through the Pearson correlation test, the public's financial sentiment tendencies and the reality financial situation (including market closing price, trading volume, and so on) have a strong correlation, which can be adopted for financial risk prediction.

The methods of financial message sentiment analysis can generally be divided into two categories: one is the traditional machine learning based and the other is deep learning based. For the methods based on traditional machine learning, the authors in [14, 15] proposed to classify the financial message through the support vector machine (SVM). In the classification process, the message is represented by the bag-of-words (BoW) approach. Traditional classifiers also include multinomial Naïve Bayes (mNB), random forest (RF), and so on [16–19]. Traditional machine learning-based methods only extract shallow features in the messages for classification, which is often not as good as that based on deep learning. For the sentiment classification method based on deep learning, a large number of labeled samples can be adopted to automatically extract more abstract and deeper level features from the financial messages, thus having more accurate sentiment estimation.

Financial message-based sentiment estimation methods adopting deep learning can be divided into three categories according to the different neural networks. The first type of method adopts the convolutional neural networks (CNNs) [20]. In the CNN, deep level feature information can be extracted through the convolutional layer, thus making it effective for a wide range of applications in deep image processing. However, as the CNN-based method cannot use the correlation or the context information in the messages, the accuracy for sentiment classification is limited. The second type of method is based on the recurrent neural network (RNN). In this type, the RNN is adopted, which can effectively make use of the context information for financial sentiment classification. The RNN-based method usually has better classification performance than the CNN-based ones. Hiew et al. [21] adopted the long-short term memory (LSTM) network to classify financial messages, which can obtain a better classification performance. The publication in [22] compared the classification results of various RNNs in detail, including the simple RNN, LSTM, stacked LSTM, gated recurrent unit (GRU), stacked GRU, bi-directional LSTM, bi-directional GRU, and so on. The following conclusions are obtained according to the comparisons: LSTM and its variations have better classification performance. The publication has also proved that in these classifiers, different optimizers, such as rmsprop, Adam, and so on, have similar accuracy. For the third type of method, financial message sentiment analysis is carried out through transfer learning. By using a large number of messages in other text domain with sentiment tags and a small number of messages in the financial domain, transfer learning can be adopted to solve the problem of lack of training samples. In [23], transfer learning was carried out by fine-tuning. The convolutional neural network is pretrained through source domain samples, and then the financial domain messages are adopted to fine-tune the parameters. In actual implementation, all parameters can be recalculated during the pretraining process. For fine-tuning,

only the high-level parameters are variable, while the low-level parameters are frozen. This can make the network quickly adapt to the samples in the financial message domain. The authors in [24] proposed to perform transfer learning through stacked denoising autoencoder (SDA). Firstly, the shared feature space in source message domains and financial message domains is extracted through the SDA. Then, the shared feature vector can be used to perform the financial message sentiment classification through other classifiers. In [25], transfer learning was carried out adopting adversarial learning. Through the gradient reversal module, the sentiment classifier and the data domain classifier are adopted for adversarial learning, so as to obtain shared feature representations of different data domains.

A new method for financial sentiment classification is proposed in this paper. The method is deep learning based and has two steps. Firstly, an abstract for the financial messages can be extracted according to the seq2seq model. During the extraction process, the situation of text input with different lengths can be dealt with. After the extraction, invalid information in financial messages can be effectively filtered out, which is favorable for the subsequent classification. The sentiment classification step is then performed through the GRU model according to the extracted abstracts. The proposed method has the advantage of coping with the situation of different message lengths. The extracted abstract can be regarded as a refinement of the raw message, which can filter out the invalid information of financial messages. As a result, it can speed up the subsequent sentiment classification process and can bring better sentiment classification accuracy. The proposed method in this paper is verified through financial message dataset from the financial social network StockTwits. By comparing the classification performances, the results show that compared with the classical SVM and LSTM methods, the proposed method in this paper can improve the accuracy of sentiment classification by 5.57% and 2.58%, respectively.

2. Methods

The overall structure of the proposed method in this paper is shown in Figure 1. It can be seen that the proposed method mainly consists of two steps. (1) Firstly, the abstract of the input financial message is extracted through the seq2seq model, noting that the extracted abstract has a fixed length. (2) The abstract is then adopted for sentiment classification through the GRU model. The results for the classification are twofold: bullish and bearish. The two-step method in this paper has the following advantages: (1) the seq2seq model can effectively be adopted to extract valid information from the financial messages, and the redundant information irrelevant to classification can be filtered out; and (2) the extracted abstract has a limited length and thus can accelerate the subsequent recognition. According to the above two steps, the proposed method will be discussed as follows.

2.1. Seq2seq Model-Based Message Abstraction. The seq2seq model has a relatively mature application in the translation task in the field of natural language processing (NLP). Since
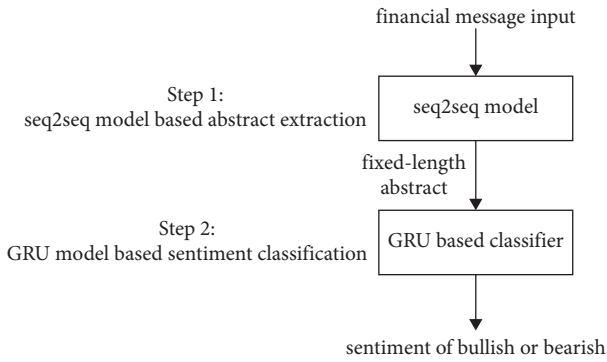
FIGURE 1: The block diagram of the proposed sentiment classification method based on financial messages.

FIGURE 2: The structure of the traditional seq2seq model.

there exists similarity between the translation task and the abstract extraction task, the seq2seq model can also be adopted for abstract extraction. In this section, the characteristics of the seq2seq model are introduced firstly, and then the application of the seq2seq model in abstract extraction is introduced in detail.

The seq2seq model was proposed in [26], which can cope with the situation of different input and output sequence lengths. Therefore, it is suitable for translation tasks in NLP. The basic structure of the seq2seq model is shown in Figure 2, which has an encoder-decoder structure. In this structure, the input of one language is encoded into a vector through the RNN, and then the vector is used as the input of the decoding RNN. Then, the output obtained is another language after translation. The biggest advantage of this model is that it can cope with the input of varying lengths and can learn the mapping relationship between two different domains. Therefore, it is also suitable for abstract extraction herein, that is, from the original financial message domain to the financial abstract domain.

In our implementation, the adoption of the seq2seq model to financial abstract extraction mainly has the following key points:

(1) Word embedding: due to the huge number of words in the financial message domain, direct one-hot encoding may result in large representation dimensions and a huge amount of training parameters, which makes it unsuitable for practical applications. Therefore, it is necessary to introduce the word embedding technology for reducing the dimensionality of the word representation vector. Traditional word embedding methods include the word2vec method, which uses words in the corpus to represent words according to a certain model, such as continuous bag of words (CBOW). After adopting the CBOW model, the obtained word has a lower dimensionality and can better reflect the similarity of the words in the word vector space [27, 28]. Figure 3 shows a schematic diagram of the training samples in the CBOW model. It can be seen that the training samples effectively contain the relevant information of the context, so that the word embedding after training can indicate the relevant information of the context.
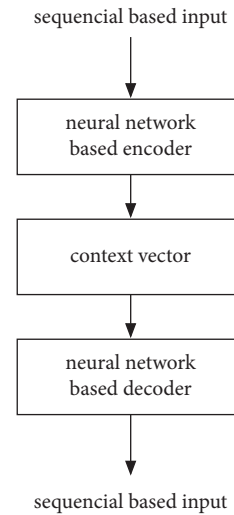
(2) Encoder-decoder structure: the basic encoder-decoder model is shown in Figure 2. In the encoder model, the output is a context vector, which has a low dimension and can effectively summarize the input sequence (in this paper, the input denotes the financial message sequence). The decoder can then output another sequence according to this vector. This sequence can make full use of the context vector information to generate a correlation with the original sequence. Generally speaking, the mentioned encoder and decoder have a similar structure and can both adopt the long-short term memory (LSTM) units to form the RNN. This can make the mentioned RNN capable of memory learning, according to its sequence input and output.

(3) Attention mechanism: the seq2seq model based on the attention mechanism herein is shown in Figure 4. The seq2seq model based on the attention mechanism has been described in detail in [29]. The attention mechanism can solve the problem of information loss in the traditional seq2seq model, that is, when the input sequence is long while the context vector dimension is limited, the context vector from the encoder can lose more information, resulting in inaccurate output.

In the above figure, a_i denotes the word embedding representation of the input financial message sequence, which is pretrained according to the prementioned CBOW model. b_i represents the output abstract word embedding representation, and c_i denotes the context vector, which can be considered as the encoded information of the input sequence. The subscript i denotes the index of the input. Among them, the input length is variable, while the length of the abstract is fixed. In the model, both the encoder and decoder only contain one LSTM layer. The key to the seq2seq model based on the attention mechanism is to calculate the new context vector c_i according to the following formula:

corpus   training samples

word A	word B	word C		

known A and C, to predict B

word A	word B	word C	word D	

known A, C and D to predict B

word A	word B	word C	word D	word E

known A, B, D and E to predict C
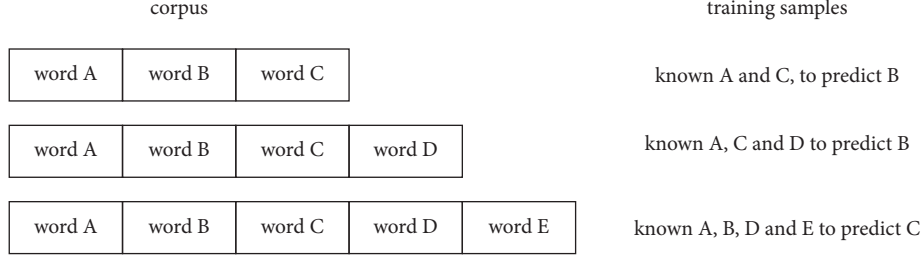
FIGURE 3: Schematic diagram of the training text in the CBOW model.
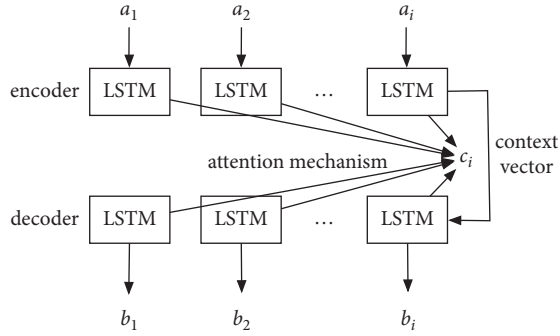
FIGURE 4: The seq2seq model based on attention mechanism.

$$c_i = \sum_{j=1}^{T} \alpha_{ij} h_j, \tag{1}$$

where h_j denotes the hidden variable corresponding to each output of the encoder, the subscript j denotes the output index, and α_{ij} denotes the attention weight, which is determined by e_{ij}:

$$\alpha_{ij} = \frac{\exp e_{ij}}{\sum_{k=1}^{T} \exp e_{ik}}, \tag{2}$$

where e_{ij} is jointly determined by the hidden variables of the encoder s_j and decoder h_j at different times, which can be expressed as

$$e_{ij} = \text{get\_score}(s_j, h_j), \tag{3}$$

where the function of get_score(.) is a scoring function, and different strategies can correspond to different scoring functions.

(4) Predication mechanism: after the training of the seq2seq model, it is necessary to predict the output of the abstract through the input. In the seq2seq model, the output of the previous state will affect the subsequent output. Therefore, if the greedy algorithm is always adopted to calculate the word with the highest output probability, errors may exist, which will affect the overall quality of the abstract. The BeamSearch algorithm [30] can be used to calculate several sequences with the largest continuous output probability to avoid accidental output errors of the decoder from affecting the overall sequence output.

2.2. GRU-Based Sentiment Classification.

The GRU-based model is essentially an instance of the RNN model. The RNN model can obtain the input correlations of sequence input by expanding the calculation graph in the time domain. In the expansion, as the input at different times is processed by the same RNN calculation unit, which shares the same weights, the network is capable of context learning. However, due to the weight sharing and the time expansion characteristics, the RNN tends to have serious gradient disappearance and gradient explosion problems. In order to solve the aforementioned problem and to make the RNN more practical, gated recurrent neural network (gated RNN) was proposed in [31]. In the gated RNN, the calculation through time expansion is redesigned, so that the network can not only accumulate the information of the previous time but also gradually forget the less important information at previous times, thereby reducing redundancy. The most commonly adopted gating unit is the LSTM gating unit. As shown in Figure 5, each LSTM gating unit contains three gates: input gate, forget gate, and output gate. The input gate processes the input data as follows:

$$i_t = \sigma\left(W_i \cdot [h_{t-1}, v_t] + b_i\right), \tag{4}$$

where h_{t-1} represents the output of the gating unit from the previous moment $t - 1$, which can also be regarded as the context vector, v_t represents the input at the current moment t, $\sigma(.)$ denotes the sigmoid activation function, and W_i and b_i denote the parameters of the cell. The output gate separately processes the output of the current unit o_t and the context vector h_t from the current unit to the next unit:

$$\begin{aligned} o_t &= \sigma\left(W_o \cdot [h_{t-1}, v_t] + b_o\right), \\ h_t &= o_t * \tanh(C_t), \end{aligned} \tag{5}$$

where C_t represents the state obtained from the forget gate. The processing of the forget gate can be expressed by the following formulas.

$$\begin{aligned} C_t &= f_t * C_{t-1} + i_t * \widehat{C}_t, \\ \widehat{C}_t &= \tanh\left(W_C \cdot [h_{t-1}, v_t] + b_C\right). \end{aligned} \tag{6}$$

The LSTM gating unit is widely adopted in RNNs because it can effectively avoid the problem of gradient disappearance or explosion. However, since one LSTM includes the above three different gate operations, the training cost is still relatively high.
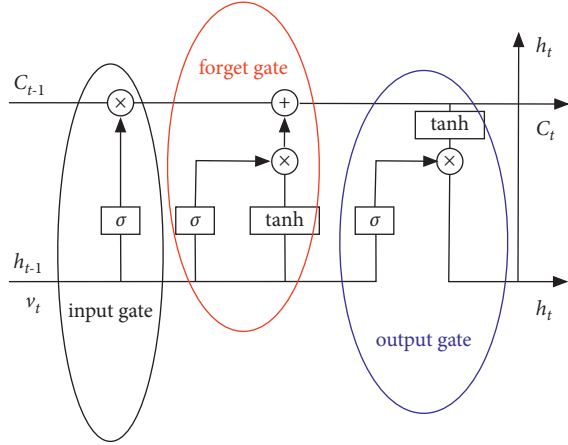
FIGURE 5: The schematic diagram of the traditional LSTM gating unit.

FIGURE 6: The schematic diagram of GRU.

In this paper, gated recurrence unit (GRU) [32] is adopted to replace the traditional LSTM unit. The GRU is an improvement to the LSTM. It contains only two gate-based processes: reset gate and update gate. Its block diagram is shown in Figure 6. The update gate of the GRU contains the functions of the input gate and the forget gate in the LSTM unit.

The restart gate:

$$r_t = \sigma\left(W_r.[h_{t-1}, x_t]\right). \tag{7}$$

The update gate:

$$z_t = \sigma\left(W_r.[h_{t-1}, x_t]\right). \tag{8}$$

The output of the GRU is

$$h_t = \left(1 - z_t\right) * h_{t-1} + z_t * \widehat{h}_t. \tag{9}$$

Among them, $\widehat{h}_t$ can be denoted as

$$\widehat{h}_t = \tanh\left(W.[r_t * h_{t-1}, x_t]\right). \tag{10}$$

3. Results and Discussion

In order to verify the effectiveness of the proposed method, the text messages from the financial social network Stock-Twits is adopted for verification. In this platform, users can post short messages and annotate the messages. There are two types of tags: bullish and bearish. In order to facilitate method comparisons, similar to the experiments in [22], messages from May to September in 2019 are selected to form the corresponding dataset, which has 12 stocks. The dataset contains a total of about 55,000 labeled samples, of which about 39,000 are classified as bullish and about 16,000 are classified as bearish. In our implementation, 80% of the samples are randomly selected to be the training dataset, and the remaining 20% are selected to be the testing dataset.

3.1. The Effects of Different Parameters over Classification Accuracy. According to the proposed method, the abstract is firstly extracted through the seq2seq model, and then sentiment classification is performed according to the extracted abstract. In the above process, different parameter choices will produce different performances. In this section, experiments are conducted on different parameters, and the impact of three different parameter selections on classification accuracy is studied. The parameters are threefold: (1) The dimension of the vector in the word embedding, the chosen dimension is 16, 32, and 64, respectively, noting that these dimensions are commonly seen vector representation dimensions for word embedding. (2) The length of the extracted abstract, the sequence length, varies from 8 to 64, with a step of 8. (3) The number of hidden units in the GRU cell. The following describes the experimental results of the three different parameter selections.

Table 1 shows the classification accuracy corresponding to the three different dimensions of the word embedding representation vectors. It can be seen that in the experiment, when the dimension of the representation vector is 64, the recognition accuracy rate can reach the maximum, which is 81.33%. Compared with the dimensions of 32 and 16, the recognition rate has improved by 0.10% and 15.43%, respectively. It can be seen that when the dimension of the representation vector changes from 32 to 64 dimensions, the effect on the classification accuracy is not obvious. Based on the comprehensive consideration of the classification accuracy and the computational burden, the dimension of the word embedding representation vector is selected as 32.

Figure 7 shows the effect of different lengths of the abstract sequence on the classification accuracy. Note that here the length of the extracted abstract herein is fixed. In the experiment, the sequence length is changed from 8 to 64, with a step of 8. As can be seen from the figure, generally speaking, the classification accuracy increases with the increase of the sequence length. When changing range is from 8 to 40, the increase in recognition accuracy is more obvious. When the dimension of the sequence length is greater than 40, the increase in recognition accuracy becomes less significant. When the sequence length is 40, compared to the sequence length of 8, the recognition rate is increased by 43.63%. When the sequence length is 64, compared to the

Table 1: The effect of different word embedding vector dimensions over the classification accuracy.

Vector dimension	Accuracy (%)
16	65.90
32	81.23
64	81.33
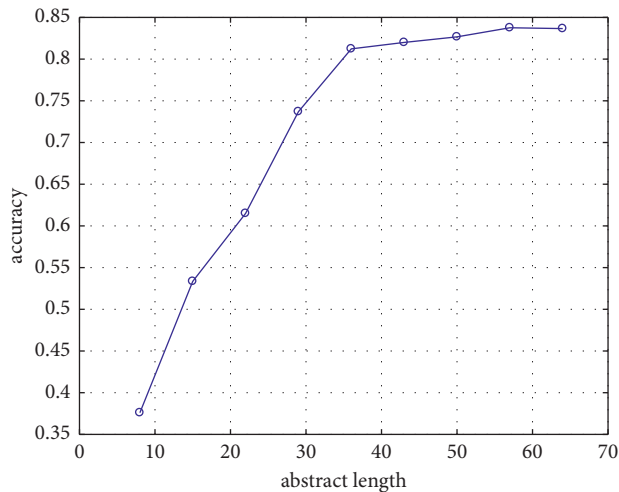
Figure 7: The impact of different abstract sequence lengths on the classification accuracy.

sequence length of 40, the recognition rate is increased by 2.43%. In this paper, considering the limited increase in the recognition rate and the increase in the computational cost, the final abstract sequence length is selected as 40.

Table 2 gives the influence of the number of hidden units in the GRU cells on the recognition accuracy. It can be seen that when the number of hidden units is 64, the recognition accuracy rate is significantly improved compared to the hidden unit number of 16 and 32. But when compared to the number of 128, the improvement in recognition rate is insignificant. Therefore, the number of hidden neurons adopted herein is selected as 64.

3.2. Method Comparisons. The proposed method extracts the abstract of financial messages, which can reduce the influence of redundant information on subsequent classification on one hand, and on the other hand, sentiment classification can be accelerated since the input length is reduced. The proposed method in this paper is compared with two representative methods. One of the methods is based on the traditional SVM classifier. The other method is based on deep learning, which has two obvious differences from the proposed method: (1) the original financial message is directly adopted as the input without abstract extraction and (2) in the sentiment classification step, the gating unit used is the LSTM unit. The correct rate of classification corresponding to the three methods is shown in Table 3. It can be seen that the proposed method has a relatively obvious improvement in the accuracy. Compared with the SVM and LSTM methods, the accuracy of sentiment classification is improved by 5.57% and 2.58%, respectively.

Table 2: The effect of different hidden neuron numbers over the classification accuracy.

Hidden neuron	Accuracy (%)
16	74.76
32	76.70
64	81.23
128	82.43

Table 3: Comparison of the classification accuracy of the proposed method, the traditional SVM method, and the method based on LSTM.

Method	Accuracy (%)
Proposed	81.23
SVM based	75.66
LSTM based	78.65

Table 4: The influence of two different strategies (strategy one: without abstract extraction; strategy two: adopting GRU instead of LSTM).

Method	Accuracy (%)
Proposed	81.23
Strategy one	78.36
Strategy two	80.98

In order to fully illustrate the effectiveness of the proposed method, it is compared with the following two methods with different strategies. In strategy one: without abstract extraction, the GRU gating unit is directly adopted for classification. In strategy two, where abstract extraction is performed, LSTM is adopted for classification instead of GRU. The results of the comparison are shown in Table 4. It can be seen that after abstract extraction, the classification accuracy has been significantly improved, which has increased by approximately 2.87%. This can fully illustrate the effectiveness of the abstract extraction strategy, which can filter out redundant information in the original financial message and extract effective information for sentiment classification. Using GRU instead of LSTM for classification can have a similar classification recognition rate. However, the model using GRU is easier to train than using LSTM. In this paper, the model training time has reduced by about 48%.

4. Conclusions

A new method for financial sentiment classification based on deep learning is proposed in this paper. The proposed method has two steps. (1) The abstract in the financial messages is extracted according to the seq2seq model. The extraction process can deal with the situation of text input with different lengths. After extraction, invalid information in financial messages can be effectively filtered out, and the subsequent classification can be accelerated. (2) After the abstract extraction, sentiment classification is performed through the GRU model according to the abstracts. The proposed method in this paper is verified through financial message dataset from the financial social network

StockTwits. The results show that compared with the classical SVM and LSTM methods, the proposed method in this paper can improve the accuracy of sentiment classification by 5.57% and 2.58%, respectively. Moreover, the effectiveness of the proposed method can be proved with the strategy of adopting abstract extraction and adopting GRU. Compared with the strategy of not extracting abstract, the recognition rate is improved by 2.87%. After adopting GRU instead of LSTM, the training time of the model has reduced by about 48%.

Data Availability

The data adopted in the paper are available on the following website: https://stocktwits.com.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] A. S. Wafi, H. Hassan, and A. Mabrouk, "Fundamental analysis models in financial markets - review study," *Procedia economics and finance*, vol. 30, pp. 939–947, 2015.

[2] P. Mohanram, S. Saiy, and D. Vyas, "Fundamental analysis of banks: the use of financial statement information to screen winners from losers," *Review of Accounting Studies*, vol. 23, no. 1, pp. 200–233, 2018.

[3] C. Abad, S. A. Thore, and J. Laffarga, "Fundamental analysis of stocks by two-stage DEA," *Managerial and Decision Economics*, vol. 25, no. 5, pp. 231–241, 2004.

[4] Kirkpatrick II, D. Charles, and J. A. Dahlquist, *Technical Analysis: The Complete Resource for Financial Market Technicians*, FT press, New Jersey, NJ, USA, 2010.

[5] M. P. Taylor and H. Allen, "The use of technical analysis in the foreign exchange market," *Journal of International Money and Finance*, vol. 11, no. 3, pp. 304–314, 1992.

[6] G. A. W. Griffioen, *Technical Analysis in Financial Markets*, Pearson Professional Education, Aminijikarai, Chennai, 2003.

[7] R. P. Schumaker, Y. Zhang, C.-N. Huang, and H. Chen, "Evaluating sentiment in financial news articles," *Decision Support Systems*, vol. 53, no. 3, pp. 458–464, 2012.

[8] R. P. Schumaker and H. Chen, "Textual analysis of stock market prediction using breaking financial news," *ACM Transactions on Information Systems*, vol. 27, no. 2, pp. 1–19, 2009.

[9] E. L. Giles, M Holmes, E McColl, F. F Sniehotta, and J. M Adams, "Acceptability of financial incentives for breastfeeding: thematic analysis of readers' comments to UK online news reports," *BMC Pregnancy and Childbirth*, vol. 15, no. 1, pp. 509–513, 2015.

[10] M. H. Pesaran, T. Schuermann, and L. V. Smith, "Rejoinder to comments on forecasting economic and financial variables with global VARs," *International Journal of Forecasting*, vol. 25, no. 4, pp. 703–715, 2009.

[11] V. Fernández-Pérez, P. E. Alonso-Galicia, M. del Mar Fuentes-Fuentes, and L. R. Ariza, "Business social networks and academics' entrepreneurial intentions," *Industrial Management & Data Systems*, vol. 114, no. 2, 2014.

[12] X. Gine, D. Karlan, and M. Ngatia, *Social Networks, Financial Literacy and index Insurance*, World Bank, Washington, DC, USA, 2013.

[13] S. C. Dow, "Cognition, market sentiment and financial instability," *Cambridge Journal of Economics*, vol. 35, no. 2, pp. 233–249, 2011.

[14] R. Moraes, J. F. Valiati, and W. P. Gavião Neto, "Document-level sentiment classification: an empirical comparison between SVM and ANN," *Expert Systems with Applications*, vol. 40, no. 2, pp. 621–633, 2013.

[15] J. Khairnar and M. Kinikar, "Machine learning algorithms for opinion mining and sentiment classification," *International Journal of Scientific and Research Publications*, vol. 3, no. 6, pp. 1–6, 2013.

[16] N. Chirawichitchai, "Sentiment classification by a hybrid method of greedy search and multinomial naïve bayes algorithm," in *Proceedings of the 2013 Eleventh international conference on ICT and knowledge engineering*, pp. 1–4, IEEE, Bangkok, Thailand, November 2013.

[17] E. Hossain, O. Sharif, and M. Moshiul Hoque, "Sentiment polarity detection on Bengali book reviews using multinomial naïve bayes," *Progress in Advanced Computing and Intelligent Engineering*, Springer, Singapore, pp. 281–292, 2021.

[18] P. Karthika, R. Murugeswari, and R. Manoranjithem, "Sentiment analysis of social media network using random forest algorithm," in *Proceedings of the 2019 IEEE International Conference on Intelligent Techniques in Control, Optimization and Signal Processing (INCOS)*, pp. 1–5, IEEE, Tamilnadu, India, April 2019.

[19] S. Sridhar and S. Sanagavarapu, "Analysis of the effect of news sentiment on stock market prices through event embedding," in *Proceedings of the 2021 16th Conference on Computer Science and Intelligence Systems (FedCSIS)*, Sofia, Bulgaria, September 2021.

[20] H. Zhang, H. Xu, J. Shi, T. Liu, and J. Ya, "Sfv-cnn: deep text sentiment classification with scenario feature representation," in *Proceedings of the International Conference on Mathematical Aspects of Computer and Information Sciences*, pp. 382–394, Springer, Gebze, Turkey, November 2019.

[21] J. Z. G. Hiew, X. Huang, H. Mou, D. Li, Q. Wu, and Y. Xu, "BERT-based financial sentiment index and LSTM-based stock return predictability," arXiv preprint arXiv:1906.09024, 2019.

[22] S. Yıldırım, D. Jothimani, C. Kavaklıoğlu, and A. Basar, "Deep learning approaches for sentiment analysis on financial microblog dataset," in *Proceedings of the 2019 IEEE International Conference on Big Data (Big Data)*, pp. 5581–5584, IEEE, Los Angeles, CA, USA, December 2019.

[23] K. M. El Hindi, R. R. Aljulaidan, H. AlSalman, and H. AlSalman, "Lazy fine-tuning algorithms for naïve Bayesian text classification," *Applied Soft Computing*, vol. 96, Article ID 106652, 2020.

[24] H. Sagha, N. Cummins, and B. Schuller, "Stacked denoising autoencoders for sentiment analysis: a review," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 7, no. 5, p. e1212, 2017.

[25] Y. Zhang, J. Li, H. Wang, and S. C. Choi, "Sentiment-guided adversarial learning for stock price prediction," *Frontiers in Applied Mathematics and Statistics*, vol. 7, p. 8, 2021.

[26] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," *Advances in Neural Information Processing Systems*, pp. 3104–3112, 2014.

[27] T. Kenter, A. Borisov, and M. De Rijke, "Siamese cbow: optimizing word embeddings for sentence representations," 2016, https://arxiv.org/abs/1606.04640.

[28] F. Mai, L. Galke, and A. Scherp, "CBOW is not all you need: combining CBOW with the compositional matrix space model," 2019, https://arxiv.org/abs/1902.06423.

[29] H. Bahuleyan, L. Mou, O. Vechtomova, and P. Poupart, "Variational attention for sequence-to-sequence models," 2017, https://arxiv.org/abs/1712.08207.

[30] S. Wiseman and A. M. Rush, "Sequence-to-sequence learning as beam-search optimization," 2016, https://arxiv.org/abs/1606.02960.

[31] G.-B. Zhou, J. Wu, C.-L. Zhang, and Z.-H. Zhou, "Minimal gated unit for recurrent neural networks," *International Journal of Automation and Computing*, vol. 13, no. 3, pp. 226–234, 2016.

[32] R. Dey and F. M. Salem, "Gate-variants of gated recurrent unit (GRU) neural networks," in *Proceedings of the 2017 IEEE 60th international midwest symposium on circuits and systems (MWSCAS)*, pp. 1597–1600, IEEE, Boston, MA, USA, August 2017.

Retraction

Retracted: Construction of College English Mobile Learning Model Based on Distributed Terminal

Scientific Programming

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope

(2) Discrepancies in the description of the research reported

(3) Discrepancies between the availability of data and the research described

(4) Inappropriate citations

(5) Incoherent, meaningless and/or irrelevant content included in the article

(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] J. Wang and L. Lei, "Construction of College English Mobile Learning Model Based on Distributed Terminal," *Scientific Programming*, vol. 2021, Article ID 2220096, 7 pages, 2021.

Research Article

Construction of College English Mobile Learning Model Based on Distributed Terminal

Jianmei Wang[1] and Linhai Lei [2]

[1]*Taiyuan University, Taiyuan 0300321, Shanxi, China*
[2]*Taiyuan Normal University, Taiyuan 0306192, Shanxi, China*

Correspondence should be addressed to Linhai Lei; leilinhai@tynu.edu.cn

The quality of English teaching cannot be separated from the students's accumulation and mastery of vocabulary, and sufficient vocabulary can ensure that students feel easy and comfortable in the language application. However, there are still some problems in teaching English vocabulary in colleges and universities, such as single presentation method, lack of real context, and other difficulties. With the rapid development of mobile Internet, the share of smart mobile devices among college students has increased, which in turn has laid the foundation for the development of new learning modes. The mobile terminal-mediated learning mode in English classroom teaching has become a hot topic of research nowadays, of which English vocabulary teaching has become a typical example. This paper is designed as a system model of edge computing, in which the user terminal device can connect wirelessly with each MECS and offload the tasks to be processed on the MECS, which helps the user terminal device to process and return the results to the user terminal device. The research in this paper also provides a basis for students' independent and lifelong learning of English words and for schools' information technology to serve English vocabulary teaching and learning.

1. Introduction

Today, we are in an era of high informationization and globalization; how to efficiently collect and process information has become a necessary skill [1]. There are three components to any language: phonology, vocabulary, and grammar. Michael Lewis said that language is made up of grammatical words, not lexicalized grammar. This shows the importance of the English vocabulary in English language teaching [2]. However, although experts and front-line English teachers have put forward many strategies to reform teaching in junior high schools over the years, such as contextual teaching method and game teaching method, it is difficult for these vocabulary teaching methods to give full play to their theoretical advantages in the context of traditional teaching and to achieve efficient English vocabulary teaching efficiency [3].

The contemporary junior high school students are in an era of network and information technology, and the communication between them has long been not. They no longer stay on paper or face-to-face mode, but a kind of instant communication mode [4]. Mobile terminals such as smartphones and tablet PCs have become popular, and mobile terminal-based English learning third-party applications have become increasingly popular [3], such as Apps for listening, speaking, reading, and writing, and have emerged in large numbers [5]. Mobile terminals have incomparable advantages of traditional vocabulary teaching, such as portability and no time and space limitations, and [6] believes that vocabulary learning is a continuous process with a lifelong nature, and the application of mobile terminals to junior high-school English vocabulary teaching is to cultivate secondary school students' lifelong ability to learn English vocabulary. What are the impacts of mobile terminals on English vocabulary teaching in junior high schools and what are the corresponding strategies to face these impacts? This is a question that needs to be addressed.

Through this study, the author hopes that the majority of English teachers will fully realize the magnitude of the impact of mobile terminals on junior high school English vocabulary teaching in the era of instant communication and given the characteristics of learning and communication of junior high school students in that era [7]. At the same time, the research in this paper provides a basis for students' independent and lifelong learning of English words, and a basis for the construction of information technology in schools to serve the teaching of English vocabulary.

The study also provides a basis for students' independent and lifelong learning of English vocabulary and for the construction of information technology in schools to serve English vocabulary teaching and learning.

The contributions of this paper are as follows:

(1) This paper is designed as a system model of edge computing. The user terminal equipment can wirelessly connect with each MECS and unload the tasks to be unloaded to MECs for processing. MECs help the user terminal equipment process and return the results to the user terminal equipment.

(2) Mobile learning and mobile education are analyzed and summarized. Learning with mobile terminal equipment is a new learning method. With the support of network communication, learners can accept learning anytime and anywhere, and then cultivate learners' awareness, attitude, and habit of lifelong learning.

(3) A large number of experiments have been performed to verify that the problems of college English vocabulary teaching are more concentrated, and show new trends and characteristics, which are summarized by the lack of attention and investment in English vocabulary teaching.

2. Development Status of Mobile Teaching and Learning at Home and Abroad

2.1. Foreign Research Status and Results. In recent years, research on the main areas of mobile learning is growing rapidly, mainly focusing on the feasibility of mobile learning, the development of mobile learning resources, the construction of WAP education sites, and short message services, and as an effective aid to lifelong and collaborative learning [8]. The main areas of research in recent years are growing rapidly, mainly focusing on the feasibility of mobile learning, the development of mobile learning resources, the construction of WAP education sites, short message services and as an effective aid to lifelong learning and collaborative learning [9].

Keegan divides distance learning into three stages according to the form and means of learning: distance learning, e-learning, and M-learning [10]. Compared with the theories of Nipper (Denmark), Bytes (UK), and Taylor (Australia) on "three developmental eras of distance learning," Keegan's theory solves the problem of coexistence of distance learning, e-learning, and mobile learning [11].

Professor Mike Sharples, President of the International Academy for Mobile Learning (IAML) [12–14] discussed how lifelong learning and mobile learning can contribute to each other, what mobile learning means for the traditional classroom, and how to use it to design classroom activities. The study in [15] proposed centralized management for the current situation of mobile learning resources in China and established quality specifications and interoperability standards.

3. Related Work

At present, domestic research is still mainly theoretical, mainly in the basic theory and an overview of mobile learning, technical support, resource development, application objects, and teaching models. In 2008, RongHuai Huang and Jyri Salomaa edited the first domestic monograph on mobile learning: "Mobile Learning - Theory - Current Situation - Trends," which comprehensively introduced the theory and characteristics of mobile learning and technology, sorted out domestic and foreign mobile learning cases and projects, and explained various application models [16]. Finally, it outlooks its future development. After that, scholars in China have published treatises on mobile learning, such as "Mobile Learning Theory and Practice" edited by [17].

In 2008, Jiaoqing Mo explored and researched the development technology of the mobile learning system, aiming to build a mobile learning system based on the JZME/JZEE platform and GPRS network, in order to extend the application of the web-based learning system in the wireless field [18].

In terms of resource development, the study in [8] designed a case study of English word learning resources based on NET mobile development technology in "Design and Research of Mobile Learning and its Resources," based on the discussion of the theory of mobile learning resources and its special features. The study in [19] in "Research on the Design of Fragmentary Learning Resources in Mobile Learning," through the analysis of a large amount of data to derive based on the clustering results, explored the feasibility and specific design strategies of fragmented learning resources in mobile learning.

In terms of the application target, it is mainly focused on two parts: one part is school students, covering students of different ages including young children, primary and secondary school students, students in higher education institutions, college students, and graduate students [20]. The study in [21] explored the impact of mobile learning on higher education and the countermeasures in teaching resources construction and teaching process management. The study in [22] conducted a survey on the current situation of mobile learning for adult learners. In terms of teaching mode, due to the characteristics of English learning, the application of mobile terminals to English learning has incomparable advantages. The study in [23] proposed four models of mobile learning in "The Application of Mobile Learning in College English Vocabulary Learning." The study in [24] discussed the content selection, teaching

process, and evaluation of mobile learning and analyzed the details of the application of mobile learning mode in high school English.

The study in [25] used the UTAUT model as the basis to empirically analyze the factors influencing college students' acceptance of mobile learning through a questionnaire survey of 593 college students at Beihang University and multiple regression. The study in [26] empirically analyzed the demand for mobile learning by sampling college students from Nanjing, Beijing, Shanghai, and other universities across China. The study in [27] studied the application of mobile learning in vocational education by taking the example of "Microcontroller Principle Technology." The study in [28] studied the application of mobile education in medical examination teaching.

3.1. Related Concept Definition

3.1.1. Mobile Learning and Mobile Education.
There is no uniform definition of mobile learning [15]. Chara and Figueiredo, in the context of distance education, define mobile learning as the ability to use any device, at anytime, anywhere, to receive learning. As an extension of online education, mobile learning not only has the basic characteristics of online education, but also has its own unique characteristics, as follows:

(1) Networked: mobile learning is realized through the access of mobile terminals to the network. So mobile learning has the characteristics of network education.

(2) Universality: with the continuous development of technology, mobile terminals such as smartphones have flooded into our lives and become popular, providing material support for the popularization of mobile learning and contributing to the realization of educational equity.

(3) Ubiquity: the so-called ubiquity is 3A (anyone, anytime, anywhere).

(4) Personalization: mobile learning can provide personalized education services according to learners' own learning styles and interests, and is also conducive to teachers' personalized teaching and hierarchical teaching [17].

(5) Hypermedia: the learning resources of mobile learning are digital contents, presented in the form of multimedia (such as text, pictures, audio, and video).

According to Professor Cui Guangzuo of Peking University and others, "Mobile education refers to the more convenient and flexible interactive teaching and learning activities by students and teachers through the use of mobile devices (such as cell phones), relying on the current relatively mature wireless mobile network, international interconnection network and multimedia technology."

In the United States and Europe, mobile learning is referred to as m-learning, short for mobile learning, or mobile education, short for M-education. Mobile education is education delivered by means of mobile learning.

3.1.2. Mobile Terminals.
Mobile terminal refers to a computer device that can be used on the move, including cell phones, laptops, and POS machines in a broad sense. But most of the time, it refers to smartphone and tablet personal computer. Smartphone users can install third-party applications such as education, shopping, and gaming applications, depending on the operating system, so that their functions can be expanded [22].

3.1.3. Third-Party Applications (Apps).
Third-party applications for smartphones (the full name is application, abbreviated as App). From system security, communication, audio-visual, photography and video, travel, and hotel to education and learning, a wide variety of Apps are available to meet the needs of cell phone users for personalized software.

3.1.4. System Scenes.
Figure 1 shows the system model of edge computing, where the user end device can connect wirelessly with each MECS and offload the tasks to be processed on the MECS, and the MECS helps the user end device to process and return the results to the user end device.

In this paper, we assume that the system consists of 1 user terminal device and n MECSs. The user terminal device has m tasks to be processed, which can be offloaded to the MECS for processing or processed locally. The set of MECS is modeled as $mecs_j = (f_j, P_j^D, rs_j)$, and the j-th MECS is modeled as f_j, where P_j^D is the processing power of the j-th MECS. The user terminal device will be modeled as $JUser = (f_1^c, P_1, P^U)$, where f_1^c represents the computing power of the user terminal device.

The task set of the user terminal device is modeled as $TASK = \{task_1, tank_2, \ldots, task_m\}$, and the i-th task is modeled as $task_i = (U_i, D_i, p, r_i^s)$, where U_i denotes the data volume of the i-th task, D_i denotes the data volume of the i-th task processing result, p denotes the number of CPU cycles required by the CPU of the user terminal device or MECS to process each bit of task, and r_i^s denotes the number of MECS resources required to process the i-th task. The offloading decision of m tasks is represented as a vector $X = \{x_1, x_2, \ldots, x_m\}$, which means that the i-th task is executed locally [25].

This paper makes the following basic assumptions:

(1) Each task of the user terminal device cannot be split, and there is no dependency between the tasks.

(2) The user terminal device can send tasks to the corresponding MECS for processing through different channels in parallel.

(3) The states of the channel during the process of uploading tasks to the MECS and returning the processing results to the user terminal device are the same.
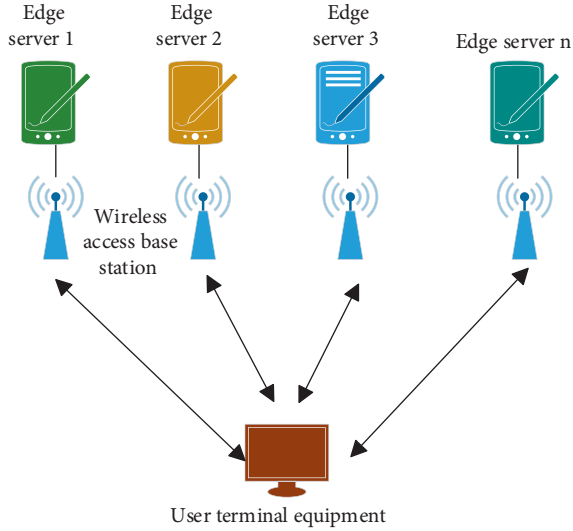
FIGURE 1: System model in edge computing scenario.

(4) The computing power of the user terminal device and the MECS remains stable during the task processing.

(5) MECS is a multithreaded server; that is, MECS can handle multiple user terminal device tasks simultaneously.

(6) The system is in a quasi-static environment; that is, the number of tasks remains constant when the user terminal device is offloading tasks, but the number of tasks of the user terminal device changes at different time periods.

3.1.5. Energy Consumption. Assume that the energy consumption of the i-th task of the user terminal device processed at the user terminal device is E_i^1, which is given as follows:

$$E_i^1 = t_i^1 * P_1. \tag{1}$$

The second category is the energy consumption generated by the user terminal device when offloading tasks to MECS for processing, which includes the transmission energy generated by the local user terminal device when the user terminal device sends tasks to MECS, the energy consumption generated by MECS when processing tasks of the user terminal device, and the energy consumption generated by MECS when returning calculation results to the user terminal device. Therefore, this paper mainly considers the energy consumption generated by the local user terminal device when uploading data to MECS and assumes that the energy consumption of the user terminal device when uploading the i-th task is E_i^U, which is given as follows:

$$E_i^U = P^U * t_i^U. \tag{2}$$

Then, the energy consumption of the system to process the i-th task E_i is the computational energy consumption of

the user terminal device to process the task and the energy consumption generated by uploading the task of MECS, which is as follows:

$$E_i = (1 - s_i) * E_i^1 + s_i * E_i^U. \tag{3}$$

Therefore, the energy consumption E generated by the system is given in the following equation:

$$E = \sum_{i=1}^{m} E_i. \tag{4}$$

3.1.6. Problem Model. In edge computing scenarios, latency and energy consumption are the two most commonly used metrics to measure the performance of offloading schemes. In this paper, delay and energy consumption are considered together, and since delay and energy consumption are of different magnitude, they need to be normalized. Assuming that the size of the particle swarm optimized for this problem is N, the delay and energy consumption of the system for the kth offloading policy to handle the user terminal device task are E_k and T_k', and the normalized results T_k' and T_E' are given in the following equations:

$$T_k' = \begin{cases} \dfrac{T_k - T_{\min}}{T_{\max} - T_{\min}}, & T_{\min} \neq T_{\max}, \\ \\ 1, & T_{\min} = T_{\max}, \end{cases} \tag{5}$$

$$E_k' = \begin{cases} \dfrac{E_k - E_{\min}}{E_{\max} - E_{\min}}, & E_{\min} \neq E_{\max}, \\ \\ 1, & E_{\min} = E_{\max}, \end{cases} \tag{6}$$

where $T_{\min}, T_{\max}, E_{\min},$ and $E_{\max}$ denote the minimum delay, maximum delay, minimum energy consumption, and maximum energy consumption generated under all offloading policies, respectively [10]. Assume that the total overhead of processing user terminal device tasks in this system under the kth offload policy is $W(X_k)$, which is given as follows:

$$W(X_k) = \lambda * T_k' + (1 - \lambda) * E_k'. \tag{7}$$

Here, the weight $\lambda \in [0, 1]$.

In this paper, the objective is to minimize the weighted sum of system delay and energy consumption, which is defined as the problem of P :

$$\begin{cases} P: & \min_{k=1} N(X_k) \\ \text{s.t.} & \sum_{i=1}^{m} \varphi_i^j r_i^s \leq rs_j, \end{cases} \tag{8}$$

where equation (8) indicates that the sum of computing resources allocated by the j-th MECS to user terminal device tasks cannot exceed the total resources owned by the j-th MECS.

3.1.7. Case Test. Mobile learning and mobile education. There is no unified definition of mobile learning and mobile education, but Finnish scholars define mobile learning as a
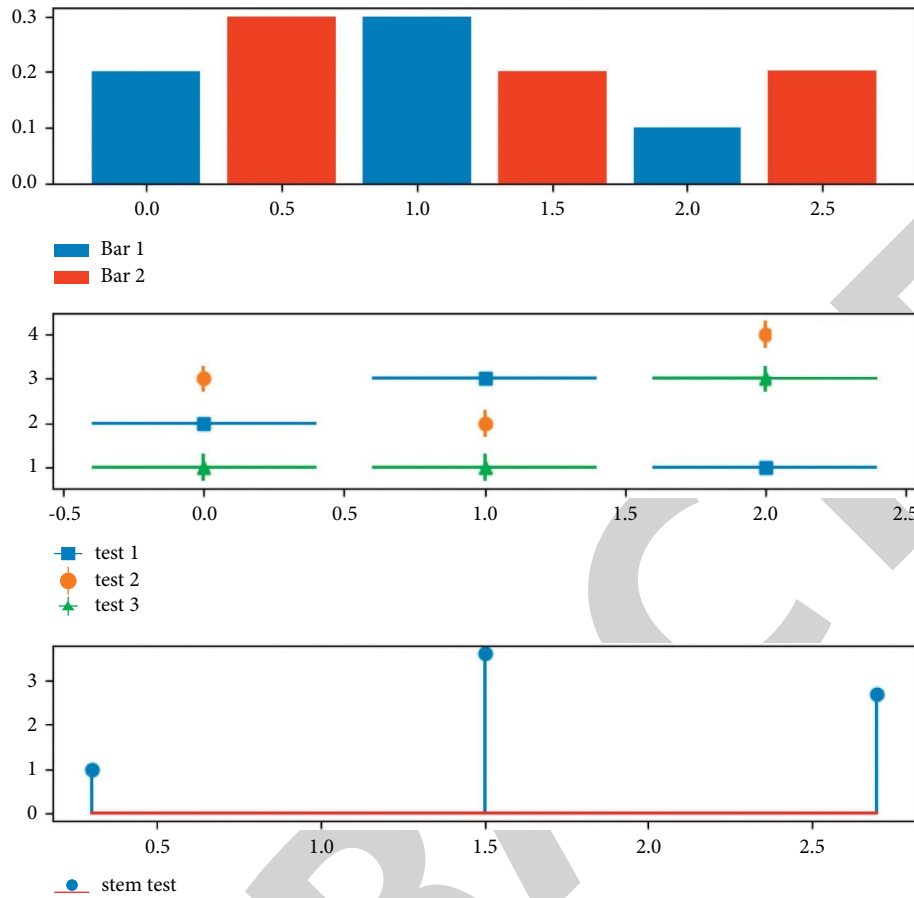
Figure 2: Different test results.

new type of learning method that uses mobile devices for learning due to the increase of people's geospatial mobility and flexible learning needs. In China, mobile learning mainly refers to the use of modern mobile devices, supported by the guarantee of network communication, so that learners can receive learning anytime and anywhere, and then develop the awareness, attitude, and habit of lifelong learning [1].

As shown in Figure 2, no physical tests, such as mobile terminals, are allowed. A mobile terminal is a computer device that can be used on the move. Broadly speaking, it includes cell phones, tablets, and laptops. In most cases, it refers to smartphones and tablets. Speakerphones and tablets can be installed with different apps according to users' needs, which can be mainly divided into three applications such as education, shopping, and games. These are also the most used and widely used programs in the study life of students and social people.

Again, as shown in Figure 3 that the existing vocabulary teaching mode is more traditional, even comparable to compulsory and high school teaching, and lacks innovation and flexibility in the selection of ways and means. In particular, the focus is on the effect of memorization, while there is no targeted response to the penetration of the comprehension link, resulting in students' fear and denigration of vocabulary learning. The monolithic teaching mechanism

not only lacks the corresponding contextual guarantee, but also is highly susceptible to forgetting [3].

As shown in Figure 4, there is a double impact on the communication between teachers and students. First of all, in the past, teachers had absolute authority in English classroom vocabulary learning, and students' main position could not be guaranteed, and generally, one-way communication between teachers and students was the main focus. In particular, it increases the opportunities for teachers to listen to students and communicate democratically, thus helping students to build their knowledge of vocabulary learning and stimulate their potential. Second, with the popularity of mobile terminals, teachers and students can effectively communicate with each other in real time by using WeChat, Weibo, QQ, and other Apps, which not only bring the psychological distance between teachers and students closer, but also personalize the learning environment and provide solutions to different chemical situations according to the students' needs, which makes the status between teachers and students equal and the students' resistance gradually disappears. Finally, instant messaging software can easily extend classroom teaching to after-school one-to-one or one-to-many, especially on the basis of text, and the use of voice tone can enter the virtual classroom scenario, which in turn plays a role in enhancing the visual effect of students, and this
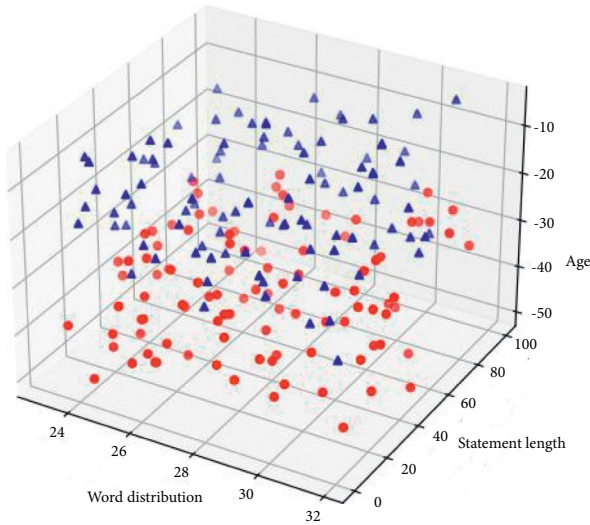
Figure 3: Learning ability at different ages.
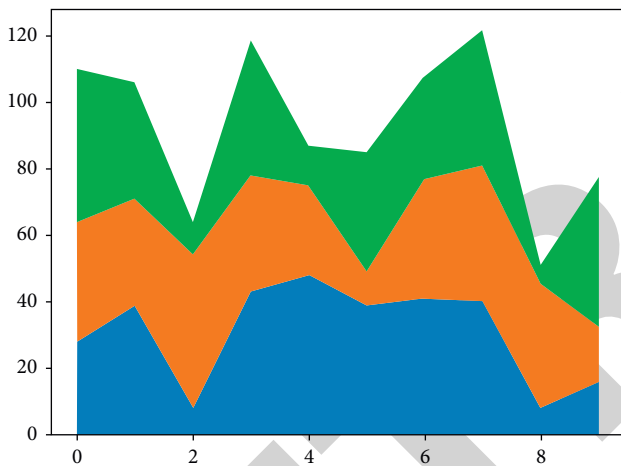
Figure 4: Distributed acceptance by different students.

emotional resonance is unmatched by previous communication mechanisms.

4. Conclusions

With the rapid development of mobile Internet, the share of smart mobile devices among college students has increased, thus laying the foundation for the development of new learning modes. The use of mobile devices as a medium for learning in English classroom teaching has become a hot topic of research, and English vocabulary teaching has become a typical example. This paper is designed as a system model of edge computing, in which the user terminal device can connect wirelessly with each MECS and offload the tasks to be processed on the MECS, which helps the user terminal device to process and return the results to the user terminal device. The research in this paper also provides a basis for students' independent and lifelong learning of English words, and schools' information technology to serve English vocabulary teaching and learning.

References

[1] X. Huang, Y. Yao, and Q. Hu, "Research on the rapid slicing algorithm for NC milling based on STL model," *AsiaSim 2012 Communications in Computer and Information Science*, vol. 325, pp. 263–271, 2012.

[2] V. L. Bond, "Music's representation in early childhood education journals," *UPDATE: Applications of Research in Music Education*, vol. 31, no. 1, pp. 34–43, 2012.

[3] J. P. Mackenbach, J. R. Valverde, M. Bopp et al., "Determinants of inequalities in life expectancy: an international comparative study of eight risk factors," *The Lancet Public Health*, vol. 4, no. 10, pp. e529–e537, 2019.

[4] S. Kreitz-Sandberg, "The influence of sociocultural and environmental factors on the adolescents' life and development: Japanese youth from a Western readers perspective," *Asia Europe Journal*, vol. 5, no. 4, pp. 499–517, 2008.

[5] P. Giardullo, "Does ""p," *Quality and Quantity*, vol. 50, no. 2, pp. 529–547, 2016.

[6] S. S. Ali, M. M. Riaz, A. Ghafoor, U. Javed, M. W. Baig, and M. U. Arif, "Takagi-sugeno fuzzy system and MTF-based ps," *Soft Computing*, vol. 20, no. 12, pp. 4695–4708, 2016.

[7] R. Marques and P. B. Lourenço, "A model for pushover analysis of confined masonry structures: implementation and validation," *Bulletin of Earthquake Engineering*, vol. 11, no. 6, pp. 2133–2150, 2013.

[8] S. Zhou and T. Zhang, "Research on the construction of flipped classroom model for English teaching based on SPOC," *Revista de la Facultad de Ingenieria*, vol. 32, no. 14, pp. 267–273, 2017.

[9] G. Fayolle, W. Levick, and R. Lajiness-O'Neill, "Amnesic syndrome following a suspected AVM in a 14 year old female: course of recovery, etiology, and MRI findings," *Archives of Clinical Neuropsychology*, vol. 27, no. 6, p. 576, 2012.

[10] C. H. Cao, Y. N. Tang, D. Y. Huang, G. WeiMin, and Z. Chunjiong, "IIBE: an improved identity-based encryption algorithm for wsn security," *Security and Communication Networks*, vol. 2021, Article ID 8527068, 8 pages, 2021.

[11] D. Wu, C. Zhang, L. Ji, R. Ran, H. Wu, and Y. Xu, "Forest fire recognition based on feature extraction from multi-view

Retraction

Retracted: Construction of Machine Learning Model Based on Text Mining and Ranking of Meituan Merchants

Scientific Programming

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope

(2) Discrepancies in the description of the research reported

(3) Discrepancies between the availability of data and the research described

(4) Inappropriate citations

(5) Incoherent, meaningless and/or irrelevant content included in the article

(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] Y. Tang, D. Liao, S. Huang, Q. Fan, and L. Liu, "Construction of Machine Learning Model Based on Text Mining and Ranking of Meituan Merchants," *Scientific Programming*, vol. 2021, Article ID 5165115, 9 pages, 2021.

Research Article

Construction of Machine Learning Model Based on Text Mining and Ranking of Meituan Merchants

Yin Tang,[1] Dongxue Liao,[1] Shuqiang Huang,[2,3] Qing Fan,[4] and Liang Liu [5]

[1]*Management School, Jinan University, Guangzhou 510632, Guangdong, China*
[2]*College of Science and Engineering, Jinan University, Guangzhou 510632, Guangdong, China*
[3]*Guangdong Provincial Key Laboratory of Public Finance and Taxation with Big Data Application, Guangzhou 510320, Guangdong, China*
[4]*School of Economics and Management, Southwest Jiaotong University, Chengdu 610031, Sichuan, China*
[5]*Venture Capital Research Center, South China University of Technology, Guangzhou 510632, Guangdong, China*

Correspondence should be addressed to Liang Liu; dongxuel@stu2018.jnu.edu.cn

In the Web 2.0 era, the problem of uneven quality and overload of online reviews is very serious, and the cognitive cost of obtaining valuable content from them is getting higher and higher. This paper explores an effective solution to address comment overload by means of information recommendation in order to improve the utilization of online information and information service quality. This paper proposes a review ranking recommendation scheme that focuses on the information quality of reviews and places more emphasis on satisfying users' personal information need. The paper's approach is used to extract and rank low-frequency keywords that appear only once in the comment set. The more useful the extracted phrases are, the more useful this review will be and the higher the usefulness votes will be, which can reflect the actual situation of this product more objectively and accurately and facilitate better consumption decisions for consumers. The experimental results show that users' satisfaction with the perceived usefulness of the reviews is jointly influenced by the information quality of Meituan's reviews and users' individual information needs; the recommendation strategy achieves the organic integration of the two, and the evaluation results under three different recommendation modes show that compared with "interest recommendation" and "utility recommendation," the satisfaction score of "fusion recommendation" is the highest

1. Introduction

With the rapid growth of the Internet and e-commerce platforms in recent years, the usefulness of online reviews has become an important influencing factor in consumer decision making [1]. Online reviews are users' evaluations and experience after experiencing commercial products and services and providing valuable information to other users. Users can learn about merchants' products and services through online reviews, which help them make better consumer decisions and reduce the reference cost of products and services. The famous Jupiter Research company, through years of research and analysis, found that 75% of consumers refer to reviews on the Internet before spending money on dining, travel, and accommodation, purchasing goods, parent-child playgrounds, and many other things. The same is true in China, with platforms such as Taobao, Jingdong, Meituan, and Where to Go [2]. Due to the openness of the Internet, the cost of posting online reviews is very low, and a lot of spam and false information make the quality of information in reviews vary, resulting in a large number of reviews, which is noisy and difficult to distinguish, and there are many ways of reviews and different language expressions, and some reviews do not bring us useful reference value [3].

"Taobao" uses whether there is a picture, whether there is a follow-up review, and the rating of the product as the filtering criteria; "public review network" blocks untrustworthy content based on user feedback; "Douban" and "Amazon" use user votes to sort reviews [4]. These filtering strategies focus on information quality and help users quickly access useful information by placing high-quality reviews at the top. Nevertheless, these filtering strategies do not focus on satisfying individual users' needs [5]. The adoption of information by individuals, besides being influenced by the quality of information, is related to individual information need, and people will care more about whether the information they receive contains content of interest to them. Especially when the amount of information exceeds one's cognitive load, people browse quickly and hope to find the content they are interested in as soon as possible.

In this paper, we propose a low-frequency keyword extraction method for review usefulness voting. The main purpose is to identify low-frequency keywords from the reviews of Meituan and to provide consumers with more choices and decisions through the study of usefulness voting, instead of just looking at the star rating given by users as the judgment index (usually five stars). Therefore, the identification and extraction of low-frequency keywords become a major difficulty for us, which mainly has the following three problems:

(1) The cohesiveness among the parts of low-frequency keywords is weak, and it is impossible to calculate the mutual information among them.

(2) Since the combination of low-frequency keywords is evaluated randomly from the perspective of probability, it is difficult to use machine learning methods by means of labeling.

(3) Low-frequency keywords also have the problem of representation, because of the low number of occurrences and the lack of contextual information. It is difficult to represent them by existing representation methods (e.g., Word2Vector).

Based on the above difficulties, there are still no more studies on the effectiveness of comment voting, which will become a key topic for our research.

2. Related Work

2.1. A Study of Reviewing Ranking and Recommendation Based on Reviewing Utility. The essence of the review ranking is to evaluate the utility of reviews and generate a Top N recommendation list based on the utility evaluation. In recent studies, [6] used fuzzy hierarchical analysis and weighted gray correlation analysis to predict the review utility, rank the reviews accordingly, and select the reviews with high information content for final recommendation. Jiang and Mccomas [7] used K-means algorithm to rank the review utility and then optimize the review ranking. Korde [8] calculated the credibility of reviews based on the number of "feature-opinion" pairs in the reviews and then invited users to evaluate the Top N reviews by questionnaire. Wen-Hsiang et al. [9] concluded that the authors' historical reviews reflect the quality of his or her published reviews and they modeled them based on the authors' previous reviews and incorporated them into the review model. It can be seen that the ranking and recommendation of reviews are mainly based on the calculation of evaluation metrics. In these studies, the evaluation metrics focus on a series of elements such as the information and content of the review, the credibility, the level of the writer, and the overall perceived utility of the reading group, which play a crucial role in identifying high-quality reviews.

A recent study, however, points out that the above evaluation indicators reflect only the quality of review information in terms of data reliability and do not emphasize the applicability of review information to the target information users [10]. Researchers argue that the evaluation of the perceived utility of online reviews is a kind of information quality assessment based on the user's perspective, which takes the user's subjective perception as the starting point to explore the utility of information and requires individuals to systematically assess the functional performance of information based on their personal experience [4, 5]. Therefore, user reviews in the online environment should not only be high-quality information that meets the standards but also focus on the degree to which the review information meets the needs and expectations of users and the value it brings to them [11]. There is no shortage of researchers who hold the same view. Hubertrajan and Dhas [12] explores product recommendations, and they argue that the validity of reviews should take consumers' individual preferences into account and look for high-quality reviews that match consumers' personal preferences. Ravi et al. [13] analyzed the quality of cloud service reviews on different online platforms to achieve review recommendations by calculating the similarity between the reviewer's personal information and the background information of the information seekers of the cloud service platform. All these studies take a personalized perspective to study the perceived value of reviews.

2.2. Research on Review-Based Recommendation Systems. Recommendation is an effective way to solve information overload, and, by probing users' information needs, recommendation systems can achieve information push oriented to personal interests and alleviate the distress caused by overloaded information [14]. The core of product recommendation system is to build an effective user and product model. Since review information is rich in users' evaluation of products, it has become a hot research topic in recent years to distill users' preferences and build user models from them and introduce them into recommendation systems. Mousavi et al. [15] classified the relevant research into three categories: lexical item recommendation, rating recommendation, and feature recommendation from the perspective of user modeling.

The lexical item-based recommendation is classified as content recommendation, which directly uses the review text

to model users and products. Seker et al. [16] extracted lexical items from users' published reviews and generates a user model with TF-IDF (term frequency-inverse document frequency) as lexical item weights, and the product model is based on the review set of the target product and finally makes recommendations based on the content similarity between the two. The literature recommendation system of [17] models the user based on the literature he has read, characterizes the lexical items with word vectors, and calculates the similarity between the user and the recommendation target (literature) up to the semantic level.

The collaborative recommendation mechanism used in rating recommendation requires the generation of a "user-rating" matrix, but the matrix sparsity problem has been a bottleneck in the performance improvement of collaborative recommendation systems. One of the solutions is to use the text data of reviews to predict users' ratings of products and then improve the "user-rating" matrix to improve the system performance. In [18], sentiment analysis was used to predict users' ratings of products based on their reviews, and a user model was built based on "predicted ratings" for product recommendation. Hiroshi [19] further improved the quality of the model by weighting the user ratings with the product theme information contained in the condensed reviews. Liu et al. [20] proposed a hybrid recommendation algorithm that integrates user ratings, sentiment, and product content and then recommended products by filling in the space "user-rating" matrix.

In summary, online reviews have been emphasized as an important information source for mining users' interests and preferences in recent research on recommendation systems. Collaborative recommendation strategies that use user reviews to generate user models or enhance the quality of the "user-rating" matrix by predicting user ratings of products are commonly adopted. These users and product models obtained from review text learning are characterized as hidden vectors, and probabilistic topic models and deep learning algorithms are widely used to improve modeling quality.

3. Model Methodology

In this paper, we discuss the identification and extraction of low-frequency keywords. The comments in the dataset are first segmented into sentences, trained by neural network model, clustered to generate the word structure of keywords, followed by word structure ranking, keyword extraction, and then the low-frequency keywords are ranked in the same phrase pattern according to the topic relevance of Meituan comments to achieve the low-frequency keywords we want to extract [21]. The specific framework is shown in Figure 1.

3.1. Word Sense Structure Generation. Word sense structure generation is based on the methods of word clustering or classification in natural language processing. The three following methods are commonly used: The first method is using external knowledge bases (e.g., WorldNet, HowNet, Cyc) to obtain semantic categories of words

directly [22]. The disadvantage of this method is that the knowledge base is difficult to build and difficult to update. The second method is using classifiers in machine learning to identify the word classes of words. This method requires a certain number of datasets to be labeled and the classifier to be trained. This method is difficult to apply when there are many classes of words. The third method is using unsupervised clustering method. This method uses a large unlabeled dataset for training and automatically clusters words into different categories using contextual information of word occurrences. The clustering method is relatively weak, but the training data is easy to obtain and the number of word categories can be chosen flexibly.

We use a word clustering approach based on natural language processing, which maps individual words in a comment to a semantic vector space. In this space, the Eulerian distances of semantically similar words are also close to each other. The Eulerian distances are then used to cluster words that belong to the same word class and are semantically similar. Each word class is represented by a label, which represents the semantic meaning of the word class in the semantic space. Then, the semantic structure of the keywords is generated by replacing all the words in the candidate keywords with the labels. The specific representation is given by the following equation:

$$y(t) = g(v\,f(U\,w(t))), \tag{1}$$

where $w(t)$ and $y(t)$ denote the input and output layers, respectively, and $s(t) = f(Uw(t\text{-RRB}))$ denotes the hidden layer.

3.2. Lexical Structure Ordering. In documents, the semantic structure has a high frequency of occurrence compared to low-frequency keywords and can be used to determine whether a semantic structure is valid or not [23]. The semantic structure of a keyword can be obtained by word structure generation, which indicates the usage pattern of the keyword. If the number of word clusters is k and the allowed semantic structure length is n, the number of semantic structures of possible parameters is k_n.

The number of occurrences of low-frequency keywords is very low in all comments, and the contextual information is sparse. Each low-frequency keyword corresponds to a semantic structure containing many keywords. The ranking of the semantic structures can be done using various ranking methods. We mainly use the number of keywords corresponding to each semantic structure as the evaluation index.

3.3. Keyword Sorting. Because the contextual information of low-frequency keywords is sparse, it is difficult to use contextual information to rank different low-frequency keywords under a single lexical structure. We use the contextual information of each word in the document set to rank the low-frequency keywords. For example, in the review of Meituan, "the peanuts in this Meituan are delicious...
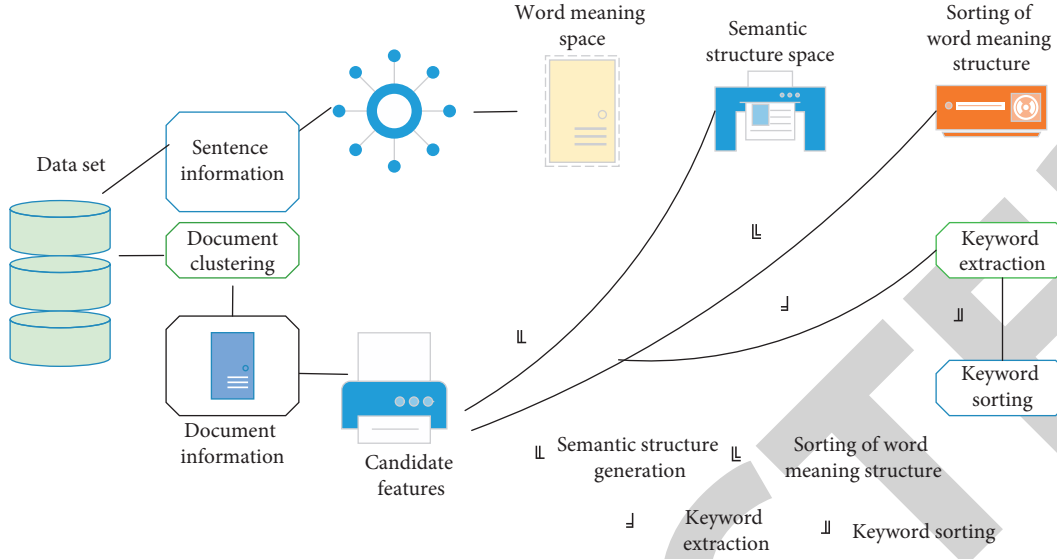
Figure 1: Framework of low-frequency keyword extraction.

and the milk tastes good." If "peanuts and milk" are a low-frequency keyword, the frequency of occurrence is low and the contextual information is sparse. However, the words "peanut" and "milk" appear more frequently in the document. Using contextual information of these words in the entire document set, the words can be ranked according to their relevance to the document topic. In order to rank the low-frequency keywords, we first generate a vector of V_i keywords, which is given by the following equation:

$$V_i = \sum w_i \in P_i \frac{V_{w_i}}{\left\|V_{w_i}\right\|_1}, \qquad (2)$$

where P_i denotes the currently ranked keyword, w_i denotes the words that form part of the keyword, and V_{w_i} denotes the vector consisting of the contextual information (word features around which the word occurs multiple times) of word w_i in the document set. Then, the rating of V_i can be given by the following equation:

$$\text{Scoring}(V_i, V_t | V_b) = \frac{|V_i - V_b|^* |V_t - V_b|}{\|V_i - V_b|_2 |V_t - V_b\|_2}, \qquad (3)$$

where V_t is the word frequency vector produced by the manually selected document clusters after document clustering, indicating the topics related to the usefulness of the USM. V_b shows the background vector generated from the word frequencies in the entire document set. The ranking of low-frequency keywords can be obtained by calculating the score of each keyword on vector V_i separately.

3.4. Commenting and User Model Building under Theme Space. In the process of LDA topic modeling [24], the "document - topic" probability distribution matrix is obtained simultaneously, and we denote θ as Review_MAX$_{i \times k}$, with i corresponding to the number of documents in the comment corpus and K the number of topics. The row vector of Review_MAX$_{i \times k}$ is the description of the probability distribution of comment r in the topic space, as in the following equation:

$$r \cdot \text{topic}_p \text{rofile} = [p_1, p_2, \dots, p_k]. \qquad (4)$$

The user model is also built on the hidden topic space. For this purpose, a set of product feature words Interest_set is used to describe the user's interest, from which the user selects the word items he/she cares about, and the algorithm maps the sequence of the selected word items to the hidden topic space. The modeling process is divided into 3 steps:

(i) Step 1: set Interest_set to generate user interest descriptions based on feature words selected by users.

(ii) Based on the LDA clustering results and the classification of cell phone features by e-commerce platform, the feature words describing the performance of cell phones are divided into 8 topics, namely, "screen effect, network signal, appearance design, photography, audio and video entertainment, operation performance, cost performance, and battery life," from which users select the features they are interested in. For example, if user u is concerned about the "appearance" and "battery performance" of the cell phone, he selects a topic descriptor from the corresponding topic to characterize u, with u.feature_profile = { battery, battery life, appearance, appearance, screen, body, size, ...}. The canonical expression is in equation (5), where Topic(f) corresponds to the set of topic words under the user's topic of interest, with mapping u.feature_profile to the LDA hidden topic space.

$$u.\text{feature}_p \text{rofile} = \{t_i | t_i \in \text{Topic}(f), f \in \text{Interest\_set},$$

$$i = 1, 2, \dots, m\}.$$

$$(5)$$

(iii) Step 2: word vector representation of user interest.

(iv) A word vector is a distributed representation of words obtained based on shallow neural network learning by representing words as an N-dimensional high-density real vector, where the word items correspond to a point in the N-dimensional space and the spacing of the points reflects the potential semantic relationships between the word items. Before mapping user interests based on feature words to the topic space, the study introduces word vectors by first converting u.feature_profile into a word vector matrix $u \cdot \text{vec}_M AX_{m \times v}$ for word vector dimensionality. The user interest model based on word vector description can convey the semantic meaning and improve the recommendation accuracy. The $u \cdot \text{vec}_M AX_{m \times v}$ matrix representation also facilitates the mapping of the user model to the topic space, where the user interest and review models are based on the same topic space; that is, they can be regarded as two points in the space, and their correlation is directly calculated by the distance formula. The word vector introduced in the study is an open-source Chinese pretraining model of Beijing Normal University [25]. The training corpus of this word vector is "Baidu Encyclopedia" with a corpus size of 4.1 G and a vector space dimension of 300.

(v) Step 3: user interest model in topic space. Topic t is expressed by the probability distribution of "topic - lexical items" generated by LDA clustering, as shown in the following equation:

$$t \cdot \text{feature}_p \text{rofile} = \{\langle f_i, w_i \rangle, \quad i = 1, 2, \ldots, n\}, \qquad (6)$$

where f_i are the feature words describing topic t, w_i are the weight of f_i, and n is the number of feature words. Correspondingly, the word vector matrix of topic t is established as $t \cdot \text{vecMAX}_{m \times v}$. Under the word vector space, the interest matrix of u is multiplied with the transpose matrix of topic t, while incorporating the topic feature word weight matrix = $W_{n \times v} = [w_1, w_2, \ldots, w_n]^T$, and finally the maximum value of the matrix operation is taken as the semantic relevance of u and t. The correlation of user u with K topics is calculated according to equation (7), and the user interest model under topic space is generated as shown in equation (8):

$$\text{Sim}_1 = \text{Max} \left(u \cdot \text{vec\_MAV}_{m \times v} \right.$$
$$\left. \times t \cdot \text{vec \_MAX}_{n \times v}{}^T \times W_{n \times l} \right), \qquad (7)$$

$$u \cdot \text{topic}_p \text{rofile\$} = [\text{Sim}_1, \text{Sim}_2, \ldots, \text{Sim}_K]. \qquad (8)$$

4. Experiment and Conclusion

4.1. Experimental Data. In this experiment, we extract data from Meituan, the largest merchant review site in China, which includes 23 areas such as restaurants, shopping centers, hotels, and travel [26]. The Meituan data contain 984,502 Meituan reviews and 584,762 non-Meituan reviews. We focus on the reviews related to Meituan in the Meituan dataset and classify them into two categories based on their usefulness: first, useful reviews, of which 449,437 reviews have a usefulness value > 0; second, useless reviews, of which 535,065 reviews have a usefulness value = 0.

4.2. Experimental Procedure. In this paper, we focus on three aspects: candidate word generation, phrase filtering, and phrase scoring. Finally, we verify the effectiveness of our experiments by determining the percentage of usefulness of the extracted low-frequency keywords in the comments and whether they are useful for users' selection and decision making. The following is a detailed introduction in three parts.

4.3. Candidate Word Generation. In modern generative linguistics, it is difficult to separate function words from content-related words. Our main work is to use function words as boundaries to form candidate words. The steps are as follows:

(1) In the document, each comment is first separated by a punctuation mark, such as { ,.; ！ ？: } .

(2) The LIWC2015 dictionary contains 19,281 discontinued words, and we use the LIWC2015 dictionary to check for separating comments, and if they are in the dictionary, we use them as boundaries to generate candidate phrases [27].

(3) Generated candidate phrases are exported to obtain the candidate phrases of the whole corpus. In order to reduce the noise and complexity of the experiment, we check whether the above problems occur by using the lexicon dictionary (the word list of lexicon dictionary contains 67,725 words) and discard the candidate phrases directly if they are not in this list [28]. By using the above two screening steps, we end up with 1,078,414 phrases in the Meituan dataset, with 31,093,419 occurrences. The distribution of phrase types is shown in Figure 2.

A represents the whole corpus, B represents the useful data comments of Meituan, and C represents useless data comments of Meituan. The percentages of candidate phrases with more than 9 occurrences are 6.27%, 6.98%, and 7.49%, respectively, while the percentages of only 1 occurrence are 71.7%, 71.12%, and 70.01%, respectively. This shows that removing low-frequency phrases will lose a lot of useful information, which is not conducive to better text information extraction and the evaluation of the usefulness of the Mission's comments.

4.4. Phrase Filter. This experiment focuses on the usefulness of the reviews of Meituan. In order to verify that low-frequency keywords contain a lot of important information and great research significance, the three following processes will be used to filter the candidate phrases [29, 30]. (1) High-frequency words can increase the accuracy of the
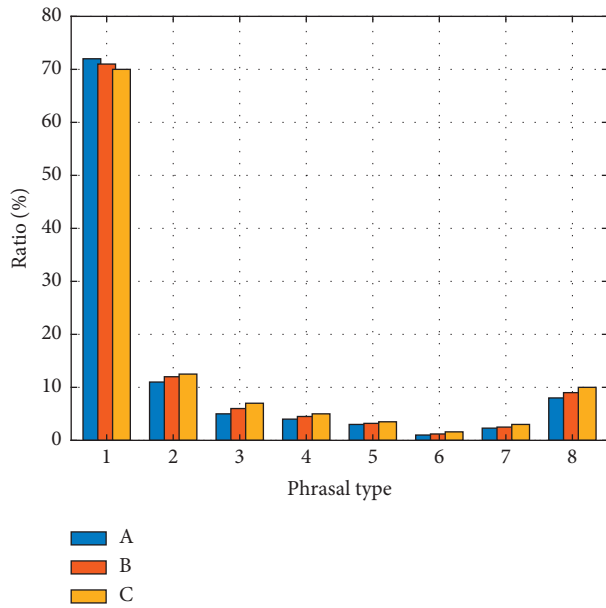
Figure 2: Distribution of phrase types.

representation. Therefore, in order to support word grouping, phrases with less than $N = 300$ word occurrences are removed. (2) In the experiment, to simplify the discussion, only the filtered comments containing phrases consisting of two words will be studied. (3) Since the goal of the experiment is to study low-frequency keywords, only phrases that occur once are discussed.

Through the above phrase filtering, there are 327, 345, 120, 828, 78, and 247 phrases left in A, B, and C datasets, respectively, and their percentages are 30.35%, 25.61%, and 23.58% respectively. The final filtering results are shown in Figure 3.

4.5. Phrase Rating.

The phrase score is very important for the whole keyword extraction. Through the above phrase filtering, we finally obtained 199,075 Meituan phrases that only appeared once in the text and contained only two words [31]. The whole Meituan phrase database is represented by a distribution of trained words, and K-means clustering is performed; that is, according to the similarity principle, data objects with high similarity are classified into the same class clusters, and data objects with high dissimilarity are classified into different class clusters, where K represents the number of class clusters and means represents the mean value of data objects in the class clusters. The clusters are divided into 200 groups, and each group is identified by the label range of "C000–C199." In order to reduce the noise, reduce the processing difficulty, and achieve better classification effect, 20,277 useful phrases and 16,362 useless phrases of Meituan were generated by replacing the extracted keywords with word labels. Since we mainly focus on the usefulness of Meituan reviews, here, we only list the usefulness categories. The details are shown in Table 1.

C15 for fruit, C155 for sweets, C51 for flavor phrases, C63 for meat or cereals, C125 for emotional adverbs, C152 for price or affect adjectives, and C149 mostly for words that describe the environment.

In this paper, we collect 2013–2014 USG usefulness reviews, and, in order to rank low-frequency words with the same phrase pattern, we define a target vector V_t, which represents the textual topic relevance of the dataset, and the identification algorithm about low-frequency keywords is shown in Table 2.

4.6. Experimental Conclusions.

From the experiment, we can get the distribution of usefulness comments of Meituan, so we can see that the usefulness votes with 5 or more occurrences only account for 6.08% of the whole Meituan comments, while those with 1 occurrences account for 52.78% of the whole usefulness votes. The low-frequency words are mostly words that objectively express the dining experience, such as "quite affordable, unforgettable, and very cold." The higher the "usefulness" vote is, the more valuable the review is and the more useful the phrases it contains; the high-frequency words are mostly words about Meituan entities, such as "steak salad, Meituan seats, cheese bread." The lower the "usefulness" vote, the lower the value of the comment and the more useless the phrases included. The distribution of "usefulness" votes is shown in Table 3.

This experiment not only shows that ignoring low-frequency keywords will lose a lot of important information but also verifies that our proposed method has made great progress in dealing with low-frequency keywords and has achieved good results in the restaurant usefulness poll, providing consumers with accurate and useful information in a more objective way.

Model parameter setting: for the LDA model, the value of the subject number K, which is related to α and β of the model, is critical. K is used as the optimization parameter and the value is determined experimentally. Figure 4 shows the clustering effects of the three modeling schemes with different K values. Overall, with increasing K, Avg_similarity tends to decrease, indicating that the intertopic similarity decreases and the stability of the clustering structure increases. On the contrary, KL dispersion increases gradually, indicating that the intertopic differences are widened and the internal cohesion is increased. With increasing K, the two metrics gradually converge. Specifically for the three modeling schemes, both sets of indicators show that the clustering effect of "synonymous feature word normalization" is significantly better than that of "noun + verb" and "feature word." Therefore, the topic clustering scheme of "synonymous feature word normalization" was adopted in the subsequent experiments. According to the experimental results (see Figure 4), the clustering model is the best. KL scatter = 8.267, Avg_similarity = 0.05, and finally K = 13.

Clustering results: Figure 5 shows the clustering results generated by pyLDAvis for K = 13. On the whole, the themes are well distributed, and most of them are clearly distinguished, with a few overlapping (themes 4 and 5, themes 1 and 2). For this reason, the following treatment was performed: for each topic clustered, the topic words were ranked in descending order of probability, and the top 8
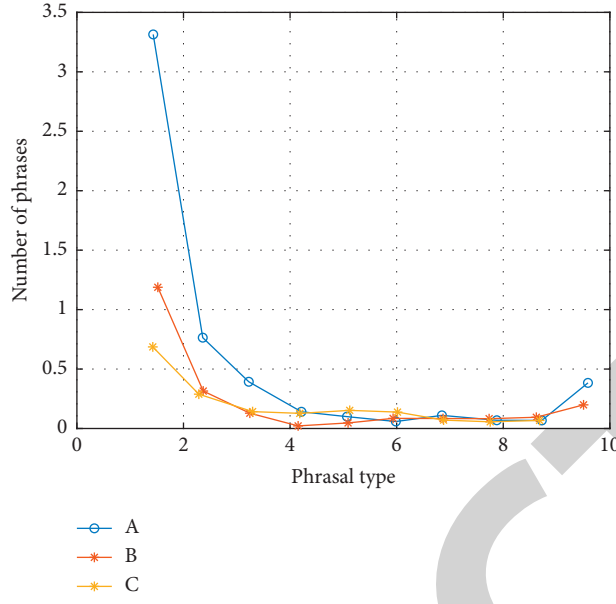
Figure 3: Phrase filtering distribution.

Table 1: Example of phrase grouping.

Phrase	Give an example
C15–C155	Cucumber frozen cake, grape mousse, cherry milkshake, peanut milk, peach crisp, almond milk
C155–C155	Cream cookies, dessert custard, cotton mousse cheese pudding, cream sundae, walnut biscuits
C15–C51	Pumpkin mustard, peanut seasoning, blackberry jam, fruity butter, cherry jam, strawberry jam
C63–C63	Pork sausage, sausage Tujia, diced chicken with vegetables, beef fried rice, pineapple corn, sausage cheese
C129–C152	Very cheap, very attractive, very bad, absolutely bad, full of taste, ridiculous
C129–C149	Very quiet, comfortable, elegant, slightly high-grade, energetic, super luxurious and very cold

Table 2: Model corresponding algorithm.

Input: a group of low-frequency phrases in the phrase pattern, all comments of the whole corpus
Output: low-frequency keyword sorting list: L_0
1) Divide the comments into restaurants and backgrounds
2) Divide restaurant comments into usefulness and uselessness
3) Generate target vector V_t and background vector V_h
4) Perform the algorithm and calculate the scoring value
5) Arrange L in ascending order L_0

Table 3: Distribution of "usefulness" votes.

Number of comments	Comment on "usefulness" low-frequency words/item	Number of comments	Comment on "usefulness" low-frequency words/item
1	237225	6	7859
2	104578	7	5247
3	46587	8	3567
4	23458	9	2549
5	13256	10	1478

words were used to describe the topic semantics. If a word appears in more than one topic at the same time, it will be assigned to the topic with the highest weight value. For example, "battery capacity" appears in both topic 4 and topic 12, but the weight value under topic 12 (0.052) is higher than that under topic 4 (0.019), so it is placed under topic 12. Clustered subject terms were adjusted to better clarify the meaning of the topics. According to the list of topic words of each topic, the 13 topics were assigned to 9 feature categories of "operation performance, screen effect, network signal,
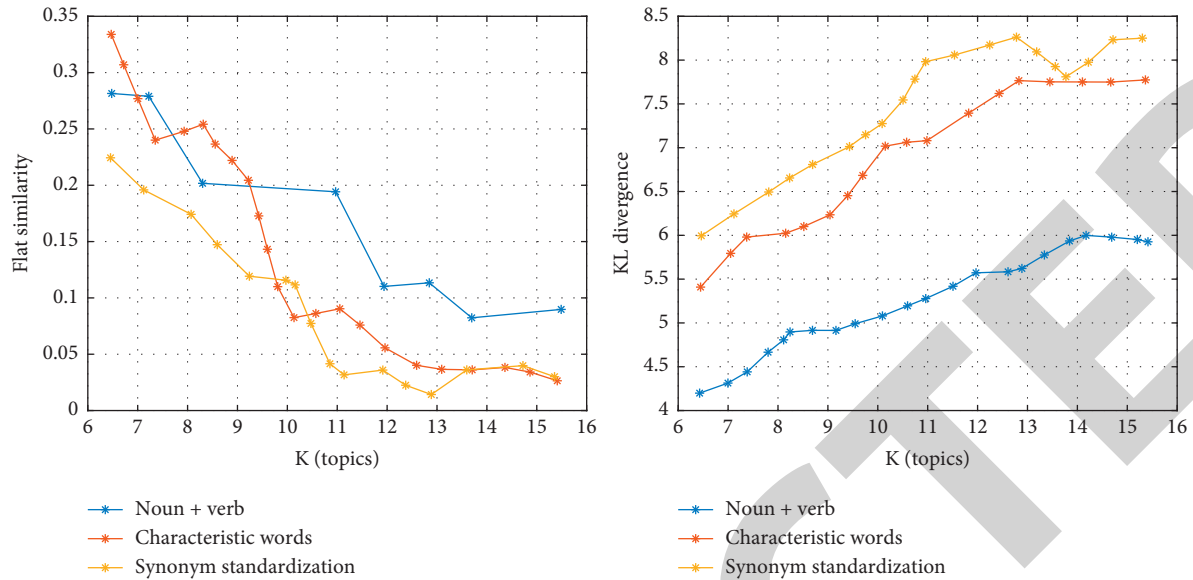
FIGURE 4: LDA topic clustering results with 3 feature modeling schemes (left: Avg_similarity; right: KL scatter in vertical coordinate).
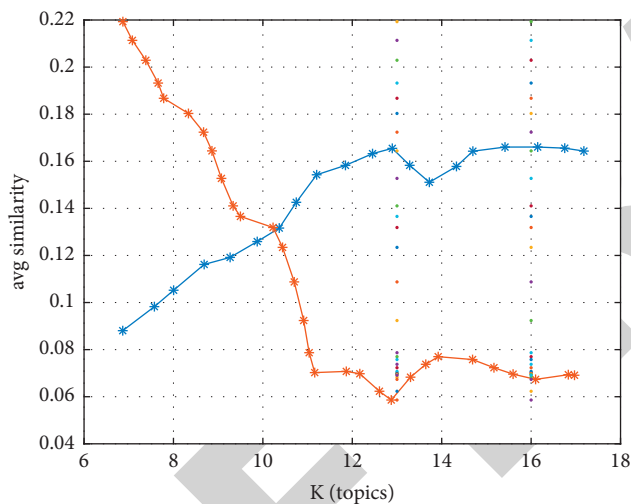
FIGURE 5: Effect of LDA topic clustering with different K values (Avg_similarity and KL scatter).

appearance design, photography, audio and video entertainment, cost performance, battery life, and others" by referring to the settings of cell phone feature indexes in digital websites, and the feature word set of user interest selection was generated accordingly, Interest_set, used for user modeling.

5. Conclusions

The study uses a probabilistic topic model to construct a user interest model in the topic space and incorporate it into the review perceived value calculation model, based on which a review recommendation strategy that integrates user interest and review utility is proposed, and the effectiveness of the recommendation strategy is tested by an online evaluation system. For the user model, the feature words characterizing user interest are treated with equal weights, but, during the testing process, it is found that users focus on product performance, and subsequent research can set weights for the feature words describing user interest to build a more refined user interest model.

The follow-up research is also prepared to introduce deep learning algorithms to explore user modeling in depth, extract user features from user comments, and improve the personalized recommendation algorithm.

Data Availability

The datasets used in this paper are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest regarding this work.

Acknowledgments

References

[1] S. Hido, S. Suzuki, R. Nishiyama et al., "Modeling patent quality: a system for large-scale patentability analysis using text mining," *Journal of Information Processing*, vol. 20, no. 3, pp. 655–666, 2012.

Research Article

Study on the Automatic Basketball Shooting System Based on the Background Subtraction Method

Yan Hu [ID][1] **and Yong Xu**[2]

[1]*Beijing Sport University & Humanities and Social Sciences Division,*
  ChengDu Sport University & School of Economics and Management, Beijing 100084, China
[2]*North China University of Technology, Beijing 100144, China*

Correspondence should be addressed to Yan Hu; xy@ncut.edu.cn

There are many drawbacks such as clustering, background updating, inaccurate testing results, and low anti-interference performance in traditional moving target detection theory. In our study, a background subtraction method to automatically capture the basketball shooting trajectory was used to eliminate the drawbacks of the fixed-point shooting system such as cumbersome installation and time and manpower consumption. It also can improve the accuracy and efficiency of moving target detection. We also synthetically compared to common methods including the optical flow method and interframe difference method. Results showed that the background subtraction method has better accuracy with an accuracy rate over about 90% than the interframe subtraction method (88%) and the optimal flow method (85%) and presents excellent robustness with considering variable speed and nonrigid objects. Meanwhile, the automatic detection system for basketball shooting based on background subtraction is built by coupling background subtraction with detection characteristics. The system detection speed built is further accelerated, and the image denoising is improved. The trajectory error rate is about 0.3, 0.4, and 0.5 for the background subtraction method, interframe subtraction method, and optimal flow method, respectively.

1. Introduction

In recent years, with the development of computer vision technology and video surveillance, the moving target detection technology (MTT), a significant part of them, has gradually attracted researchers' attention and also has been widely applied in different domains such as defense and security monitoring [1–3]. MTT is a technology which can separate variable parts from a video image on the basis of image segmentation with geometric and statistical features. Results from MTT can provide interesting regions for an object's identification, track, position, and behavior analysis. MTT has been regarded as a novel technology with high work efficiency and manpower-saving [4]. Multiobject information can be obtained from the video image sequence compared to the stationary image sequence when using MTT to test an object. In a stationary scenario, the image

subtraction method has been regarded as one of the best efficient methods because we can easily obtain the background reference model with a net and nonvariable target in the stationary scenario. Meanwhile, a real-time update background reference model used the current frame image so that the background model can adapt scenario variation. As for the moving scenario, both the moving compensatory procedures of background and detection methods applied in the stationary scenario are needed to test the moving target in the image due to its complicated variation [5].

Currently, MTT methods have been improved and/or innovated with the development of computer vision technology [6]. The development regular for MTT can be generalized as follows: (1) the optimal flow method on the basis of optimal flow technology, the interframe difference method on the basis of subtraction image, and the background subtraction method. Background difference and

interframe difference have been known as the most popular detection methods in the application visual image procedure system; (2) continuous improvement in traditional detection methods. As for drawbacks for traditional methods, some methods based on improvement make results more stable, more practical, and more real time. On the other hand, combining different MTTs, shortcomings from the individual method can be made up to each other; (3) a new algorithm will appear continuously. With the progress and innovation of digital image procedure technology and computer vision technology, some more efficient, advanced, and more robust algorithms on the basis of old methods have been presented. These methods can adapt environmental variation, noise, and shadow interference. Also, these algorithms can promote the progress of the application of MTT in practice; (4) more robust, easier, real-time detection results and common applications. We should have knowledge that there are still more challenges in applying these MTTs [7]. For example, if we are urging to reach its real-time function, its accuracy may decrease, and vice versa. The reason of cause of this contradictory phenomenon may be owing to common knowledge that high accuracy usually accompanies with a large calculation procedure which caused worse real-time performance; similarly, good real time will decrease its accuracy due to low anti-interference skill. These contradictory attributes of MTTs lead to the individual method making sense only in a particular scenario. Coping with these contradictory problems of MTTs has been a challenge for researchers in the theory research field.

There are an increasing number of MTTs with the improvement of theory on the basis of traditional MTTs. For instance, the optimal flow method [8], interframe difference method [9], and background subtraction method [10] have been widely used in different research fields due to their respective advantage. The optimal flow method tests objects more accurately and is reasonable for testing a moving target. The most typical algorithm of optimal flow is the Lucas–Kanade (L–K) and Horm–Schunck (H–S) algorithm. With the increasing calculation skill of computers in recent years, a large number of optimal flow technologies have been appeared. For example, Liu et al. [11] presented an optimal flow feature-based robust gait-characterizing method, and they obtained a reasonable conclusion. Zhu et al. [12] provided an L–K improvement-based method and applied it into mileage calculation. Lee et al. [13] measured cable elongation at break based on the improved L–K optical flow method. Interframe subtraction can test and segment a moving target according to video sport information between neighboring frame images because image information between neighboring frames from a video image contains much moving target information. However, we should acknowledge that the original interframe method still has drawbacks. For example, when grayscale and texture in neighboring frames are similar with each other, this method can only obtain the target's edge contour, target moving information, however, which cannot be completely detected using the original method. Moreover, when the target moves at a high speed, the background occlusion variable area will become large between neighboring frames, which will lead to misjudgment of the occlusion background as the target, which further impacts, to some extent, the feature parameter extraction of target and moving target segmentation. Therefore, many researchers improved the original interframe subtraction method. For instance, Yuan Hang and Wang [14] used 3 interframe subtractions to investigate a moving target. Zheng et al. [15] detected and tracked the human body by combining 3 interframe subtractions with the mean-shift method. The theory of the background subtraction method is similar with the interframe mentioned above; the difference between them is that the background subtraction method need not use neighboring frames, but builds a background reference model which is subtracted from the current image's frame to detect the moving target. In other words, the selection of the background reference model is vital for completely segmenting moving regions; meanwhile, due to its real-time performance in the video image sequence, the background subtraction method is the most popular method to detect moving target information.

After basketing, the capture of the trajectory brings many difficulties due to the particularity of its movement characteristics. In our study, we used the background subtraction method coupled with an automatic capture system to detect the basketball trajectory for the offsetting efficiency mentioned above; meanwhile, we also applied the optimal flow method and interframe approach to detect basketball trajectory information to compare them with the background subtraction method. The basketball shooting trajectory is parabolic, and the background subtraction method can meet the conditions of basketball trajectory capture in any scene. Traditional methods of capturing the trajectory of basketball shots usually lead to unsatisfactory effects due to improper sample selection. Therefore, in our study, the Gaussian mixture model background difference method is used to improve the traditional method and the accuracy of capturing motion trajectories in complex scenes. The method can effectively improve the accuracy of the automatic capture of basketball shooting trajectories and also the adaptability to complex scenes and improve the limitations of the capture method.

2. Theory of the Algorithm of Background Subtraction, Interframe, and Optimal Flow Methods

2.1. Background Subtraction Method. The core theory of background subtraction is matching the current frame with the reference image from the background model and then calculating the similarity value between the image point and that in the background model. The mathematical expression is as follows:

$$p_{\text{target}} = \begin{cases} 1 \left| p_{\text{current}}(x, y) - p_{\text{background}}(x, y) \right| \geq T, \\ 0 \left| p_{\text{current}}(x, y) - p_{\text{background}}(x, y) \right| < T, \end{cases} \quad (1)$$

where point (x, y) is the pixel value in any position, $x = 0, \ldots, H-1, y = 0, \ldots, V-1$, where H and V represent the horizontal

and vertical resolution, respectively, $p_{\text{current}}(x, y)$, $p_{\text{background}}(x, y)$ is the pixel value in the current frame and the pixel value in the background model in the point (x, y), respectively; and T represents the segmentation threshold. When p_{target} is 1, it means presports attractions are in point (x,y) and 0 means at the background point. If the grayscale value of the background reference point is over that of an unknown point, then it can be regarded as presports attraction, and the background point otherwise.

In order to extract quickly and accurately the position and outlook feature of the basketball from the video/image sequence, a binarization procedure is used for the grayscale image. In this procedure, the selection of segmentation threshold is first in the process of binarization. Currently, there are many methods to accomplish the selection of threshold (Table 1).

2.1.1. Background Model.

As we all know, the scenario environment will change with time moving, so it is necessary to the real-time update background model. In recent years, a lot of methods have been used to update the model (see Table 2). These methods were usually improved and optimized on the basis of original theory.

After constructing the background model and then subtracting each pixel in the video image sequence from the background model built, if the pixel value exists in the same location between the image sequence and background model, the pixel point is regarded as a background point, and as a moving target otherwise. In our design procedure, we first build the background model based on the former m-1 frame from image sequence information and then subtract the current image from the image sequence in the background model, and finally, final image information over the threshold can be obtained. However, the image sequence over the threshold is not complete due to noise; therefore, we need to eliminate its noise using morphology theory. The detailed procedure is shown in Figure 1.

2.2. Interframe Subtraction Algorithm.

The interframe subtraction method subtracts pixels between continuous 2 or 3 frames from the video image sequence and then compares those with the threshold preset to extract moving regions from image information. In common situations, excepting for an interested moving target, other objects are static in scenarios of the image sequence. Therefore, variation of parts in the image is only caused by moving parts. However, in fact, the moving target usually exists in a complicated environment, where much noise exists; therefore, we need to eliminate noise like the background subtraction method. The process of the method is shown in Figure 2.

In the detection procedure, $D_m(x, y)$ is obtained by subtracting $(m-1)$-th from the m-th frame according to equation (2), and then, if its value is over the threshold, the image value is 1; otherwise, it is 0. The details of the mathematical expression are as follows:

$$D_m(x, y) = \left| F_x(x, y) - F_{m-1}(x, y) \right|,$$

$$B_m(x, y) = \begin{cases} 1, D_m(x, y) > T, \\ 0, D_m(x, y) \le T, \end{cases} \tag{2}$$

There are many studies combining the background with interframe to detect moving target information.

2.3. Optimal Flow Method.

Chen et al. [19] proposed the optimal flow concept in the 1950s; they projected an object in a three-dimensional space on a two-dimensional plane; once the object moved, an optimal flow field was formed in the scenario, where the location of the target can be judged by comparing the variation of pixels between neighboring frames. If the grayscale value of image sequence between neighboring frames in the projected image is unchanged and the image in each frame sequence is continuous, it satisfies the following equation:

$$I(x, y, t) = I(+dx, y + dy, t + dt), \tag{3}$$

where $I(x, y, t)$ represents the grayscale value at position (x, y) and time t.

Taylor's expansion of equation (3) is

$$I_x \frac{dx}{dt} + I_y \frac{dy}{dt} + I_t = 0, \tag{4}$$

where I_x, I_y, and I_t represent the partial differential forms of the grayscale value in position (x, y) at time t, respectively. Equation (4) is the moving status of the target we have aimed. The method is applied in many fields by researchers. Zhou et al. [20] deeply analyzed and improved the problem of detecting moving targets by the L–K optical flow method based on the optical flow algorithm.

Some hypotheses for the optimal flow method should be pointed: (1) constant brightness: the brightness value (pixel gray value) of a pixel changed with time, and the color of adjacent frames can remain unchanged. This is the basic setting of the optical flow method; all optical flow methods must meet it; (2) continuous time: continuous "small movement," time changes will not cause drastic changes in the location. In this way, the gray value change caused by the position change between adjacent frames can be used to obtain the deviation of the gray value to the position. All optical flow methods must meet it; and (3) spatial consistency: the pixels of the same subimage have the same motion.

Of course, we should point these three methods have their own advantages and disadvantages (see Table 3). Comparing the other two methods, the background subtraction method is priorly suggested by a large number of scholars. In our study, we compared the three methods' performance in basketball shooting trajectory detection to judge the best detection methods for basketball shooting trajectory.

TABLE 1: Selection methods' comparison of segmentation threshold for the binarization procedure.

Names	Description	Mathematical expression	Reference
Maximum entropy	Entropy maximum between presports attraction and background point	$T = \arg\max[H_f(T) + H_b(T)]$	Kapur et al. [16]
Mean grayscale	All of the pixel point values are added up in the image and then divided by the number of all of pixels, and the average value is considered as threshold	$T = \sum_{n=0}^{n=L} f(i, j)/N$	Timo P Kaivosoja et al. [17]
Maximum intraclass variation	Maximum variation for the average grayscale value between the foreground region and whole image	$\delta^2 = P_0(u_0 - u)^2 + P_1(u_1 - u)^2$	Mhenni et al. [18]

TABLE 2: Methods used to update background models.

Method name	Description	Characteristics
Median	Median value between continuous multiple-frame sequences as the grayscale value of pixel in fixed time scales	Inconveniently obtains time sequence
Mean	As to the "median" method but the average value of frame sequence as the pixel value	Sensitive to light variation in the environment and dynamic background
Kalman filter	Predicts image transform results on the basis of Kalman filter theory	Long time to eliminate noise and uncontrollable procedure process
Single Gauss	Takes each grayscale value of pixel as the stochastic variable, and the whole process follows Gauss distribution	Convenient calculation process, but bad performance in a complicated scenario
Multiple Gauss	Superimposes a single Gauss process, multimodal situation in a complicated scenario	More model to superimpose and complies with a complicated scenario
Nuclear density	A nonparameter method estimates the current pixel value in a certain moment using nuclear density function	Prior distribution is not needed to know before calculating the density function of the sample
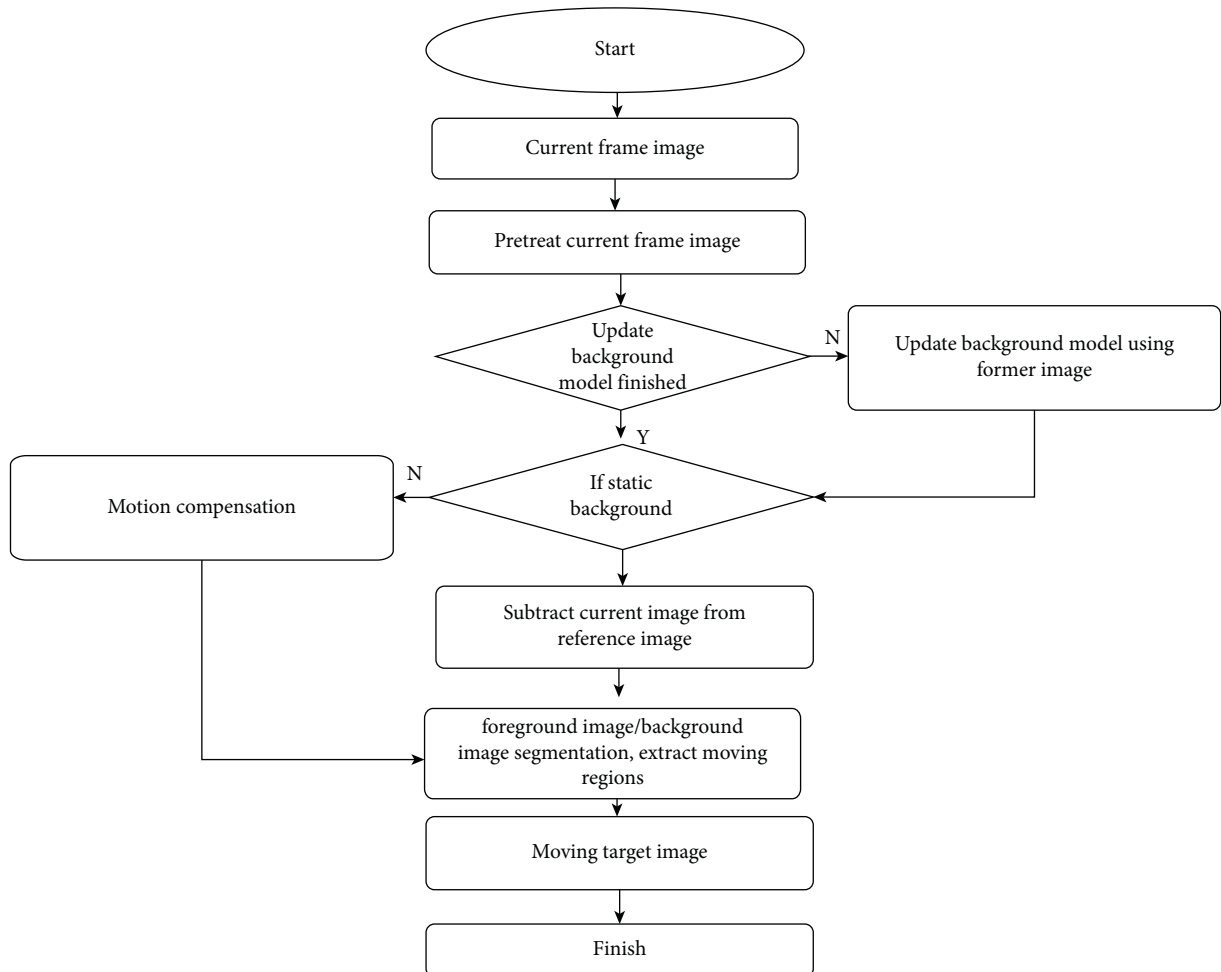
FIGURE 1: Background algorithm procedure for noise elimination.
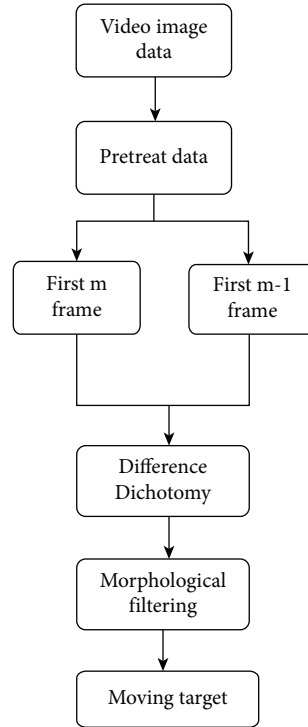
FIGURE 2: Noise eliminating procedure of the interframe method.

TABLE 3: Three detection algorithm methods' comparison between each other.

	Optimal flow	Interframe subtraction	Background subtraction
Detection results	Whole region	Target contour	Whole region
Algorithm complexity	Big	Small	Decided by the background model
Applicable scene	Camera position can change	Camera position must be fixed	Camera position must be fixed
Robustness	Poor	Good	Little good
Advantages	Extent applicable field	Easy to compute	Complete segmentation and low complexity
Disadvantages	Slow computing speed	Incomplete detection target	Need to update the background model

3. Characteristics of Basketball Shooting

In order to detect the basketball shooting trajectory accurately, we first need to know about the feature of basketball shooting; after identifying the characteristics of shooting, the position of the capture device can be further set. The feature can be concluded as follows:

(1) The shot will spin after the shot or when the board is scratched

(2) The direction of shooting on the board may be right, middle, left, and line of shooting in any positions

(3) Under conditions of the shooting goal, the diameter of basketball changes from big to small

(4) When the basketball touches the hoop or backboard, the ball will be shaped

3.1. Device Installation. The selection of a capture device is extremely important to accurately detect the basketball shooting trajectory. There are a lot of devices to choose for obtaining data such as cameras and sensor devices. In our study, we chose sensors which were set on the athlete body and board to obtain data of the basketball shooting trajectory. Combining the comparison results with the basketball sports environment and sports characteristics, we chose the sexual sensor to complete the design. It is highly adaptable for environmental factors and low cost, suitable for the basketball environment. Because of its dependence on smaller environmental factors, its robustness is also more stable. Therefore, we used this sensor to complete the design. Among inertial sensors, MEMS sensors have the advantages of wireless transmission, low cost, superior trajectory capture effect, and convenient operation, which are widely used. Therefore, in this design, the MEMS inertial sex sensor completes our design. There are 3 parts in the trajectory capture framework: the trajectory collector, repeater, and server. According to the abovementioned framework, a fixed trajectory collector was installed on the basketball stadium, collecting the trajectory of the basketball movement to complete the capture work. To ensure the stability of the process of the trajectory capture, we designed the repeater. The repeater is responsible for the track processing and forwarding of trace data. In the design process of this

repeater, charging is used for power supply to the ARM1176-S core processor, Bluetooth communication transmitter set up, and wireless network communication interface.

Samsung ARM1176 core is used to process the data received by the repeater. The trajectory collector, repeater, and server were connected to form a network of shooting trajectory capture equipment. After this network is set up and fixed, adopting a one-to-many topology, the entire network is divided into 2 layers to ensure the real time and continuity of the capture process.

3.2. Data Preprocessing. In this design process, in order to ensure the real-time capture of trajectory images, inertial sensors were used to capture the trajectory of the shooting on the basis of shooting action. In addition, to install a fixed collector on the backboard, a miniature inertial sensor was installed on the body to ensure the accuracy of trajectory acquisition. According to human kinematics, the corresponding human skeleton model was established on the basis of shooting motion capture. Based on the waist and legs, all the human body postures of the joint take the root node as the origin, and the inertial sensor was placed on the athlete's forearm, and the corresponding position and relative posture of the human body were used to capture the shooting action. Reading the sensor data of the human body model through the abovementioned settings and according to the movement data of these nodes, the collection time of the backboard collector was adjusted. After a person wears the sensor, in order to ensure the accuracy of motion capture, it is adjusted through the server interface. After setting data, collection results were uploaded in the inertial sensor, driving the repeater, transferring sensor information, and exercise information into the txt file format. The abovementioned data were stored in the server for processing the data.

In the process of basketball sports collection, even the standardized sports process and sophisticated collection equipment cannot directly obtain the data of the shooting trajectory. The data obtained in the abovementioned steps mainly include the real data of human body movement, the data generated by the gravity and material force of basketball, the noise in the sports environment, and the transmission sensor's inherent zero drift random noise. Methods of noise extraction mainly include field average and median value filter. In our study, we used the field average method to eliminate noise.

Finally, after the preliminary work was carried out, the basketball shooting trajectory can be finished by using three moving target detection methods (background subtraction, interframe subtraction, and optimal flow). The details of the workflow are presented in Figure 3.

3.3. Construction of the Automatic Identification System. In our study, we combined the automatic identification algorithm with the MATLAB application to design the automatic detection system for basketball shooting. Firstly, we identified hardware and software equipment of the system using the environment of software and hardware of
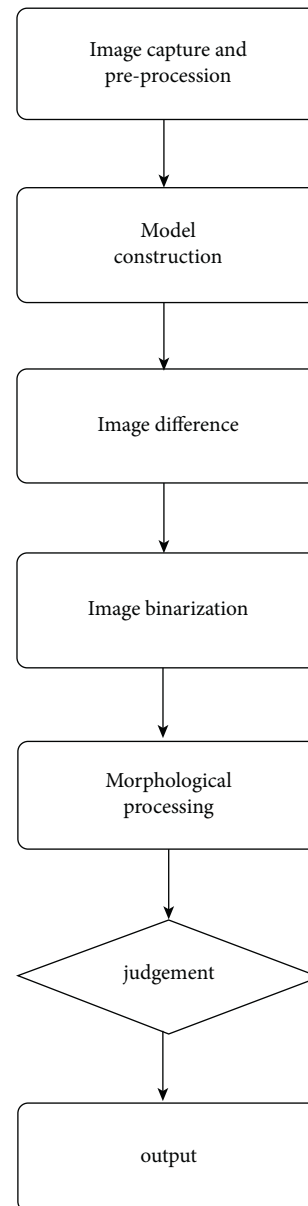
Figure 3: Workflow of the basketball shooting trajectory detection system.

the system and then constructed an interactive interface using the GUI tool of MATLAB. Lastly, we evaluated the performance of the automatic identification system constructed using twenty data groups.

Hardware devices mainly include industry cameras and computers. The software environment mainly includes the development environment and operating environment. These are all built through the MATLAB environment.

4. Results and Discussion

4.1. Accuracy Evaluation of the System. We evaluated the accuracy of the system through the shooting goal which contains false detection rate, missed detection rate, and accuracy rate. We totally shoot the basketball 20 times;

TABLE 4: Statistics of shooting a basketball for the background subtraction model/interframe subtraction model/optimal flow model.

Serial number	Shooting number	Shooting goal	Judging the goal
1	21	9	10/10/9
2	19	7	6/6/6
3	21	6	6/5/5
4	21	5	5/4/3
5	20	7	7/7/7
6	22	4	4/2/2
7	18	7	7/5/6
8	20	8	8/6/5
9	21	9	7/6/6
10	23	6	6/4/3
11	18	7	6/6/4
12	19	6	6/5/5
13	17	8	7/6/5
14	23	7	6/5/4
15	22	6	6/5/5
16	21	2	2/1/0
17	21	5	4/3/3
18	20	7	6/5/6
19	23	6	5/4/4
20	22	4	3/2/2
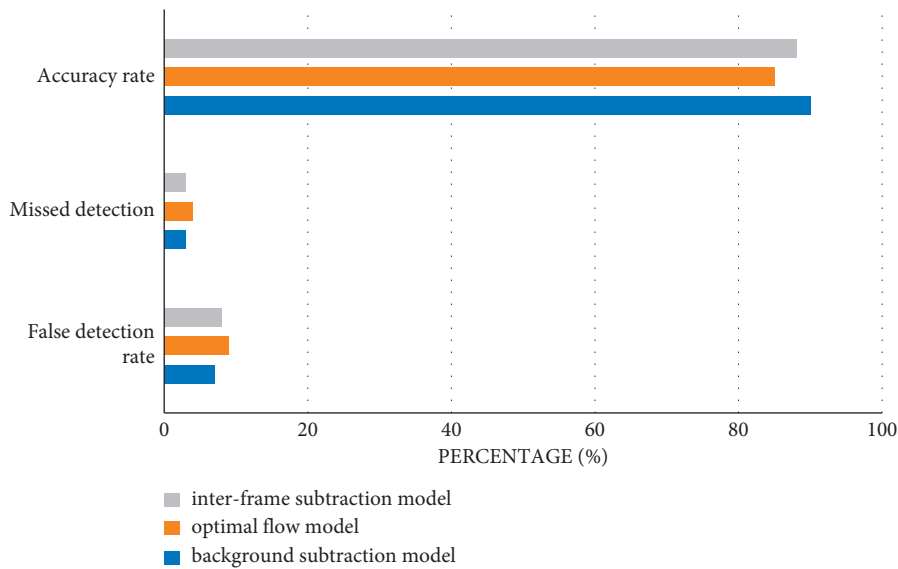
FIGURE 4: Accuracy evaluation of three different models (blue: background subtraction model; orange: optimal flow model; and gray: interframe subtraction model.

the results can be found in Table 4. Then, we evaluated the accuracy of the system we built through comparing with the actual condition. Results from Figure 4 show that only the background subtraction model's accuracy rate reached 90%; both the other two models' accuracy rates are over 85% (88% for the interframe subtraction model and 85% for the optimal flow model). Therefore, we can infer the ranking of performance from high to low in the basketball shooting trajectory in the background subtraction model, interframe subtraction model, and optimal flow model.

4.2. Simulation Evaluation. In order to ensure the effectiveness of our design and whether it can solve the related problems about the original method, the realization environment is constructed, and the research on its capture effect is completed. A form of comparative experiment is used to verify the capture accuracy of the designed automatic capture method and the original trajectory capture method.

1000 images are acquired through image video sequence acquisition. Region segmentation and feature reorganization of the image and video sequence were performed; the intensity of noise interference received is −8 dB in the progress
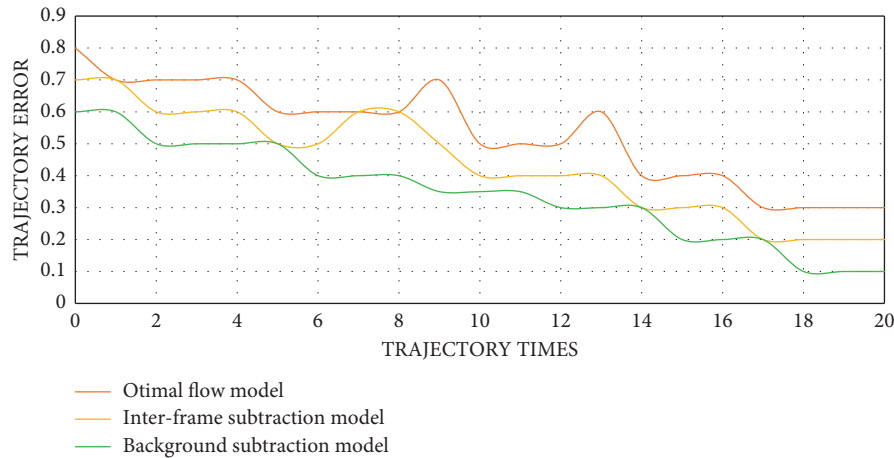
FIGURE 5: Trajectory error comparison between the background subtraction, interframe subtraction, and optimal flow model. Training was performed 20 times.

of image acquisition. To carry out the basketball flight trajectory tracking simulation experiment, we first collected the basketball trajectory data and then eliminated their noise. Lastly, juxtaposed experiment was conducted with trajectory error (Figure 5).

Results from Figure 5 show that the best performance between the 3 models was of the background subtraction model with mean 0.3 error, followed by the interframe subtraction model (0.4) and optimal flow model (0.5).

Computer vision technology is hybrid by image procession, artificial intelligence, and machine vision. In our study, we designed an automatic detection system for basketball trajectory using computer vision technology. However, we should acknowledge that our system still needed to improve due to time and other objective factors; for example, there are false detection and missed detection in our experiment. We only consider one basketball in our experiments; a multitarget problem is needed to consider. DSP and FPGA technology should be considered into the automatic detection system to reduce costs in future work.

5. Conclusions

In our study, we constructed an automatic identification system on the basis of the GUI tool of MATLAB application for basketball shooting trajectory detection based on the background subtraction method. Meanwhile, we also compared the model with the other two models, the interframe subtraction model and optimal flow model. We first performed a comparative experiment: the basketball was practically shot 20 times, and the shooting number and the shooting number of shooting goal were recorded. Industry cameras, sensors, and computers were used to collect data about basketball shooting. The basketball shooting trajectory was judged using the automatic detection system on the basis of the background subtraction method compared to the practical experiment. Moreover, we performed a simulation experiment in order to meet the validity and accuracy of the automatic detection system we have built. Accuracy evaluation with false detection rate, missed rate,

and accuracy rate was compared to practical experiment of 20 times shooting. In our simulation experiment, we evaluated the trajectory error rate on the basis of trajectory data sequence from 1000 images which were used to extract the gray pixel feature value. The results show that compared to the other two methods (interframe subtraction method and optimal flow method), the background subtraction method has better accuracy (average 95%) and real-time performance for trajectory detection, is more robust for detecting the target with uncertain moving speed, and is nonrigid. In the simulation experiment, the average trajectory error rate reached 0.3, 0.4, and 0.5 for the background subtraction method, interframe subtraction method, and optimal slow method, respectively.

Data Availability

The experimental data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare no conflicts of interest regarding the present study.

References

[1] H. Shogo, H. Kei, C. Suzuki, R. Sakurai, T. Suda, and K. Yoshioka, "Estrus detection using background image subtraction technique in tie-stalled cows," *Animals*, vol. 11, no. 6, p. 1795, 2021.

[2] J. Wang and J. Chen, "An improved background subtraction method for adaptive rate compressive sensing," *Journal of Physics: Conference Series*, vol. 1914, no. 1, Article ID 012024, 2021.

[3] I. Hoffmann, "Data analysis and background subtraction in neutron spin echo spectroscopy," *Frontiers in Physics*, vol. 8, 2021.

[4] A. J. Lipton and H. Fujiyoshi, "Moving target classification and tracking from real-time video," in *Proceedings Fourth*

IEEE Workshop on Applications of Computer Vision. WACV'98 (Cat. No.98EX201), NJ, USA, October 1998.

[5] G. L. Foresti, "Object recognition and tracking for remote video surveillance," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 9, no. 7, pp. 1045–1062, 1999.

[6] I. Iszaidy, R. Ngadiran, R. B. Ahmad, N. Ramli, M. I. Jais, and V. Vijayasarveswari, "An analysis of background subtraction on embedded platform based on synthetic dataset," *Journal of Physics: Conference Series*, vol. 1755, no. 1, Article ID 012042, 2021.

[7] D. Pankaj, "Real-time surveillance for critical activity detection in ICUs," in *Proceeding of the Second International Conference on Computer and Communication Technologies*, March 2016.

[8] P. Anandan and M. J. Black, "A framework for the robust estimate of optimal flow," in *Proceeding of the Fourth International Conference on Computer Vision*, pp. 231–236, Berlin, Germany, May 1993.

[9] Ye Tian, C. Yu, F. Xie, S. Gao, and Ru Manhui, "Research on video detection method of moving target oriented to substation," *IOP Conference Series: Earth and Environmental Science*, vol. 804, no. 3, Article ID 032011, 2021.

[10] A. M. Hamad and N. Tsumura, "Background subtraction based on time-series clustering and statistical modeling," *Optical Review*, vol. 19, no. 2, pp. 110–120, 2012.

[11] S. Liu, K. Luo, N. Ye, C. Wang, W. Jue, and B. Zeng, "OIFlow: occlusion-inpainting optical flow estimation by unsupervised learning," *IEEE Transactions on Image Processing*, vol. 30, pp. 6420–6433, 2021.

[12] J. Zhu, Y. Wu, and X. Shao, "Two-step phase extraction and random phase shift estimation in phase-shifting profilometry based on least-squared optical flow method," *Optics Communications*, vol. 499, Article ID 127270, 2021.

[13] C. Lee, T. Lee, T. Nonomura, and K. Asai, "Evaluating the applicability of a phase-averaged processing of skin-friction field measurement using an optical flow method," *Journal of Visualization*, vol. 23, no. 5, pp. 773–782, 2020.

[14] C. Yuan Hang and J. Wang, "A motion image detection method based on the inter-frame difference method," *Applied Mechanics and Materials*, vol. 490-491, pp. 1283–1286, 2014.

[15] H. Zheng, A. Wen Ju, and Z. Li, "A motion vehicle detection method based on self-adaptive background subtraction with cumulative inter-frame difference," *Advanced Materials Research*, vol. 655-657, pp. 890–894, 2013.

[16] J. N. Kapur, P. K. Sahoo, and A. K. C. Wong, "A new method for gray-level picture thresholding using the entropy of the histogram," *Computer Vision, Graphics, and Image Processing*, vol. 29, no. 85, pp. 273–285, 1985.

[17] T. P. Kaivosoja, S. Liu, J. Dijkstra, J. Dijkstra, T. Sheth, and O. A. Kajander, "Comparison of visual assessment and computer image analysis of intercoronary thrombus type by optical coherence tomography in clinical patients," *Interventional Cardiology*, vol. 10, no. 3, 2018.

[18] A. Mhenni, E. Cherrier, C. Rosenberger, and N. E. B. Amara, "Analysis of Doddington zoo classification for user dependent template update: application to keystroke dynamics recognition," *Future Generation Computer Systems*, vol. 97, pp. 210–218, 2019.

[19] T. D. Chen, J. Hu, C. Lu, and Z. J. He, "Moving target tracking using sparse optical flow method," *Advanced Materials Research*, vol. 718-720, no. 1, pp. 2335–2339, 2013.

[20] T. Zhou, Y. Song, J. Qin, J. Wu, and Y. Hui, "Improved L-K optical flow method to detect moving targets," *Fujian Computer*, vol. 36, no. 8, pp. 10–13, 2020.

Retraction

Retracted: Application Research on Roller Skater Detection, Tracking, and Trajectory Prediction Based on Video Stream

Scientific Programming

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] S. Duan, L. Meng, D. Ma, and L. Mi, "Application Research on Roller Skater Detection, Tracking, and Trajectory Prediction Based on Video Stream," *Scientific Programming*, vol. 2021, Article ID 2702272, 9 pages, 2021.

Research Article

Application Research on Roller Skater Detection, Tracking, and Trajectory Prediction Based on Video Stream

Shaolou Duan, Lingfeng Meng, Delong Ma, and Liangyu Mi [ID]

Hebei Sport University, Shijiazhuang 050041, Hebei, China

Correspondence should be addressed to Liangyu Mi; miliangyu@hepec.edu.cn

With the continuous progress of science and technology, the sport of roller skating has developed rapidly and the technical level of the game has become higher and higher. Its sport performance has been rapidly improved. However, China's roller skating is relatively late, and there is still a certain gap compared with many Western developed countries. In order to improve the performance of China's roller skating, this study takes the representative Chinese and foreign excellent speed skaters as the research object and compares the sprinting technology of Chinese and foreign excellent speed skaters by using image measurement and image analysis to obtain the kinematic parameters and data of the athletes' sprinting technology in the competition state. In view of the problem that the current video target tracking algorithm is easy to follow multiple targets, a video multiobject detection and tracking algorithm with improved tracking learning detection (TLD) is studied with the skater in the video as the research object. For the lost target, the prediction function of Kalman filter algorithm is used to track the trajectory of the typical target in the video, and the trajectory tracked by Kalman filter algorithm is used to compensate the lost part of TLD algorithm, so as to obtain the complete trajectory of the typical target in the video to improve the accuracy of video multiobject tracking. Since the existing trajectory prediction algorithms have the limitation of poor accuracy, a social-long short-term memory (Social-LSTM) network-based video typical target trajectory prediction algorithm is proposed to predict the trajectory sequences of typical targets to be detected by incorporating the contextual environment information and the interaction relationship between multiple target trajectories into the Social-LSTM network. The simulation results show that the proposed trajectory prediction algorithm outperforms the traditional LSTM algorithm, Hidden Markov Model Algorithm, and Hybrid Gaussian model algorithm, which is helpful to improve the accuracy of video roller skater target trajectory prediction, and the tracking success rate is 0.98.

1. Introduction

"Roller skating is also called roller skating; it is an ice sport in the land auxiliary training process gradually evolved in the formation of sport." With the continuous development and growth of sports, roller skating has also been flourishing, and the new projects of modern roller skating have emerged and diversified, developing different forms of sports such as speed skating, roller skating ball, extreme roller skating, and figure skating, which are popular among people [1]. The International Roller Derby Federation was officially admitted as a member of the International Olympic Committee (IOC) at the IOC Session held in Prague, Czechoslovakia, in 2007 [2]. Speed skating is a sport that uses roller skates or skating tools on the field or road to determine victory or defeat by speed, and as a component of roller skating, it is the sport that best reflects the characteristics of roller skating, such as competitiveness, skill, and fierce confrontation [3]. Speed skating is a relatively young sport, but in many Western countries, speed skating has been popularized.

At present, the world of speed skating has reached a fairly high level of competition, and sport performance is constantly improving [4]. In China, the sport of roller skating is in a maturing stage, and there is a big gap compared to many countries, especially in Western Europe. Although, in recent years, China's speed skating project has been developed rapidly, especially in the long-distance

project with very good results, such as a gold and a silver medal won by China's players in the World Roller Skating Championships marathon competition, relative to some high level countries, our gap is still very obvious [5, 6]. For example, in the men's individual time trial of the World Roller Skating Championships in 2007, our athlete He Xin only ranked first and did not enter the final. Therefore, it is very meaningful to study the short-distance project of speed skating in China, which can, on the one hand, provide certain training theoretical reference for China's speed skating coaches, so as to improve the technical level of China's speed skaters in the short-distance project and improve the sports performance, on the other hand, provide help for the popularization of China's speed skating technology, and can promote the further development of China's speed skating sport. It can also promote the further development of speed skating in China and play an indirect role in the implementation of the latest national fitness program in China [7, 8].

Because the individual time trial start has a certain uniqueness and the project is a competitive race project to determine the winner of the race, the start is particularly important and the individual time trial start is characterized by speed roller skaters not listening to the gunshot of the starter to start [9], but according to their own preparation in the coincidental seconds of time to decide to start, when the electronic timer infrared of the electronic timer is cut off by the runner's wheel, the timer will automatically calculate time [10]. This way, it is very advantageous for the runner, and there is sufficient time to self-adjust and to adjust the body to the best state, so that it is more conducive to start after the maximum impulse to start, with the shortest time to get speed. The sprint is an acceleration process after the start; the purpose is to quickly reach the dynamic from the static state, in the shortest time and the most energy-efficient way to get the best speed. It has a very important role in the generation of initial speed, so the study of sprinting technology is very important [11]. The technical movements in the sprinting process are very complicated, and the details of each technical movement are changed in the short-sprinting stage, such as body posture, pedal angle, stride frequency, stride length, and joint angle. These elements are the key to the sprinting technique, and the changes of these elements will directly affect the athletic level of the runner, and they interact with each other and are interrelated, the chain reaction of several factors.

2. Related Work

2.1. Starting Technique.
"The kinematics of the knife starting technique of China's excellent male speed skating athletes research" in [12] reveals the athletes in the starting phase of the lower limbs of the ice stirring process contained in the kinematic law, to enrich the technical theory of speed skating, for teaching, training, and scientific research to provide reference basis short-track speed skating starting technology analysis; the article introduced the short-track speed skating is China's dominant project, and the work in [13, 14] introduces that short-track speed skating is the dominant sport in China, especially the female athletes have long been the world leaders in this sport. Their advantage lies in the starting phase, which is highlighted in the reaction speed, explosive power, and starting technique. In the training through continuous summary, the starting technology should have a deeper understanding. In [15], through the analysis of the domestic and foreign excellent mobilization starting technology, an in-depth description of the short-track speed skating starting technology essentials and its training methods is given. Combined with the competition rules of the project, it puts forward rationalized suggestions to give full play to its characteristics and advantage [16]. The Key Points of Starting Technique and Training for Short-Distance Projects of Youth Speed Skaters "sums up that youth athletes should seize the key technical points in training practice and constantly improve training methods and means in order to improve the starting technique level continuously. A Trial on Starting Technique of Claypool Ice Skating" predicts that, with the improvement of technical proficiency and mastery, there is potential for the Claypool Ice Skaters to improve their starting performance again [17].

2.2. Sports Biomechanics Research.
"Biomechanical Analysis of Starting Techniques of Excellent Speed Skaters" in [18] is a comparative analysis of the old and new ice skate starting techniques of domestic excellent speed skaters using photomechanical methods, and the following conclusions are drawn. At present, the distance from the center of gravity projection to the front support point of the new type of ice skate standing start by our athletes is large at the moment of starting, and the forward leaning angle and back pedaling angle of the new type of ice skate are generally large. Angle and backstroke angle are generally large, which leads to the instantaneous horizontal speed drop of the athletes' sprint phase. Athletes are using high-frequency skating to quickly complete the sprint phase and into the gliding phase [19]. This article starts from the comparative photomechanical analysis of the old and new ice skating starting techniques of domestic outstanding speed skaters, trying to find out the photomechanical parameters and respective characteristics of both, in order to improve the speed skating starting technique and improve the speed skating training level. "A Comparative Study on the Starting Movements of Chinese and Japanese Men's Speed Skaters in Meters" in [20] aims to reveal the characteristics of the starting movements of Chinese and Japanese athletes. The authors came to the result that the Chinese athletes' starting preparatory posture has small support point spacing and small front stability angle and their center of gravity is more forward, which creates favorable conditions for completing the transition from the static state to the fast starting stirring state, and their stability is slightly lower than that of Japanese athletes. Kinematic analysis of an excellent speed skater Yu Fengtong's starting technique in [21] shows that a higher stride frequency in the starting phase is an important factor for Yu Fengtong's fast starting speed. In the starting process, the center of gravity is less fluctuating, and the angle of stirrups

is closer to the angle of forward leaning of the upper body so that a larger center of gravity speed is obtained, and the starting stirrups are completed with a higher frequency [22].

2.3. Algorithm of This Paper.

In this paper, we propose a video multiobject detection and tracking algorithm based on improved TLD to obtain the complete trajectories of moving targets. The moving trajectory data are analyzed, and a Social-LSTM-based video typical target prediction algorithm is proposed to improve the accuracy of prediction by combining contextual feature information.

2.4. Improved TLD-Based Video Multiobject Detection and Tracking Algorithm.

Video target trajectory acquisition mainly includes video target detection and video target tracking, detecting the position of video targets [23]. Then, the video targets are tracked and the position of video targets in each image frame is marked to form the motion trajectory of targets in the video. Video target trajectory acquisition methods can be mainly classified into tracking methods based on deep learning combined with correlation filtering, tracking methods based on prediction, and tracking methods based on optical flow [24]. The classical algorithms corresponding to the three methods are efficient convolution operator ECO [25] algorithm for tracking combined with depth features and correlation filtering [13], Kalman filtering algorithm, and Lucas–Kanade (LK) [26] optical flow algorithm. Although these three algorithms can detect video target trajectories in a specific environment well, there are still some problems in detecting video target trajectories.

Robust tracking algorithms need to be designed in complex scenes to solve the unavoidable problems in practical applications such as lighting changes and obstacle occlusion. Depth features obtained by deep learning methods are more accurate than traditional manual features, but the tracking real-time performance is low. Correlation filtering transforms the computation to the frequency domain, which substantially improves the tracking speed. Combining deep learning with correlation filtering can fully utilize the advantages of both methods.

When there are multiple targets in a video scene, effective tracking of multiple targets is not feasible by relying solely on one of the current tracking algorithms [15]. It is necessary to select a reasonable tracking algorithm from the current mainstream tracking algorithms and combine several different tracking algorithms to complement the advantages and disadvantages to study and design the moving target trajectory tracking algorithm.

The TLD algorithm is a kind of tracking-while-detecting algorithm; when there are occluding objects in the video, the target will not be detected and lost; when there are no occluding objects, the algorithm is a stable and effective algorithm. In order to improve the accuracy of video target tracking, this paper proposes a video multiobject trajectory acquisition algorithm with improved TLD to track and obtain the running behavior trajectory of the target in the case of target movement. The TLD algorithm is used as the basic tracking algorithm, and the Kalman filtering algorithm

is used as the auxiliary tracking algorithm. The trajectory obtained by the Kalman filtering algorithm is used to supplement the missing part of the trajectory obtained by the TLD algorithm, and the complete trajectory of the video target is obtained. The specific steps of the algorithm are as follows:

Step 1: video is input.

Step 2: TLD algorithm and Kalman filtering algorithm are initialized.

Step 3: the video target is tracked and detected using the TLD algorithm, and it is determined whether the target profile is detected in the current video frame; if so, Step 4 is executed; otherwise, the state equation of the Kalman prediction stage is updated, and Step 6 is executed.

Step 4: it is determined whether the detected target contour is complete; if so, Step 5 is executed; otherwise, the equation of state in the Kalman prediction phase is updated and Step 6.1 is executed.

Step 5: the training parameters are updated in the TLD learning model.

Step 6: the location of the video target is recorded, and it is determined whether the video is finished. If the video is over, the video target trajectory data are output; otherwise, Step 1 is executed.

2.5. Video Target Trajectory Prediction Algorithm Incorporating Contextual Features.

Motion trajectories of video typical targets have temporal dependencies between adjacent frames, while different motion targets may have mutual effects in the same frame. If these two problems cannot be solved reasonably, the accuracy of video typical target motion trajectory prediction will be affected. In order to solve the abovementioned problems, this paper proposes a Social-LSTM-based algorithm for video target trajectory prediction [27]. The algorithm uses LSTM to process the temporal relationship between the same type target motion trajectory in different video frames. When dealing with multiple typical targets, the LSTM is used to predict the trajectory of each target separately, and a social-pooling layer is added between multiple LSTMs to deal with the interaction between multiple targets, so as to predict the trajectory of the typical targets of the video more accurately.

2.6. Social-LSTM Construction Method.

Social-LSTM is an improvement of LSTM, which is a network model that adds an S-Pooling to the LSTM structure to predict the trajectory of a target considering the influence of neighboring targets in the same frame on its trajectory. This S-Pooling allows the LSTMs of spatial proximal sequences to share the hidden states of each network with each other. The Social-LSTM model effectively considers the influence of video typical target running rules and other targets in the same scene and jointly predicts the motion trajectories of all typical targets in that scene [28]. The structure of the Social-LSTM network is shown in Figure 1. In Figure 1, h_i ($i = 1, \ldots, n$) denotes the
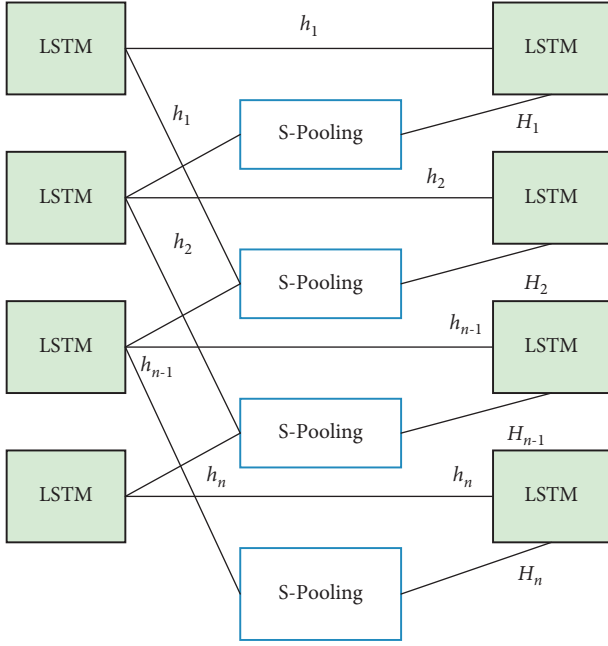
Figure 1: Social-LSTM network structure.

predicted trajectory of the LSTM corresponding to the ith video target in the same scene and n is the number of video targets in the same scene and is the hidden state of the ith video target in the same scene.

Figure 1 represents an LSTM neural network, and each LSTM is used to predict the trajectory of a typical target in the video. Since LSTM networks use unsupervised learning, each LSTM needs to learn historical trajectory data to obtain its weight parameters and use the network to predict the coordinate trajectory of the video target at future moments. Since the trajectory of a target in a video may be affected by the motion of its neighboring targets, this paper connects the LSTM network of the target to be detected with the LSTM network of its neighboring targets through an S-Pooling layer to form a new network, namely, Social-LSTM, and uses this network to predict the trajectory of the target to be detected.

The role of S-Pooling is to translate the influence of its neighboring targets on the motion of the target to be predicted (i.e., contextual information), H_i^t and H_i^t the formula is given by

$$H_i^t(m, n, :) = \sum_{j \in N_i} I_{mn}(x_j^t - x_i^t, y_j^t - y_i^t) h_j^{t-1}, \quad (1)$$

where H_i^t represents the hiding state of the ith video target corresponding to the LSTM network at time t; $I_{mn}(x, y)$ represents the indication function. If the trajectory coordinates (x, y) are in the (m, n) neurons of the grid, the $I_{mn}(x, y)$ value is 1; (x_i^t, y_i^t) represents the trajectory coordinates of the ith video target at time t, (x_i^t, y_i^t) represents the trajectory coordinates of the jth video target at time t, and $(x_j^t - x_i^t, y_j^t - y_i^t)$ represents the difference between the trajectory coordinates of the jth and ith video at time t; N_i represents the motion attribute of the ith adjacent video target; and h_j^{t-1} represents the LSTM prediction trajectory corresponding to the jth video target at time T-1.

After calculating the hidden state of the S-Pooling layer, the hidden state of the target to be detected and its own motion state are fused with the S-Pooling layer to calculate the trajectory of the target to be detected, and the calculation formula is shown as follows:

$$\begin{cases} e_i^t = \phi\left(x_i^t, y_i^t; \mathbf{W}_e\right), \\ a_i^t = \phi\left(H_i^t; \mathbf{W}_a\right), \\ h_i^t = \mathrm{LSTM}\left(h_i^{t-1}, e_i^t, a_i^t; \mathbf{W}_{\mathrm{LSTM}}\right), \end{cases} \quad (2)$$

where $\mathbf{W}_e$, $\mathbf{W}_a$ is the weight matrix of embedding; $\mathbf{W}_{\mathrm{LSTM}}$ is the weight matrix of LSTM; a_i^t and e_i^t are the pooled social hidden state tensor and coordinates, respectively; and h_i^t denotes the predicted LSTM trajectory of the ith video target at time t.

The Social-LSTM-based video target trajectory prediction algorithm uses LSTM to deal with the temporal relationship of the same target motion trajectory between different frames. When dealing with multiple targets, the LSTM is used to predict the trajectory of each target separately, and an S-Pooling layer is added between multiple LSTMs to deal with the interaction between multiple targets, so that the trajectory of a typical target of the video can be predicted more accurately [29].

Social-LSTM can take into account the influence factors between neighboring target trajectories and add the influence factors to S-Pooling. This network model can take into account the video target's own trajectory pattern and the influence of other neighboring targets on its trajectory and fuse the two to predict the trajectory of the video target, which helps to improve the accuracy of typical target trajectory prediction.

2.7. Contextual Feature Extraction Methods. The biggest difference between video and image is the possession of temporal dimensional information. Contextual features, i.e., the dependence of video motion targets in the two frames before and after the video image and the influence of other motion targets in the same frame on the trajectory of the target, are to be detected. In order to accurately predict the trajectory of a typical target and eliminate the influence of other neighboring targets on its trajectory, this paper adds an S-Pooling to LSTM to extract the contextual features of a typical target in the same frame, and the network can take into account the influence of other moving targets in a grid on its trajectory in the same frame.

Since the memory unit in the LSTM network can store the operational status of the typical target in the previous frames, the operational status of the typical target in the current frame can be taken into account by the self-looping unit in the LSTM network. In this paper, we use the memory unit in the LSTM network to extract the dependency relationship between the typical target in the previous and the next two frames.

2.8. Social-LSTM-Based Algorithm for Predicting the Typical Target Trajectory of Video. In this paper, we propose a Social-LSTM-based video typical target trajectory prediction

algorithm to predict the trajectories of typical targets in videos. Since LSTM uses unsupervised learning, the weight parameters in the network can only be obtained by learning the training data. To calculate the weight parameters in this network, a bivariate Gaussian distribution with mean $\mu_i^{t+1} = (\mu_x, \mu_y)_i^{t+1}$, standard deviation $\sigma_i^{t+1} = (\sigma_x, \sigma_y)_i^{t+1}$, and correlation coefficient ρ_i^{t+1} is used to predict the trajectory coordinates of the video target at $t + 1$ using the hidden state, and the mean, standard deviation, and correlation coefficient in the bivariate Gaussian distribution can be found by a $5 \times D$ order weight matrix $\mathbf{W}_p$. The calculation of the trajectory of the video target at time t predicted by the hidden state is shown as follows:

$$(\widehat{x}, \widehat{y})_i^t \sim N\left(\mu_i^t, \sigma_i^t, \rho_i^t\right), \tag{3}$$

where $(\widehat{x}, \widehat{y})_i^t$ denotes the trajectory coordinates of the typical target of the video at time t predicted by using the hidden state; $N(\mu_i^t, \sigma_i^t, \rho_i^t)$ denotes the bivariate Gaussian distribution. The mean, standard deviation, and correlation coefficients of the bivariate Gaussian distribution are calculated as follows:

$$\left(\mu_i^t, \sigma_i^t, \rho_i^t\right) = \mathbf{W}_p h_i^t. \tag{4}$$

In order to find the weight parameters in this network, the negative log-likelihood loss function is defined in this paper, and the unknown weight parameters in this network are obtained by finding the minimum value of the negative log-likelihood loss function, and the negative log-likelihood loss function is defined in the following form:

$$\text{loss}_i\left(\mathbf{W}_e, \mathbf{W}_1, \mathbf{W}_p\right) = -\sum_{t=T_s}^{T_e} \ln\left(P\left(x_i^t, y_i^t \mid \mu_i^t, \sigma_i^t, \rho_i^t\right)\right), \tag{5}$$

where loss_i is the loss function of each video target trajectory; T_s is the predicted start time; T_e is the predicted end time; and $P(x_i^t, y_i^t, \mu_i^t, \sigma_i^t, \rho_i^t)$ is the conditional probability of the predicted trajectory with respect to a bivariate Gaussian distribution [30].

Since the added S-Pooling does not introduce additional unknown parameters, the network only needs to find the weight parameters in the LSTM structure. Based on the historical trajectory data of the video target and the minimum value of the loss function, the unknown $\mathbf{W}_1$, $\mathbf{W}_e$, and $\mathbf{W}_p$ weight parameters in the LSTM structure can be found.

After the network learning is completed, the test data of the video target trajectory is input, and the predicted trajectory of the video target is derived according to equation (4); using the predicted trajectory, the hidden state of the video target is derived according to equation (3), and the trajectory of its neighboring targets in the video scene is obtained.

3. Data Preprocessing

Canny edge detection is a very popular edge detection algorithm, which was proposed by [31]. It is a multistage algorithm, i.e., it consists of multiple steps. First is image noise reduction. We know that the gradient operator can be used to enhance an image, essentially by enhancing the edge contours; i.e., it is possible to detect the edges. However, they are all heavily affected by noise. Then, our first step is to think of removing the noise first because noise is where the gray scale changes a lot, so it is easy to be identified as a pseudoedge. The purpose of doing this is to keep only strong edge contours then. Some edges may not be closed and need to be supplemented from the point that meets between low and high to make the edges as closed as possible, as shown in Figure 2.

4. Roller Derby Forecast

4.1. Analysis of the Correlation between the Sprint Phase and Final Performance. As the individual time trial is a segment distance project and it is a project to decide the winner of the race by the time, the initial speed has a relatively important impact on the final result [23]. Sprint technology is mainly reflected in the distance of the stage after the start and before entering the skating run. The distance is around the new rules of the World Skating Federation from the starting point to the distance of entering the curve in meters. The correlation coefficient can directly show that the speed skating sprint is significantly correlated with the performance of the meter, which directly affects the performance of the meter. The correlation coefficient of this study can be calculated by the sprint performance and the meter performance, and the correlation coefficient, therefore, shows that the sprint performance in speed skating has a significant correlation with the final performance of an individual time trial, which can directly affect the performance of this project. It can be seen through Table 1 that our speed skaters' sprint speed in the first meter is significantly lower than that of the world's best athletes, the average performance of the world's best speed skaters in the sprint stage in meter, and the average performance of our speed skaters in the sprint, which is different from the world level. The best short-distance skater in China, He Xin, has a worse spring than the world average, and Liu Zhicong has a worse sprint than the world average, so there is still some room for improvement. Through the improvement of technology, body strength, and flexibility, it is possible to improve the sprint speed, which is more conducive to the improvement of the meter performance.

4.2. Analysis of Stride Frequency, Stride Length, and Sprint Speed

4.2.1. Comparative Study of Step Frequency. Step frequency is one of the important factors to determine the speed, and step frequency is the number of times the two legs alternate in a unit of time, mainly by the impact of factors such as pedal time and support time; in addition, the athlete's body coordination, strength, flexibility, height, and many other factors have a certain impact on the step frequency [14].

From Table 2, we can see that speed skaters are using high-frequency isochronous techniques in the sprint phase, which is in line with the technical characteristics of speed skating sprint. Our skaters' stride frequency is in
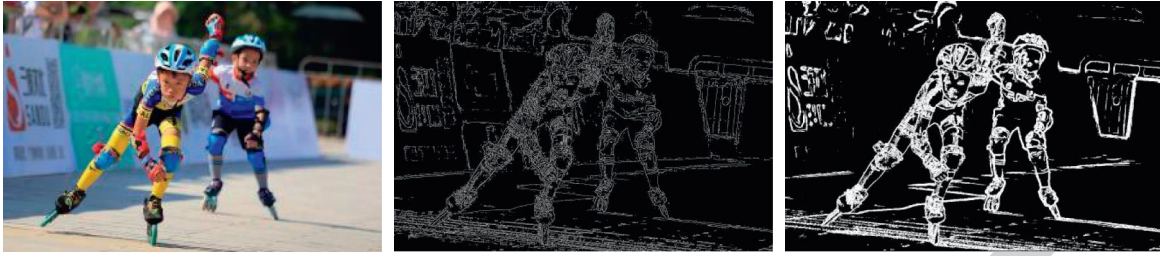
Figure 2: Different treatment effects.

Table 1: Six male athletes from China and abroad before the results and gamma statistics.

Full name	8 m	300 m
Joey	1.25	24.250
Perdo	1.22	24.265
Song Qingyang	1.27	24.499
He Xin	1.37	25.395
Liu Zhicong	1.42	25.666
Yuan Jianjin	1.40	25.622

Table 2: World's best athletes' sprint phase step frequency statistics in subseconds.

Full name	Step frequency
Joey	3.22
Perdo	3.25
Song Qingyang	3.46
$\overline{x}$	3.31
S	0.26

subseconds, and the standard deviation is in subseconds. The analysis of the table shows that the difference between our speed skaters and the world average is in subseconds. Although the stride frequency is related to the innate individual conditions, however, from another aspect, it reflects that there is a certain gap between the body strength and coordination of speed skaters in China, and the step frequency of Song Qingyang in the same Asian region can reach subseconds, and the height is in meters.

So, in comparison, our athletes are still lacking in this area, which can be improved through the training of the body's first power, and there is a lot of room for improvement, as shown in Table 3. However, attention should be paid to the intensity of the load in training; too much load will affect the speed of completion of the action, if the weight is too small, it will be difficult to realize the rapid power, and athletes should try to experience the maximum force and maximum speed of the feeling.

4.3. A Comparative Study of Stride Length. The stride length is mainly influenced by the force and speed of the stirrup, while the height, leg length, and flexibility of the athlete are all factors that affect the stride size. The sport effect of sprinting is mainly reflected through the stride speed. A reasonable combination of stride frequency and stride length can improve the performance of athletes in sprinting.

Table 3: Domestic outstanding athletes' sprint phase step frequency statistics in subseconds.

Full name	Step frequency
He Xin	2.77
Liu Zhicong	2.79
Yuan Jianjin	2.65
$\overline{x}$	2.73
S	0.13

Domestic and foreign outstanding athletes sprinting stage stride length statistical comparison is as shown in Table 4. The stride length and stride frequency are comprehensive factors that determine the stride speed in the sprinting process, which can directly reflect the effect of the athlete's sprinting and is one of the important indicators to evaluate whether the speed skater's sprinting technique is reasonable. Excellent athletes can highlight the characteristics of stride frequency and stride length in the competition through training and their own physical conditions.

As shown in Figure 3, the large stride length in the game can seize the opportunity, but from the speed skating start after the sprint phase to analyze, the first step should not be too large, and if the stride length is too large, it will certainly cause the body's center of gravity to be too high and the pedal angle becomes large, resulting in the reduction of the horizontal force of the pedal. The initial speed and acceleration of the sprint are affected, which in turn affects the starting speed. The first step of the start should be small; in the starting stirrup, the stirrup angle is small to try to make the body thrust forward, and stirrup reaction force in the horizontal direction of the force will be larger, conducive to the start of a greater thrust.

In the speed skating sprint process, the position of the foot support point is always changing; when the foot hits the ground after the start of the stirring process, it changes to forward skating, and the direction of power propulsion is not backward but lateral; although inertial skating can increase the stride length, but because there is not enough inertia after the start, speed skaters in the sprint process increase the friction coefficient between the wheel and the track, so as to obtain a relatively high speed. However, it is necessary to accelerate the frequency to overcome the frictional resistance generated by the contact between the wheels and the track, so that, in the speed skating sprint process, the appropriate way to reduce the stride length is more conducive to the improvement of sprint speed. Therefore, if you keep the same pedaling force per step in the

Table 4: Domestic and foreign outstanding athletes sprinting stage stride length statistical comparison table (m/s).

	First step	Step 2	Step 3	Step 4	Step 5	Step 6
Domestic athletes	0.42	0.55	0.69	0.93	1.25	1.32
Foreign athletes	0.33	0.46	0.84	1.22	1.38	1.46

Figure 3: Tracking error curve of the arm coordinate position by different methods.

Figure 4: Images of the skating process of the athlete.

fixed camera, and the marker points were put on the roller skating arms within 5 meters of the filing range to get a complete roller skating action of each athlete, totaling 50 images. The simulation is mainly implemented in the MATLAB software environment under the Windows7 system using different methods to detect and track the skating image targets during the skating process of skaters [7, 17].

Firstly, the error curves of the skating arm coordinate position and the real position of the skating arm during the skating process under the method of this paper, the template update method, and the least-squares line method are given, as shown in Figure 3.

Analysis of Figure 3 shows that the corresponding coordinate positions of the template update method and the least-squares linear method deviate from the real position, while the corresponding coordinate position tracking error curve of the skating process of the skating arm of this paper has a smaller error, indicating that the method of this paper is closer to the real skating process of the skating arm trajectory and the tracking effect is better.

The tracking success rates of the three different methods are compared in Table 5. The tracking success rates (%) of the three methods were compared, and the results are shown in Table 5, where the tracking success rate = total number of pixels accurately tracked N/total number of pixels in the target area of the skating arm target.

Analysis of Table 5 shows that the tracking success rate of the skater's skating arm target is much higher than that of the template update method and the least-squares linear method, which is mainly because this method first extracts the outline of the skater's skating arm target, constructs the global matching approximation function of the skating arm target based on the position and size of the skater's skating arm target, and thus, determines the skating arm target tracking success rate. This makes the tracking success rate of the roller skater's roller skating arm target higher.

The trajectory tracking experiments with the skating arm motion target during the skating process were conducted by this paper, the template update method and the least-squares linear method, respectively. The track time (s) of the three methods is compared, and the results are shown in Figure 5.

sprinting phase, you can control the sprinting speed by changing the pedaling frequency.

From a comparative analysis of Chinese and foreign speed roller skaters' sprint phase stride frequency, stride length, and the first six steps of the sprint, it can be seen that China's outstanding male speed roller skaters have a certain gap in the sprint phase stride frequency and stride length compared to foreign outstanding athletes; in the first two steps of the sprint phase after the start, the average value of our athletes is the first step and the second step is meters, while the first two steps of the sprint phase of foreign athletes are relatively small. The first two steps of the sprint phase are relatively smaller than the first two steps of foreign athletes. Therefore, the main factors that lead to the difference of speed skaters' performance in the sprint phase are stride frequency and stride length.

4.4. Simulation Proof. In order to prove the effectiveness of the proposed algorithm of the roller skating process arm skating image trajectory tracking method, it is necessary to conduct a simulation; the experimental object is selected from 15 male excellent athletes of our school of physical education, with professional training time of about 8 years; these athletes in the experimental stage are without sports injuries and are proficient in roller skating basic action essentials and are a good representative, as shown in Figure 4. The entire roller skating process was filmed with a

Table 5: Comparison of the target tracking success rate of different methods of roller skating arm.

Method	Number of image pixels (piece)	Total number of target pixels (piece)	Number of tracking pixels (piece)	Tracking success rate (%)
Paper method	500	357	350	98
Template update method	500	357	284	79
Least-squares linear method	500	357	268	76
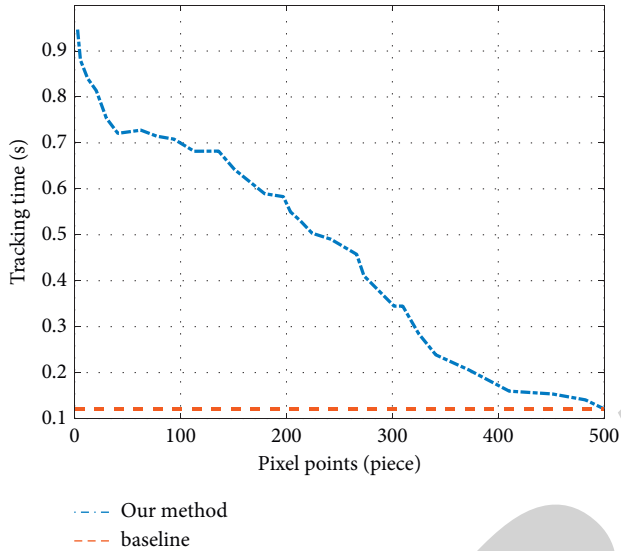
Figure 5: Time spent tracking the trajectory of different methods of roller skating arm movements.

This is mainly because this method extracts the target trajectory of the skater's skating arm by estimating the Gaussian model parameters of the skating arm differential image and introducing the boundary detection operator to construct the skating arm motion boundary image and then extracts the target trajectory of the skater's skating arm during the skating process. Based on this, the morphological operator is used to calculate the body ratio and tightness of the target region of the skating arm. This process solves the contradiction between the amount of computation and the amount of information to a certain extent, which makes this paper's method to track the target trajectory of the skating arm during the skating process of the skater in a shorter time. Athletes should strengthen leg training and other links.

5. Conclusions

In this paper, we use video measurement and image analysis to obtain kinematic parameters and data of athletes' sprinting techniques in competition and conduct a comparative study on the sprinting techniques of excellent speed skaters in China and abroad. In this paper, video multi-objective detection and tracking algorithm with improved TLD is investigated with the skaters in the video. The experimental results show that the proposed trajectory prediction algorithm is better than the traditional LSTM algorithm, Markov model algorithm, and hybrid Gaussian model algorithm, which is helpful to improve the accuracy of target trajectory prediction of video skaters, and the tracking success rate is 0.98.

Data Availability

The dataset used in this paper is available from the corresponding author upon request.

Conflicts of Interest

The authors declare no conflicts of interest regarding this work.

References

[1] M. Archana and G. Kalaisevi, "Object detection and tracking based on trajectory in broadcast tennis video - ScienceDirect [J]," *Procedia Computer Science*, vol. 58, pp. 225–232, 2015.

[2] W. Marynowsky, S. Ferguson, A. Fraietta, and O. Bown, "The ghosts of roller disco', a choreographed, interactive performance for robotic roller skates," in *Proceedings of the Fourteenth International Conference on Tangible, Embedded, and Embodied Interaction*, pp. 631–637, New YorkNYUnited States, February 2020.

[3] K. Kardiyono, S. Soegiyanto, and T. Rahayu, "Roller skating activities as a reflection of the community life style in semarang municipality[J]," *The Journal of Educational Development*, vol. 6, no. 2, pp. 265–271, 2018.

[4] X. Yu, H. W. Leong, C. Xu, and Q. Tian, "Trajectory-based ball detection and tracking in broadcast soccer video," *IEEE Transactions on Multimedia*, vol. 8, no. 6, pp. 1164–1178, 2006.

[5] C. Zhang, T. Xie, K. Yang et al., "Positioning optimisation based on particle quality prediction in wireless sensor networks," *IET Networks*, vol. 8, no. 2, pp. 107–113, 2019.

[6] Z. Huang, J. Wang, L. Pi, X. Song, and L. Yang, "LSTM based trajectory prediction model for cyclist utilizing multiple interactions with environment," *Pattern Recognition*, vol. 112, p. 107800, 2021.

[7] D. Bronzi, Y. Zou, F. Villa, S. Tisa, A. Tosi, and F. Zappa, "Automotive three-dimensional vision through a single-photon counting SPAD camera[J]," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 3, pp. 782–795, 2015.

Research Article

Research on Learning State Based on Students' Attitude and Emotion in Class Learning

Dong Huang (ID) **and WeiXin Zhang**

The College of Education, Yunnan Normal University, Kunming 650000, Yunnan, China

Correspondence should be addressed to Dong Huang; 2011010118@st.btbu.edu.cn

In basic education, timely and accurate grasp of students' classroom learning status can provide real-time information reference and overall evaluation for teachers and managers, which has a very important educational application value. At present, a lot of information technology is applied in the analysis of classroom student behavior state, and the state analysis technology based on a classroom video has the characteristics of strong timeliness, wide dimension, and large capacity, which is especially suitable for the analysis and acquisition of students' classroom state, and attracts the attention of major educational technology companies. However, the current student state acquisition technology based on video analysis lacks large scenes and has low practicability, and finally, the video-based student classroom behavior state analysis technology mainly focuses on a single behavior feature, which cannot fully reflect the student's classroom behavior state. In view of the above problems, this study introduces the face recognition algorithm based on a student classroom video and its implementation process, improves the hybrid face detection model based on a traditional model, and proposes the neural network algorithm of student expression recognition based on a visual transformer. The experimental results show that the proposed algorithm based on students' classroom videos can effectively detect students' attention and emotional state in class.

1. Introduction

It has always been difficult for teachers and administrators to keep track of all students' classroom learning. In traditional education, in order to better educate students, teachers conduct after-class analysis through traditional methods such as teaching diaries, teaching files, watching videos, and homework analysis, and then provide solutions according to the results [1]. However, the traditional after-school analysis method not only increases the burden of teachers but also makes it difficult to ensure the comprehensiveness, objectivity, and real time of the analysis results. With the rapid development of economy and the continuous promotion of education informatization in all countries around the world, teachers are in urgent need of an intelligent classroom learning state analysis tool to help teachers get the learning state of all students in class, deal with and display the overall state of the classroom after class, and then reflect on and improve their teaching process [2]. Intelligent analysis means have important practical significance for improving teachers' professional level and students' learning effect.

In recent years, education in most countries has shifted from elite education to mass education. The widespread popularization of education aims at improving the quality of education in the whole society. The evaluation system of classroom learning has always been accompanied by mass education, but students' test scores are often taken as the absolute criterion for evaluation. As a result, we only know the quality of the results, but do not know the causes of the results, and we cannot find appropriate adjustment schemes to improve the quality of education [3]. The students took notes seriously in class and actively communicated with the teacher, which reflected that the teacher's teaching content was attractive. On the contrary, students bow their heads or look out of the window for a long time with negative facial expressions. These states indicate that they do not understand the content of

the class or do not concentrate on the class, which indirectly indicates that the way of teaching cannot attract students well. Usually, school administrators do not consider students' performance in class and only rely on students' scores and leaders to check the classroom situation. Such unilateral evaluation of teachers' teaching quality is not accurate, nor can it help teachers to understand the real learning situation of students. Therefore, in this case, the analysis of the classroom teaching process of students listening to the status of education is crucial [4].

There are many reasons why students' learning efficiency is not good in the process of classroom teaching. As the subject of receiving knowledge in the whole process, students' learning status in class can be studied as an important evaluation index of students' learning efficiency in class. At the same time, it is also the key factor to realize the effective teaching of teachers. Teachers hope to master each student's classroom learning situation comprehensively and accurately in the classroom teaching process, so as to make corresponding adjustments to the teaching content and teaching progress, so as to achieve more efficient teaching purpose. Generally speaking, if a student listens to the teacher carefully, most of his attention is directed toward the teacher during class; that is, he looks up and listens attentively to the teacher, and his emotional state is concentrated.

The existing classroom surveillance cameras are basically installed in the front and back of the classroom. By analyzing the images obtained from the cameras in front of the classroom, students' listening status can be judged. The detection of students' learning state in the classroom scenario is divided into two steps. First, the head position of the students is detected, and then, the head state of the students is further identified to see whether the head is looking up to listen to the teacher or looking down at the mobile phone or doing other things [5]. However, at present, many teachers understand students' situation in classroom learning through classroom observation and questioning, which often leads to the lag and one-sidedness of classroom information transmission and feedback. In particular, with the popularity of electronic devices such as smartphones and tablet computers, a large number of "phubbers" have emerged in classroom teaching [6]. Therefore, the combination of statistics and analysis of students' "head-up rate" in class and intelligent algorithm analysis of students' emotional state can judge students' class concentration to a certain extent, thus helping teachers effectively improve classroom teaching efficiency [7].

2. Related Work

Student's state of learning is an important index of students' classroom learning efficiency. The state of students' classroom learning generally refers to whether students look up at the teaching content for a long time and actively communicate with the teacher, whether they take notes carefully, and whether their facial expressions are in a positive or negative state [8]. Wearable devices are invasive to some extent and will inevitably have a certain influence on the subjects. There is a gap between the data obtained and the real state of students' classroom behavior. In addition,

wearable devices are expensive, large in size, and complicated in the process of wearing, so it is difficult to popularize them in practical classroom education [9]. With the rapid development of smart devices, another method to collect video, image, voice, and other digital signals through cameras, microphones, and other devices has become widely popular. This method extracts information from these digital signals, such as students' facial expressions, natural language, and body posture, and finally processes, analyzes, and integrates this information to get students' classroom behavior state. The classroom student state analysis technology based on video images only needs to use the classroom camera system, which is of low cost and less invasive and has almost no influence on the learning process of students [10, 11]. Through the artificial intelligence algorithm, students' learning status can be analyzed in real time, comprehensively and multidimensionally. Through the above analysis, students' classroom behavior, emotional state, learning state, and other situational information in the class can be captured by intelligent monitoring equipment and mobile learning devices. Therefore, the current intelligent classroom obtains students' physiological signals and behavioral state data through various devices, and then collects and analyzes these data to get the current students' classroom behavior state, so as to better grasp the classroom situation and timely adjust the teaching strategy to improve the teaching effect. [12].

With the deepening of the research, the scholars have analyzed the research status of the analysis of students' classroom behavior state from two aspects, physiological signals and visual images, according to the different ways of collecting students' characteristics. Physiological signals refer to when people's inner emotions change, the body or brain will send out one or more physiological signals; through the collection and analysis of these physiological signals, such as EEG signals, EMG signals, skin temperature, and eye movement, one can know the current student's mood and psychological state [13]. Nourbakhsh et al. [14] proposed to detect the cognitive load level of students by analyzing the skin signals in the time domain and frequency domain. They collected the skin signals generated by learners in the process of completing learning tasks of different difficulties in experiments and then analyzed these skin signals in the time domain and frequency domain. By comparing the spectral features of skin signals in different difficult learning tasks, we found that the frequency domain features of skin signals had better identification ability for emotional cognition categories. Zhan [15] combined pupil size, blink frequency, and blink frequency with facial expression to construct the recognition framework of learners' emotional state. An intelligent teaching agent evaluated students according to the arousal dimension, interest dimension, and pleasure dimension in the framework and then made corresponding cognitive feedback, such as knowledge point proposal and learning suggestion. This combination of learners' eye movement tracking and facial expression recognition can enable the intelligent teaching agent to more accurately identify the emotional state and cognitive state of distance learners.

Sinha et al. [12] proposed the use of brainwaves and other physiological signals to track and detect learners' cognitive and emotional states during learning. This method uses electroencephalogram (EEG) wave signal to estimate learners' difficulty in understanding the learning content and uses heart rate variability and RGS signals collected by skin electrodes to estimate learners' emotional state, which is compared with learners' current academic performance. Zhu et al. [16] use smart wristbands of wearable devices to extract physiological signals of learners, collect and analyze students' handwriting status and heart rate activity through smart wristbands, and then obtain learners' current cognitive status. This method adopts the method of multisignal synthetic judgment, so the result is relatively accurate. All of the above studies need to obtain the physiological data of learners through wearable collection devices and analyze the classroom behavior state of students by using different physiological performances of people in specific states. Due to the accuracy and specificity of physiological signals, very accurate analysis results can be obtained. However, due to the use of complex wearable devices, the testers will establish a psychological preset and know that they are in a tested state, which will affect the objectivity of the results of physiological signal analysis [17]. Moreover, the cost, size, and deployment requirements of wearable acquisition devices make it difficult for such classroom behavior state analysis methods to be widely used.

The development of image recognition depends on the progress of image equipment (the progress of intelligent equipment such as HIGH-DEFINITION cameras has promoted the progress of image recognition research). The whole process has a lot of work and many links, so the final results are often biased or even wrong. With the deepening of research, intelligent analysis methods based on visual images are increasing. In this method, video images are first captured by the camera, and then, the data are input into the algorithm to identify, record, and analyze the students' expressions, head posture, and other explicit actions, and finally, the current classroom behavior state of the students is given. According to the different behavior characteristics of students, there are mainly four methods based on face detection, head posture estimation, facial expression recognition, and multiple action recognition, which are discussed next.

2.1. Methods Based on Face Detection. Fujisawa and Aihara [13] estimated learners' interest in learning by detecting the transformation of face size. In the experiment, the face detection algorithm uses the front face detector in the OpenCv [19] open-source vision library to conduct face detection through the camera directly above the computer. The experiment proves that the number of times the face is close to the screen and the entertainment of the material are closely related to the learner's interest. OpenCv visual open-source library contains a large number of tools for computer vision, image processing, behavior recognition, and other related fields. Hou et al. [20] proposed the application of face detection technology to the quality assessment of students' lectures in 2016. Haar-like face features that have been trained in OpenCv open source visual library were selected in the experiment, and these features were applied to the AdaBoost cascade algorithm for face detection, and the classroom head-up rate of students was calculated by detecting the number of faces, and the average classroom head-up rate of students in a fixed time was obtained.

2.2. Method Based on Head Posture Estimation. Rahman et al. [21] proposed to track the learning state of learners according to their head posture and the distance between learners and Kinect. In the experiment, a Kinect motion camera was used to obtain the information of the learner's head posture angle and distance depth, and then interest expression function was constructed. This method used a lot of physical knowledge to calculate and then tracked the learner's interest.

2.3. Methods Based on Facial Expression Recognition. Psychological research shows that positive emotions promote cognitive activities during learning, whereas negative emotions hinder cognitive activities. The research results of psychologist Mehrabian [22] show that emotional information consists of 7% language, 38% voice, and 55% facial expression, so students' emotional states can be obtained through the recognition and analysis of facial expression. Feng et al. [23] used 16 Haar-like features to extract face features, learned and trained classifiers with the AdaBoost algorithm, and cascaded strong classifiers to form the final expression classifier. Facial expression recognition is carried out through the facial expression classifier, and facial expression recognition technology is brought into the remote classroom, which realizes the facial expression recognition and emotion judgment system under the network environment for the first time, and improves the efficiency of online teaching and user satisfaction. Cheng et al. [24] selected 34 feature points to define facial geometric features. After marking feature points, Gabor wavelet was used to extract facial feature information, and SVM (support vector machine) was used to classify expressions to obtain expression classifier. The structure model of the intelligent teaching system based on expression recognition and sight tracking technology is proposed. Sun et al. [25] obtained facial expression classification by combining facial AU unit and third-order tensor. In the experiment, AU facial unit was used to eliminate the influence of individual differences on facial expression recognition effect and improve the accuracy of facial expression recognition. By separating facial features from personal facial features, the function of facial recognition and emotional intervention can be realized with high precision. Jiang et al. [26] used a variety of algorithms to identify and study the "confused" expressions of students in the learning process and concluded that the random forest algorithm has the best effect on identifying the confused expressions of students.

2.4. Multifeature-Based Analysis Method. With the development of deep learning, Whitehill et al. [27] marked the degree of participation of students as four levels and collected AU units, hand movements, and head posture information of students' faces by a gaussian wave filter and a support vector machine. Through the continuous training of the audit network, it is concluded that the movements of head lowering, side head, mouth, and eyes have great weight in judging students' participation. Han et al. [28] used AAM (active appearance model) to mark face feature points and then marked key points in the training set and utilized principal component analysis (PCA), extracting average shape by dimensionality reduction as shape model. The researchers studied the tilt of the head and the position of the lip and eye features during class and obtains the data of students "listening," "understanding," "doubt," "resistance," and "disdain." To examine each state of the head posture and to validate a specific analysis of students' expression, a classroom assessment of analysis of facial expression and head posture was conducted. Chen et al. [29] established a random forest model to identify students' head posture and facial expressions and used the teacher-student interaction platform to record the learning interaction between teachers and students in class. Although scholars have conducted a lot of studies, most of them are still based on traditional networks, and there are few recent applications of deep learning neural networks. Based on this, this study integrates traditional methods with the latest methods.

3. Face Recognition Based on Hybrid Architecture

Face detection is a mature aspect in the field of image processing. Scholars have proposed various algorithms for different data sets. Therefore, this study proposes a face detection algorithm under the hybrid architecture based on the characteristics of the row and column distribution of the classroom. The algorithm uses an algorithm with high detection accuracy as the fine detection algorithm and an algorithm with high detection speed as the rough detection algorithm. The algorithm calls the fine detection algorithm and the rough detection algorithm according to different face conditions.

3.1. Description of Algorithm. Face detection algorithm is to use the detection window on the image-intensive multiscale sliding and then determine whether the image in the detection window is a face. The goal of face detection is to find the corresponding positions of all the faces in the image. The output of the algorithm is the coordinate of the outer rectangle of the face in the image and may also include posture information such as tilt angle. The face detection algorithm should first have a large number of samples, then extract face features in the positive samples for learning, and

then put into the model for training and finally through the verification results.

3.2. Data Preparation. Before face detection, we need a lot of data to preset rules, telling the machine that images with certain features are human faces, whereas those with other features are nonhuman faces. The diversity of positive samples can make the algorithm have correct answers in different scenarios, whereas negative samples can make the algorithm more accurate to exclude other nonface answers that are very close to human faces. The diversity of datasets can make the algorithm be used in different scenarios and ensure the robustness of the algorithm. The widely used ones are shown in Table 1.

3.3. Feature Extraction. The early template matching algorithm is based on the geometric features of the face to determine whether it is a face; with the improvement of the algorithm, more and more facial feature methods are proposed. The Haar features proposed by Papageorgiou et al. [34] are trained as face features by traversing images with different Haar rectangular frames. Haar features have good modeling ability in uneven lighting scenarios. Zhu et al. [35] proposed the histogram of oriented gradient (HOG), which is resistant to light changes by calculating and counting the histogram of gradient direction in local areas of the image to constitute features. Local binary pattern (LBP) proposed by T. Ojala et al. describes the local texture features of images through operators; it has significant advantages such as rotation invariance and gray invariance. In addition, features such as scale-invariant feature transformation and integral channel feature are used for face detection.

3.4. Model Training. Model training is an important part of face detection, is through the algorithm to input the face features of the training model, and then can directly call the model for face detection. At present, the algorithm commonly used for face detection is the support vector machine algorithm, and its main principle is the sample vector through function transformation mapping to the high-dimensional space and then in the high-dimensional space mapping to find the maximum interval of the interface. Freund proposed the AdaBoost algorithm, which is a classical iterative optimization algorithm. The core idea of its application in target detection is to take the target feature as a weak classifier, combine several weak classifiers into a strong classifier according to certain rules, and finally connect the strong classifier in series to carry out target detection and classification. Then, the classical convolutional neural network was proposed. The convolutional neural network could learn the features of the detected target independently and classify the detected target in the output layer after passing through the convolutional layer and pooling layer.

TABLE 1: Face recognition dataset.

Data set	Basic components
LFW [30]	The dataset includes the name of the image, the location information of the boundary box, and the location information of the key feature points of the face in the image
FDDB [31]	It is a sample picture of a face rotated and occluded by different lighting and resolution. The 2,845 images included 5,171 annotated faces.
AFLW [32]	It is a large-scale face database including multipose and multiview
COCO [33]	COCO dataset is a large, rich object detection, segmentation, and subtitle dataset, mainly taken from complex everyday scenes.

The hybrid face detection algorithm relies on the steadiness of the student's position and calls the fine and rough face detection algorithms according to different face conditions. At the beginning of the class, the students' face location information was collected for the first time through the rough detection algorithm, and using this information, students' seat, that is, the static position of students, is drawn. In the following detection process, we use the face rough detection algorithm for the first face detection, the detection of the face position information, and the student static position area for comparison. If there is no face detected in the student static position area, the second layer of face fine detection algorithm is called for face detection in the student static position area. If a student static position area uses the fine detection algorithm to detect no face many times, it is considered that the position of the student disappeared and the student static position coordinates are deleted. After that, students' basic state was judged by gesture recognition, and the algorithm structure is shown in Figure 1.

3.5. Classroom Behavior Status Assessment.

The facial expression and head posture are combined to analyze the classroom behavior state of the students. The students' expressions obtained by using the expression recognition model of the convolutional neural network are divided into positive emotions and negative emotions. The head posture estimation algorithm was used to divide the recognized head posture of the students into nine directions, and the attention of the students was judged by the difference between the head posture of the target student and the head posture of the surrounding students. In the process of classroom teaching, students need to read books, take notes, and answer questions constantly. Therefore, head posture cannot completely represent students' learning status and can only assist other algorithms to make more detailed judgments.

Head pose estimation (HPE) usually refers to the identification of head position and direction parameters in a spatial coordinate system. The direction parameter refers to the degree of rotation in the three coordinate axes of the spatial coordinate system. The three direction parameters are yaw, pitch, and roll. Head pose estimation is to calculate the head direction parameters by comparing the face feature points in the digital image with the corresponding feature points in the general 3D model. The face feature points in the digital image are obtained by using the Dlib68 feature point

detector on the basis of face positioning, and the flow is given next:

3.5.1. Data Acquisition and Processing.

The classroom teaching videos of natural environment are obtained through the camera (model), and the images are extracted by OpenCv. The width and height of images are W and H. Then, camera calibration is carried out, where the internal parameter matrix is to transform 3D camera coordinates to 2D homogeneous image coordinates. OpenCv camera calibration function is used to calibrate the camera. The internal parameters of the camera are as follows:

$$A = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}, \qquad (1)$$

where f_x, f_y is the focal length of the camera. Generally, the focal length of the camera selects the width and height of the image, which c_x, c_y represents the offset of the camera's optical axis in the image coordinate system. Generally, the center point of the image is selected.

3.5.2. Feature Detection and Head Pose Estimation.

The mixed face detection was used to detect the face, and then, the public Dlib68 feature point detector was used to obtain the coordinate information of six feature points of the face: the outer corner of the left eye and the outer corner of the right eye, the tip of the nose, the left lip angle, the right lip angle, and the tip of the chin. The 2D/3D mapping was obtained by solving the PnP (perspective-n-point) problem, and the rotation and translation vectors of the head pose were output. Then, the student's head pose value (X, Y, Z) is obtained by converting the flip vector into the Euler angle.

3.6. Attention Judgment.

People's visual attention refers to the object or gaze direction of people's eyes, and students' attention can be judged by the difference in their eyes. First, the nose tip coordinates obtained by the feature point detector were used as the starting point, and the nose tip coordinates in the 3D coordinates of the face were used as the ending point to draw the students' line of sight. Although the students' line of sight is different, but all the students are
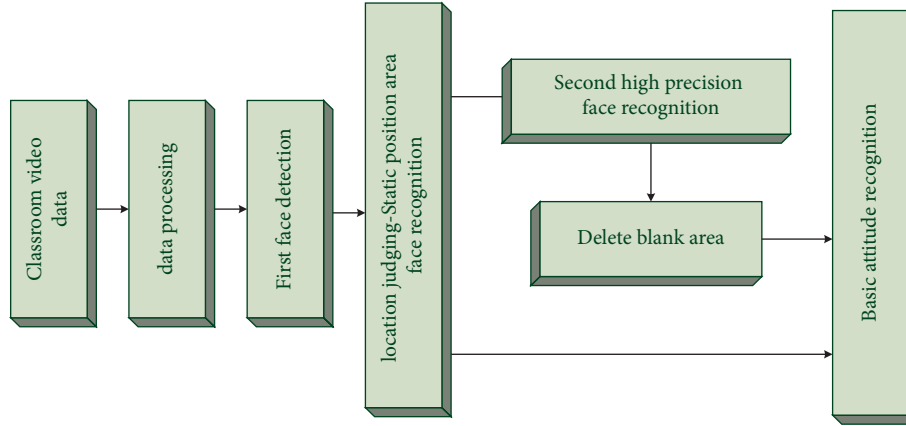
FIGURE 1: Face recognition algorithm structure diagram.

facing the blackboard, and only a very few students' line of sight is toward other positions in the classroom, the students' special behavior can be obviously judged according to the students' line of sight, and the students are in a state of inattention at this time with great probability. The students' head posture was divided into nine directions, and the attention of the students was judged by comparing the line of sight between the target student and the surrounding students.

3.6.1. Partitions. Get the three-dimensional coordinate of student's frontal head posture [X, Y, Z].

3.6.2. Euclidean Distance. The face coordinate of the current student is taken as the starting point, and the face coordinate of other students in the face dataset is taken as the ending point:

$$D_i = \sqrt{(x_0 - x_i)^2 + (y_0 - y_i)^2}. \tag{2}$$

3.6.3. Screening and Judgment. Since all the students in the class are facing the blackboard, the eyes of the students on both sides are not the same, so the eyes of the students on the left side have no reference significance to the right side, and the eyes on the right side are the same. Set the aisle distance as A, delete the faces of the students whose D value is greater than A in the face dataset, sort the faces of the remaining students according to the size of D_i, and get A set whose distance from the target student is from near to far.

To judge the state of a single student, this study selected a group of 9 people and judged the attention of the target student by the difference between the facial orientation of the 8 students and the target student. Eight students are selected from the set of faces in ascending order obtained in Step 3 for comparison. If the value of the target student's face orientation is the same as that of the target student's face direction, then the target student's attention value is increased by one to compare one by one and save the last attention value. It is considered that the attention value of students is greater than 5 and the attention rate of students in the class is obtained by calculating the value of all students and comparing it with the number of students.

3.7. Facial Expression Recognition. Research by psychologist Mehrabiai shows that emotional messages consist of 7 percent of words, 38 percent of voices, and 55 percent of facial expressions. In this study, the convolutional neural network is used to train the facial expression recognition model and the recognized facial expressions are divided into positive emotions and negative emotions according to psychology. The expression recognition model based on the convolutional neural network can get rid of the traditional algorithm to extract the display features of each expression. By combining the extraction of each expression feature with the fuzzy classification of the network, the expression recognition model can improve the performance and generalization ability.

The network is mainly optimized based on the ResNet network structure, because the ResNet network structure can well extract the features of the image, and the optimization of the network structure can make the network better extract the features of the facial region. The overall structure of the model can be divided into three parts: feature extraction, relationship modeling, and expression classification. The optimized ResNet is used as the backbone network to extract features. The weight of the extracted feature is calculated by a layer of self-attention mechanism, and the weight obtained is multiplied by the feature matrix to obtain the final feature matrix. The eigenmatrix is then flattened and projected onto specific dimensions as input to the transformer. Then a transformer encoder is used to model the relationship between face regions. The network eventually calculates the expression of the input image through a simple Softmax function. The facial expression recognition neural network established in this study is shown in Figure 2.
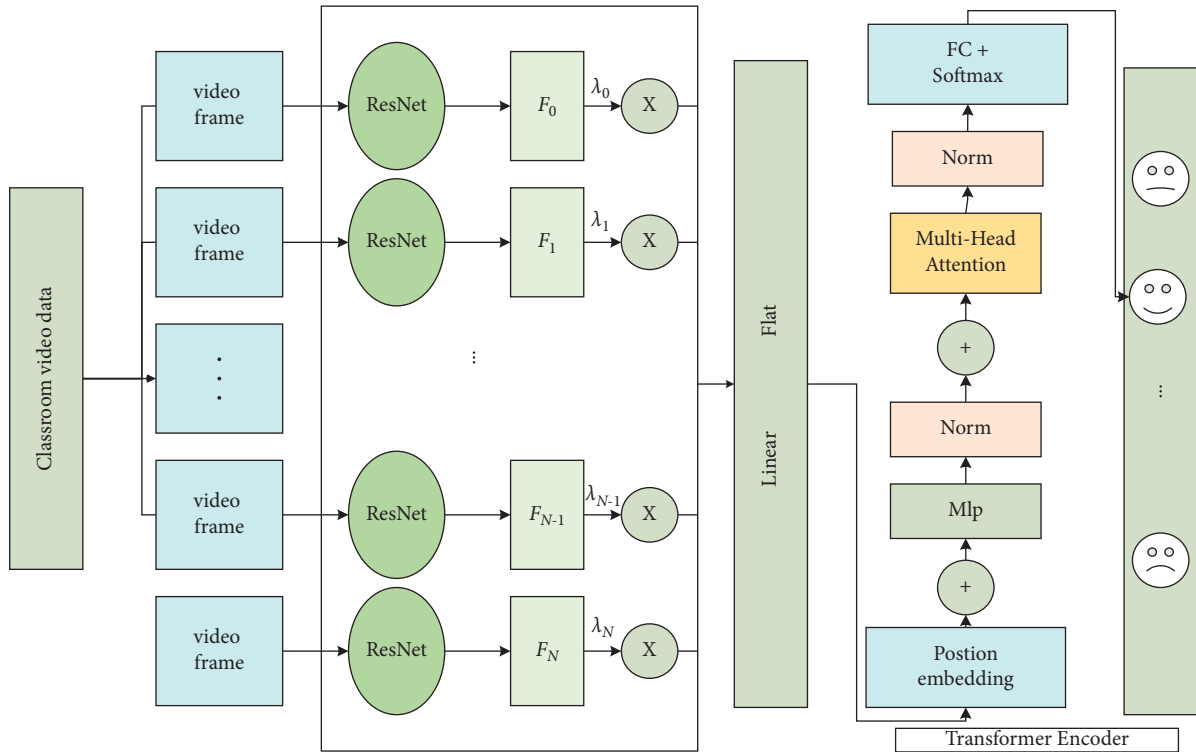
FIGURE 2: Facial expression recognition neural network.

(1) Image cutting: The MTCNN [21] model is used for face location. According to the output results of the MTCNN model, the face is cropped. In order to model the relationship between nonocclusion face regions, the cropped face region needs to be segmented into uniformly segmented images with a size of 20×20 pixels. These segmented images are sent to the feature extraction network for feature extraction. In order to avoid the loss of boundary information in the segmentation process, the image is divided into overlapping image blocks in this study. Each image block has repeated pixels with the surrounding blocks, increasing the correlation between image blocks. Then, these image blocks are sent into the feature extraction network.

(2) Feature extraction: ResNet introduces identity mapping into the network, which can solve the problem of network model degradation and gradient disappearance with the increase in network depth [18], thus improving network performance. ResNet works as follows: suppose the input is x and a certain network layer is set as H. The original network is learning output $h(x)$. After the identity mapping is introduced, the original input x is transmitted to the output through shortcut connections. At this point, the network only needs to learn the residual $f(x) = h(x) - x$ of input and output, and the problems of previous models can be solved through residual learning. Feature extraction is the small-size image after processing. If the maximum pooling operation is still used, some global features may be lost, so

SoftPool [22] is adopted in this study to replace maximum pooling. Compared with other pooling operations, SoftPool can retain both global and local information during pooling. Recognition works better. The calculation method of SoftPool is shown in the following formula:

$$\widetilde{a} = \sum_{i \in R} \frac{e^{a_i} * a_i}{\sum_{j \in R} e^{a_j}}. \tag{3}$$

(3) In order to better study the facial expressions of students, this study does not directly use Softmax for classification, but also needs to send the extracted features into the transformer for further feature extraction. Therefore, the full connection layer of the last layer is modified, and the original full connection layer is changed into two full connection layers, 512 and 100 dimensions. Finally, the obtained 100-dimensional feature vectors are sent into the transformer as tokens for training, which can retain more nonlinear features and effectively reduce the occurrence of overfitting phenomenon. Through experiments, feature extraction using SpResNet can effectively improve the accuracy of recognition.

(4) Vision Transformer is a model proposed by Google in 2017. Originally used in natural language processing tasks, a transformer relies on the attention mechanism and can make the network pay attention to certain words selectively. Later, Carion [13] introduced a transformer into the field of computer vision and proposed an end-to-end target detection

model, DETR. By combining CNN and transformer, the predicted results are finally output. Google proposed a new Vision Transformer(ViT) [19], which migrates the transformer originally used for NLP task into visual classification task out of the box, using transformer instead of CNN. Finally, excellent results have been achieved in large-scale datasets.

Unlike a traditional transformer, which receives serialized tokens as input, Vision Transformer's input is a 3D image, Therefore, the original 3D image data $x \in R^{H \times W \times C}$ need to be divided into image blocks and then the picture is expanded into a one-dimensional vector $x_p \in R^{N \times (p^2 c)}$, where (H, W) is the resolution of the original image, C is the number of channels of the image, and $(p^2 c)$ is the size of each image sequence. Finally, these vectors are flattened to the model size and the output $x_p e$ is embedded, This is followed by the addition of an additional classification header to the sequence, which is a learnable embedding vector through which classification is ultimately performed. Since each image block has a certain position in the un-cropped image, location coding needs to be added to the sequence to retain location information. The calculation method is shown in the following formula:

$$z_0 = \left[x_0, x_p^1 e, x_p^2 e, \ldots, x_p^n e\right] + e_{\text{pos}}, \qquad (4)$$

where e_{pos} is position embedding and z_0 is the initial input to the transformer. The transformer consists of multiheaded self-attention mechanisms and MLP blocks, each of which is followed by a LayerNorm (LN) layer. The calculation method for the transformer is shown as follows:

$$
\begin{aligned}
z'_\ell &= \text{MSA}\left(\text{LN}\left(z_{\ell-1}\right)\right) + z_{\ell-1}, \\
z_l &= \text{MLP}\left(\text{LN}\left(z'_\ell\right)\right) + z'_\ell, \\
y &= \text{LN}\left(z_L^0\right),
\end{aligned}
\qquad (5)
$$

where L is the number of image blocks and $\ell = (1, \ldots, L)$. Multiheaded self-attention (MSA) is the core mechanism of transformer. It is composed of single-headed attention mechanism, namely, self-attention (SA). The calculation method of single-headed attention is shown in the following formula:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V, \qquad (6)$$

where Q is a query, K is a key, and V is a value. They are linear variations of input tokens. d_k is the dimension of K. By calculating the dot product, the similarity between different tokens can be calculated, thus obtaining the global long-term concern, which is conducive to modeling the relationship between clipping regions. Multiheaded attention mechanism is a series of K single-headed attention output, and the calculation method is as follows:

$$\text{MSA}(z) = [SA_1(z), \ldots, SA_k(z)]U_{\text{msa}}, \qquad (7)$$

where $U_{\text{msa}} \in R^{kd_k \times D}$ and $SA_1(z)$ is the single-headed attention mechanism.

4. Experimental Analyses

4.1. Experimental Environment. In this study, experiments were carried out under the configuration of artificial intelligence computer, and the model proposed was trained and tested using NVIDIA Tesla V100 GPU. In the experiment, the MTCNN model was used to conduct face alignment and face region cropping for all images in the dataset, and then, they were adjusted to 224×224 size, mainly using ResNet18 as the baseline experiment. Adam was used to optimize the model, and the initial learning rate was set at 0.001. The latest face rough detection algorithm and fine detection algorithm, as well as the hybrid face detection algorithm of this article, were selected to detect 90 seconds of real classroom teaching video from the detection speed and accuracy analysis. The number, accuracy, and speed of faces detected by each algorithm in the detection process of 15 seconds, 30 seconds, 45 seconds, 60 seconds, 75 seconds, and 90 seconds are selected for analysis and comparison. A represents the model of this study, B represents the TRADITIONAL FINE detection model of CNN, and C represents the traditional AdaBoost coarse detection model.

4.2. Analysis of Experimental Results. From Figure 3, as time goes by, the total number of faces detected by all models does not change linearly, and the maximum number of faces detected by all models decreases first and then increases slowly. Because we randomly selected the time period, it proved the objectivity of the experiment from the side, and this phenomenon was consistent with real life. At the beginning of the class, the students sat upright and, as the class went on, some students began to pay attention and accompanied by various small movements, which led to the fluctuation of face detection. The total number of faces detected by the hybrid model proposed in this study is higher than that detected by the coarse and fine models in all time periods, which proves that our model is still reliable with the change of students' posture and the passage of time. From the results, the accuracy of the coarse detection model is the lowest, and the precision detection is in the middle. The model in this article absorbs the advantages of the two models and gets better results.

From Figure 4, the overall change of the accuracy detection result is inconsistent with the experimental result of total number of faces, among which coarse detection and fine detection have some similarity. The change of the algorithm in this study is small as time goes by, and it is proved from the side that the detection time length randomly selected will not affect the mixed model too much. From the perspective of the first 60 seconds, the detection accuracy of the model in this study decreases slightly with time, whereas that of the other two models decrease linearly. The only difference is that coarse detection drops to the lowest point and begins to rise linearly after 45 seconds, whereas fine detection moves backward by about 15 seconds compared with coarse detection. From the point of view of the minimum detection accuracy, the minimum detection accuracy of the model in this study is close to the maximum accuracy
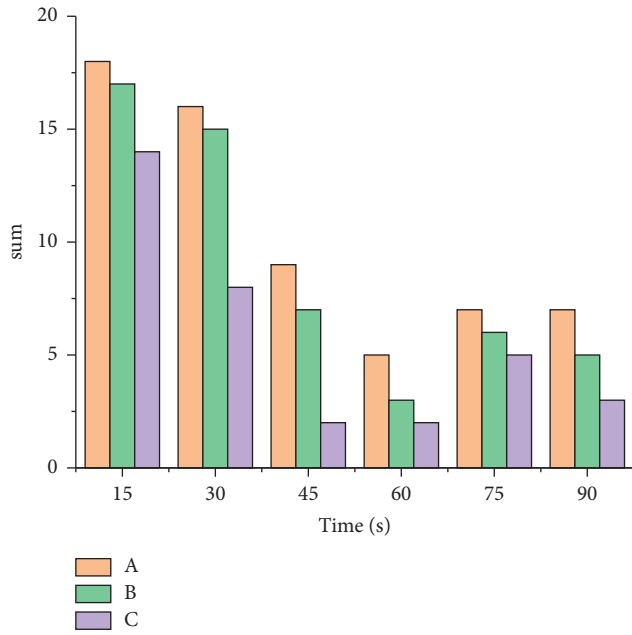
FIGURE 3: Face detection number result comparison.
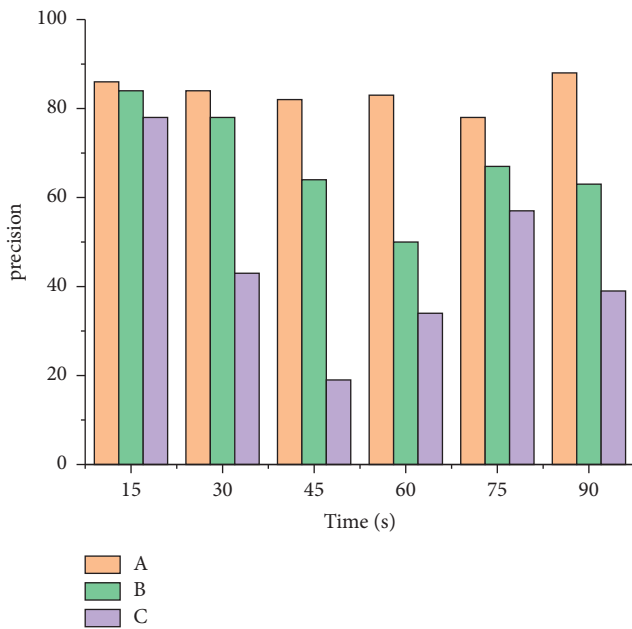
FIGURE 4: Comparison of face detection accuracy results.

FIGURE 5: Experimental results of the students' attention rate in the class.

of rough detection, which is about 78%, and 25% higher than the minimum accuracy of fine detection. By comparing the above experimental results from multiple angles, the model in this study not only achieves a good total number of faces but also has relatively stable detection accuracy.

As can be seen from Figure 5, in the first 1/3 time points randomly selected, the students' attention rate is the highest corresponding to assassin students' concentration. After that, the attention rate of students began to decline and began to fluctuate slightly after a certain time. It is worth noting that both algorithms get the lowest attention rate at

45 seconds; that is, when the class time is halfway through, students' attention is the lowest. In actual classroom teaching, it is not recommended that the teacher explain the most important content during this period, and the main content should be concentrated in the first half of the class. As the class size generally does not exceed 50, the number of individual objects is small for the budget algorithm, so the highest detection rate of the two algorithms is similar. However, from the average result, the model in this study is about 10% higher than the fine detection model. More importantly, the method of judging the students' attention rate by the students' position will reduce the accuracy due to the different directions of students' eyes. However, students' vision can be adapted to more scenes, such as large conference rooms and studio rooms, based on the difference between target students' and surrounding students' vision, without extra manual marking of students' locations. The hybrid model integrates the advantages of various algorithms and can be used in various situations.

Facial expression is the most direct reflection of students' psychological emotions. In order to better test the effectiveness of the model, we trained the model on a large number of public facial expression datasets and put 3000 test pictures from FER20L3 into the trained facial expression recognition model for testing. The model got seven expressions, but anger and fear were rarely seen in the classroom. According to the psychological dichotomies (positive and negative aspects), this paper divides seven kinds of expressions into two categories. The latitude of negative emotion represents the emotional experience of the individual showing negative or angry emotion, while the dimension of positive emotion reflects the individual showing positive emotion. So we classify happiness, surprise, and neutrality as positive emotions, anger, disgust, fear, and sadness as negative emotions.
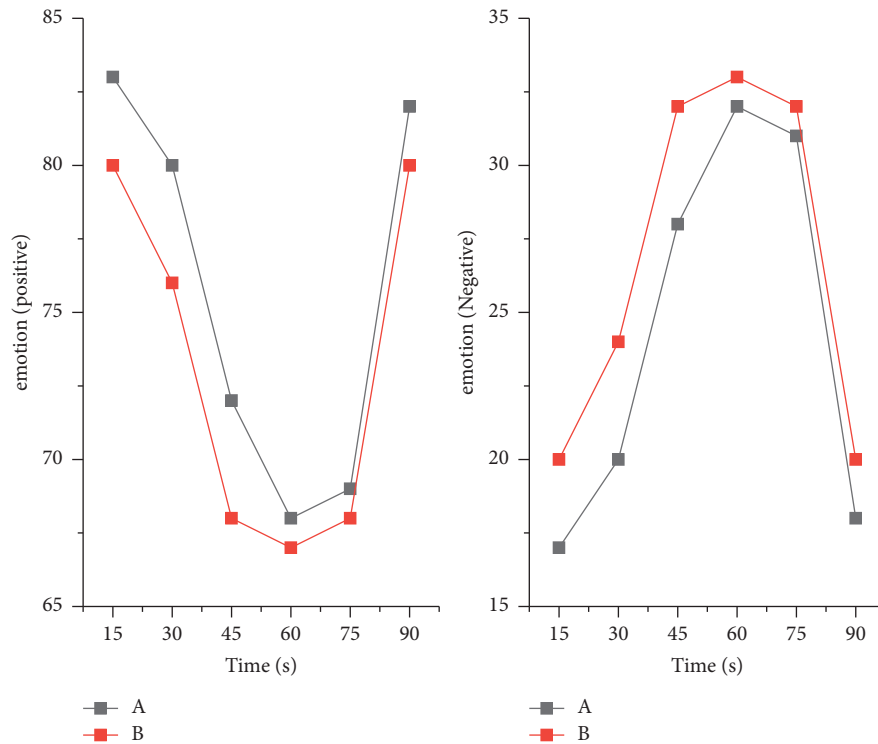
FIGURE 6: Student facial expression recognition results.

The experimental results are shown in Figure 6. By comparing the experimental results of algorithm and manual marking, it can be concluded that the accuracy rate of facial expression recognition model for students in natural class is 90%. More importantly, students' positive emotions continued to decline in the first 2/3 of the period from the beginning of class, and their negative emotions reached the highest level in the period from 1/2 to 5/6. The experimental results and analysis are in line with the actual situation, so school administrators should timely adjust the distribution of the classroom content according to these results, so as to improve the learning efficiency of students in a limited time.

5. Conclusions

In education, the classroom has always been the most important occasion for students and teachers to study and communicate, and the behavior state of students in the classroom has also attracted academic attention and research. This study collects students' classroom behavior state based on video images and analyzes the evaluation methods and shortcomings of classroom teaching at present. The basic methods of classroom behavior state are reviewed, face detection algorithm and expression recognition algorithm are introduced, and the implementation details are described. Based on the traditional algorithm, the hybrid face detection algorithm is improved, and a face expression recognition model based on visual t is established for students, and the feasibility and accuracy of the system are verified in the public dataset. It proves that the intelligent analysis of students' classroom videos will help teachers and other school administrators to make teaching scientific and improve teaching quality. Along with the continuous development of intelligent devices such as sensors, the physiological signal and other modal information will be added to the more video intelligent analysis of college students, through a variety of modal signal complement each other, will all aspects of the analysis of students' comprehensive emotional state, to help teachers and school administrators scientific teaching, further improve the quality of teaching.

Data Availability

The dataset used in this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declared that they have no conflicts of interest regarding this work.

References

[1] C. Zhou, "Teacher: organizer and manager of student learning -- discussion on the role orientation of teachers in the new era," *Education*, vol. 000, no. 6, p. 1, 2018.

[2] W. Qiao, "Evaluation and analysis of teachers' online guidance behavior based on ahp and neural network," *Journal of Tianjin TV University*, vol. 3, no. 4, pp. 23–27, 2007.

[3] W. Gao, *Construction of Evaluation Standard of Classroom Teaching Behavior Oriented to Promoting Students' Active Learning -- Research Based on Delphi Survey Method*, Educational Research and Experiment, 2013.

[4] X. Xu, "Research on illegal problems and countermeasures of teacher's," *Classroom Teaching*, vol. 4, pp. 7–9, 2021.

[5] J. Jiao, "Camera and face recognition technology in classroom," *China Information Technology Education*, vol. 19, 2019.

[6] H. Xu and W. Liu, "Design and practice exploration of student classroom supervision system based on "internet +" thinking," *Communication and Copyright*, no. 11, pp. 145–147, 2017.

[7] S. Tang, "Research on application of improved neural algorithm in classroom face recognition situation analysis," *Information & Systems Engineering*, vol. 311, no. 11, pp. 124-125, 2019.

[8] N. A. N. Zhou, *Research on Indoor/outdoor Scene Recognition Based on Wearable Device*, Shanghai Jiaotong University, Shanghai, China, 2017.

[9] G. Qian and W. Yu, "Application status of intelligent wearable devices in senile patients with dementia," *Chinese Journal of Nursing*, vol. 25, no. 14, p. 4, 2018.

[10] Z. Wang, *Research and Development of Student Class Status System Based on Video Analysis*, Xinjiang University, Ürümqi, China, 2019.

[11] L. Jia, Z. Zhang, and X. Zhao, "Classroom student status analysis based on ARTIFICIAL Intelligence video processing," *Modern Educational Technology*, vol. 29, no. 12, p. 7, 2019.

[12] A. Sinha, R. Gavas, D. Chatterjee, R. Das, and A. Sinharay, "Dynamic assessment of learner's mental state for an improved learning experience," in *Proceedings of the Frontiers in Education Conference*, El Paso, TX, USA, October 2015.

[13] K. Fujisawa and K. Aihara, *Estimation of Interest from Physical Actions Captured by Familiar User Device*, Springer, London, UK, 2011.

[14] N. Nourbakhsh, Y. Wang, F. Chen, and R. A. Calvo, "Using galvanic skin response for cognitive load measurement in arithmetic and reading tasks," in *Proceedings of the Australian Computer-human Interaction Conference*, pp. 420–423, Melbourne, Australia, November 2012.

[15] Z. Zhan, "Emotional and cognitive Recognition model of distance Learners based on intelligent Agent–Coupling between Eye movement tracking and facial expression recognition technology," *Modern Distance Education Research*, vol. 6, no. 5, 2013.

[16] Z. Zhu, S. Ober, and R. Jafari, "Modeling and detecting student attention and interest level using wearable computers," in *Proceedings of the IEEE International Conference on Wearable & Implantable Body Sensor Networks*, IEEE, Eindhoven, Netherlands, May 2017.

[17] L. Shen, "Characteristics of wearable devices and NXP solutions," *Electronics World*, vol. 28, no. 5, p. 1, 2021.

[18] Y. Cheng, *Research on Classroom Teaching Behavior Analysis Method Based on Video*, Central China Normal University, Wuhan, China, 2015.

[19] G. R. Bradski and A. Kaehler, *Learning OpenCV - Computer Vision with the OpenCV Library: Software that sees*, DBLP, Germany, 2008.

[20] L. Hou, Y. Wang, and S. Zhang, "The application of face detection technology in teaching evaluation," *Electronic World*, no. 24, pp. 37-38, 2016.

[21] M. Rahman, M. Debnath, S. Sharmin, L. Alam, S. Arefin, and M. Hoque, "Designing an empirical framework to measure the level of interest of Human," in *Proceedings of the International Conference on Electrical Information & Communication Technology*, IEEE, Khulna, Bangladesh, December 2016.

[22] A. Mehrabian, "Communication without words," *Journal of University of East London*, vol. 2, no. 4, pp. 53–55, 1968.

[23] F. Mantang, Q. Ma, and R. Wang, "Research on intelligent network teaching system based on facial expression recognition," *Computer Technology and Development*, vol. 17, no. 6, pp. 193–196, 2011.

[24] M. Cheng, M. Lin, and Z. Wang, "Research on intelligent teaching system based on expression recognition and sight tracking," *Distance Education in China*, vol. 7, no. 5, p. 6, 2013.

[25] B. Sun, Y. Liu, and J. Chen, "Emotional analysis based on facial expression in intelligent learning environment," *Modern Distance Education Research*, vol. 5, no. 2, pp. 96–103, 2015.

[26] B. Jiang, W. Li, and Z. Li, "Automatic recognition of learning confusion based on facial expression," *Open Education Research*, vol. 11, no. 4, pp. 101–108, 2018.

[27] J. Whitehill, Z. Serpell, Y.-C. Lin, A. Foster, and J. R. Movellan, "The faces of engagement: automatic recognition of student engagementfrom facial expressions," *IEEE Transactions on Affective Computing*, vol. 5, no. 1, pp. 86–98, 2014.

[28] L. Han, L. I. Yang, and Z. Zhou, "Research on modern distance education," vol. 2, no. 4, pp. 97–103, 2017, in Chinese.

[29] L. Chen, Z. Luo, and R. Xu, "Intelligent analysis tn of students' learning interest in classroom teaching environment," *Electronic Education Research*, vol. 39, no. 8, pp. 93–98, 2018.

[30] G. B. Huang, M. Mattar, T. Berg, and E. Learned-Miller, *Labeled Faces in the Wild: A Database forStudying Face Recognition in Unconstrained Environments*, USA, 2008.

[31] V. Jain and E. Learned-Miller, *A Benchmark for Face Detection in Unconstrained Settings*, USA, 2010.

[32] M. Kostinger, P. Wohlhart, P. M. Roth, and H. Bischof, "Annotated facial landmarks in the wild: a large-scale, real-world database for facial landmark localization," in *Proceedings of the IEEE International Conference on Computer Vision Workshops*, IEEE, Barcelona, Spain, November 2012.

[33] T. Y. Lin, M. Maire, and S. Belongie, "Microsoft COCO: common objects in context," in *Proceedings of the European Conference on Computer Vision*, Springer International Publishing, Zurich, Switzerland, September 2014.

[34] C. P. Papageorgiou, M. Oren, and T. Poggio, "A general framework for object detection," in *Proceedings of the Sixth International Conference on Computer Vision (IEEE Cat. No.98CH36271)*, Bombay, India, January 1998.

[35] Q. Zhu, M. C. Yeh, K. T. Cheng, and S. Avidan, "Fast human detection using a cascade of histograms of oriented gradients," in *Proceedings of the IEEE Computer Society Conference on Computer Vision & Pattern Recognition*, IEEE, New York, NY, USA, June 2006.

Research Article

Building Virtual Scene Construction and Environmental Impact Analysis Based on Image Processing

Fan Ying [1,2] **and Zhou Bo** [1]

[1]*Post Doctoral Research Station of Environmental Science and Engineering (Human Settlements) of Sichuan University, Chengdu, China*
[2]*Department of Environmental Art, Academy of Fine Arts, Sichuan Conservatory of Music, Chengdu, China*

Correspondence should be addressed to Fan Ying; 20904033@bit.edu.cn

With the rapid development of computer technology, the virtual scene construction technology of image processing has gradually become a hotspot in computer research. The application of virtual scene construction technology in the construction industry is also expanding, with the intelligentization of architectural design and construction. With the rapid development of the industry, virtual technology can better provide users with diversified services and experiences. The main content of the article is as follows: (1) The article introduces image construction technology and analyzes the three-dimensional presentation of virtual technology in construction design and the diversified application of virtual technology in construction design. (2) The article introduces two virtual scene construction techniques: one is to combine the real scene obtained by taking pictures with the virtual scene by image technology processing to construct a natural virtual scene; the other is the use of geometric construction by architects. The virtual building model is drawn by the model method, and the virtual building model is thus obtained. (3) The article chooses two representative buildings as cases. The virtual platform monitors and records the movement trajectory of volunteers in the virtual environment in real time. The experimental results show that the number and shape characteristics of the escalators in the building all have a little impact on volunteers' awareness. (4) The article sets up three control groups of normal mode, fixed skin, and variable skin. It analyzes the environmental impact of buildings from three different aspects of wind environment, thermal environment, and light environment and proposes related control measures.

1. Introduction

With the advent of the information age, the construction industry has also entered an intelligent field; the application of virtual scene construction technology in the construction industry is also expanding. The use of virtual reality technology when designing buildings can not only visually display the art of the building. The effect is also a great advantage in saving energy and labor. Huaqin et al. [1] discussed a remote virtual platform based on VRML-JAVA language, which is constructed and scene-oriented with realistic and interactive information. Jin et al. [2] used virtual reality 3D holographic projection technology to realize the 3D image imaging scene of the building suspended in the air. The article combines image technology with the real estate industry, analyzes the building's modeling and imaging steps, improves the expressiveness of the building, and also brings great value and development prospects to the construction industry. Hou and Li [3] proposed a collision detection strategy based on enclosure and enclosure hierarchy to meet the performance requirements of virtual hoisting collision detection. R. Joseph and Perera [4] introduced the mechanism of applying internal and external design structures after converting 2D drawings into 3D building information models (BIMs). In the early stage of construction, it is very useful to create a 3D model according to the customer's requirements. The article introduces the method of 3D model creation. Pruss et al. [5] introduced a computer-implemented method for generating building block instructions for building block models. Guo and Chen

[6] discussed the role and significance of AR in airport construction projects. In the early stage of building construction, designers have to apply many advanced technologies, such as image processing technology and virtual model construction to generate virtual building models with computers. Zhao et al. [7] produced a hierarchical model to describe the 3D dynamic geographical process. Describing the geographical factors that change with time during the construction process and simulating visual scenes with computers are of great significance to the management and decision-making of the building. In [8], large eddy simulation (LES) was performed on the airflow around various types of block arrays to estimate the pedestrian wind environment. The experiment selected five different arrays for simulation. The experimental results show that the area ratio is an important indicator for estimating the wind environment. We can use the area ratio to simulate the height change of the building. Lang [9] proposed a model for estimating solar radiation based on existing weather records and determining typical meteorological months and interpolation procedures for missing data. Kajikawa et al. [10] discussed the benefits, limitations, and future directions of the evaluation framework [11]. Do [12] proposed an analysis method for building environmental service plans that are verified throughout the year. The service plan of building environment service is a very important part of the whole design. The designed plan must meet the requirements of the indoor environment. The article proposes an analysis method of building service that can meet most of the requirements. Yun and Won [13] uses machine learning algorithms to establish a built environment model, which can be used to evaluate the comfort of occupants. We have proposed an energy control measure, which can effectively record the energy consumption of the room and record the comfort of the residents, so as to provide a better accommodation environment for the owners. Ivanova [14] focused on using anisotropic sky-viewing angle factors to estimate the background component of incident diffuse solar radiation on building facades under the orthogonal obstructed sky. Because the Earth's available resources are very limited, we have to design buildings that can use solar energy. Xueping [15] tested the 4 most commonly used heating methods and discussed indoor thermal environmental conditions for different heating methods. The temperature in the north is relatively low in winter, and many citizens choose certain heating measures in winter. The experiment compares common heating methods and provides a reference for citizens to choose heating methods.

2. Virtual Reality Technology

2.1. Three-Dimensional Presentation of Virtual Reality Technology in Architectural Design.
Architectural design work is very complicated. In the design work, designers will inevitably encounter some problems, such as the size of the building being not clear enough and the structural drawings of the building having some problems. But, the reason is that building virtual scene construction technology can solve these problems very well. The design drawings are presented in three-dimensional graphic construction technology, which largely avoids architectural errors, facilitates designers to find problems, and improves work efficiency.

When faced with a single two-dimensional or three-dimensional overprofessional drawing model, the construction personnel cannot make a correct understanding of the construction operation of the building, causing deviations between the design and the actual building, making the construction impossible and causing serious losses from many parties. Participants can immerse in the virtual space and intuitively understand the designer's description of the main structure, spatial layout, and details of the building or building group, so as to grasp the detailed requirements in the work and avoid repeated construction problems that cause irreparable losses to subsequent construction. The three-dimensional sensory presentation allows participants to better complete the construction requirements and tasks during construction and finally realizes the application, economy, solidity, and beauty of the building.

Architectural designers use the immersion, interactivity, and conception of virtual reality technology to fully integrate with the architectural concept [16]. First, the building participants have an immersive feeling during use. Second, the building participants can interact with most of the objects in the virtual environment by using sensing devices. This makes the complex and abstract architectural design become concrete with the use of virtual reality technology. The greater advantage is that designers can communicate with other architectural participants remotely, thereby reducing time costs, breaking geographical restrictions, and making architectural design more wide-area.

2.2. Application of Virtual Reality Technology in Architectural Design.
The technical personnel input the engineering information and design parameters of the architectural design into the three-dimensional modeling software through the computer, and the three-dimensional modeling software builds the three-dimensional architectural model and supporting database based on the input data. Technicians issue operating commands to the model drive system through the computer, adjust the building model information, display the progress of the building project design plan in real time, follow up the project, and adjust the model parameters in time. Building participants are immersed in the virtual simulation space created by virtual reality technology and feel the reality of the lighting changes, architectural pattern, modeling structure, and surrounding environment in the virtual simulation space [17]. Compared with the method of drawing building models from ordinary drawings, the use of virtual technology not only is environmentally friendly and cost-saving but also allows more people to participate in the process of architectural design and better understand architectural details and specific construction conditions. It provides convenience for designers to modify the architectural design plan in the later stage of construction, and at the same time, it is convenient to save the architectural design, which is conducive to the comparison and update of future designs [18]. The virtual building model is the basis

for constructing the virtual real scene. The size ratio between the measurement unit of the model and the real scene and the parameter values of the entire model all affect the final presentation effect of the virtual real scene.

2.3. Analysis of Building Environmental Impact.

Environmental performance is reflected in the park design as the external and internal environments of the park. The external environment of the park includes ecological environment, traffic environment, outdoor functional space created by the building, and environmental impacts around the park. According to different time domains, environmental quality assessment can be divided into environmental quality review assessment, environmental quality status assessment, and environmental quality forecast assessment. Environmental quality assessment of construction projects mainly refers to the assessment of environmental quality status. Environmental postassessment refers to the systematic investigation and evaluation of the actual impact degree of the environment after the development and construction activities are carried out, so as to test the implementation degree and effectiveness of reducing the impact and test the true credibility of the environmental assessment results. To evaluate the effectiveness of the implementation of the proposed environmental protection countermeasures, and to analyze and discuss the influencing factors that are not fully understood in the environmental assessment, check the implementation degree and implementation effect of reducing the impact and test the true credibility of the environmental assessment results. After evaluating the effectiveness of the proposed environmental protection measures, the fully understood environmental impact problems are classified and studied, so as to improve the technical methods and levels of environmental assessment and implement remedial measures, thus achieving the effect of reducing the impact.

3. Theory and Technology Based on Image Construction

3.1. Image Deformation Technology.

The image deformation technology can realize the excessive viewpoint, while interpolating the shape and texture of the image [19], as shown in Figure 1.

A pair of line segments can define the mapping relationship between the source image and the target image [20]; the specific algorithm is as follows: calculate PQ based on line segment u, v, where

$$
\begin{aligned}
u &= \frac{(X - p) \cdot (Q - P)}{\|Q - P\|^2}, \\
v &= \frac{(X - P) \times \text{perpendicular}(Q - P)}{\|Q - P\|}.
\end{aligned}
\tag{1}
$$

Then, find u, v based on $P'Q'$ and line segment X'; when using multiple pairs of line segments, it is necessary to perform a weighted average on the multiple offsets obtained, as shown in Figure 2.

The specific algorithm is as follows, for each pixel on the target image:

$$
\begin{aligned}
\text{Dis\_Sum} &= (0, 0), \\
\text{Weight\_Sum} &= 0.
\end{aligned}
\tag{2}
$$

Calculate the offset of the line segment as

$$
D = X_i' - X_i'.
\tag{3}
$$

Calculate the shortest distance from X to $P_i Q_i$ as

$$
\begin{aligned}
\text{weight} &= \left(\frac{\text{length}^p}{(a + \text{dist})} \right)^b, \\
\text{Dis\_Sum}+ &= D_i * \text{weight}, \\
\text{Weight\_Sum}+ &= \text{weight}.
\end{aligned}
\tag{4}
$$

Finally,

$$
X' = X + \frac{\text{dis\_Sum}}{\text{Weight\_Sum}}.
\tag{5}
$$

3.2. Projection Reconstruction Algorithm.

The Euclidean motion theory is used to explain the geometric relationship between the two-dimensional image and the three-dimensional image captured using the same camera. All points $p \in E^3$ in the space can be represented by a certain three-dimensional coordinate point in R^3:

$$
X = [X_1, X_2, X_3]^T = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} \in R^3.
\tag{6}
$$

The coordinates of p and q in space are X and Y, respectively, by the vector formed by p and q:

$$
v = Y - X \in R^3.
\tag{7}
$$

The inner product of two vectors $v, u \in R^3$ in space is

$$
\langle u, v \rangle = u^T v = u_1 v_1 + u_2 v_2 + u_3 v_3 \quad \forall u, v \in R^3.
\tag{8}
$$

The vector product is

$$
u \times v = \begin{bmatrix} u_2 v_3 - u_3 v_2 \\ u_3 v_1 - u_1 v_3 \\ u_1 v_2 - u_2 v_1 \end{bmatrix} \in R^3.
\tag{9}
$$

We define a matrix $\hat{u} \in R^{3\times3}$:

$$
\hat{u} = \begin{pmatrix} 0 & -u_3 & u_2 \\ u_3 & 0 & -u_1 \\ -u_2 & u_1 & 0 \end{pmatrix} \in R^{3\times3}.
\tag{10}
$$

The vector product can be expressed as

$$
u \times u = \hat{u} v.
\tag{11}
$$
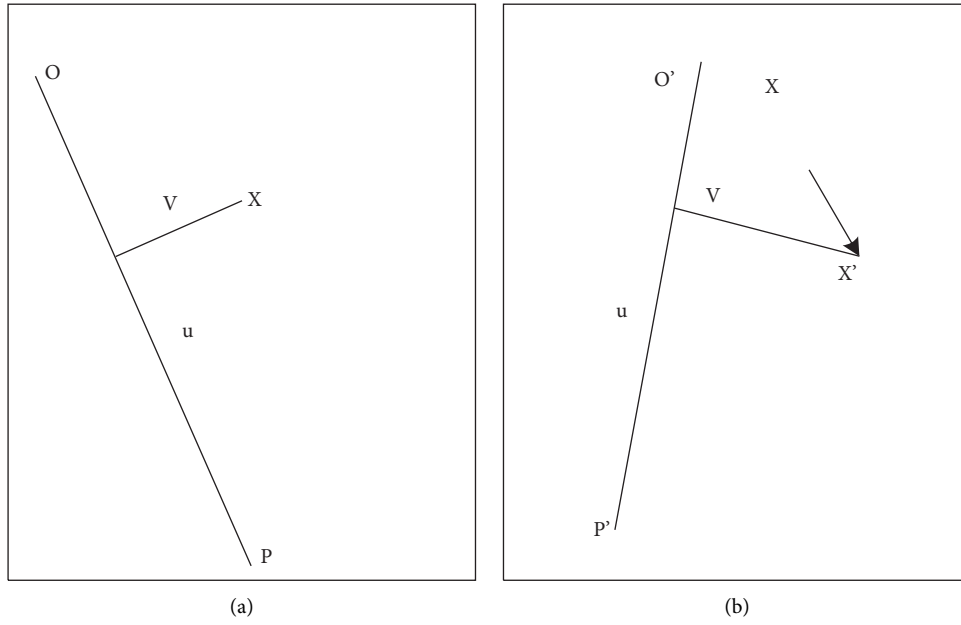
FIGURE 1: Deformation of a pair of corresponding line segments. (a) Target image. (b) Source image.
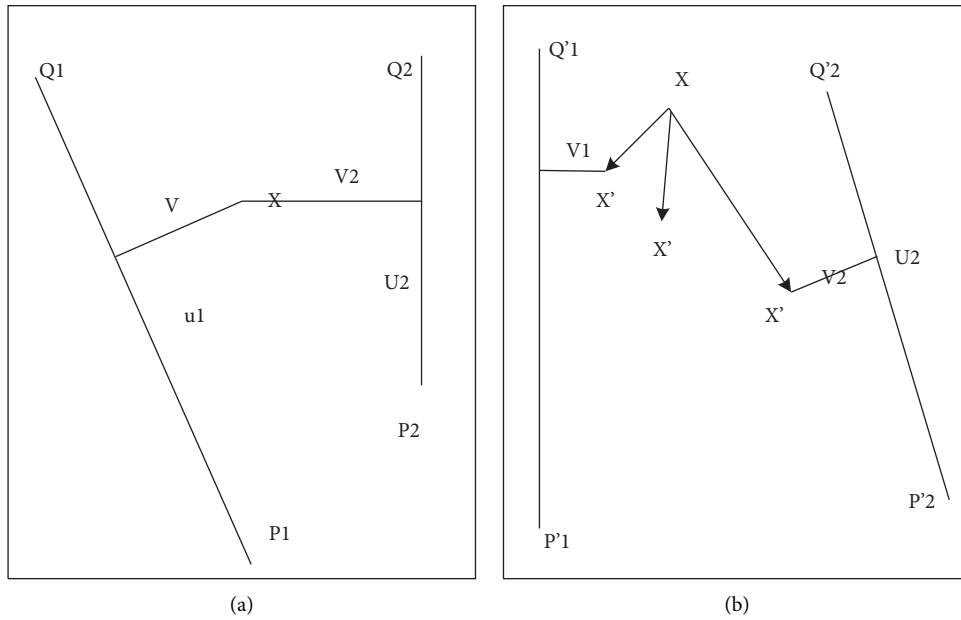
FIGURE 2: Deformation of multiple pairs of corresponding line segments. (a) Target image. (b) Source image.

It is also necessary to introduce rigid body motion theory, as shown in Figure 3.

The formula can be expressed as

$$X_w = R_{wc}X_c + T_{wc}. \tag{12}$$

Figure 3 coordinates of point P in the camera coordinate system are (X, Y, Z), and the two-dimensional coordinates on the plane are $X = \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \in R^3$ [21]:

$$x = \begin{bmatrix} x \\ y \end{bmatrix} = \frac{f}{Z}\begin{bmatrix} X \\ Y \end{bmatrix}. \tag{13}$$

The homogeneous coordinate form is as follows:

$$Z\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{pmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{pmatrix}\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}\begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}, \tag{14}$$
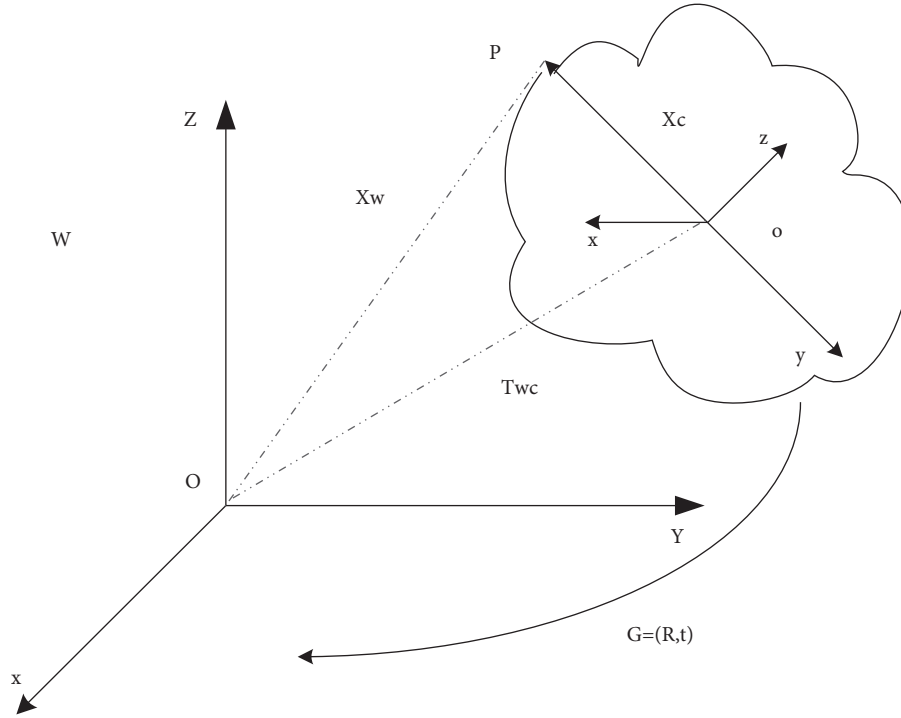
FIGURE 3: Rigid body motion between the world coordinate system and its own coordinate system.

where

$$K_f = \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & f \end{bmatrix} \in R^{3\times3},$$

$$\prod_0 = K_f = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \in R^{3\times4}. \tag{15}$$

Let $\lambda \dot{=} Z$, and the formula can be written as

$$\lambda x = K_f \prod_0 X \tag{16}$$

3.3. Uncertain Measurement of Environmental Impact.
The environmental impact of the construction process includes building materials, construction machinery, resources, and labor, which can be abstracted into mathematical formulas [22]:

$$E_{\text{total}} = \sum_{i=0}^{I} E_i. \tag{17}$$

Here, E_{total} is the total emission of environmental pollutants [23]:

$$E_i = E_m + E_e + E_r + E_l. \tag{18}$$

The formula for environmental pollutant emission per unit amount of environmental influencing factors is

$$E = e \times Q. \tag{19}$$

During the construction process, the simulation module provides dynamic data under the influence Q_v [24]:

$$Q_v = f(Q|\prod). \tag{20}$$

According to the environmental pollution impact assessment method, the damage value of the human body, ecology, and natural resources can be calculated:

$$p_k = \sum_k \left(\sum_j \left(E_j \times EF_j \right)_k \times CF_k \right). \tag{21}$$

4. Experimental Simulation

4.1. Virtual Building Scene Construction Technology.
Computer graphics technology is one of the methods of constructing virtual scenes. We usually use the technology of "combination of virtual and real" to construct the virtual environment required by users [25]. The nonexistent scene is the virtual scene. The combination of the real scene and the virtual scene creates the most natural virtual scene, as shown in Figure 4.

The virtual model can also be drawn by the architect using the method of geometric modeling. It must be drawn according to the real size of the building, and then, a virtual model is constructed based on the geometric model and the real map drawn by the architect, as shown in Figure 5.

This section is mainly on the virtual technology and real technology of building analysis, through the combination of

(a)

(b)

(c)

Figure 4: Building scene diagram combining virtual and real scenes. (a) Real scene image. (b) Virtual scene image. (c) Architectural scene combining virtual and real scenes.

virtual and real ways to effectively integrate the two styles, forming a more perfect building simulation map.

4.2. Simulation Experiment.
This study selects the more representative architectural model 1 and architectural model 2 of our country's construction projects as cases. Architecture 1 and architecture 2 are widely distributed in our country and have a large amount of construction; we took photos and measurements of the case on the spot, and then, according to the experiment, it is required to construct a virtual simulation scene. In order to avoid the interference of environmental factors on the experiment, when constructing the virtual scene, the dimensions, position, and other factors must be consistent with the real environment, as shown in Figure 6.

We selected 30 volunteers to participate in the experiment. The average age of volunteers is 20–40 years, and the proportion of males and females is not equal. The virtual experiment platform is used to collect volunteers' pathfinding trajectories in building 1 and building 2, in order to reduce certain volunteers. The experiment error is caused by the proficiency of the scene, so before the formal start of the experiment, each volunteer needs to enter the venue in advance to get familiar with the interactive process of the virtual scene. The specific data are shown in Table 1.

The experiment monitors and records the movement trajectory of volunteers in the virtual environment in real

(a)

(b)

(c)

FIGURE 5: Image after combining virtual and real scenes. (a) Virtual geometric model. (b) Real image. (c) Building mockup.

time through the virtual platform. Combined with the experimental results, we can find that the volunteers' route selection in building 2 is more consistent, because the number of escalators in the visible range of the volunteers is different. When there are a large number of escalators within the visible range, volunteers' choices are more scattered, such as building 1; each volunteer has a different strategy for seeking roads. The local characteristics of building 1 and building 2 are shown in Table 2.

Volunteers conducted research on the overall form cognition and local feature cognition of the scene layout. Volunteers' cognitive biases were mainly concentrated on the turning angle of the horizontal channel. The specific data are shown in Table 3.

From Table 3, we can see that there are significant differences in the volunteers' awareness of the atrium content, shape, and size characteristics of the two buildings.

Building 2 is superior to Building 1 in the number of escalators in the atrium. In terms of information, the information of Building 1 is obviously better than that of CapitaLand, in terms of atrium shape recognition, building 1 and building 2 do not show a significant difference, but building 1 has a higher score than building 2.

4.3. Environmental Performance Analysis

4.3.1. Environmental Analysis. In order to verify the impact of building energy consumption on the indoor environment, we have adopted a series of measures such as ventilation and lighting to simulate and analyze the building. The simulated building is in an area where the sunshine is strong and the average temperature is medium and high. Details are set out in Table 4, and the rate of indoor personnel in the room is shown in Figure 7.

(a)

(b)

Figure 6: Continued.
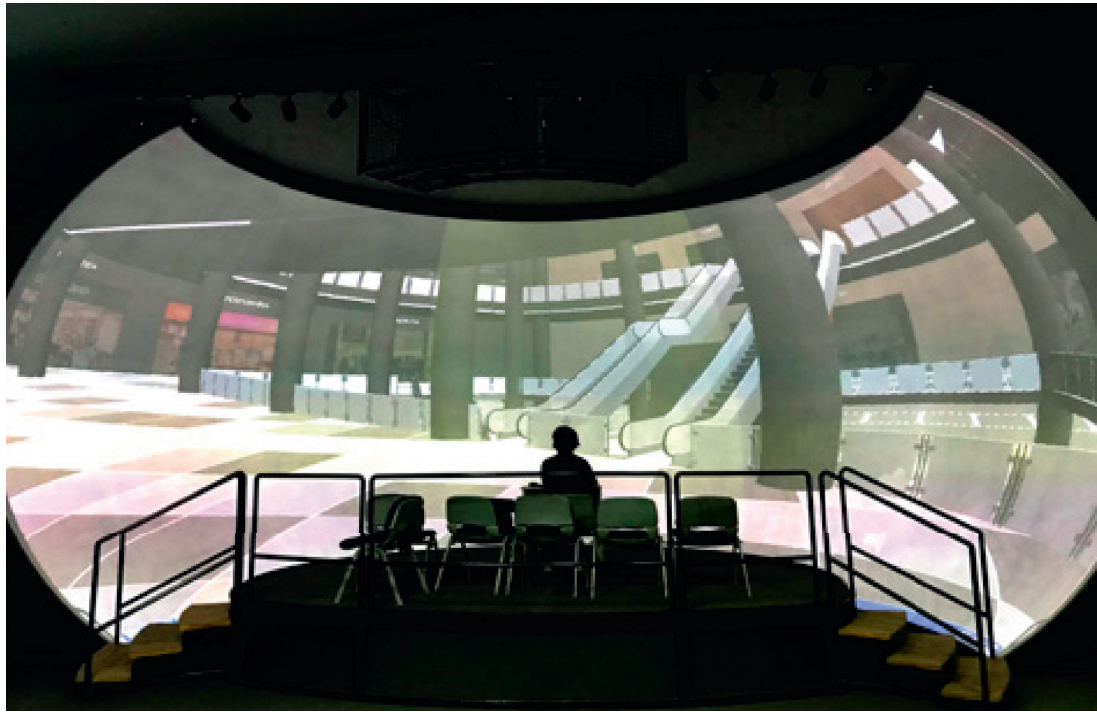
(c)

FIGURE 6: Building virtual map. (a) Building 1. (b) Building 2. (c) Virtual simulation scene.

TABLE 1: Virtual data types and their variable contents.

Category	Variable content	Data sources
Personal attributes	Number, age, gender, virtual environment experience	Experimental questionnaire
Path trajectory	Path selection, path length	Coordinate data, process record
Spatial cognition	Overall layout: plane configuration (geometric form, path turning angle) Local features: atrium space (number of escalators, atrium shape, size) Initial entrance (location label)	Cognitive map

TABLE 2: Local characteristics of building 1 and building 2.

Case study	Building 1	Building 2
Ground floor plan		
Standard floor plan		
Overall arrangement	Linear layout There are angle changes in the horizontal channel, with two obvious turning points	Circular layout The horizontal paths intersect at an angle, forming a unequal triangle
Atrium features	Atrium (including escalators): 5 : 3 squares, 2 round/ellipse; squares are arranged along the main passage; round/ellipse is at the turning point	4 atriums (including escalators): all oval 3 are located at the vertices of the triangle and 1 is located on the longer side

TABLE 3: Analysis of the test results of the atrium space awareness under different layout situations.

Atrium features	Building 1 mean	Standard deviation	Building 2 mean	Standard deviation	F value	Significance (P)
Number of escalators	0.52	0.51	0.85	0.37	6.90	0.011
Traits	3.36	0.99	3.15	1.19	0.45	0.506
Dimensional characteristics	3.00	1.29	1.81	0.90	14.79	0.000
Cognitive total score	6.88	1.79	5.81	1.44	5.58	0.022

TABLE 4: Building simulation situation.

Name	Design building model
Building orientation	Zhengnan
Number of indoor personnel	3 people
Air conditioning system COP	Cooling 3.2, heating 1.7
Indoor temperature (°C)	Summer 26, winter 18
Air changes (mechanical ventilation ach)	1
Indoor illumination (lux)	The illuminance of the bedroom, living room, and kitchen is 300, and the rest is 150
Lighting power density (W/m^2)	5
Glare index (DGI)	The bedroom, living room, and kitchen are less than 27, and the rest of the rooms are less than 28
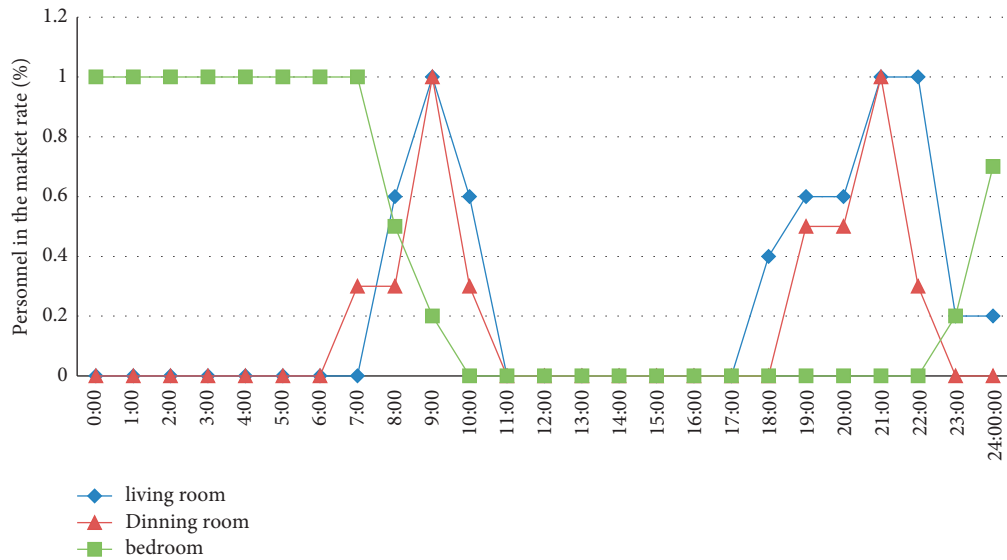
FIGURE 7: Statistic chart of staff in-room rate.

4.3.2. Experimental Verification. In order to test the impact of building energy consumption and the built environment, we set up three control groups: normal mode, fixed skin, and variable skin. The normal mode is only the building itself, and the variable skin is the additional skin in addition to the building itself. The building skin will have a certain impact on the indoor thermal environment and building energy consumption. We compared these modes, analyzed the three environmental performances of wind environment, thermal environment, and light environment, and recorded their design elements and related control measures; the specific data are shown in Table 5.

According to Figure 8, we can conclude that compared with the normal mode, the variable skin mode has good ventilation and relatively less energy consumption. It saves energy to a certain extent and can also highlight its own architectural characteristics.

Thermal environment influencing factors and control measures are shown in Table 6.

To verify the thermal environmental impact and total energy consumption data of the building under three different modes, we adopt three modes of lighting, cooling, and heating to analyze the thermal environmental impact, as shown in Table 7.

From Figure 9, we can see that the total energy consumption of buildings in the cooling mode is the largest. In terms of lighting, fixed skin and variable skin will increase certain energy consumption. The normal mode is the most energy efficient. By comprehensive comparison, variable skin has the highest energy-saving efficiency.

TABLE 5: Wind environment influencing factors and control measures.

Environmental performance	Design phase	Design elements	Control measures
Wind environment	Park planning	Landscape	The dominant wind direction in winter sets trees and shrubs facing the wind
		Site design	The opening direction of the open space faces the dominant summer wind direction
		Longitudinal design	The setting of the sunken square is conducive to controlling the winter wind field
		Layout of the building	Staggered arrangement of dominant winds is seen in winter Ventilation channels are reserved for the dominant wind direction in summer
	Park planning	Building space layout	The staggered height of the building is conducive to ventilation, and the height difference should be controlled not to be too large to avoid sudden wind
		Architectural form	The sharp corners of the building's facade should be reduced on the windward side in winter and should be replaced with a streamlined design
	Architectural design	Main entrance design	Roundabout design is to be done, to avoid the dominant wind direction in winter
		Landscape design	The actual situation of architectural design is combined to optimize the wind environment of landscape design
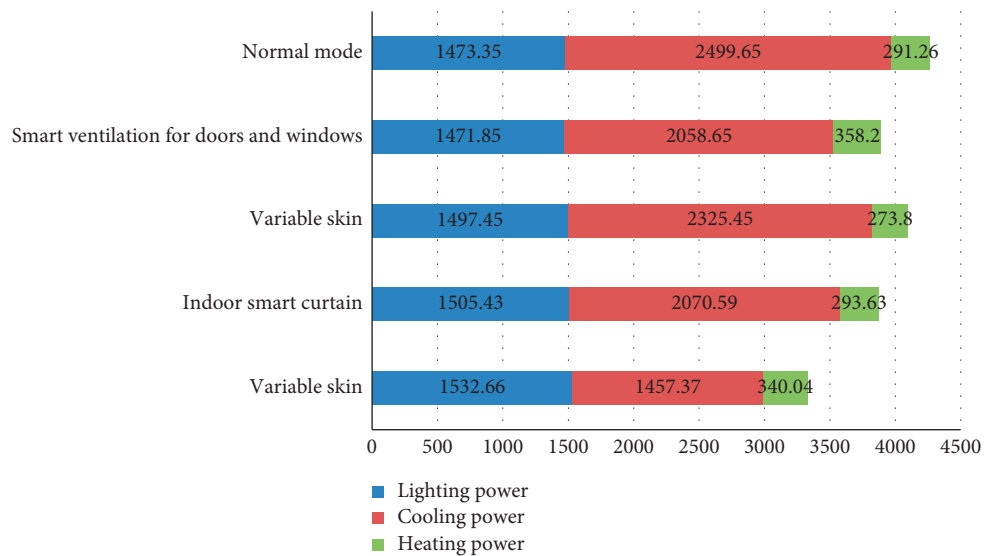
FIGURE 8: Data statistics of different modes.

Factors affecting the light environment and control measures are shown in Table 8.

We introduced natural light and recorded the sunshine in winter and summer at different locations of the building under three different conditions, as shown in Figure 10.

We can see from Figure 10 that compared with the normal mode, the variable skin can effectively reduce the glare time in the room to a certain extent, can reduce the energy consumption caused by architectural lighting, and increase the comfort of natural lighting.

TABLE 6: Thermal environment influencing factors and control measures.

Environmental performance	Influencing factors	Design phase	Design elements	Controlling factor
Thermal environment	Sun radiation	Park planning	Landscape	Increase greening rate Design landscape water bodies Increase shade measures
		Traffic design		Adopt light color series for road laying materials
		Architectural design	Facade design	Control the ratio of windows to walls Take sun protection measures Optimize sunroof settings
	Urban heat island	Park planning	Layout of the building	Optimize building layout In addition to meeting the sunshine requirements, the building spacing should also consider the sunshine requirements of the green space
		Site design	Landscape garden	Buffer isolation zones such as water body, green space, and vertical greening are set up with the surrounding strong heat island environment
	Personnel activities	Park planning	Traffic design	The human-shaped area is far away from the heat dissipation facilities such as the outdoor unit of the air conditioner
		Architectural design	Electromechanical design	Cooling towers, outdoor units of air conditioners, and other heat dissipation measures take heat isolation measures

TABLE 7: Thermal environment energy consumption data table under different modes.

Architecture	Normal mode	Fixed skin	Variable skin
Lighting (kWh)	1473.35	1831.93	1532.66
Refrigeration (kWh)	2499.65	1793.94	1457.37
Heating (kWh)	291.26	541.91	340.04
Total energy consumption analysis (kWh)	4264.26	4167.77	3330.07
Energy-saving efficiency		2.26%	21.91%
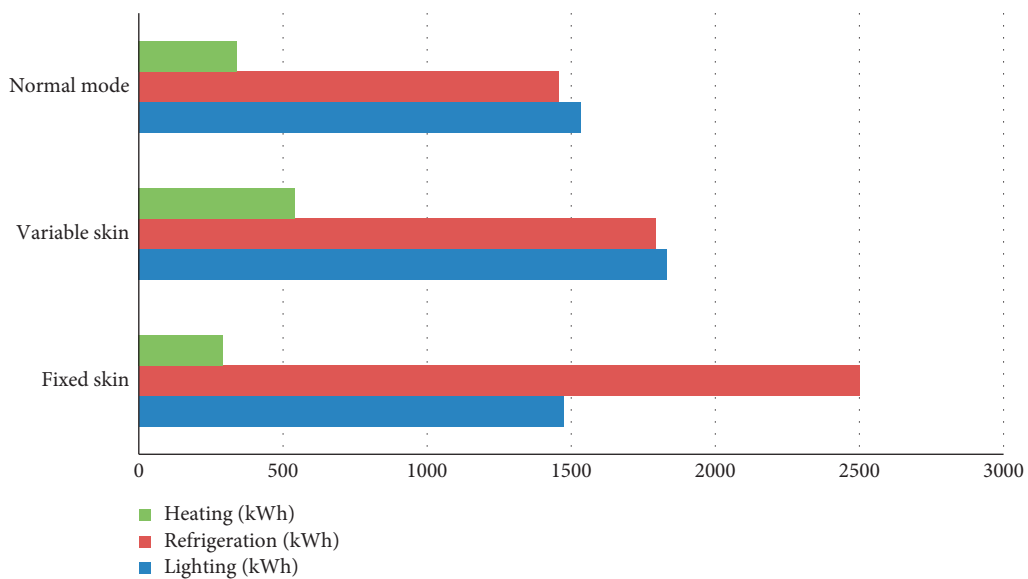
FIGURE 9: Thermal environment impact statistics under different modes.

TABLE 8: Factors affecting the light environment and control measures.

Environmental performance	Design phase	Design elements	Control measures
Light environment	Park planning	Construction layout	Sunshine analysis in the park considers the impact of green space Park sunshine analysis appropriately increases the sunshine duration of outdoor functional spaces
		Longitudinal design	The setting of the square is conducive to the lighting of the underground space
	Architectural design	Graphic design	Control the depth of the room and optimize the architectural pattern
		Facade design	Optimize the ratio of windows and walls
		Curtain wall design	The curtain wall adopts low-reflectivity materials to control glare
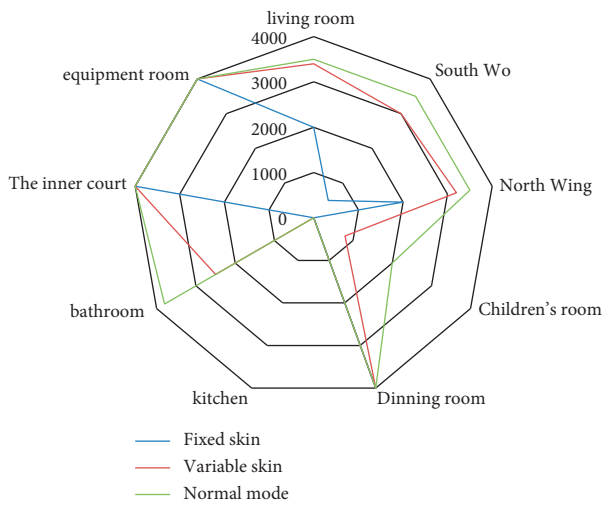
FIGURE 10: The glare exceeds the standard in each room.

5. Conclusion

Starting from virtual reality technology, this paper analyzes the influence of buildings in the environment from the perspective of architectural model. The main conclusions are as follows: (1) This paper introduces two kinds of virtual scene construction technologies: one is to combine the real scene obtained by taking pictures with the virtual scene by image technology, so as to construct a natural virtual scene; the other is that architects draw virtual building models by using geometric modeling methods, thus obtaining virtual building models. (2) This paper chooses two representative buildings as cases and monitors and records the moving track of volunteers in the virtual environment in real time through the virtual platform. Through the experimental results, it can be concluded that the number and shape characteristics of escalators in buildings have a little influence on the cognitive degree of volunteers. (3) This paper sets up three control groups, common mode, fixed skin, and variable skin, analyzes the environmental impact of buildings from three different aspects of wind environment, thermal environment, and light environment, and puts forward relevant control measures.

Data Availability

The experimental data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest regarding this work.

References

[1] Z. Huaqin, D. Shenghui, and Z. Liangqin, "Design and development of long-distance virtual teaching system based on VRML and JAVA languages," in *Proceedings of the ISECS International Colloquium on Computing, Communication, Control, and Management*, no. 6, pp. 22–36, IEEE, Sanya, China, Aug. 2009.

[2] Y.-q. Jin, W.-j. Lu, and W. juan, "A research on the construction and realization of the indoor building model suitable for holographic projection," *DEStech Transactions on Computer Science and Engineering*, pp. 110–121, 2018.

[3] X. T. Hou and C. H. Li, "Research on collision detection for virtual hoisting construction scene," *Journal of System Simulation*, no. 20, pp. 25–62, 2014.

[4] R. Joseph and U. Perera, "Build-IT -- an interactive web application for 3D construction, interior and exterior design," *IEEE* in *Proceedings of the 2014 5th International Conference on Intelligent Systems, Modelling and Simulation (ISMS)*, no. 23, pp. 114–146, Langkawi, Malaysia, January 2014.

[5] M. Pruss, "Automatic generation OF building instructions for building block models," *Geo-Spatial Information Science*, vol. 6753, no. 110, pp. 110–145, 2007.

[6] Y. Guo and J. Chen, "Augmented reality technology and application in aerodrome building," *Geo-Spatial Information Science*, vol. 6753, Article ID 675324, 2007.

[7] W. Zhao, W. Jiang, M. Zhao, and L. Wei, "3D temporal scene building of airport construction," *Journal of Geomatics*, no. 7, pp. 11–15, 2019.

[8] A. Abd Razak, A. Hagishima, N. Ikegaya, and J. Tanimoto, "Analysis of airflow over building arrays for assessment of urban wind environment," *Building and Environment*, vol. 59, no. 1, pp. 56–65, 2013.

[9] S. Lang, "Research and development of weather data for building energy analyses," *Nuantong Kongtiao/HV and AC*, vol. 32, no. 4, pp. 22–50, 2002.

[10] Y. Kajikawa, T. Inoue, and T. N. Goh, "Analysis of building environment assessment frameworks and their implications for sustainability indicators," *Sustainability Science*, vol. 6, no. 2, pp. 233–246, 2011.

[11] J. Jaakkola, "The office environment model: a conceptual analysis of the sick building syndrome," *Indoor Air*, vol. 8, no. S4, pp. 7–16, 2010.

[12] B. Y. Do, "Building environment design simulation software DeST(6): simulation and analyses of building environmental service scheme," *Hv & Ac*, no. 20, pp. 50–78, 2004.

[13] J. Yun and K. H. . Won, "Building environment analysis based on temperature and humidity for smart energy systems," *Sensors*, no. 10, pp. 22–45, 2012.

[14] S. M. Ivanova, "3D analysis of the incident diffuse irradiance on the building's surfaces in an urban environment," *International Journal of Low Carbon Technologies*, vol. 10, no. 9, pp. 395–397, 2015.

[15] L. I. Xueping, "Test and analysis of indoor thermal environment of urban heating residential building in hanzhong area," *Journal of Landscape Research*, no. 01, pp. 104–107+115, 2017.

[16] X. Meng, R. Du, J. F. Jaja, and A. Varshney, "3D-Kernel foveated rendering for light fields," *IEEE Transactions on Visualization and Computer Graphics*, no. 99, p. 1, 2020.

[17] D. Oktavia, R. Rangkuti, and N. Yusuf, "Elements OF narrative and function OF narrator IN jun CHIU'S crop circles," *LANGUAGE LITERACY Journal of Linguistics Literature and Language Teaching*, vol. 4, no. 2, pp. 368–374, 2020.

[18] Y. Lai, X. Lan, Y. Liu, and N. Zheng, "An efficient depth image-based rendering with depth reliability maps for view synthesis," *Journal of Visual Communication and Image Representation*, vol. 41, pp. 176–184, 2016.

[19] X. Tong and K. Zhou, "Special issue:deep learning in computer graphics," *IEEE Computer Graphics and Applications*, vol. 39, no. 2, p. 25, 2019.

[20] J. Weisman, "Evaluating architectural legibility," *Environment and Behavior*, vol. 13, no. 2, pp. 189–204, 2016.

[21] A. Baskaya, C. Wilson, and Z. Ozcany, "Wayfinding in an unfamiliar E nvironment different spatial settings of two polyclinics," *Environment and Behavior*, vol. 36, no. 6, pp. 839–867, 2004.

[22] Bo. Zhou, "Research on innovation of display design based on virtual reality technology," *Modern Decoration (Theory)*, no. 9, pp. 45–48, 2011.

[23] W. Zhang, "The application of virtual reality technology in future architectural design," *Science Chinese*, no. 5, pp. 114–178, 2014.

[24] M. Liao and L. Yu, "Comparative study on the construction models of foreign eco-industrial parks," *Journal of Suzhou Education College*, no. 2, pp. 169–173, 2016.

[25] J. Wang and Z. Lu, "Research on the application ideas and strategies of passive design and performance tools in green building design," *Huazhong Architecture*, no. 8, pp. 88–95, 2014.

Research Article

Research on Multimodal Music Emotion Recognition Method Based on Image Sequence

Zhao Yu [ID]

Guangxi Arts University, College of Music Education, Department of Keyboard Instruments, Nanning 530022, Guangxi Province, China

Correspondence should be addressed to Zhao Yu; 20100010@gxau.edu.cn

The work of music performance system is to control the light change by identifying the emotional elements of music. Therefore, once the identification error occurs, it will not be able to create a good stage effect. Therefore, a multimodal music emotion recognition method based on image sequence is studied. The emotional characteristics of music are analyzed, including acoustic characteristics, melody characteristics, and audio characteristics, and the feature vector is constructed. The recognition and classification model based on neural network is trained, the weight and threshold of each layer are adjusted, and then the feature vector is input into the trained model to realize the intelligent recognition and classification of multimodal music emotion. The threshold of the starting point range of a specific humming note is given by the center clipping method, which is used to eliminate the low amplitude part of the humming note signal, extract the short-time spectral structure features and envelope features of the pitch, and complete the multimodal music emotion recognition. The results show that the calculated kappa coefficient k is greater than 0.75, which shows that the recognition and classification results are in good agreement with the actual results, and the classification and recognition accuracy is high.

1. Introduction

Music is an art form that takes sound as a means of communication and then produces emotional experience. Music can communicate emotion directly in the form of sound movement. The essence of music is emotion. The specific form of music sound wave vibration is directly related to human emotion. According to this connection, music can be used to describe people's emotional activities in detail. All music activities obey and reflect the fluctuations of people's inner world, whether the creators and performers vent their emotions or the listeners accept the emotional connotation of the music. Nowadays, digital music technology has brought great changes to music, a traditional and classic way of emotional communication. The development of computer science has brought revolutionary progress to the creation, communication, storage, and release of music works. Especially with the continuous enrichment of computer music materials, it has become an urgent scientific research topic to study the emotional information of music works by using

intelligent information analysis and processing methods, so as to make the computer have the ability to recognize and express music emotions like people. Music appeared earlier than language. When human beings did not use language to express their feelings, they had learned to use music [1, 2]. It can be said that music plays an important role in human history, and music has been integrated into all aspects of human life [3]. With the continuous development of science and technology, the creation, storage, and dissemination of music have been greatly changed. Music is an art form that takes sound as a means of communication and then produces emotional experience. Music can directly carry out emotional communication in the form of sound movement [4, 5]. It can be said that the essence of music is emotion. The specific form of music acoustic vibration is directly related to human emotion. According to this connection, music can be used to describe people's emotional activities in detail [6]. All music activities obey and reflect the fluctuations of people's inner world, whether the creators and performers vent their emotions or the listeners accept the emotional

connotation of the music. Nowadays, digital music technology has brought great changes to music, a traditional and classic way of emotional communication. The development of image sequence has brought revolutionary progress to the creation, communication, storage, and release of music works. Generally, image sequence noise is an unpredictable random signal. Noise is very important for image sequence processing. It affects all links of input, acquisition and processing of image processing, and the whole process of output results [7, 8]. In particular, the input of image and the suppression of acquisition noise are very key problems. If the input is accompanied by large noise, it will inevitably affect the whole process and output results. Therefore, a good image sequence processing system, whether analog processing or digital processing by computer, takes reducing the noise of the first level as the main target [9, 10]. In particular, with the continuous enrichment of computer music materials, it has become an important research content to use the image sequence intelligent information analysis and processing method to study the emotional information of music works, so as to make the computer have the ability to recognize and express multimodal music emotions like people.

In this regard, relevant scholars have proposed many studies. Reference [11] proposed the common neural mechanism of emotion processing in music and vocalization and compared the neural mechanisms involved in vocalization and music processing, so as to observe their possible similarities in emotional content coding. Positive and negative emotional sounds (such as laughter and crying) and violin music stimuli extracted by numbers are used as stimuli, which have common melody contour and main pitch/frequency characteristics. Reference [12] proposed that the semantic and episodic memory of music are provided by different neural networks, and the extraction of brain semantic memory and episodic memory is completed by different neural networks. It is basically obtained through language and visual space materials. Two delay identification tasks are constructed, one containing only familiar items and the other only unfamiliar items. For each recognition task, the general extraction target is presented in the previous semantic task. By comparing two perceptual control tasks with another perceptual control task, the situational task and semantic task are compared. Based on the above analysis, a multimodal music emotion recognition method based on image sequence is proposed. The music emotion features including acoustic features, melody features, and audio features are analyzed, and the feature vector is constructed. The recognition and classification model based on neural network is trained, the weight and threshold of each layer are adjusted, and the feature vector is input into the trained model to realize the intelligent recognition and classification of multimodal music emotion. The threshold of the starting point range of a specific humming note is given by the center clipping method, which is used to eliminate the low amplitude part of the humming note signal, extract the short-time spectral structure features and envelope features of the pitch, and complete the multimodal music emotion recognition. The recognition and expression of multimodal music emotion enable users to realize emotional human-computer interaction through music, which enriches the research content of human-computer interaction technology.

2. Multimodal Music Emotion Recognition and Classification Based on Image Sequence

In addition to the necessary music itself, a perfect music performance is a complementary live atmosphere. In music performance, the contrast of the on-site atmosphere is mainly realized by lighting, which is often changed with the emotional factors expressed in the music to assist the music to create a good stage effect. In this context, in order to better control the light, multimodal music emotion recognition is very important [13–15]. Therefore, aiming at multimodal music emotion, a classification and recognition model is constructed to complete the research on intelligent recognition and classification of multimodal music emotion in music performance system.

2.1. Analysis of Emotional Characteristics of Multimodal Music. The realization of multimodal music emotion recognition is based on multimodal music emotion features, so multimodal music emotion feature extraction is the first link of this research [16, 17]. In the previous multimodal music emotion classification, most of them take a music feature as the classification basis. Although they can also complete the classification task, their accuracy cannot be guaranteed. In order to solve the above problems, in this study, a variety of music features are extracted and fused based on image sequences and then classified and recognized based on fusion features. The principle of image sequence is shown in Figure 1.

In order to identify the emotional characteristics in music [18], it is necessary to understand the composition of music. Among them, the music related factors that can obviously show emotional characteristics include acoustic characteristics, melody characteristics, and audio characteristics.

2.1.1. Acoustic Characteristics. Acoustic feature refers to the physical quantity that represents the acoustic characteristics of multimodal music speech. It is also a general term for the acoustic performance of many elements of sound, for example, the energy concentration area, formant frequency, formant intensity, and bandwidth representing the timbre of multimodal music, as well as the duration, fundamental frequency, and average voice power representing the prosodic characteristics of multimodal music speech. For the classification of multimodal music speech, the traditional method is to study the characteristics of pronunciation organs, such as the tongue position of vowels, front and back, and the pronunciation position of consonants. Now, with the progress of science and technology, further fine research can be made according to the acoustic characteristics.
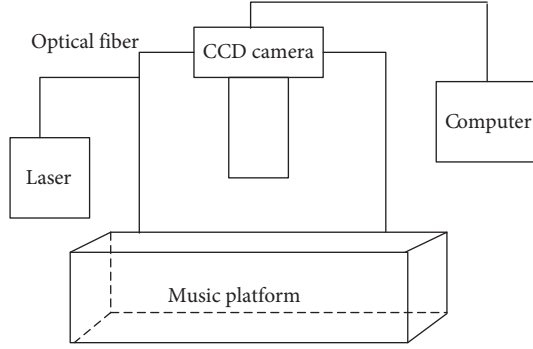
FIGURE 1: Schematic diagram of image sequence.

Acoustic factor is the most basic component of music [19, 20]. Music with different emotions shows different acoustic characteristics, and the basic corresponding relationship is shown in Table 1.

2.1.2. Melody Characteristics. Melody features are also called melody features; that is, the lines composed of high and low tones with different lengths are the soul of music and the melody of music. The tones are organized according to certain laws [21–23]. The extracted features include five aspects.

(1) Balance parameter Y_1: Balance refers to the proportional value of the volume in the left and right channels. The calculation formula is as follows:

$$Y_1 = P_{an}(k), k = 1, 2, \ldots, 16. \tag{1}$$

(2) Volume parameter Y_2: Volume refers to the loudness of the sound that can be heard by the human ear. The calculation formula is as follows:

$$Y_2 = \frac{V_{olume}}{127}. \tag{2}$$

(3) Pitch parameter Y_3: Pitch refers to the vibration frequency of the fundamental frequency of a note. Fast paced music has fast vibration frequency; on the contrary, it has slow vibration frequency. The calculation formula is as follows:

$$Y_3 = \frac{1}{n} \sum_{i=1}^{n} P_{itch}. \tag{3}$$

(4) Average strength parameter Y_4: Strength refers to the strength of the power generated by music. Soothing music has weak strength, while more shocking music has strong strength [24, 25]. The calculation formula is as follows:

$$Y_4 = \sum_{i=1}^{N} \frac{V_{el(k,i)}}{N}, k = 1, 2, \ldots, 16. \tag{4}$$

(5) Note energy parameters Y_5: Note energy refers to the sum of the product of note pitch and length. The calculation formula is as follows:

$$Y_5 = \sum_{i=1}^{n} (P_{ij} \times D_{ij}), j = 1, 2, \ldots, 16. \tag{5}$$

In the formula, $P_{an}(k)$ represents the balance value of left and right channels, and its value range is 0–127; V_{olume} represents the volume of the track, with a range of 0–127; P_{itch} stands for note pitch; n represents the number of notes in the track; $V_{el(k,i)}$ represents the intensity value of the i note in the k track; k indicates track number; N represents the number of notes in the k track; P_{ij} and D_{ij} represent the pitch and length of i notes in the j track channel.

2.1.3. Audio Features. Audio feature is an important condition for recognizing and identifying multimodal music emotion. Different music emotion is expressed through different audio features. Audio is one of the important influencing factors in music, which affects the rhythm of music. The faster the rhythm, the more obvious the audio, and the happier the multimodal music emotion expressed. On the contrary, multimodal music emotion is more dull or depressing [26, 27]. The description of audio features based on image sequences can be carried out from two aspects, real-time domain features and frequency domain features [28].

(1) Time domain characteristics

The time domain characteristics of audio refer to the time domain parameters of each frame calculated from the music signal, mainly including zero crossing rate and amplitude [29–31]. The following is a specific analysis.

(1) Zero crossing rate Z_n: Zero crossing rate refers to the frequency at which the audio signal waveform passes through the zero level. Generally speaking, the zero crossing rate in the high-frequency band of a piece of music will be relatively high; on the contrary, the zero crossing rate will be relatively low. Through this parameter, we can well distinguish between voiced and unvoiced sounds in music. Generally, unvoiced sounds are mostly used in cheerful music, while voiced sounds are often used in slow and deep music. The calculation formula of zero crossing rate is as follows:

$$Z_n = \frac{\sum_{m=1}^{N} \text{sgn}[s_n x(m)] - \text{sgn}[s_n x(m-1)]}{2N}. \tag{6}$$

In the formula, $s_n x(m)$ represents the symbol function of the audio signal $x(m)$; N represents the effective width of the window; n represents the time position of the window.

(2) Range M_n: Amplitude refers to the width expanded by the waveform vibration of audio signal [32–34]. The more passionate the music, the greater the audio amplitude. The more soothing the music, the smoother the audio amplitude. The audio amplitude is described as follows:

TABLE 1: Corresponding relationship between acoustic characteristics and multimodal music emotion.

Acoustic characteristics	Happy	Hate	Anger	Sadness	Fear
Pronunciation	Normal	Normal	Tighten	Vague	Clear
Pitch mean	Very high	Very low	Very high	Slightly lower	Very high
Pitch range	Very wide	Slightly wider	Very wide	Slightly narrow	Very wide
Pitch change	Smooth, curved up	Wide, downward bending	Stress mutation	Bend down	Normal
Tone quality	Breathing sound, singing sound	Mumble, chest sound	Breathing sound	Resonance sound	Sharp voice
Speed of speech	Fast or slow	Very fast	Slightly faster	Slightly slower	Soon
Strength	High	Low	High	Low	Normal

$$M_n = \sum_{m=n-(N-1)}^{n} |x(m)w(n-m)|$$
$$= \sum_{m=n-(N-1)}^{n} |x(m)|w(n-m). \tag{7}$$

In the formula, $w(n-m)$ represents the moving window function.

(3) Frequency domain characteristics: The frequency domain characteristics of audio include two: spectral centroid C_t and spectral flux F_t. The calculation formula is as follows:

$$\text{Spectrum centroid } C_t = \frac{\sum_{n=1}^{N} M_t[n] \times n}{\sum_{n=1}^{N} M_t[n]},$$
$$\text{Spectral flux } F_t = \sum_{n=1}^{N} \left(N_t[n] - N_{t-1}[n]\right)^2. \tag{8}$$

In the formula, $M_t[n]$ represents the amplitude of the short-time spectrum of the t frame at the frequency point n; $N_t[n]$ and $N_t[n-1]$ represent the normalized amplitude of the spectrum of the t frame and the $t-1$ frame at the frequency point n, respectively.

Based on the above three categories and 14 multimodal music emotional features, a feature vector is formed, which is used to describe the emotional factors of a piece of music. It is described as follows:

$$U = \{U_1, U_2, U_3\}. \tag{9}$$

In the formula, U_1 represents acoustic characteristics; U_2 represents melody characteristics; U_3 represents audio characteristics. The audio feature structure is shown in Figure 2.

2.2. Construction of Multimodal Music Emotion Recognition Classification Model.

Based on the emotional features contained in the above music, a classification and recognition model is established to realize multimodal music emotion recognition and classification, and a neural network is used to construct the model [35, 36]. BP neural network is an intelligent algorithm invented by simulating the working principle of human brain neural network. The neural network mainly includes three
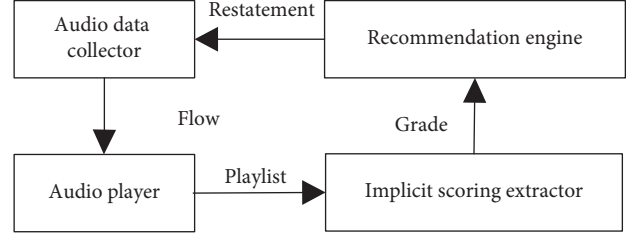
FIGURE 2: Audio feature structure.

layers, and the classification processing is realized through the operation of each layer. The classification and recognition model constructed by this algorithm is shown in Figure 3.

In Figure 3, training is the key in model construction, and the specific process is as follows. First, enter the choice of training samples, and after operation of hidden layer and output layer, you will get results, and then to compare the results with the expected results, when the difference between them is less than the set threshold, the training is completed; otherwise, there will be back propagation, difference from the output to the input, and repetitive process, until you reach the optimal weight and threshold. The purpose of BP neural network training is to adjust and optimize the weights and thresholds connected at every two levels in the model. Therefore, the formula is given as follows.

(1) Adjustment formula of connection weight w_{ij} and threshold θ_j between input layer and hidden layer:

$$\begin{cases} w_{ij}(N+1) = w_{ij}(N) + \beta \cdot \mu_j^k \cdot c_i \\ \theta_j(N+1) = \theta_j(N) + \beta \cdot \mu_j^k \\ i = 1, 2, \cdots, n \\ j = 1, 2, \cdots, p \\ 0 < \beta < 1 \end{cases} \tag{10}$$

In the formula, μ_j^k represents the error value in the hidden layer; c_i represents the input eigenvector; N represents the number of iterations; k represents the number of training samples; n represents the number of neurons in the input layer; p represents the number of neurons in the hidden layer.

(2) Adjustment formula of connection weight v_{jt} and threshold γ_t between hidden layer and output layer:
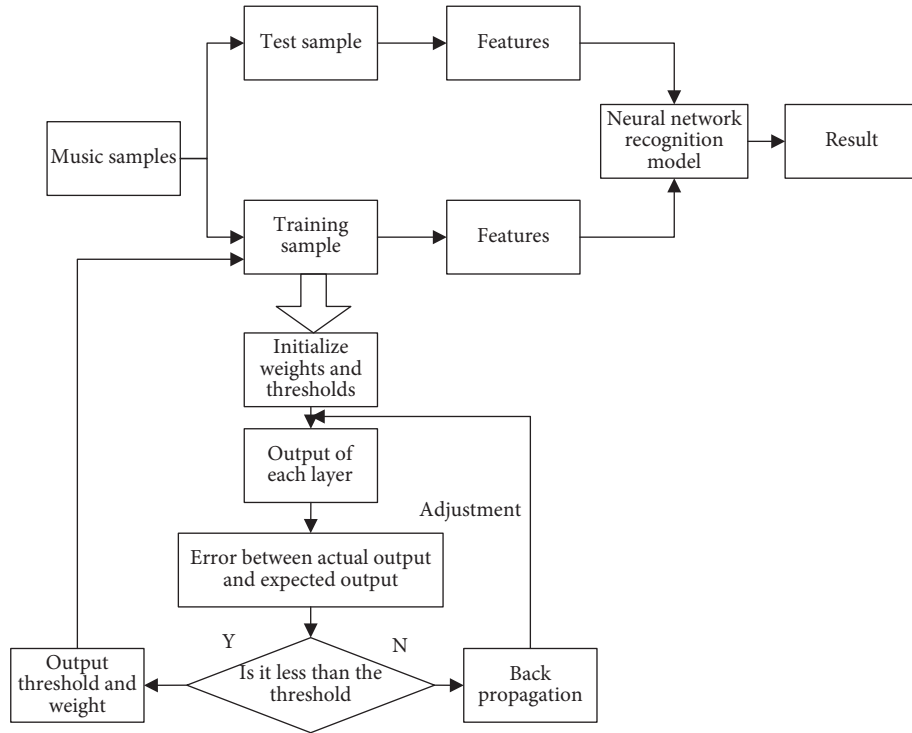
FIGURE 3: Classification and recognition model based on BP neural network.

$$\begin{cases} v_{jt}(N+1) = v_{jt}(N) + \alpha \cdot d_t^k \cdot y_j \\ \gamma_t(N+1) = \gamma_t(N) + \alpha \cdot d_t^k \\ j = 1, 2, \cdots, p \\ t = 1, 2, \cdots, q \\ 0 < \alpha < 1 \end{cases} \quad . \quad (11)$$

In the formula, d_t^k represents the error value between the target eigenvector and the actual output vector; y_j represents the output of the hidden layer.

The trained model based on BP neural network can realize multimodal music emotion classification by inputting test music samples.

2.3. Intelligent Recognition of Note Starting Point Based on Clipping

2.3.1. Calculation of Correlation Function between Note Signals. In the process of intelligent optimization and recognition of the note starting point of feature tone retrieval, the initial note signal is preprocessed based on the image sequence to filter the noise of the high-frequency part. The random note signal is divided into short-term stationary signals based on the image sequence, the similarity between different phonetic waveform signals is calculated, and the cross-correlation function between each note signal is obtained. The design of recognition framework based on image sequence is shown in Figure 4.

Pervasive environment combines network technology and mobile technology and designs a customer-oriented
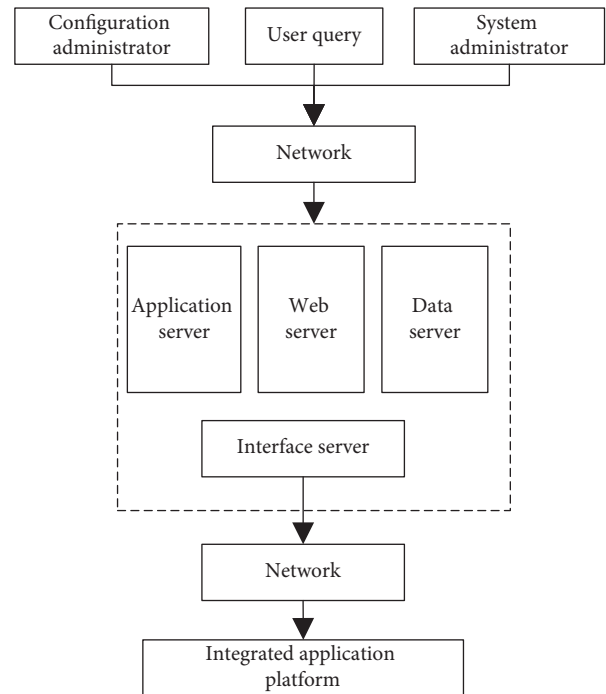
FIGURE 4: Frame diagram of recognition based on image sequence.

adaptive recommendation structure. Pervasive environment is composed of network devices, including computers, mobile phones, and various network connected appliances, and network services including computing, management, and control. In this environment, the network can collect query, configuration, and management information from

users and administrators, transfer these pieces of information to each server port, and then apply them to the comprehensive platform through the network to provide the basis for the design of the recommendation system.

The specific steps are detailed as follows.

Assuming that n represents the note frame length and N represents the sampling points in the frame, each humming note signal in the feature tone retrieval is windowed and framing processed by formula (11), so as to make each humming note signal short-term stable:

$$W(n) = \frac{x(n) \times E(n)}{N \times n}. \tag{12}$$

In the formula, $x(n)$ represents any humming note signal and $E(n)$ represents the short-time energy of $x(n)$.

Assuming that $s(k)$ represents the current sampling value of short-time humming note signal, $s(k)$ is defined as the linear combination of historical sampling value and excitation signal, which is expressed by the following formula:

$$s(k) = \frac{s(n) \times e(n) \times v(n)}{a_i \times p} \times G. \tag{13}$$

In the formula, a_i represents the prediction coefficient of the image sequence, p represents the prediction order of the image sequence, G represents the gain factor of the image sequence, $e(n)$ represents the glottic pulse excitation of the image sequence, and $v(n)$ represents the channel response value of the image sequence.

$x(n)$ is judged as the result of glottic pulse excitation $e(n)$ filtered by channel response $v(n)$, and $e(n)$ is a short-time humming note signal with periodic characteristics.

Assuming that $R_{\text{corss}}(t)$ represents a function with the same period, the similarity between waveform signals of different humming notes is calculated by the following formula:

$$R_{\text{auto}}(t) = \frac{1}{N} \sum_{n=1}^{N} x(n)x(n+t). \tag{14}$$

The similarity between the waveform signals of different humming notes mainly has two states: Cross and Jiugong grid, as shown in Figures 5 and 6.

Regular squares are used to represent the similarity between different humming note waveform signals. Generally, the image sequence value is 0 or 1. The two-dimensional space is formed by a large number of image sequences. The adjacent elements are the subelements to be studied, and their shape is mainly square. $y(n)$ represents $x(n)$ and signals with the same period T, and the mathematical expression for discrete-time signals is given by the following formula:

$$R_{\text{cross}}(t) = \frac{1}{N} \sum_{n=1}^{N} x(n) \times y(n)^T. \tag{15}$$

Based on the image sequence, the center clipping method is used to give the threshold of the starting point range of specific humming notes, which is expressed as follows:
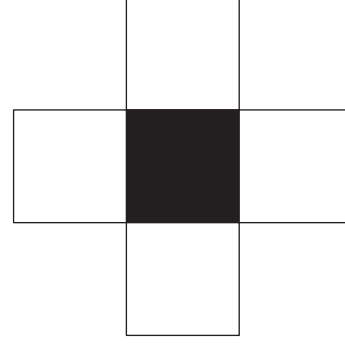
Figure 5: Cross structure.
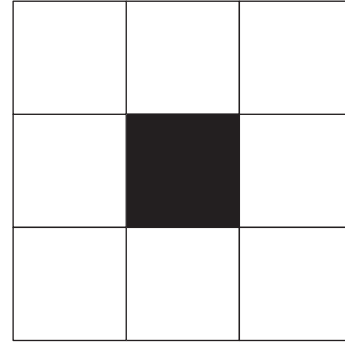
Figure 6: Jiugong lattice structure.

$$x''(n) = \frac{s(n) + g(n)}{y(n) \times x(n)}. \tag{16}$$

In the formula, represents Additive White Noise Gaussian independent of $s(n)$. $y(n)$ is the third-order level signal of humming note searched by $x(n)$ through clipping method, eliminates the low amplitude part of humming note signal, and calculates the correlation function between humming note starting point signals.

To sum up, it can be explained that, in the process of intelligent optimization and recognition of note starting point of feature tone retrieval, the initial note signal is preprocessed, the similarity between different note waveform signals is calculated, and the cross-correlation function between each note signal is obtained, which lays a foundation for intelligent optimization and recognition of note starting point of feature tone retrieval.

2.3.2. Intelligent Optimization Recognition of Note Starting Point Based on Starting Point Feature. Because the melody pitch feature extraction is a key link in the intelligent optimization and recognition of the note starting point of the feature tone retrieval and directly affects the quality of the feature tone retrieval, in the process of recognition, the short-term spectral structure features and envelope features of the melody pitch are extracted based on the correlation function between the obtained note starting point signals based on the image sequence. Based on the feature transformation and fusion of each melody pitch starting point, the intelligent optimization recognition of note starting

point is completed. The flowchart of image sequence feature extraction is shown in Figure 7.

According to Figure 7, firstly, the input multimodal music audio signal is prefiltered to convert the input analog audio into a digital audio signal within the sound frequency range that can be received by the human ear. Secondly, according to the short-time stability of the audio signal, the preweighted audio signal is processed into frames, and the Hamming window is used to window the signal of each frame to reduce the influence of Gibbs effect. The short-time Fourier transform converts the time domain signal into the frequency domain signal, which is convenient for the triangular window filtering of the subsequent Mel filter. Then, the logarithm of the filtered signal is taken, and the discrete cosine transform is carried out to remove the correlation between the signals of various dimensions, and the signal is mapped to the low dimensional space. Finally, the Mel cepstrum coefficient is obtained by spectral weighting, cepstrum mean subtraction, and difference processing. Because the lower order parameters of cepstrum are easily affected by the characteristics of speaker and channel, the recognition ability is improved.

The specific steps of intelligent optimization identification are detailed as follows.

Assuming that $\partial(o)$ represents the smoothing parameters of the pitch trajectory, based on the obtained $R'_{\text{cross}}(t)$, the short-time spectral structure features of the extracted humming melody pitch represented by BN and the envelope features represented by MFCC are extracted by the following formulae:

$$\text{BN} = \frac{R'_{\text{cross}}(t) \times \partial(o)}{\varpi_{(j)} \times \varepsilon(h)}, \tag{17}$$

$$\text{MFCC} = \frac{R'_{\text{cross}}(t) \times \partial(o)}{\varpi_{(j)} \times \text{BN}}. \tag{18}$$

In the formula, $\varpi_{(j)}$ represents the number of starting points of humming notes and $\varepsilon(h)$ represents the offset vector.

A set of transformation matrices for the starting points of humming melody pitch is obtained by discrimination training. Based on the image sequence, each transformation matrix in the set corresponds to a region in the feature space division of the starting points of humming notes, which is transformed with the transformation matrix corresponding to the region to which the feature vector belongs. It is assumed that $o(t)$ represents the input feature of time t, A_i represents the transformation matrix corresponding to the i domain, and the characteristic transformation of the s melody pitch segment is described by the following formula:

$$o'_s(t) = R \sum_{i=1}^{A_i} S \times o(t) \times x_{i,s}. \tag{19}$$

In the formula, R represents the starting paragraph of melody pitch after domain division and $x_{i,s}$ represents the weight coefficient corresponding to the selected feature transformation matrix A_i.
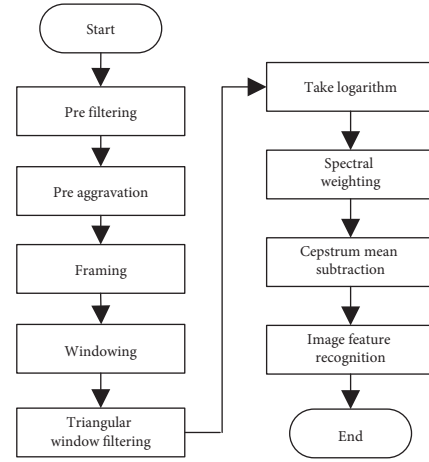
FIGURE 7: Flowchart of image sequence feature extraction.

Assuming that h represents the excitation signal of the BN layer humming melody pitch node, the transformation matrix features represented by M^{BN} and M^{RDLT} are fused by the following formula:

$$y_{\text{con}}(t) = \frac{\left[M^{\text{BN}}, M^{\text{RDLT}}\right]}{h \times xo(t)}. \tag{20}$$

In the formula, $xo(t)$ represents the regularization function.

Assuming that the estimated value of β noise spectrum is used, the parameters of the fused transformation matrix feature $y_{\text{con}}(t)$ are optimized by the following formula:

$$H_{\text{MPE}}(Y_{\text{con}}) = \frac{y_{\text{con}}(t) \times M^{\text{BN}} \times M^{\text{RDLT}}}{M^O}. \tag{21}$$

In the formula, M^O represents the transformation matrix corresponding to the nonzero coefficient term. Based on the results calculated by formula (21), the intelligent recognition of note starting point in feature tone retrieval can be effectively completed, so as to complete the research of multimodal music emotion recognition method based on image sequence.

3. Experimental Analysis

In order to test the application effect of the multimodal music emotion recognition method based on image sequence, MATLAB software is used as the algorithm operation platform, and a specific example is selected for simulation test and analysis. The experimental environment settings are shown in Table 2.

The samples selected in the test are from the emotional corpus. According to the selected samples and the emotions to be expressed, they are divided into five categories. The specific distribution of samples is shown in Table 3.

Kappa coefficient is selected as the index to evaluate the intelligent recognition and classification of music emotion. Kappa coefficient is used for consistency test and classification accuracy. Its calculation formula is as follows.

TABLE 2: Experimental parameter setting.

Parameter	Numerical value
Node	10
CPU	2
Core frequency	1.9 GHz
Memory	8 GB

TABLE 3: Distribution of test samples.

Emotion type	Music clip name	Characteristic dimension	Sample properties
Happy	Red head rope fragment	455	Training sample
	Love in the rain	355	Training sample
	Carmen fragment	784	Test sample
	Trout fragment	232	Test sample
Sadness	Schindler list theme song clip	534	Training sample
	Liang Zhu fragment	454	Training sample
	Pathetique fragment	234	Test sample
	Parting fragment	545	Test sample
Tender	Lullaby fragment	215	Training sample
	Blue Danube segment	313	Training sample
	Little star clip	534	Test sample
	To Alice	341	Test sample
Anger	Polish dance pieces	132	Training sample
	Destiny fragment 1	431	Training sample
	International song clip	453	Test sample
	Empty madness	315	Test sample
Fear	Gloomy Sunday clip	341	Training sample
	Ghost call clip	345	Training sample
	Thirteen pairs of eyes	422	Test sample
	Step by step press clip	244	Test sample

TABLE 4: Recognition results of some music clips.

Music clip name	Emotion type	Paper method	Reference [11] method	Reference [12] method
Little star clip	Tender	Tender	Sadness	Sadness
Blue Danube segment	Tender	Tender	Sadness	Happy
Liang Zhu fragment	Sadness	Sadness	Tender	Tender
Love in the rain	Happy	Happy	Tender	Tender
Trout fragment	Happy	Happy	Happy	Tender
Ghost call film	Fear	Fear	Fear	Sadness
Segment 2	Sadness	Sadness	Sadness	Fear

$$k = \frac{p_o - p_e}{1 - p_e}. \tag{22}$$

In the formula, p_o is the observation consistency rate and p_e represents the expected consistency rate. The larger the values of $k \in [-1, 1]$ and the larger the k value, the more consistent the two results. When $k \geq 0.75$, the results are consistent and the classification recognition is more accurate. If $k < 0.4$, it indicates lack of consistency and poor classification and recognition accuracy.

Input the test samples in Table 3 into the trained neural network model, test the samples, count the sample test results, and calculate the kappa coefficient. The results are as follows:

$$k_{\text{Happy}} = 0.855,$$
$$k_{\text{Sadness}} = 0.870,$$
$$k_{\text{Tender}} = 0.912, \tag{23}$$
$$k_{\text{Anger}} = 0.825,$$
$$k_{\text{Fear}} = 0.811.$$

The kappa coefficient k values calculated above are greater than 0.75, indicating that the recognition and classification results are in good agreement with the actual results, and the classification and recognition accuracy is high, which has achieved the research purpose. The

multimodal music emotion recognition method is used to identify the music fragments in the test set, and some test samples and their discrimination results are intercepted, as shown in Table 4.

It can be seen from Table 4 that the identification results of the same test sample by different methods are different. Rhythm and melody characteristics have a great influence on the recognition of music emotion. On the premise that the image sequence is unchanged, selecting the appropriate music feature input vector will improve the accuracy of multimodal music emotion recognition to a certain extent.

4. Conclusion and Prospect

4.1. Conclusion. Multimodal musical emotion is a breakthrough in the field of artificial intelligence. It has become a new research feature of computer science, cognitive science, neuroscience, brain science, psychology, behavioral science, and other interdisciplinary fields. Multimodal musical emotion understanding is an important branch of emotion computing and has a broad development prospect. The multimodal music emotion recognition method based on image sequence verifies the performance of the algorithm through an example. The kappa coefficient proves that the classification recognition accuracy of the algorithm is high, which achieves the research goal. Meanwhile, the rhythm and melody characteristics have a great influence on the recognition of music emotion.

4.2. Prospect. The possible future research direction is to apply deep learning method to music emotion recognition. Deep learning is a kind of based on feature hierarchical structure, characteristics of unsupervised learning learning method, has a lot of the hidden layer of all the excellent characteristics of artificial neural network learning ability, learning and to the characteristics of the characterization of the nature of the data more through millions of music is used to study characteristics. Thus, let the machine independently choose better music features to describe the relationship between the music and the emotion.

Data Availability

The raw data supporting the conclusions of this article will be made available by the author, without undue reservation.

Conflicts of Interest

The author declares that there are no conflicts of interest regarding this work.

References

[1] I. Peretz, J. Ayotte, R. J. Zatorre et al., "Effects of vocal training in a musicophile with congenital amusia," *Neuron*, vol. 33, no. 2, pp. 185–191, 2020.

[2] A. P. Montgomery, A. Mousavi, M. Carbonaro, and D. Hayward, "Using learning analytics to explore self-regulated learning in flipped blended learning music teacher education," *British Journal of Educational Technology*, vol. 9, no. 10, pp. 114–127, 2019.

[3] S. Hawkins, "Situational influences on rhythmicity in speech, music, and their interaction," *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, vol. 369, no. 1658, Article ID 20130398, 2019.

[4] M. Bigliassi, C. I. Karageorghis, G. K. Hoy, and G. S. Layne, "The Way You Make Me Feel: Psychological and cerebral responses to music during real-life physical activity," *Psychology of Sport and Exercise*, vol. 41, no. 3, pp. 211–217, 2019.

[5] L. Bochen, X. Liu, K. Dinesh, and Z. Duan, "Creating a Multitrack classical music performance Dataset for multimodal music analysis: Challenges, Insights, and applications," *IEEE Transactions on Multimedia*, vol. 12, no. 14, pp. 159–164, 2019.

[6] S. Nag, S. Sanyal, A. Banerjee, R. Sengupta, and D. Ghosh, "Music of brain and music on brain: a Novel EEG Sonification approach," *Cognitive Neurodynamics*, vol. 13, no. 4, pp. 13–31, 2019.

[7] X. Wang, G. Soumitra, and G. Sun-Wei, "Quantitative quality control in microarray image processing and data acquisition," *Nucleic acids research*, vol. 29, no. 15, pp. 75–80, 2019.

[8] L. Reichel and U. O. Ugwu, "Tensor Krylov subspace methods with an invertible linear transform product applied to image processing," *Applied Numerical Mathematics*, vol. 166, no. 8, pp. 186–207, 2021.

[9] W. Z. Liang, I. Possignolo, X. Qiao, and K. DeJonge, "Utilizing digital image processing and two-source energy balance model for the estimation of evapotranspiration of dry edible beans in western Nebraska," *Irrigation Science*, vol. 204, no. 39, pp. 617–631, 2021.

[10] M. Talaat, M. Tayseer, and A. El-Zein, "Digital image processing for physical basis analysis of electrical failure forecasting in XLPE power cables based on field simulation using finite-element method," *IET Generation, Transmission & Distribution*, vol. 14, no. 26, pp. 6703–6714, 2020.

[11] A. M. Proverbio, F. Benedetto, and M. Guazzone, "Shared neural mechanisms for processing emotions in music and vocalizations," *European Journal of Neuroscience*, vol. 51, no. 5, pp. 1987–2007, 2020.

[12] H. Platel, B. Jean-Claude, B. . Desgranges, F. Bernard, and F. Eustache, "Semantic and episodic memory of music are subserved by distinct neural networks," *NeuroImage*, vol. 20, no. 1, pp. 244–256, 2019.

[13] S. Hizlisoy, S. Yildirim, and Z. Tufekci, "Music emotion recognition using convolutional long short term memory deep neural networks," *Engineering Science and Technology an International Journal*, vol. 24, no. 3, pp. 760–767, 2020.

[14] K. W. Cheuk, Y. J. Luo, B. Balamurali, and G. Roig, "Regression-based music emotion prediction using triplet neural networks," in *Proceedings - ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 21, no. 1, pp. 15–21, 2020.

[15] S. Chowdhury, V. Praher, and G. Widmer, "Tracing back music emotion predictions to sound Sources and Intuitive perceptual Qualities," vol. 14, no. 6, pp. 45–52, 2021.

[16] D. . Zheng, "Music emotion recognition classification algorithm based on forward neural network," *Information & Technology*, vol. 43, no. 12, pp. 57–61, 2019.

[17] X. Tang, C. X. Zhang, and L. I. Jiang-Feng, "Music emotion recognition based on deep learning," *Computer Knowledge and Technology*, vol. 15, no. 11, pp. 232–237, 2019.

[18] F. Pan, L. Zhang, Y. Ou, and X. Zhang, "The audio-visual integration effect on music emotion: behavioral and

physiological evidence," *PLoS One*, vol. 14, no. 5, Article ID 0217040, 2019.

[19] T. L. Nguyen, B. L. Trieu, Y. Hiraguri, M. Morinaga, T. Morihara, and T. Yano, "Effects of changes in acoustic and non-acoustic factors on Public Health and Reactions: Follow-up Surveys in the Vicinity of the Hanoi Noi Bai International Airport," *International Journal of Environmental Research and Public Health*, vol. 17, no. 7, pp. 2597–2530, 2020.

[20] S. H. Park and P. J. Lee, "Reaction to floor impact noise in multi-storey residential buildings: the effects of acoustic and non-acoustic factors," *Applied Acoustics*, vol. 150, no. 7, pp. 268–278, 2019.

[21] J. L. Tian and S. N. University, "Dynamic Visualization method of music melody based on MATLAB," *Modern Computer*, no. 33, pp. 3–6+10, 2019.

[22] I. Goienetxea, I. Mendialdua, I. Rodríguez, and B. Sierra, "Statistics-based music generation approach Considering Both rhythm and melody Coherence," *IEEE Access*, vol. 7, no. 2, Article ID 183365, 2019.

[23] Y. Yu and S. Canales, "Conditional LSTM-GAN for melody generation from Lyrics," *Computer Science*, vol. 15, no. 8, pp. 26–35, 2019.

[24] M. Farzaneh and R. Mahdian Toroghi, "Music generation using an interactive Evolutionary algorithm," *Pattern Recognition and Artificial Intelligence*, vol. 18, no. 12, pp. 207–217, 2019.

[25] S. Swaminathan and E. G. Schellenberg, "Musical ability, music training, and language ability in childhood," *Journal of Experimental Psychology Learning Memory and Cognition*, vol. 46, no. 12, pp. 2340–2348, 2019.

[26] D. Herremans and T. Bergmans, "Hit Song prediction based on early Adopter data and audio features," *Sound*, vol. 16, no. 10, pp. 148–153, 2020.

[27] B. Gong, M. Kaya, and N. Tintarev, "Contextual Personalized Re-Ranking of music recommendations through audio features," *Information Retrieval*, vol. 6, no. 9, pp. 89–95, 2020.

[28] A. Kmb, A. Tb, A. Ds, and Z. Zhao, "Contributions of MIR to soundscape ecology. Part 3: Tagging and classifying audio features using a multi-labeling k -nearest neighbor approach," *Ecological Informatics*, vol. 51, no. 5, pp. 103–111, 2019.

[29] Z. Wang, L. Wang, and H. Huang, "Joint low rank embedded multiple features learning for audio-visual emotion recognition," *Neurocomputing*, vol. 388, no. 5, pp. 324–333, 2020.

[30] X. Guo, W. Zhong, L. Ye, L. Fang, and Q. Zhang, "Affective Video content analysis based on two Compact audio-visual features," *Communications in Computer and Information Science*, vol. 16, no. 2, pp. 355–364, 2020.

[31] E. Han and H. Cha, "Audio feature extraction for effective emotion classification," *IEIE Transactions on Smart Processing & Computing*, vol. 8, no. 4, pp. 100–107, 2019.

[32] Z. H. Wang, G. J. Horng, T. H. Hsu, and A. Aripriharta, "Heart sound signal recovery based on time series signal prediction using a recurrent neural network in the long short-term memory model," *The Journal of Supercomputing*, vol. 76, no. 1, pp. 8373–8390, 2019.

[33] F. Huang, J. Zeng, Y. Zhang, and W. Xu, "Convolutional recurrent neural networks with multi-sized convolution filters for sound-event recognition," *Modern Physics Letters B*, vol. 34, no. 23, Article ID 2050235, 2020.

[34] R. Haeb-Umbach, S. Watanabe, T. Nakatani et al., "Speech processing for digital Home Assistants: Combining signal processing with deep-learning Techniques," *IEEE Signal Processing Magazine*, vol. 36, no. 6, pp. 111–124, 2019.

[35] R. S. Da, Özbey, Muzaffer, Çatlı, and A. Burak, "A transfer-learning approach for Accelerated MRI using deep neural networks," *Magnetic Resonance in Medicine*, vol. 84, no. 3, pp. 663–685, 2020, https://pubmed.ncbi.nlm.nih.gov/?term=%C3%87ukur+T&cauthor_id=31898840.

[36] Y. Zhang, T. S. Lee, M. Li, F. Liu, and S. Tang, "Convolutional neural network models of V1 responses to complex patterns," *Journal of Computational Neuroscience*, vol. 46, no. 1, pp. 33–54, 2019.

Research Article

Large Margin Graph Embedding-Based Discriminant Dimensionality Reduction

Yanjia Tian [ID] [1,2] **and Xiang Feng** [ID] [1,3]

[1]*Department of Computer Science and Engineering, East China University of Science and Technology, Shanghai, China*
[2]*School of Electronics and Information, Shanghai DianJi University, Shanghai, China*
[3]*Shanghai Engineering Research Center of Smart Energy, Shanghai, China*

Correspondence should be addressed to Xiang Feng; xfeng_ecust@163.com

Discriminant graph embedding-based dimensionality reduction methods have attracted more and more attention over the past few decades. These methods construct an intrinsic graph and penalty graph to preserve the intrinsic geometry structures of intraclass samples and separate the interclass samples. However, the marginal samples cannot be accurately characterized only by penalty graphs since they treat every sample equally. In practice, these marginal samples often influence the classification performance, which needs to be specially tackled. In this study, the near neighbors' hypothesis margin of marginal samples has been further maximized to separate the interclass samples and improve the discriminant ability by integrating intrinsic graph and penalty graph. A novel discriminant dimensionality reduction named LMGE-DDR has been proposed. Several experiments on public datasets have been conducted to verify the effectiveness of the proposed LMGE-DDR such as ORL, Yale, UMIST, FERET, CMIU-PIE09, and AR. LMGE-DDR performs better than other compared methods, and the corresponding standard deviation of LMGE-DDR is smaller than others. This demonstrates that the evaluation method verifies the effectiveness of the introduced method.

1. Introduction

Dimensionality reduction (DR) is more important in most fields such as machine learning and pattern recognition [1–4]. It aims to resolving the curse of dimensionality by achieving relevant low-dimensional representations of high-dimensional datasets. Linear discriminant analysis (LDA) and principal component analysis (PCA) are the most representative methods [5, 6]. PCA obtains low-dimensional space by maximizing variance. LDA can use label information to project the feature space to distinguish categories by maximizing the interclass distance and minimizing the intraclass distance. However, LDA cannot capture the local structure of data. As is known, the local structures of high-dimensional data are very important for data representation.

K near neighbor graph can better characterize the local structure of data [7]. Thus, over the past years, graph embedding-based dimensionality reduction methods have sprung up [7, 8], such as LLE [9], Isomap [10, 11], and Laplacian eigenmap [12]. However, these manifold learning methods do not directly process the new samples because they do not obtain any mapping function, which is known as the 'out-of-sample' problem [13]. Therefore, to solve the problem, a more effective method is presented to obtain the explicit projection mapping. Locality preserving projections (LPPs) are to preserve the local structure of data in the low-dimensional space, which is a famous method [2]. For its simplicity and effectiveness, its variants have been proposed [14, 15]. However, LPP performs worse in classification since it does not fully use label information, which is an unsupervised method [16]. Neighborhood preserving projection

(NPP) preserved the local neighborhood information on the data manifold [17].

To further improve the classification performance, discriminant graph embedding-based methods have gradually become a popular research topic by using label information, which aims to preserve the within-class geometrical structure while, at the same time, maximizing the between-class distances of different manifolds [18]. Thus, recently, more and more discriminant graph embedding-based methods have been studied. Marginal fisher analysis (MFA) constructs two adjacency graphs to maximize the separability between pairwise marginal data points [19]. Local discriminant embedding (LDE) [20] utilized the label information and proposed the nearest neighbor-based embedding. However, it suffers from the so-called small-sample-size (SSS) problem that it cannot directly be applied to high-dimensional data [20]. Considering the local intraclass attraction or interclass repulsion, discriminant neighborhood embedding (DNE) was proposed to make data points in the same class compacted, whereas the gaps between classes become wider in a low-dimensional subspace [21]. However, DNE does not always set the edges with its neighbors of different classes, which would reduce the interclass distance in the new space and will deteriorate the classification [22]. Thus, Ding et al. constructed double adjacency graphs to link their homogeneous and heterogeneous neighbors and introduced a more effective version of DNE termed DAG-DNE [22]. Inspired by DAG-DNE, some discriminant analysis-based methods have been proposed over the past few years [23–33].

Most dimensionality reduction methods can be unified in the graph-embedding framework [19]. The ways to construct the similarity graph and the penalty graph among these methods are different [34]. Therefore, the graph-embedding-based methods are sensitive to the weight matrix, whereas they endow the same weight for each sample (including marginal samples) in the same way. However, as stated in [35], these marginal samples located in the class margin in the high-dimensional space have been treated to achieve maximum between-class hypothesis margin and good classification performance, which is more crucial in the classification performance. Therefore, large hypothesis margins between near neighbors of these marginal samples can improve the discriminating power of embedding features and should be treated separately. In this study, for the marginal samples, the nearest neighbors' hypothesis margin of the marginal sample has been considered and maximized to improve the discriminant power, in addition to constructing double adjacency graphs. In this study, a novel large margin graph embedding-based discriminant dimensionality reduction named LMGE-DDR has been introduced. Most experimental results confirm the effectiveness of the proposed LMGE-DDR on several public datasets.

2. Methods

Firstly, the common notations in this study are presented. The high-dimensional data are denoted as $X = [x_1, x_2, \ldots, x_n] \in R^{d \times n}$ with n samples in d-dimensions and include C classes with class $c_i \in 1, 2, \ldots, C$. $y_i = \mathbf{P}^T x_i$ denotes that the sample x_i is transformed by the matrix $\mathbf{P} = [p_1, p_2, \ldots, p_r] \in R^{d \times r}$, where $d \gg r$, $p_i \in R^d$ is any one column vector. $N_k^+(x_i)$ ($N_k^-(x_i)$) and $N_k(x_i)$, respectively, denote the k neighbors with the same class (different class) and k neighbors of sample x_i.

2.1. DNE.
Discriminant neighborhood embedding (DNE) considered the local intraclass attraction and interclass repulsion and learned the intrinsic graph F^w and penalty graph F^b as follows:

$$
\begin{aligned}
F_{ij}^w &= \begin{cases} 1, & x_i \in N_k(x_j) \text{ or } x_j \in N_k(x_i), \quad c_i = c_j, \\ 0, & \text{otherwise,} \end{cases} \\
F_{ij}^b &= \begin{cases} 1, & x_i \in N_k(x_j) \text{ or } x_j \in N_k(x_i), \quad c_i = c_j, \\ 0, & \text{otherwise.} \end{cases}
\end{aligned} \tag{1}
$$

The objective function can be denoted as follows:

$$
\begin{aligned}
\max \theta(P) &= \Psi(P) - \Phi(P), \\
&\text{s.t. } P^T P = I.
\end{aligned} \tag{2}
$$

Herein,

$$
\begin{aligned}
\Phi(P) &= \sum_{x_i \in N_k^+(x_j)} \sum_{x_j \in N_k^+(x_j)} {P^T x_i - P^T x_j}^2 F_{ij}^w, \\
&= 2P^T X (D^w - F^w) X^T P,
\end{aligned} \tag{3}
$$

where $D_{ii}^w = \sum_j F_{ij}^w$.

$$
\begin{aligned}
\Psi(P) &= \sum_{x_i \in N_k^-(x_j)} \sum_{x_j \in N_k^-(x_j)} {\mathbf{P}^T x_i - \mathbf{P}^T x_j}^2 F_{ij}^b, \\
&= 2P^T X (D^b - F^b) X^T P,
\end{aligned} \tag{4}
$$

where $D_{ii}^b = \sum_j F_{ij}^b$.

The constraint $P^T P = I$ can preserve the local structure and reinforce the discriminant ability [36].

The objective in (2) can be rewritten by the formal of trace as follows:

$$
\begin{aligned}
\theta(P) &= \Psi(P) - \Phi(P), \\
&= 2tr P^T X (D^b - F^b) X^T P - 2tr P^T X (D^w - F^w) X^T P \\
&= 2tr P^T X (D^b - F^b - D^w + F^w) X^T P \\
&= 2tr P^T X S X^T P,
\end{aligned} \tag{5}
$$

where $S = D^b - F^b - D^w + F^w$. Therefore, the objective function (2) can be rewritten as follows:

$$
\begin{cases} \max & \theta(P) = tr\{P^T XSXP\}), \\ \text{s.t.} & P^T P = I. \end{cases} \tag{6}
$$

The projection matrix P can be found by resolving the following eigenvector problem:

$$XSX^T P = \lambda P, \tag{7}$$

where λ_i is the eigenvalues, $i = 1,...,d$, and P_i ($i = 1,...,d$) is the corresponding eigenvector. Assume $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_d$ and $P = [P_1, P_2, \ldots, P_r]$. The details are presented in [21].

2.2. DAG-DNE. Double adjacency graph-based discriminant neighborhood embedding termed DAG-DNE constructed double adjacency graphs to propose a more effective version of DNE. In DAG-DNE, $\mathbf{F}^b$ and $\mathbf{F}^w$ can be defined as follows:

$$F_{ij}^w = \begin{cases} 1, & x_i \in N_k(x_j) \text{ or } x_j \in N_k(x_i), \\ 0, & \text{otherwise,} \end{cases}$$
$$F_{ij}^b = \begin{cases} 1, & x_i \in N_k(x_j) \text{ or } x_j \in N_k(x_i), \\ 0, & \text{otherwise.} \end{cases} \tag{8}$$

The projection matrix $\mathbf{P}$ can be solved as in DNE as follows:

$$\begin{cases} \max \theta(\mathbf{P}) = \Psi(\mathbf{P}) - \Phi(\mathbf{P}), \\ s.t. \; \mathbf{P}^T P = \mathrm{I}. \end{cases} \tag{9}$$

3. Proposed Method

It is revealed that the weights in the adjacency matrix have been endowed in the same way for each sample including the marginal sample, which cannot further improve the between-class hypothesis margin and deteriorates the classification performance. In this study, the marginal sample is defined in Definition 1. The hypothesis margin was studied as in [37–39].

Definition 1 (marginal sample). x_i is regarded as a marginal sample if $\exists x_j \in N_{k_0}(x_i)$ and $c_i \neq c_j$.

The marginal samples in this study are the ones located in class margin. Figure 1 is the k_0 near neighbors' graph and shows the marginal samples (i.e., {5, 6, 7, 8}).

Definition 2 (hypothesis margin). As is shown in [37], the hypothesis margin can be defined as follows:

$$H(x) = \frac{1}{2} x - \text{near miss}(x) - x - \text{near hit}(x), \tag{10}$$

where nearhit(x) and nearmiss(x) denote the nearest neighbors of sample x with the same class and different class, respectively. $\| \cdot \|$ represents the L_2 norm. The sample x can be accurately recognized by 1NN classifier (the nearest neighbor) when $H(x) > 0$, as illustrated in Figure 2.

Definition 3 (heterogeneous near neighbors' hypothesis margin). A marginal of sample x_1 is shown in Figure 3 to illustrate the heterogeneous near neighbors' hypothesis margin of x_1, which is defined as follows:

$$N_{k1}^+(x_1) = \{x_2, x_3, x_4\}. \tag{11}$$

Herein, $N_{k_2}^+(x_1) = \{x_2, x_3, x_4\}$, $N_{k_1}^-(x_1) = \{x_5, x_6, x_7\}$.
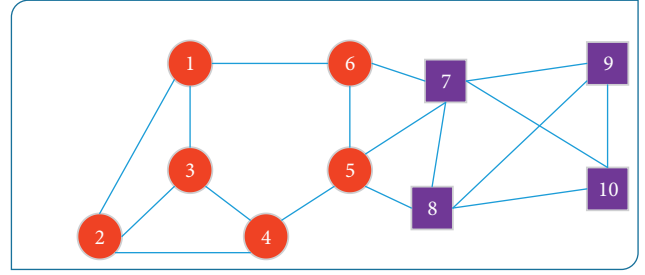
FIGURE 1: An illustration of k_0 near neighbors graph and marginal samples. Different style corresponds to different classes ($k_0 = 3$). Nodes {5 6 7 8} are the marginal samples.
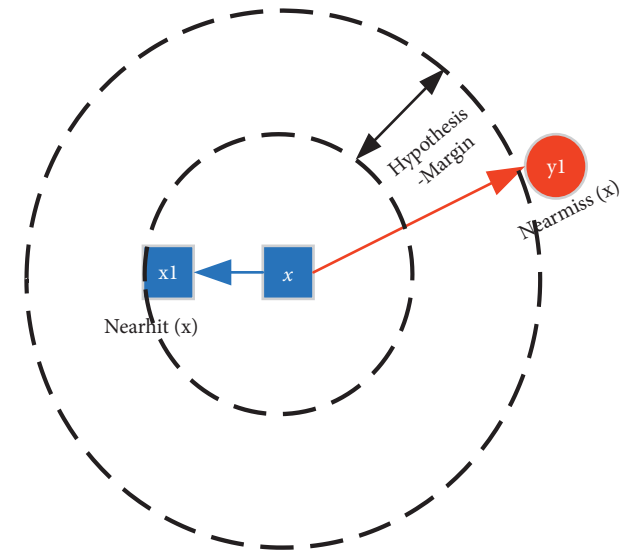
FIGURE 2: Illustration of the hypothesis margin.
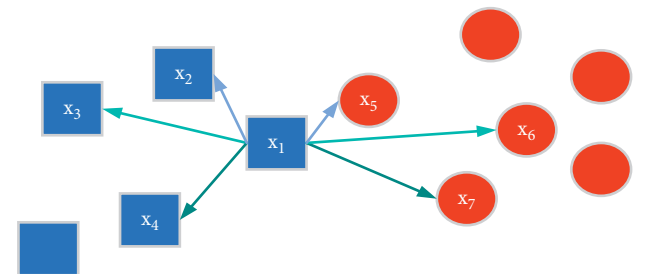
FIGURE 3: Illustration of heterogeneous near neighbors' hypothesis margin ($k_1 = k_2 = 3$).

As shown in (11), it can keep the heterogeneous samples separated and achieve a large margin between heterogeneous near neighbors when all the expressions in brackets are larger than zero, which means it can be correctly classified by the 1NN classifier.

4. LMGE-DDR

On the basis of DAG-DNE, the marginal samples in high-dimensional space are additionally treated separately by

maximizing the heterogeneous near neighbors' margin, which can improve the discriminant power. LMGE-DDR can be proposed as follows:

The interclass weight F_{ij}^b and F_{ij}^w are the same as in DAG-DNE, the objective function of LMGE-DDR can be denoted as

$$\begin{cases} \max \theta(P) = \Psi(P)\Phi(P) + M(P), \\ s.t. \ P^T P = \mathrm{I}. \end{cases} \tag{12}$$

$$\begin{aligned} M(P) &= \sum_{x_i \in MS} \sum_{x_t \in N_{k1}^-(x_i)} \sum_{x_i \in MS} \sum_{x_j \in N_{k1}^-(x_i)} \left((1-\alpha)\left\| P^T x_i - P^T x_t \right\| - \alpha \left\| P^T x_i - P^T x_j \right\| \right), \\ &= \sum_{x_i \in MS} \left((1-\alpha) \sum_{x_t \in N_{k1}^-(x_i)} k2 \left\| P^T x_i - P^T x_t \right\| - \alpha \sum_{x_j \in N_{k2}^+(x_i)} k1 \left\| P^T x_i - P^T x_j \right\| \right), \\ &= \sum_{x_i \in MS} \left((1-\alpha) \sum_{x_t \in N_{k1}^-(x_i)} k2 (x_i - x_t) - PP^T (x_i - x_t) - \alpha \sum_{x_j \in N_{k2}^+(x_i)} k1 (x_i - x_t) - PP^T (x_i - x_t) \right). \end{aligned} \tag{13}$$

The solution of (12) is easily obtained by solving the maximum eigenvalue problem.

$$XSX^T P + \sum_{x_i \in MS} \left((1-\alpha)G_i - \alpha H_i \right) P = \lambda P. \tag{14}$$

Here, $G_i = \sum_{x_t \in N_{k1}^-(x_i)} k_2 (x_i - x_t)^T (x_i - x_t)$

$$H_i = G_i = \sum_{x_j \in N_{k1}^-(x_i)} k_1 (x_i - x_t)^T (x_i - x_t), \tag{15}$$

where λ_i is the eigenvalues, $i = 1,...,d$, and P_i is the corresponding eigenvector, $i = 1,...,d$. Assume $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_d$ and $\mathbf{P} = [P_1, P_2, \ldots, P_r]$.

The details of LMGE-DDR can be seen in Algorithm 1.

5. Analysis of LMGE-DDR

In this section, LMGE-DDR will be analyzed to illustrate the effectiveness in preserving the geometrical and discriminant structures.

Although LMGE-DDR is similar to DAG-DNE in constructing an adjacency graph, for the marginal samples in high-dimensional space, LMGE-DDR maximizes the heterogeneous near neighbors' hypothesis margin to achieve a large between-class margin in low-dimensional subspace and discriminate the local structure of neighbors, improving the discriminant power compared to DAG-DNE.

The performances of LMGE-DDR in a Toy data are illustrated in Figure 4.

As shown in Figure 4(a), for the sample x_1, nearhit(x_1) is x_2 and nearmiss(x_1) is x_3. Thus, based on (12), the hypothesis margin of x_1 is denoted as follows:

Here, $\Psi(P), \Phi(P)$ are the same as in DAG-DNE. MS denotes the marginal samples set in the high-dimensional space. α is a trade-off parameter and here $\alpha \in [0, 1]$.

This objective function is transformed into two parts as follows:

$\Psi(P) - \Phi(P) = 2tr P^T XSX^T P$ based on (5).

$$\begin{aligned} H(x_1) &= \frac{1}{2} (x_1 - \text{nearmiss}(x_1) - x_1 - \text{nearhit}(x_1)), \\ &= \frac{1}{2} (x_1 - x_3 - x_1 - x_2) = -0.83. \end{aligned} \tag{16}$$

Based on Definition 2, the sample x_1 will be recognized by mistake because its hypothesis margin is less than zero.

The embedded results and hypothesis margins in one-dimensional space are illustrated in Figures 4(b)–4(e)). It can be seen that the hypothesis margins of sample x_1 in the low-dimensional space are less than zero in MFA and DAG-DNE, which is the opposite situation in MNMDP, DNE, and LMGE-DDR. In LMGE-DDR, the hypothesis margin of sample x_1 is larger ($H(x_1) = 0.39$) than that in DAG-DNE, which is useful for the classification.

Overall, maximizing the heterogeneous near neighbors' hypothesis margin of marginal samples can further improve the discriminant power in low-dimensional space.

6. Experiments

In this section, compared with several popular methods such as DAG-DNE, DNE, MNMDP, and MFA, LMGE-DDR is conducted on several experiments systematically to verify its effectiveness. Specifically, the performance of LMGE-DDR is illustrated on the experiments of face recognition and 2-dimensional visualization. The randomly selected l images from each person constitute the training data, and the remaining are the testing data. The nearest neighbor parameters k, k_1, and k_2 in constructing adjacency graphs are set as l-1 for all methods as in [40]. PCA is taken to reduce

Input: a training set $\{(x_i, c_i)\}_{i=1}^{N}$, α, k_1, k_2, k and the dimensionality of discriminant subspace r.
Output: projection matrix P;
(1) Construct the intraclass adjacency graph $\mathbf{F}^w$ by:
$$F_{ij}^w = \begin{cases} 1, & x_i \in N_k(x_j) \text{ or } x_j \in N_k(x_i) \\ 0, & \text{otherwise} \end{cases}$$
and interclass adjacency graph $\mathbf{F}^b$ by:
$$F_{ij}^b = \begin{cases} 1, & x_i \in N_k(x_j) \text{ or } x_j \in N_k(x_i), \\ 0, & \text{otherwise}, \end{cases}$$
(2) computing the MS, G_i, H_i based on (15)
(3) Eigendecompose the matrix, where $XSX^T + \sum_{x_i \in MS}((1-\alpha)G_i - \alpha H_i))$
(4) Choose the r largest eigenvalues corresponding eigenvectors: $\mathbf{P} = [P_1, P_2, \ldots, P_r]$.

ALGORITHM 1: LMGE-DDR.

(a)

(b)

(c)

(d)
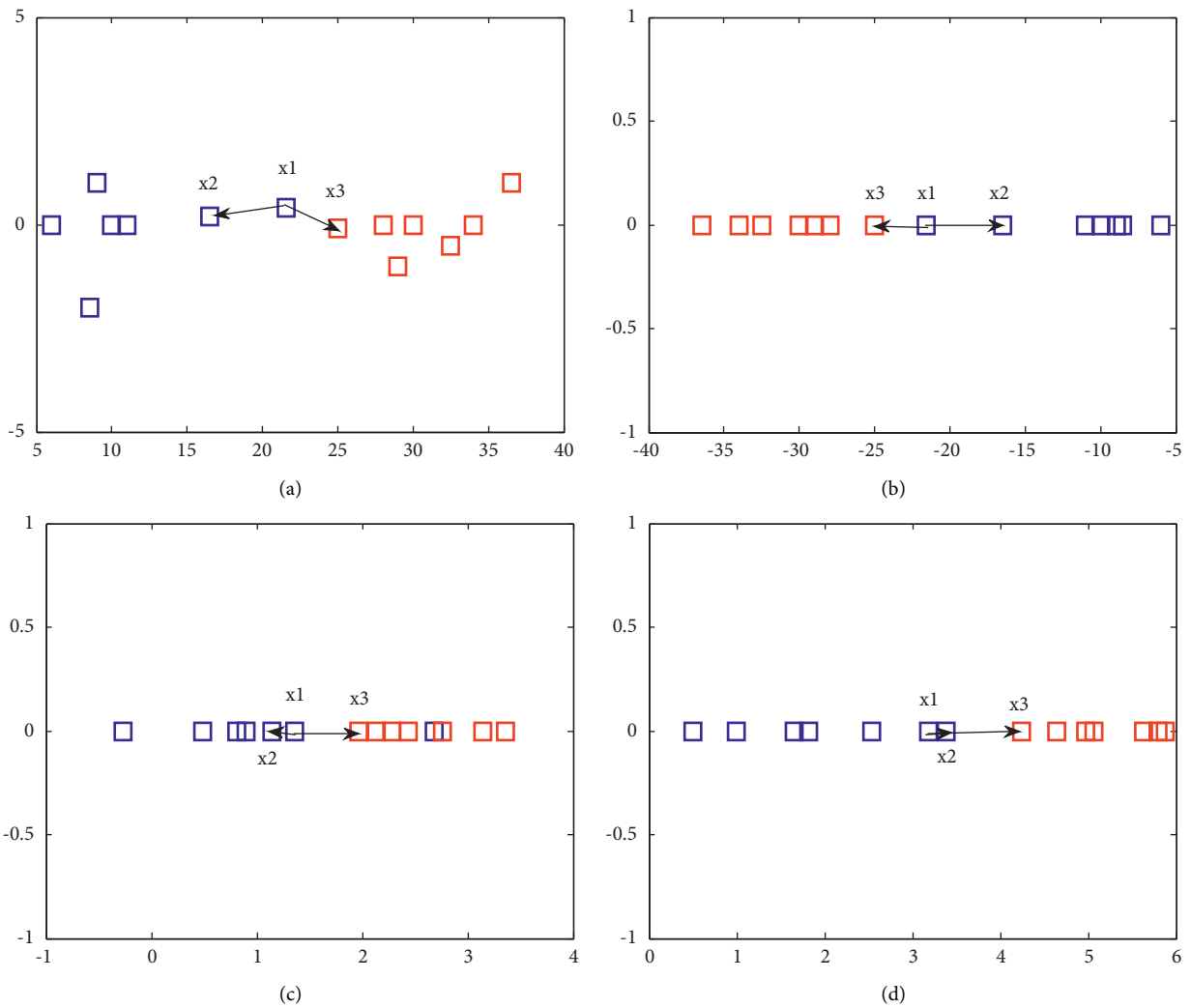
FIGURE 4: Continued.

(e)

(f)
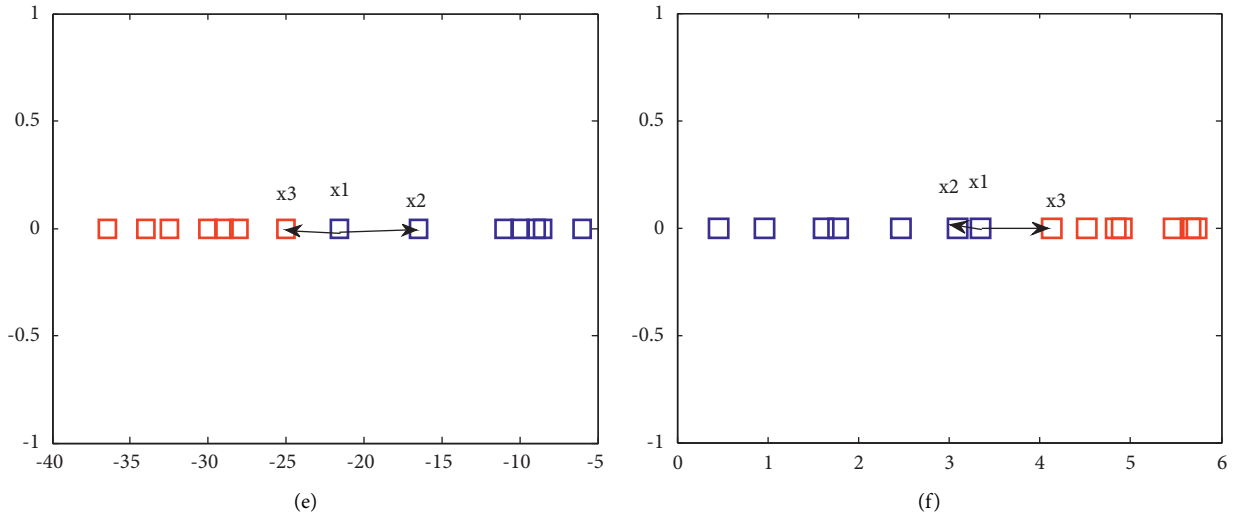
Figure 4: Illustration of embedding space and hypothesis margins: (a) Toy data ($H(x_1) = -0.83$)); (b) MFA ($k_1 = k_2 = 5$, $H(x_1) = -0.84$; (c) MNMDP ((k) = 5, $H(x_1) = 0.19$); (d) DNE ((k) = 5, $H(x_1) = 0.43$); (e) DAG-DNE ((k) = 5, $H(x_1) = -0.84$); (f) LMGE-DDR ((k) = $k_1 = k_2 = 3$, (a) = 0.8, $H(x_1) = 0.39$).

(a)

(b)

(c)

(d)

Figure 5: Continued.

(e)
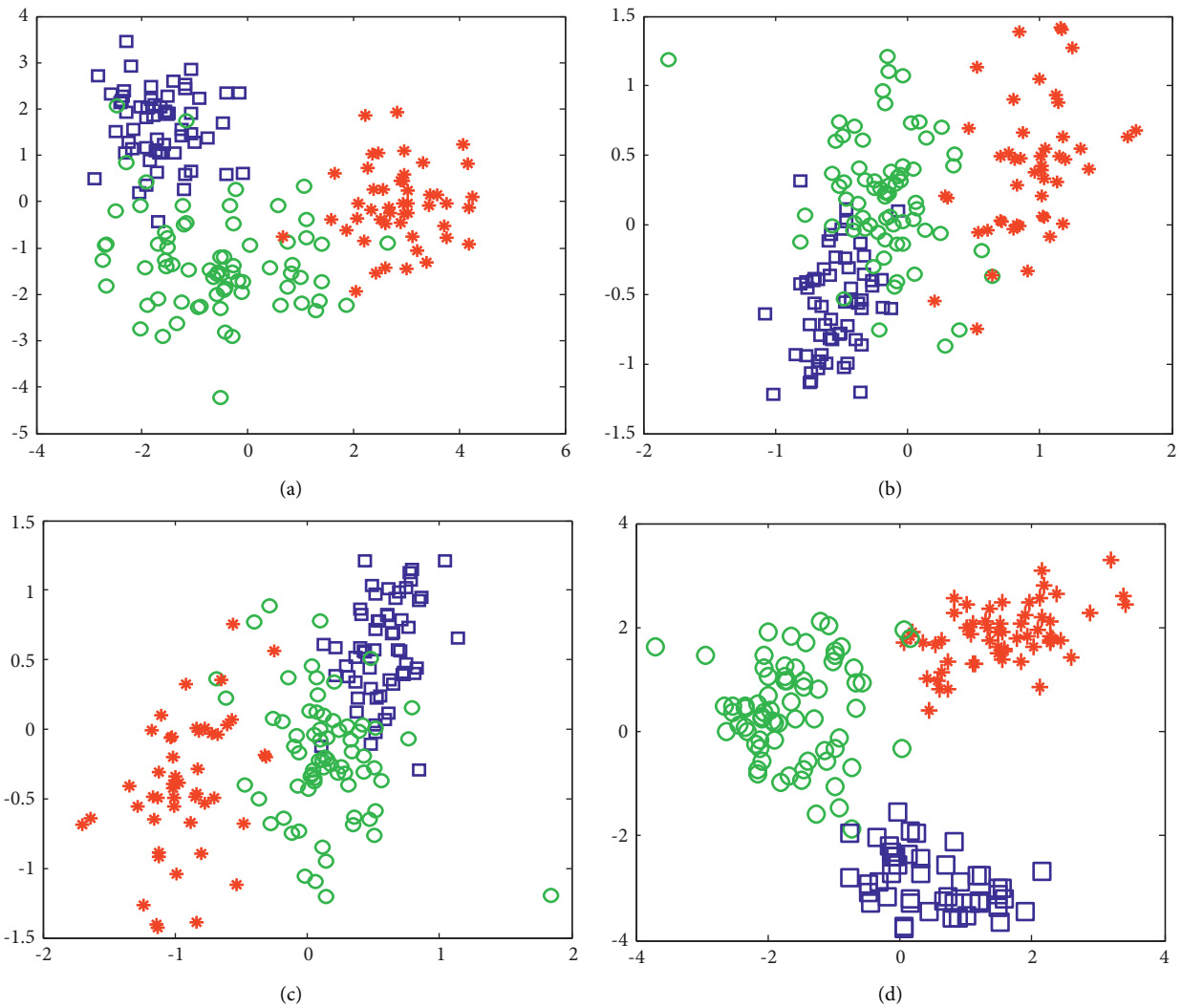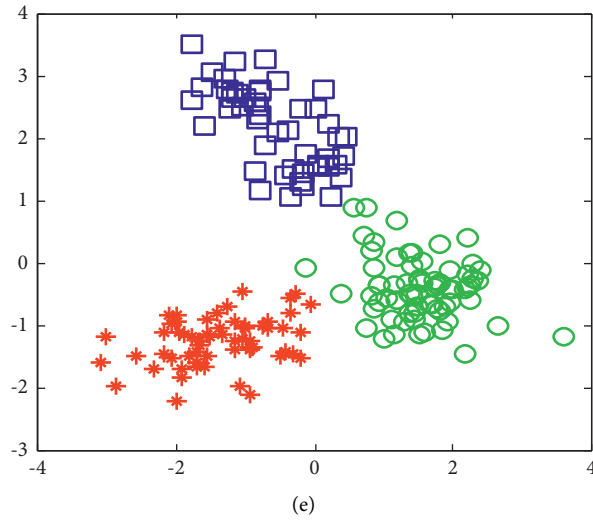
FIGURE 5: Illustration of visualization on wine dataset: (a) MFA ($k_1 = 15$, $k_2 = 25$); (b) DNE ((k) = 15); (c) MNMDP ((k) = 15); (d) DAG-DNE ((k) = 15); (e) LMGE-DDR ((k) = $k_1 = k_2 = 15$, $k_0 = 9$, (a) = 0.8).

(a)

(b)

FIGURE 6: Continued.

(c)

(d)
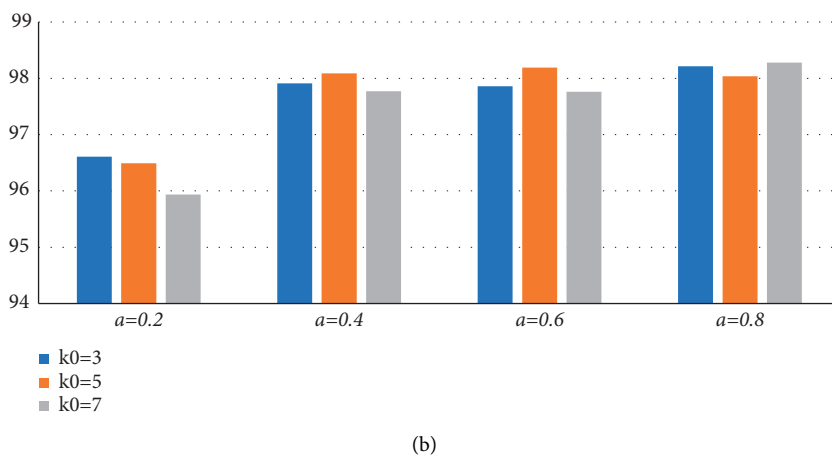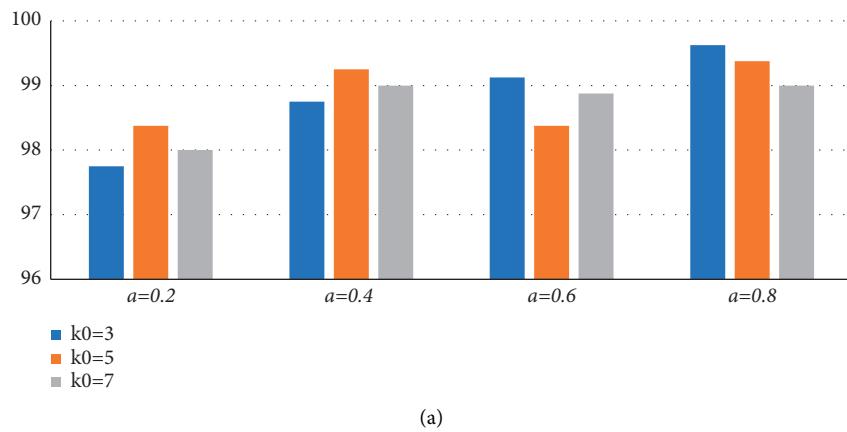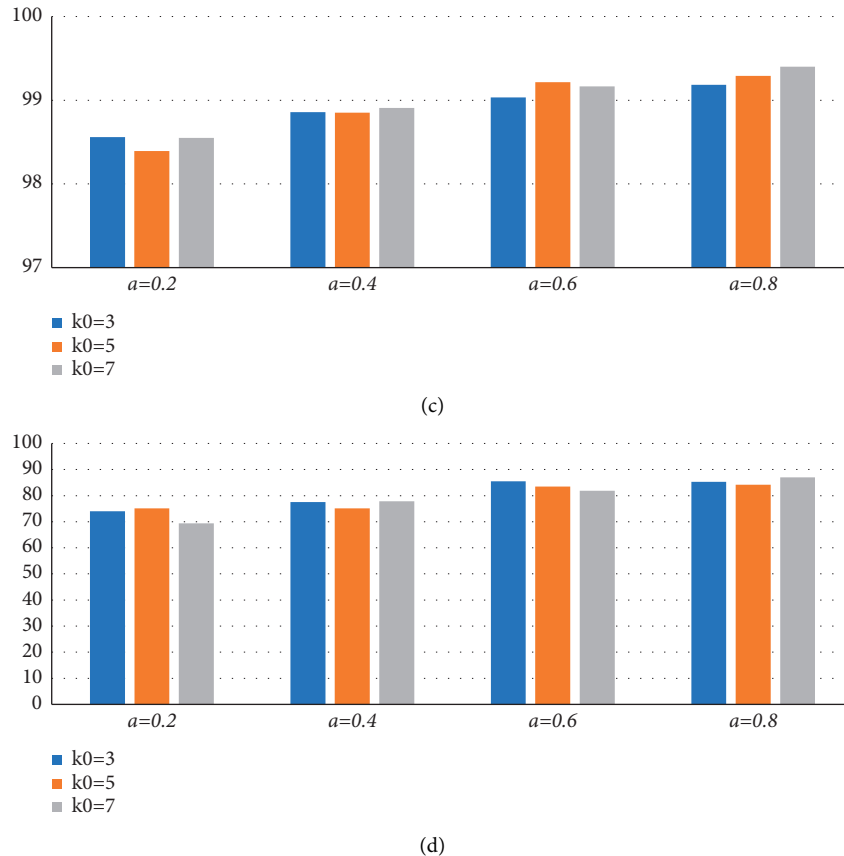
Figure 6: Recognition accuracy (%) of LMGE-DDR with variations of a, k0: (a) ORL (1 = 9); (b) UMIST (1 = 10); (c) AR (1 = 9); (d) Yale (1 = 9).

the dimensions by nearly 98% of the image energy. The 1NN classifier is applied to perform the classification. The average result of 20 runs is regarded as the classification result.

6.1. 2D Visualization.

Wine dataset is taken to perform the 2D visualization as shown in Figure 5 [41]; from Figure 5, it can be seen clearly that the sample points in the low-dimensional space learned by LMGE-DDR are separated compared to DAG-DNE.

6.2. Face Recognition.

LMGE-DDR is evaluated on the ORL (http://www.cad.zju.edu.cn/home/dengcai/Data/FaceData. html), FERET [42], AR [43], Yale†, UMIST (https://www. sheffield.ac.uk/eee/research/iel/research/face), and CMU-PIE09† face datasets to evaluate the classification performance and it has been systematically compared with several popular methods such as MFA, MNMDP, DNE, and DAG-DNE.

6.2.1. Parameter Analysis.

The sensitivity of parameter k_0, a in LMGE-DDR is analyzed on several face datasets when parameters k, k_1, and k_2 have been set as l-1. Figure 6 presents the best recognition rates of LMGE-DDR with the different values of k_0, a. The results in Figure 6 reveal that the recognition accuracy of LMGE-DDR fluctuates up and

down. In total, the best recognition accuracy can be achieved when a and k_0 are larger. The reason is that large a can make marginal samples tightly clustered toward the class center. The large k_0 is, the more marginal samples are. That is to say, heterogeneous near neighbors' margin of more marginal samples can be maximized and achieve large between-class margin, which is favorable for classification. Thus, the values of k_0 and a in LMGE-DDR on different datasets are adopted by cross-validation in face recognition experiments.

6.2.2. Experiments Results.

In this section, several experiments on public datasets have been conducted to verify the effectiveness of the proposed LMGE-DDR, such as ORL, Yale, UMIST, FERET, CMIU-PIE09, and AR, whose example images are shown in Figure 7. Each image in ORL is first aligned and cropped to 32 × 32. Each image in Yale is first aligned and cropped to 32 × 32. Each image in UMIST is first aligned and cropped to 40 × 50. All the images in FERET are cropped to 80 × 80. All the images in CMIU-PIE09 are cropped to 64 × 64. All the images in AR are cropped to 50 × 40. Tables 1–6 are the best recognition results on different datasets. Figure 8 are the recognition results on different dimensions.

As shown in Figure 8 and Tables 1–6, we can see that in most experiments, LMGE-DDR performs better than other
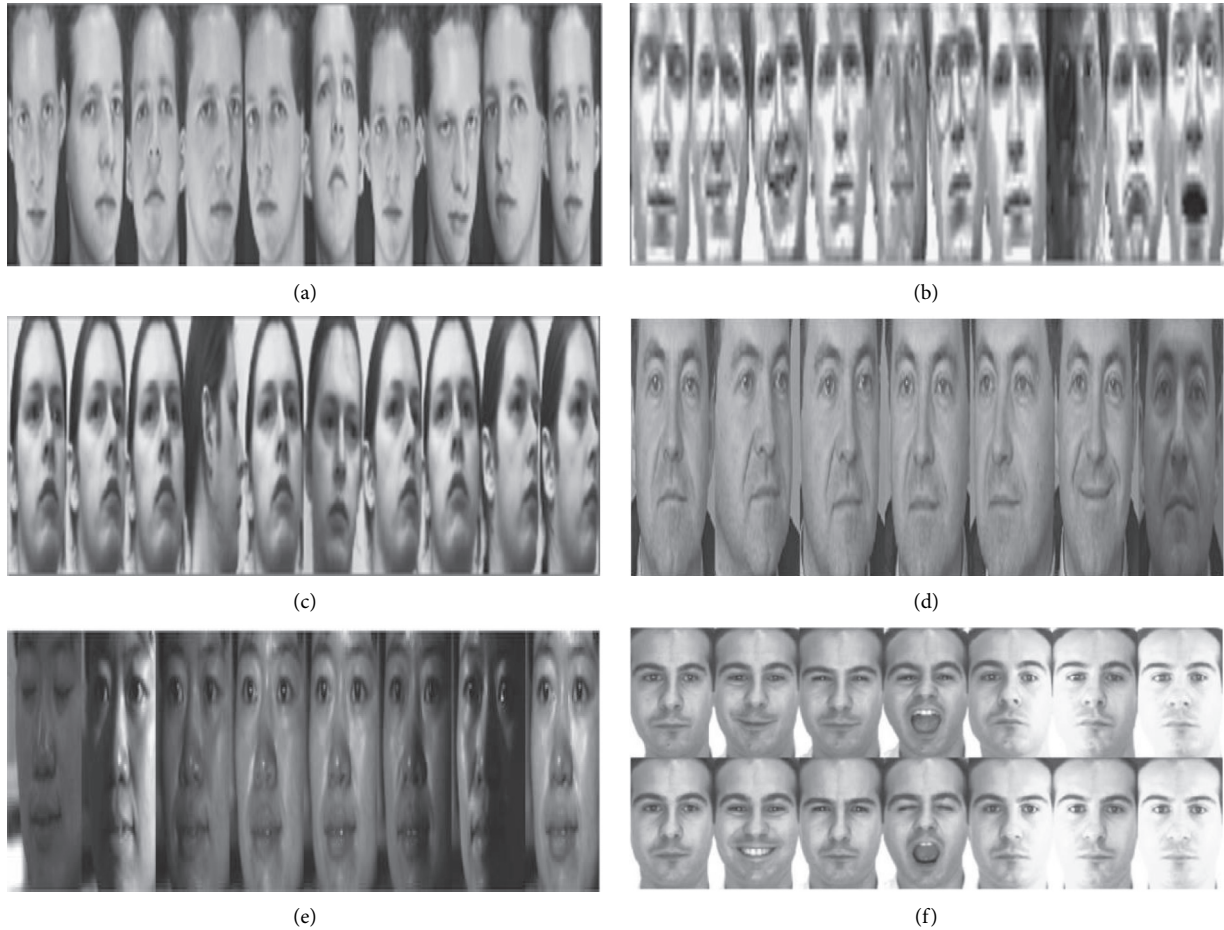
FIGURE 7: Illustration of some images on different datasets: (a) ORL; (b) Yale; (c) UMIST; (d) FERET; (e) CMU PIE; (f) AR.

TABLE 1: The classification performance on the ORL dataset (%).

Method	$l = 5$ ($k_0 = 3$, $a = 0.8$)	$l = 6$ ($k_0 = 3$, $a = 0.8$)	$l = 7$ ($k_0 = 5$, $a = 0.8$)
MFA	88.10 ± 2.36 (55)	91.19 ± 1.98 (61)	92.79 ± 1.78 (54)
DNE	89.08 ± 1.88 (39)	91.53 ± 1.21 (43)	93.46 ± 1.76 (39)
MNMDP	95.45 ± 1.51 (20)	96.01 ± 1.21 (27)	97.61 ± 1.11 (31)
DAG-DNE	96.67 ± 0.94 (40)	97.66 ± 0.97 (59)	98.29 ± 1.16 (34)
LMGE-DDR	$\mathbf{96.95 \pm 1.14}$ **(24)**	$\mathbf{98.03 \pm 0.89}$ **(22)**	$\mathbf{98.88 \pm 0.98}$ **(30)**

TABLE 2: The classification performance on the Yale dataset (%).

Method	$l = 7$ ($k_0 = 3$, $a = 0.8$)	$l = 8$ ($k_0 = 5$, $a = 0.8$)	$l = 9$ ($k_0 = 7$, $a = 0.8$)
MFA	62.34 ± 4.24 (45)	63.63 ± 5.61 (50)	68.23 ± 5.71 (45)
DNE	67.36 ± 4.17 (21)	69.23 ± 5.33 (26)	73.23 ± 5.63 (28)
MNMDP	71.23 ± 3.62 (26)	72.20 ± 5.39 (21)	76.23 ± 5.29 (21)
DAG-DNE	77.58 ± 4.76 (21)	77.87 ± 5.56 (18)	81.17 ± 5.26 (16)
LMGE-DDR	$\mathbf{83.17 \pm 3.21}$ **(40)**	$\mathbf{83.78 \pm 4.15}$ **(28)**	$\mathbf{86.17 \pm 4.55}$ **(22)**

compared methods and the corresponding standard deviation of LMGE-DDR is smaller than others.

6.2.3. Time Cost Analysis. In this section, the time cost of different methods is evaluated on several datasets including ORL, Yale, UMIST, and FERTET. In Table 7, it is computed by the time of running one time where $l = 5$, $d = 20$.

It can be concluded that LMGE-DDR is comparable with other methods in time cost; however, some perform better than others.

Table 3: The classification performance on the UMIST dataset (%).

Method	$l = 10(k_0 = 7, a = 0.8)$	$l = 11 (k_0 = 7, a = 0.8)$	$l = 12(k_0 = 7, a = 0.8)$
MFA	95.81 ± 1.42 (35)	96.54 ± 1.01 (40)	95.81 ± 1.42 (35)
DNE	97.45 ± 1.11 (12)	97.98 ± 1.17 (10)	97.45 ± 1.11 (12)
MNMDP	98.16 ± 0.84 (12)	98.52 ± 1.01 (10)	98.16 ± 0.84 (12)
DAG-DNE	98.43 ± 0.68 (26)	98.91 ± 0.76 (18)	99.13 ± 0.62 (24)
LMGE-DDR	**98.85 ± 0.51 (24)**	**99.04 ± 0.54 (22)**	**99.35 ± 0.59 (24)**

Table 4: The classification performance on the FERET dataset (%).

Method	$l = 4 (k_0 = 3, a = 0.8)$	$l = 5 (k_0 = 3, a = 0.8)$	$l = 6 (k_0 = 3, a = 0.8)$
MFA	36.68 ± 1.55 (60)	60.00 ± 2.32 (65)	44.53 ± 2.79 (60)
DNE	69.47 ± 1.43 (22)	56.58 ± 1.97 (12)	60.38 ± 2.57 (16)
MNMDP	99.03 ± 0.34 (76)	72.47 ± 1.66 (72)	77.33 ± 2.78 (42)
DAG-DNE	84.75 ± 1.28 (30)	87.88 ± 1.04 (36)	89.63 ± 1.76 (48)
LMGE-DDR	**88.05 ± 0.86 (32)**	**89.71 ± 0.67 (40)**	**90.73 ± 1.68 (36)**

Table 5: The classification performance (%) on the CMIU-PIE09 dataset.

Method	$l = 10 (k_0 = 5, a = 0.8)$	$l = 8 (k_0 = 5, a = 0.8)$	$l = 6 (k_0 = 5, a = 0.8)$
MFA	72.45 ± 2.67 (60)	63.47 ± 2.35 (60)	53.17 ± 1.65 (60)
DNE	72.60 ± 1.79 (60)	64.56 ± 1.81 (60)	54.89 ± 1.67 (60)
MNMDP	93.24 ± 0.89 (60)	91.11 ± 1.01 (60)	87.51 ± 0.97 (60)
DAG-DNE	93.92 ± 0.73 (56)	92.65 ± 1.02 (59)	90.56 ± 0.88 (59)
LMGE-DDR	**94.29 ± 0.61 (60)**	**93.31 ± 0.83 (52)**	**91.35 ± 0.83 (54)**

Table 6: The classification performance (%) on the AR dataset.

Method	$l = 5 (k_0 = 3, a = 0.8)$	$l = 7 (k_0 = 3, a = 0.8)$	$l = 9 (k_0 = 7, a = 0.8)$
MFA	76.59 ± 1.67 (200)	83.38 ± 1.19 (201)	87.15 ± 1.65 (196)
DNE	78.19 ± 1.58 (190)	86.13 ± 1.49 (194)	95.04 ± 1.15 (126)
MNMDP	97.55 ± 0.75 (72)	99.03 ± 0.34 (76)	**99.51 ± 0.19 (64)**
DAG-DNE	97.47 ± 0.84 (196)	98.81 ± 0.37 (132)	99.02 ± 0.39 (56)
LMGE-DDR	**97.81 ± 0.72 (60)**	**99.07 ± 0.22 (56)**	99.42 ± 0.27 (38)

(a)   (b)
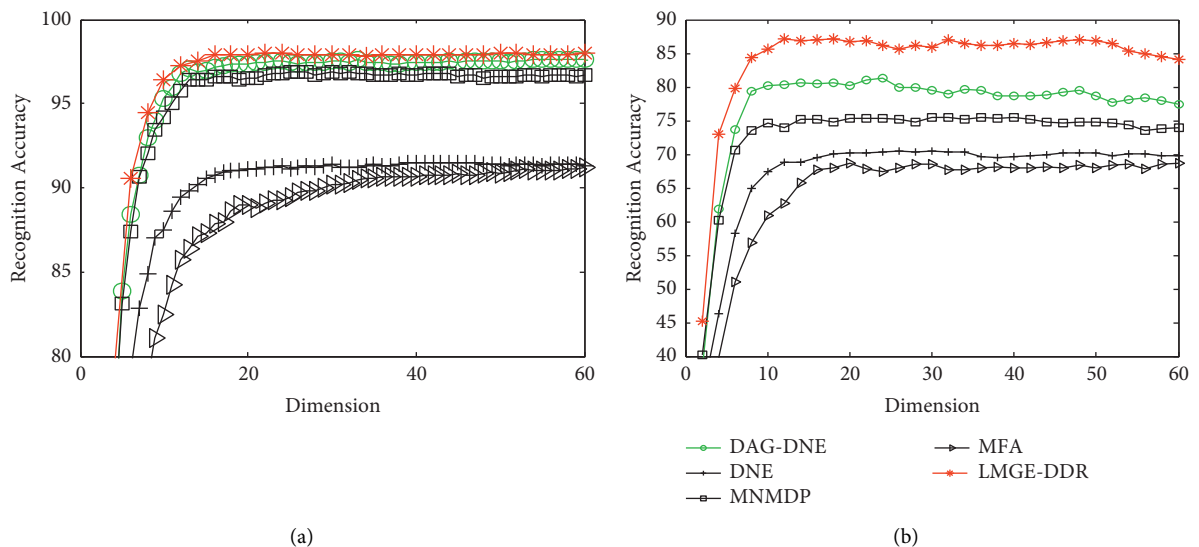
Figure 8: Continued.
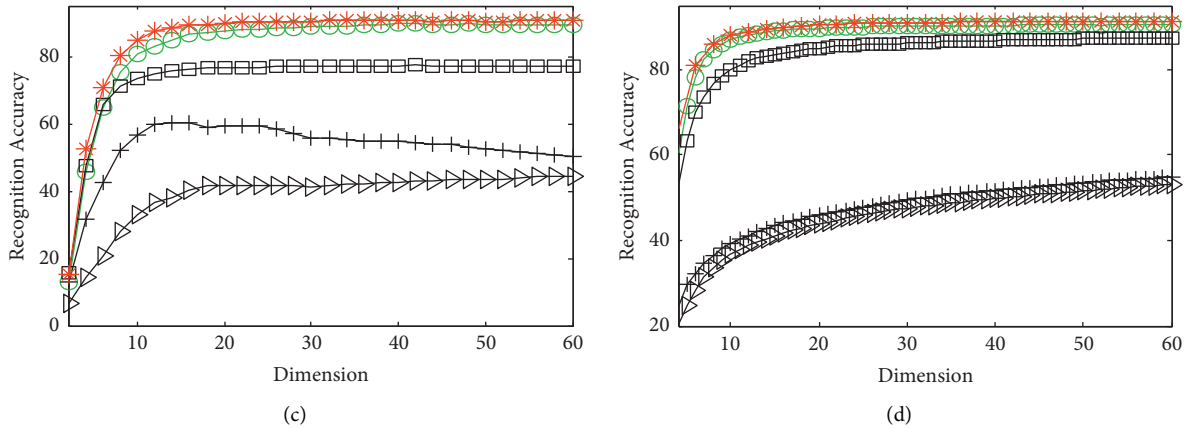
FIGURE 8: Recognition rate curves of different methods versus dimensions: (a) ORL ($l = 6$); (b) Yale ($l = 9$); (c) FERET ($l = 6$); (d) CMU PIE ($l = 6$).

TABLE 7: The time cost (seconds) on different datasets for different methods.

Method	ORL	Yale	UMIST	FERTET
LMGE-DDR	0.0529	0.0285	0.0471	0.6080
DAG-DNE	0.0448	0.0171	0.0325	0.6118
MNMDP	0.0495	0.0139	0.0428	0.7186
DNE	0.0613	0.0153	0.0491	0.9115
MFA	0.0950	0.0464	0.1410	1.6723

7. Conclusions and Future Works

In this study, we propose a novel graph embedding-based dimensionality reduction approach named LMGE-DDR, which is based on heterogeneous near neighbors' hypothesis margin. Different from other discriminant learning methods, for marginal samples in high-dimensional space, we additionally maximize the heterogeneous near neighbors' hypothesis margin to achieve a large between-class margin, excluding learning two kinds of adjacency graphs for each same equally This is very crucial for classification of the experiment results. Experimental results illustrate the effectiveness of LMGE-DDR. In this paper, we also employed several evaluation methods to evaluate the proposed model. The results show that on several public datasets such as ORL, Yale, UMIST, FERET, CMIU-PIE09, and AR, the proposed model outperformed other benchmark models. However, in constructing adjacency graphs and marginal samples, it will be influenced by the noise, which is not completely avoided. In the future works, how to evaluate the reliability of neighborhood will be studied by introducing an adaptive adjacency factor as in [44].

Data Availability

The experimental data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest regarding this work.

References

[1] M. A. Turk and A. P. Pentland, "Face recognition using eigenfaces," *Computer Vision Pattern Recognition*, vol. 27, no. 3, pp. 586–591, 1991.

[2] X. F. He, S. C. Yan, Y. X. Hu, P. Niyogi, and H. J. Zhang, "Face recognition using Laplacianfaces," *Pattern Anal. Mach. Intell.*vol. 27, pp. 328–340, 2005.

[3] Y. Y. Lin, T. L. Liu, and H. T. Chen, "Semantic manifold learning for image retrieval," *Proc. ACM Int. Conf. Multimedia*, pp. 249–258, 2005.

[4] X. F. He, D. Cai, and J. W. Han, "Learning a maximum margin subspace for image retrieval," *Knowl. Data Eng.*vol. 20, pp. 189–201, 2008.

[5] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of Neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.

[6] S. Deng, C. Wang, Z. Fu, and M. Wang, "An intelligent system for insider trading identification in Chinese security market," *Computational Economics*, vol. 57, no. 6, 2020.

[7] H. Cai, V. Zheng, and K. Chang, "A comprehensive survey of graph embedding: problems, techniques and applications," *IEEE Transactions on Knowledge and Data Engineering*, vol. 30, no. 9, pp. 1616–1637, 2017.

[8] P. Goyal and E. Ferrara, "Graph embedding techniques, applications, and performance: a survey," *Knowl-Based Syst*, vol. 151, pp. 78–94, 2018.

[9] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, pp. 2323–2326, 2000.

[10] J. B. Tenenbaum, V. de-Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, 2000.

[11] H. C. Qu, L. Li, Z. N. Li, and J. Zheng, "Supervised discriminant isomap with maximum margin graph regularization for dimensionality reduction," *Expert Systems with Applications*, vol. 180, no. 15, Article ID 115055, 2021.

[12] M. Belkin and P. Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation," *Neural Computation*, vol. 15, no. 6, pp. 1373–1396, 2003.

[13] Y. Bengio, J. Paiement, P. Vincent, O. Delalleau, N. L. Roux, and M. Ouimet, "Out-of sample extensions for LLE, isomap, MDS, eigen-maps, and spectral clustering," in *Proceedings of the International Conference Neural Information Processing Systems*, pp. 177–184, Whistler, Canada, December 2004.

[14] D. Lu, C. Ding, J. Xu, and S. Wang, "Hierarchical discriminant analysis," *Sensors*, vol. 18, no. 1, p. 279, 2018.

[15] X. Yang, G. Liu, Q. Yu, and R. Wang, "Stable and orthogonal local discriminant embedding using trace ratio criterion for dimensionality reduction," *Multimedia Tools and Applications*, vol. 77, no. 3, pp. 3071–3081, 2018.

[16] W. Yu, X. Teng, and C. Liu, "Face recognition using discriminant locality preserving projections," *Image and Vision Computing*, vol. 24, no. 3, pp. 239–248, 2006.

[17] X. He, D. Cai, S. Yan, and H. J. Zhang, "Neighborhood preserving embedding," in *Proceedings of the 2005, International Conference on Computer Vision*, vol. 2, pp. 1208–1213, ICCV, Cagliari, Italy, September 2005.

[18] X. F. He, D. Cai, S. C. Yan, and H. J. Zhang, "Neighborhood preserving embedding," in *Proceedings of the IEEE International Conferenece Computer Vision (ICCV'05)*, vol. 1, pp. 1208–1213, Beijing, China, October 2005.

[19] S. Yan, D. Xu, B. Zhang, H. J. Zhang, Q. Yang, and S. Lin, "Graph embedding and extensions: a general framework for dimensionality reduction," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 1, pp. 40–51, 2007.

[20] H. T. Chen, H. W. Chang, and T. L. Liu, "Local discriminant embedding and its variants," in *Proceedings of the International Conference Computer Vision and Pattern Recognition*, San Diego, CA, USA, June 2005.

[21] W. Zhang, X. Y. Xue, H. Lu, and Y. F. Guo, "Discriminant neighborhood embedding for classification," *Pattern Recognition*, vol. 39, pp. 2240–2243, 2006.

[22] C. T. Ding and L. Zhang, "Double adjacency graphs-based discriminant neighborhood embedding," *Pattern Recognition*, vol. 48, pp. 1734–1742, 2015.

[23] Q. You, N. Zheng, S. Du, and Y. Wu, "Neighborhood discriminant projection for face recognition," *Pattern Recognition Letters*, 2007.

[24] M. Wan, Z. Lou, and Z. Jin, "Margin maximum embedding discriminant (MMED) for feature extraction and classification," in *Proceedings of the IEEE Chinese Conf. Pattern Recognition (CCPR)*, pp. 1–5, Nanjing, China, November 2009.

[25] M. Sugiyama, "Dimensionality reduction of multimodal labeled data by local Fisher discriminant analysis," *Journal of Machine Learning Research*, vol. 8, pp. 1027–1061, 2007.

[26] D. Cai, X. He, K. Zhou, J. Han, and H. Bao, "Locality sensitive discriminant analysis," in *Proceedings of the Joint Conference Artificial Intelligence*, pp. 1713–1726, Gold Coast, Australia, December 2007.

[27] P. Huang, C. Chen, Z. Tang, and Z. Yang, "Feature extraction using local structure preserving discriminant analysis," *Neurocomputing*, vol. 140, pp. 104–113, 2014.

[28] J. P. Gou, Y. Z. Zhan, M. Wan, X. Shen, J. Chen, and L. Du, "Maximum neighborhood margin discriminant projection for classification," *Science World Journal*, vol. 16, 2014.

[29] J. Gou, Y. Yang, Z. Yi, J. Lv, and Y. Zhan, "Discriminative globality and locality preserving graph embedding for dimensionality reduction," *Expert Systems with Applications*, vol. 144, Article ID 113079, 2019.

[30] S. Wang, C. Ding, C. H. Hsu, and F. Yang, "Dimensionality reduction via preserving local information," *Future Generation Computer Systems*, vol. 108, no. 22, 2018.

[31] J. P. Gou and Z. Yi, "Locality-based discriminant neighborhood embedding," *The Computer Journal*, vol. 9, pp. 1063–1082, 2013.

[32] H. S. Hu, D. Z. Feng, and Q. Y. Chen, "A novel dimensionality reduction method: similarity order preserving discriminant analysis," *Signal Processing*, vol. 182, Article ID 107933, 2020.

[33] Z. Liu, K. Shi, K. Zhang, W. Ou, and L. Wang, "Discriminative sparse embedding based on adaptive graph for dimension reduction," *Engineering Applications of Artificial Intelligence*, vol. 94, Article ID 103758, 2020.

[34] F. Luo, B. Du, L. Zhang, L. Zhang, and D. Tao, "Feature learning using spatial-spectral hypergraph discriminant analysis for hyperspectral image," *IEEE Transactions on Cybernetics*, vol. 49, pp. 2406–2419, 2018.

[35] G. Zhao and Y. Wu, "Efficient large margin-based feature extraction," *Neural Processing Letters*, vol. 50, no. 12, 2019.

[36] R. Duda, P. Hart, and D. Stork, *Pattern Classification*, John Wiley & Sons, Hoboken, NJ, USA, 2nd edition, 2000.

[37] R. G. Bachrach, A. Navot, and N. Tishby, "Margin based feature selection-Theory and algorithms," in *Proceedings of the Of the 21st International Conference on Machine Learning*, pp. 43–50, ACM Press, Banff, Canda, July 2004.

[38] K. Crammer, R. Gilad-Bachrach, A. Navot, and N. Tishby, "Margin analysis of the lvq algorithm," in *Advances in Neural Information Processing Systems*, MIT Press, Cambridge, MA, USA, 2003.

[39] G. D. Zhao, Z. Y. Zhou, and J. M. Zhang, "Theoretical framework in graph embedding-based discriminant dimensionality reduction," *Signal Processing*, vol. 189, Article ID 108289, 2021.

[40] J. Yang, D. Zhang, J. Yang, and B. Niu, "Globally maximizing, locally minimizing: unsupervised discriminant projection with applications to face and palm biometrics," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 4, pp. 650–664, 2007.

[41] Wine Database: http://www.resset.cn/databases(Accessed on 2021-7-01).

[42] P. J. Phillips, "The facial recognition technology (FERET) database," 2004, http://www.itl.nist.gov/iad/humanid/feret/feret_master.html>.

[43] Q. Gao, J. Liu, H. Zhang, X. Gao, and K. Li, "Joint global and local structure discriminant analysis," *IEEE Transactions on Information Forensics and Security*, vol. 8, no. 4, pp. 626–635, 2013.

[44] Y. Gao, Y. Zhang, J. Pan, S. Luo, and C. Yeng, "Discriminant analysis based on reliability of local neighborhood," *Expert Systems with Applications*, vol. 175, Article ID 114790, 2021.

Research Article

Trajectory Tracking Method of Volleyball Player's Arm Hitting Image Based on D-P Algorithm

Hao Qin (ID)

Taiji Martial Arts Institute of Jiaozuo University, Jiaozuo, Henan 454000, China

Correspondence should be addressed to Hao Qin; zq04021@jzsz.edu.cn

Aiming at the problems of poor image tracking effect, low precision, and long time in the process of image tracking of volleyball player's arm hitting, a volleyball player's arm hitting image tracking method based on D-P algorithm is proposed. This paper analyzes the basic concept, basic principle, and basic equation of D-P algorithm and collects the arm stroke trajectory image of volleyball players under the three-dimensional visual model. Using wavelet multiscale decomposition method, the arm stroke trajectory of volleyball players is filtered, and the edge contour feature points of the arm stroke image of volleyball players are extracted. Using the gray histogram feature extraction method, the gray information of volleyball player's arm hitting trajectory image is enhanced. Combined with pixel adaptive enhancement technology, the key action feature points of volleyball player's arm hitting image trajectory are located. Based on D-P algorithm, the volleyball player's arm hitting image trajectory is adjusted and modified to realize the correct tracking of volleyball player's arm hitting image trajectory. The experimental results show that the trajectory tracking effect of volleyball player's arm hitting image is better, which can effectively improve the tracking accuracy and shorten the tracking time.

1. Introduction

At present, with the rapid development of computer technology, image processing technology, and image recognition technology, human motion trajectory recognition technology has been widely used in different industries such as entertainment and user interface, and its application in the field of sports training is becoming more and more extensive [1–3]. Applying image processing technology to sports training is the key to improving the correction ability of sports movements and the effect of sports training [4]. In volleyball competition, whether the athlete's arm movement is scientific and reasonable determines the quality of spiking. The volleyball hitting movement is complex and the technical difficulty is high. It is necessary to carry out real-time analysis and normative correction of the technical movement to improve the ability of sports planning [5]. In order to master the spiking technique better, it is necessary to track the arm trajectory of volleyball players. Volleyball arm stroke is the key to scoring. Tracking the image trajectory of

volleyball arm stroke, combined with image processing technology for trajectory analysis, is of great significance to improve the accuracy of volleyball arm stroke [6, 7]. In this case, how to effectively recognize the arm trajectory of volleyball players has become the main problem to be solved in this field.

At present, scholars in related fields have made some progress in motion trajectory tracking. Piao and Kim [8] proposed a trajectory tracking method based on a backward semi-Lagrangian method. The backward semi-Lagrangian method is used to track the trajectory required to solve the guidance center model. The discrete system numerical solution of Cauchy problem is designed, and the required trigger point is found by interpolation method. Based on the estimated value of the improved physical quantity, the interpolation solution is calculated at the starting point to solve the trajectory tracking required by the guidance center model. This method has certain reliability. Zhang et al. [9] proposed a semiglobal finite time trajectory tracking method for disturbed nonlinear systems based on high-order sliding

mode. Through the high-order sliding mode integral finite time disturbance feedforward decoupling process, a non-recursive design framework is proposed to simplify the gain adjustment mechanism. An inherently nonsmooth control law is constructed from the system information. By proposing a semiglobal tracking control target, the trajectory tracking without restrictive nonlinear growth constraints is realized. This method is simple and effective. However, the above methods have the problems of poor trajectory tracking effect, low accuracy, and long time.

Aiming at the above problems, a trajectory tracking method of volleyball player's arm hitting image based on D-P algorithm is proposed. The edge contour feature points of the trajectory are extracted by filtering the arm stroke trajectory image of volleyball players. The gray information of the track image is enhanced to locate the key action feature points of the track. Based on D-P algorithm, the trajectory of volleyball player's arm hitting image is adjusted and corrected to realize correct trajectory tracking. The tracking effect of this method is good, which can effectively improve the tracking accuracy and shorten the tracking time.

2. Basic Theory of D-P Algorithm

2.1. Basic Concepts of D-P Algorithm. Dynamic programming (D-P) algorithm is an optimization method applied to solve multistage decision-making problems [10–12]. At present, D-P algorithm has been widely used in resource theory, chemical engineering, variational method, economics, and optimal control theory. It is used to solve some optimal problems. This kind of problem refers to such a kind of activity, which can be divided into several interrelated stages, and decisions need to be made in each stage. This decision requires that the whole process achieve the best activity effect [13, 14]. Here, the selection of decision-making in each stage is not determined arbitrarily. Its determination depends on the current state, and it affects its subsequent state. Assuming that it is divided into 10 stages, when the decisions in these 10 stages are determined, these decisions form a decision sequence. An activity route of the whole process can be determined through the decision sequence, which solves this problem. The multistage decision-making problem is shown in Figure 1.

As can be seen from Figure 1, these interrelated stages can facilitate the solution of the problem. In the face of different processes, the problem can be divided into different stages, which is not invariable. Usually a variable is used to describe the stage. This variable is called the stage variable. In most cases, this variable is discrete, denoted by k. The first stage is from point A to point B, the second stage is from point B to point C, the third stage is from point C to point D, and the fourth stage is from point D to point E. Point A in Figure 1 is usually called a state, and the variables used to describe each state are called state variables [15]. This state variable does not have to be a number; it can also be a set of numbers, and even a vector can be used to describe the state. In order to facilitate the understanding of the state variable in the k stage when solving the problem, it is often
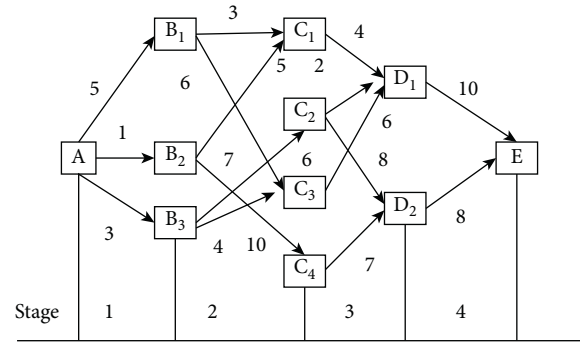
Figure 1: Multistage decision-making problem.

represented by x_k. One state of the first stage is represented by a, and three states of the second stage are represented by $B1$, $B2$, and $B3$. Similarly, $C1$, $C2$, $C3$, and $C4$ represent the state of stage 4, respectively, and $D1$, $D2$, and e represent the state of stages 4 and 5, respectively. When the state of a certain stage is known, it is necessary to choose from this state of this stage to the state of the next stage, and this choice is a decision. For ease of understanding, $u_k(x_k)$ is usually used to represent the decision variable, where x_k represents the k state of the x stage. The same as state variables, a number, a group of numbers, or even a vector can also be used as decision variables to describe decision-making. The strategy is for decision-making. The strategy can be said to be a group of decision sequences at various stages. In most problems, the transition from one state to another state is not random. It follows a certain criterion, which is called the transfer equation [16]. According to the state transition equation, when the state variable x_k of the k stage and its decision variable u_k are known, state x_{k+1} of the $k + 1$ stage can be obtained through calculation.

2.2. Basic Principle of D-P Algorithm. The basic idea of D-P algorithm is to find the point on the curve with the maximum distance from the connecting line between the points at both ends of the curve according to a certain tracing route and judge whether the maximum distance is greater than the given threshold. The principle of D-P algorithm is shown in Figure 2.

AB is the given curve. In order to find the inflection point on the curve, traverse the curve AB from a to B, calculate the distance from the point on the curve to the connecting line of AB, and find out the maximum value of the distance, which is obviously point $C1$. If the distance is greater than the set threshold, the curve is divided into two sections with $C1$ as the boundary, that is, curves AC1 and C1B. In these two sections of curves, the inflection points are found according to the same method as above. When setting the appropriate maximum distance threshold, two inflection points, that is, $C1$ and $C2$, can be found. It can be seen from the search process that the determined inflection points do not appear in sequence according to the contour curve. Because the feature points do not appear in sequence according to the contour curve, after obtaining the feature points, in order to fit the curve, the feature points should be
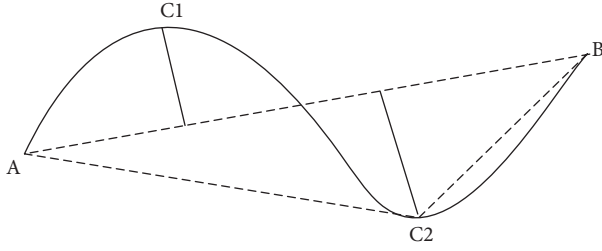
Figure 2: D-P algorithm principle.

stored in the form of binary tree, and each feature point should be obtained in the first order when storing and in the middle order when using.

(1) Optimization principle: the optimal solution of a multilevel decision problem must include the optimal solution of its subproblems. Decision-making is the choice made at each stage, and multilevel selection constitutes a decision strategy [17]. In the shortest path problem in Figure 1, the shortest path from A to E and its subpath from the intermediate

node to the end point e must also be the shortest path from the current intermediate node to the end point E. In this problem, decision-making is the selection of the next location at each point from point A. The current location is called the state. The path from the selection made in each state to the final point is a decision strategy.

(2) State transition principle: D-P algorithm cannot be directly applied to the target state, so the premise of realizing D-P algorithm is to discretize the target state space and change the continuous state into discrete state [18–20]. At this time, the target space will be divided into units of size $\Delta \times \Delta$. Similarly, the speed space of the target will also be divided into units of $\Delta x \times \Delta y$, where $\Delta x = \Delta y = \Delta v$ defines 1 resolution unit/frame. In order to distinguish it from the continuous state $(x, \dot{x}, y, \dot{y})$, the discrete state is represented as $(\overline{x}, \overline{\dot{x}}, \overline{y}, \overline{\dot{y}})$. For state $(\overline{x}, \overline{\dot{x}}, \overline{y}, \overline{\dot{y}})$ of the k frame, the relationship between the $k + 1$ frame and the k frame is expressed as

$$\left\{ x(k+1) \in [(\overline{\dot{x}} + \overline{x} - 2)\Delta, (\overline{\dot{x}} + \overline{x})\Delta], \dot{x}(k+1) \in [(\overline{\dot{x}} - 1)\Delta v, \overline{\dot{x}}\Delta v], y(k+1) \in [(\overline{\dot{y}} + \overline{y} - 2)\Delta, (\overline{\dot{y}} + \overline{y})\Delta], \dot{y}(k+1) \in [(\overline{\dot{y}} - 1)\Delta v, \overline{\dot{y}}\Delta v]. \right. \tag{1}$$

When the target is maneuvering, the number of state transitions q will be severely affected. Increased target mobility will lead to an increase in state transitions q, because only by increasing the number of state transitions can the expansion of the target state brought about by changes in target speed be satisfied. Assuming that each state at time k is known as $(\overline{x}, \overline{\dot{x}}, \overline{y}, \overline{\dot{y}})$, there are q possible valid states $x(k + 1)$ corresponding to it at time $k + 1$.

(3) Reverse order recurrence principle: reverse order recurrence is the specific implementation method of D-P algorithm to solve the optimal strategy; that is, it deduces forward from the final state and records the feasible decisions and corresponding costs in the previous stage for the next decision to evaluate the overall decision cost [21]. In this way, the final feasible decision set is obtained by recursion in reverse order until the initial state, and the optimal strategy is selected according to the recorded cost value, and then the whole decision order is traced back. In the shortest path problem, the overall cost is the path length, and the decision cost is the distance weight between two nodes.

The idea of using D-P algorithm to solve multistage decision-making problem is to regard a shortest path of the final required solution as an optimal strategy of multilevel decision-making [22, 23]. The D-P algorithm divides the multilevel decision-making problem into a combination of multiple subproblems and then obtains

the optimal strategy through the optimality principle and inverse recurrence principle of the D-P algorithm. The D-P algorithm has the following advantages:

(1) It can reduce the amount of calculation: Using D-P algorithm to solve this problem requires 31 addition times and 19 comparison times, which greatly reduces the amount of calculation. When the number of segments is large and the selection of multiple segments is large, the advantage of D-P algorithm is more prominent; and D-P algorithm is convenient to use mainframe computer to calculate the iterative method.

(2) The calculation results are rich: Using D-P algorithm for calculation, we can not only get the shortest path and corresponding distance from A to E but also get the shortest path and corresponding distance from any point to E. It can be seen that the whole optimal decision can be obtained based on D-P algorithm, so D-P algorithm is an effective method to solve the multistage decision-making problem.

2.3. Basic Equation of D-P Algorithm. The D-P algorithm problem can be boiled down to selecting a K stage strategy $\{u(x_1), u(x_2), \ldots, u(x_K)\}$ so that the value function reaches the optimal expression as

$$f_k(x_k) = \max_{\{u_1, \ldots, u_k\}} v(x_1; u_1, \ldots, u_k). \tag{2}$$

In formula (2), $f_k(x_k)$ represents the optimal value function of the entire strategy [24]. The setting function can be expressed as

$$v(x_1; u_1, u_2, \ldots, u_k) = \sum_{j=1}^{k} w_j(x_j, u_j). \qquad (3)$$

In formula (3), $w_j(x_j, u_j)$ represents the cost of the j stage. The optimality principle of the D-P algorithm has the following relationship:

$$f_k(x_k) = \max_{\{u_k\} \in U} [w_k(x_k, u_k) + f_{k-1}(x_{k-1}), \quad k = 2, 3, \ldots, M]. \qquad (4)$$

The initial cost is assumed to be

$$f_1(x_1) = w_1(x_1, u_1). \qquad (5)$$

Formula (4) and formula (5) are the equations used to calculate the strategy value function, which are called the basic equations of D-P algorithm. The recursive form of formula (6) can be obtained by combining the first two formulas:

$$f_k(x_k) = \max_{\{u_k\} \in U} [h_k(x_k)]. \qquad (6)$$

In formula (6), $h_k(x_k)$ represents the stage value function, and there is

$$h_k(x_k) = w_k(x_k, u_k) + \max_{\{u_{k-1}\} \in U} [h_{k-1}(x_{k-1})]. \qquad (7)$$

The initial condition is assumed to be

$$h_1(x_1) = w_1(x_1, u_1). \qquad (8)$$

Formula (7) is the general form of the basic equation of D-P algorithm, and the optimal strategy can be expressed as

$$\{\widehat{x}_k\} = \left\{ \widehat{x}_k : \left(\sum_{k=1}^{K} h_k(x_k) \right)_{\max} \right\}. \qquad (9)$$

In formula (9), $\widehat{x}_k = \{\widehat{x}(1), \widehat{x}(2), \ldots, \widehat{x}(K)\}$ is the state corresponding to each stage of decision-making in the optimal strategy recorded by the system in the recursive process.

3. Tracking Method of Arm Hitting Image Trajectory of Volleyball Players

3.1. Collect the Arm Stroke Trajectory Image of Volleyball Players. In order to realize the volleyball player's arm stroke image trajectory tracking based on the D-P algorithm, firstly, it is necessary to collect the digital features of the volleyball player's arm stroke trajectory, use the digital imaging equipment to track the volleyball player's arm stroke trajectory visually, and combine the computer three-dimensional vision acquisition method [25, 26] for image acquisition and feature analysis.

The computer three-dimensional vision acquisition frame difference that defines the volleyball player's arm shot

trajectory is $g\prime$, the discrete sampling rate is $\nabla_x = [1, -1]$, and the block pheromone of the edge contour of the single frame volleyball player's arm shot trajectory tracking image is

$$\begin{cases} \min & F(x) = [f_1(x), f_2(x), \ldots, f_m(x)]^T, \\ & g\prime \leq 0, i = 1, 2, \ldots, q, \\ s.t. & h_j = 0, j = 1, 2, \ldots, p. \end{cases} \qquad (10)$$

The pheromone obtained by formula (10) is combined into an image, and in the imaging sequence acquisition, it is assumed that $\nabla_x = [1, -1]^T$ is the high-frequency part $y = [\nabla_x g\prime, \nabla_y g\prime]$ of the tracking image of the volleyball player's arm hitting the ball. In the computer three-dimensional imaging space, using the spatial invariant feature decomposition method [27, 28], the obtained binary image of the volleyball player's arm shot trajectory is

$$\min_{x,k} \lambda \|x \otimes k - y\|_2^2 + \frac{\|x\|_1}{\|x\|_2} + \beta \|k\|_1. \qquad (11)$$

According to the binary image results, the texture information transmission model of action image in volleyball player's arm stroke trajectory tracking is constructed, which is described as follows:

$$c(x, y) = \sum_W [I(x_i, y_i) - I(x_i + \Delta x, y_i + \Delta y)]^2. \qquad (12)$$

In formula (12), $(\Delta x, \Delta y)$ is the probability density function of the position distribution of the volleyball player's arm shot trajectory image in the air, and (x_i, y_i) is the coordinate point of the volleyball player's arm shot trajectory tracking.

According to the visual information feature transmission structure, the pixel space of the above coordinate points is reconstructed, and the volleyball background image B and the foreground image I are scaled and information is fused to obtain the noise distribution model of the volleyball player's arm hitting trajectory. The volleyball player's arm hitting trajectory distribution module is divided into $(W/2) \times (H/2)$ subblocks, and the image information fusion tracking equation is expressed as

$$\begin{aligned} \dot{x} &= V \cos \theta \cos \phi_V, \\ \dot{y} &= V \sin \theta, \\ \dot{z} &= -V \cos \theta \sin \phi_V, \\ \dot{\theta} &= \omega_y \sin \gamma + \omega_z \cos \gamma. \end{aligned} \qquad (13)$$

In formula (13), x, y, z is the distribution feature quantity of the local information feature points of the volleyball player's arm hitting trajectory. Through the above steps, the volleyball player's arm shot trajectory image acquisition is thus realized.

3.2. Pretreatment of Volleyball Player's Arm Hitting Trajectory. Based on the image acquisition of volleyball player's arm stroke trajectory, the wavelet multiscale decomposition method is used to filter the volleyball player's

arm stroke trajectory. Blind noise separation is carried out in the imaging area of the image, and the wavelet analysis method is used to obtain the multicolor difference kernel matrix of volleyball player's arm hitting trajectory tracking, which is marked as [29–31]

$$
\begin{cases}
f(x_1, x_2) = r_1 x_1 \left(1 - \dfrac{x_1}{N_1} - \sigma_1 \dfrac{x_2}{N_2}\right) = 0, \\[4mm]
g(x_1, x_2) = r_2 x_2 \left(1 - \sigma_2 \dfrac{x_1}{N_1} - \dfrac{x_2}{N_2}\right) = 0.
\end{cases}
\tag{14}
$$

$$
\mathbf{G}_{m,n} = \begin{bmatrix} g_{(m,n)}(1,1) & g_{(m,n)}(1,2) \\ g_{(m,n)}(2,1) & g_{(m,n)}(2,2) \end{bmatrix},
\tag{15}
$$

We have

$$
g_{(m,n)}(u, v) = I_{(k)_g}[2(m-1) + u, 2(n-1) + v].
\tag{16}
$$

In formula (16), $u \in \{1, 2\}$; $v \in \{1, 2\}$ represents the scale of the block fusion of the volleyball player's arm shot trajectory. The wavelet multiscale decomposition method is used to filter the volleyball player's arm shot trajectory, and the one-dimensional wavelet transform is used to obtain the regular feature distribution pixel set of the volleyball player's arm shot trajectory as [35, 36]

$$
x_i(t) = \left[\omega_{i1}^{lk}, \dots, \omega_{in}^{lk}\right][x_1(t-k), \dots, x_n(t-k)]^T.
\tag{17}
$$

In formula (17), $[\omega_{i1}^{lk}, \dots, \omega_{in}^{lk}]$ is the surface block feature quantity of the volleyball player's arm hitting trajectory.

Through the above steps, the volleyball player's arm shot trajectory filtering process is thus realized.

3.3. Extracting the Trajectory Feature of the Volleyball Player's Arm Shot Image.

Based on the wavelet multiscale decomposition method for filtering the arm stroke trajectory of volleyball players, the arm stroke trajectory of volleyball players is tracked. This paper proposes a trajectory tracking method of volleyball player's arm hitting image based on D-P algorithm, extracts the edge contour feature points of volleyball player's arm hitting image trajectory, and gives the mother wavelet function of volleyball player's arm hitting image trajectory filtering as follows:

$$
H(z) = \frac{P_1 \sum_{K=1}^{H} p_{(k)_g}(i, j)}{P_2 \sum_{K=1}^{H} p_{(k)*g}(i, j)}.
\tag{18}
$$

For k adjacent points, the graph decomposition of volleyball player's arm hitting image trajectory is carried out in the basis function of mother wavelet. Taking pixel (θ^e, p^e) as the parent feature point of the trajectory of the volleyball player's arm hitting image based on wavelet transform, the multiscale wavelet decomposition method is used to dynamically filter the trajectory of the volleyball player's arm

In formula (14), r_1 represents the state correlation estimated feature value of the volleyball player's arm trajectory, r_2 represents the correlation coefficient, σ_1 represents the feature matching degree, and N_1 is the linear component of the volleyball player's arm trajectory. Set h as the edge pixel set of volleyball player's arm hitting trajectory, use the adaptive block feature matching method to segment the image contour of volleyball player's arm hitting trajectory [32–34], divide the chromatographic image into $M \times N$ subblocks $\mathbf{G}_{m,n}$ of 2×2, and obtain the feature distribution matrix of volleyball player's arm hitting trajectory tracking as follows:

$$
m = 1, 2, \dots, M; \; n = 1, 2, \dots, N.
$$

hitting image, and the gray pixel set of the volleyball player's arm hitting image trajectory is obtained:

$$
\sigma(Z; D_x) = \sum_{i > j} \left| d_{ij}(Z) - d_X(x_i, x_j) \right|^2.
\tag{19}
$$

In formula (19), $d_{ij}(Z)$ is the Euclidean distance of the boundary pixels of the volleyball player's arm shot image trajectory, and $d_X(x_i, x_j)$ is the active contour component of the local area of the image in the 4×4 subgrid. The edge contour feature points of the volleyball player's arm hitting image trajectory are extracted, and the gray histogram feature extraction method is used to enhance the gray information of the volleyball player's arm hitting trajectory image. The center pixel set and edge pixel set of the volleyball player's arm hitting image trajectory in $N \times N$ window ω_i are expressed as follows:

$$
\begin{aligned}
I_{if}(x, y) &= I * G(x, y, \sigma_i), \\
I_{iv}(x, y) &= I * \text{stdfilt}(x, y, \omega_i), \\
S_{gif}(x, y) &= -\log\left(P_{if}(x, y)\right).
\end{aligned}
\tag{20}
$$

In formula (20), $G(x, y, \sigma_i)$ represents the similarity feature. Combined with wavelet multiscale decomposition method for multiscale feature optimization, the boundary geometric feature of volleyball player's arm hitting image trajectory tracking can be obtained, so as to realize the feature extraction of volleyball player's arm hitting image trajectory.

3.4. Tracking the Image Track of Volleyball Player's Arm Hitting.

According to the above extracted trajectory characteristics of volleyball player's arm hitting image, the trajectory algorithm of volleyball player's arm hitting image is designed. σ_0 is set as the edge information scale of the volleyball player's arm shot image trajectory, and using the RGB decomposition method, the pixel feature component $J(x)$ of the volleyball player's arm hitting image trajectory is obtained as follows:

$$J(x) = \frac{I(x) - A}{\max(t(x), t_0)} + A * \sigma_0. \tag{21}$$

In formula (21), t_0 represents the edge autocorrelation feature amount of the volleyball player's arm shot image trajectory, and $t(x)$ represents the edge autocorrelation feature function. Combined with the pixel adaptive enhancement technology, the key action feature points of the volleyball player's arm shot image trajectory are located, and the output is

$$F = \widetilde{p}(x, y) = p(x, y)\left(\frac{v(x)}{y(x)}\right)^{1/2}. \tag{22}$$

The D-P algorithm is used to track the trajectory of the volleyball player's arm hitting image. The optimization process of the D-P algorithm is as follows:

$$E_{\text{int}(v_i)} = \frac{S_c}{2}\left(\alpha_i \big| d - |v_i - v_{i-1}|\big|^2 + \beta_i \big| v_{i-1} + 2v_i + v_{i+1}\big|^2\right). \tag{23}$$

We have

$$d = \frac{1}{n}\sum_{i=0}^{n-1}|v_i - v_{i-1}|. \tag{24}$$

The D-P algorithm is used for block fusion processing of volleyball player's arm stroke trajectory image, and the output volleyball player's arm stroke image trajectory tracking map is as follows:

$$\omega(i, j) = \frac{E_{\text{int}(v_i)}}{f_R(z)Z(i)}\exp\left(-\frac{d(i, j)}{h^2}\right). \tag{25}$$

In formula (25), $Z(i)$ represents the symbolic distance function extracted from the volleyball player's arm shot image trajectory feature. Through the above analysis, the D-P algorithm is used to adjust and correct the error of the tracked volleyball player's arm hitting image trajectory, so as to realize the correct tracking of the volleyball player's arm hitting image trajectory. The implementation process of the algorithm is shown in Figure 3.

Through the above steps, the volleyball player's arm stroke trajectory image is collected under the three-dimensional visual model, the volleyball player's arm stroke trajectory is filtered by wavelet multiscale decomposition method, and the edge contour feature points of the volleyball player's arm stroke image trajectory are extracted. The gray histogram feature extraction method is used to enhance the gray information of the volleyball player's arm hitting trajectory image. Combined with the block feature matching technology, the key action feature points of the volleyball player's arm hitting trajectory image are located.
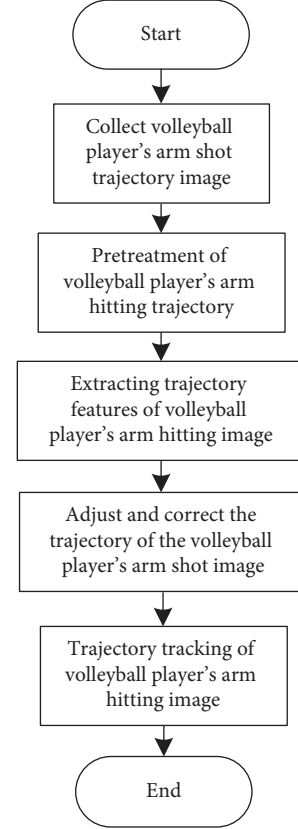
Figure 3: Flow chart of algorithm implementation.

D-P algorithm is used to adjust and modify the image trajectory of volleyball player's arm stroke, so as to realize the correct tracking of volleyball player's arm stroke image trajectory.

4. Experimental Simulation and Analysis

4.1. Setting the Experimental Environment. In order to verify the effectiveness of the trajectory tracking method of volleyball player's arm hitting image based on D-P algorithm, a simulation experiment is carried out in MATLAB. The volleyball player's arm shot trajectory image acquisition and scanning frequency is 16 kHz, and functions such as receive in MATLAB are used to collect the volleyball player's arm shot trajectory image. The feature matching of the volleyball player's arm hitting trajectory is carried out in a 5×5 block mode, and the three-dimensional visual collection sample set of volleyball is 1000 images. According to the above simulation environment and parameter settings, the simulation experiment of volleyball player's arm stroke image trajectory tracking is carried out. Firstly, the computer three-dimensional vision acquisition

Figure 4: Acquisition results of arm hitting image of original volleyball player.

Figure 5: Trajectory tracking effect of volleyball player's arm hitting image based on the proposed method.

of volleyball player's arm stroke trajectory image is carried out, and the original image acquisition results are given in Figure 4.

4.2. Trajectory Tracking Effect of Volleyball Player's Arm Hitting Image. In order to verify the trajectory tracking effect of volleyball player's arm hitting image, the trajectory tracking of volleyball player's arm hitting image is realized. It is necessary to filter the acquisition results of the original volleyball player's arm hitting image, extract the trajectory features of the volleyball player's arm hitting image, and locate the key action feature points. Using the proposed method, the trajectory tracking of volleyball player's arm hitting image is realized, and the trajectory tracking effect of volleyball player's arm hitting image of the proposed method is shown in Figure 5.

According to Figure 5, the proposed method can effectively track the volleyball player's arm hitting image trajectory, capture the volleyball player's arm hitting image trajectory, and locate the key action feature points of the volleyball player's arm hitting image trajectory. It can be seen that the trajectory tracking of volleyball player's arm hitting image based on the proposed method is better.

4.3. Trajectory Tracking Accuracy of Volleyball Player's Arm Hitting Image. In order to further verify the trajectory tracking accuracy of the volleyball player's arm hitting image of the proposed method, the root mean square error (RMSE) of position is taken as the evaluation index. The smaller the RMSE value, the higher the tracking accuracy of the method. The calculation method is as follows:

$$
\mathrm{RMSE} = \frac{1}{CK} \sum_{c=1}^{C} \sum_{k=1}^{K} \sqrt{\left(\tilde{p}_{x,k}^{c} - p_{x,k}^{c}\right)^2 + \left(\tilde{p}_{y,k}^{c} - p_{y,k}^{c}\right)^2}.
$$

(26)

In formula (26), C represents the number of experiments, K represents the number of processed frames, and $\tilde{p}_{x,k}^{c}$ and $\tilde{p}_{y,k}^{c}$ represent the coordinates of the target in the k

frame of the c experiment. Using the method of [8], the methods of [9], and the proposed method, the arm stroke image trajectory of volleyball players is tracked, and the comparison results of tracking accuracy of arm stroke image trajectory of volleyball players with different methods are given in Figure 6.

It can be seen from Figure 6 that with the gradual increase of the number of experiments, the RMSE value of arm stroke image trajectory tracking of volleyball players with different methods increases. When the number of experiments is 40, the RMSE value of volleyball player's arm stroke image trajectory tracking in the method of reference [8] is 0.46, the RMSE value of volleyball player's arm stroke image trajectory tracking in the method of reference [9] is 0.57, while the RMSE value of volleyball player's arm stroke image trajectory tracking in the proposed method is only 0.2. Therefore, the RMSE value of the proposed method is small, which can effectively improve the trajectory tracking accuracy of the volleyball player's arm hitting image.

4.4. Trajectory Tracking Time of Volleyball Player's Arm Hitting Image. On this basis, the trajectory tracking time of volleyball player's arm hitting image of the proposed method is verified. The trajectory of volleyball player's arm hitting image is tracked by using the method of [8], the method of [9], and the proposed method, respectively. The comparison results of trajectory tracking time of volleyball player's arm hitting image of different methods are given in Figure 7.

It can be seen from Figure 7 that, with the increasing number of experiments, the trajectory tracking time of arm hitting image of volleyball players with different methods increases. When the number of experiments reaches 40, the trajectory tracking time of volleyball player's arm stroke image in the method of [8] is 43.2 s, and the trajectory tracking time of volleyball player's arm stroke image in the method of [9] is 55.6 s, while the trajectory tracking time of volleyball player's arm stroke image in the proposed method is only 24.3 s. It can be seen that the trajectory tracking time of volleyball player's arm hitting image is short.
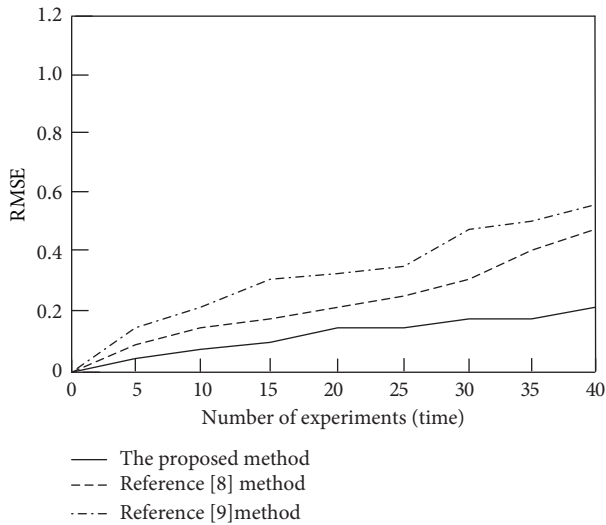
FIGURE 6: Comparison results of trajectory tracking accuracy of volleyball player's arm shot images with different methods.
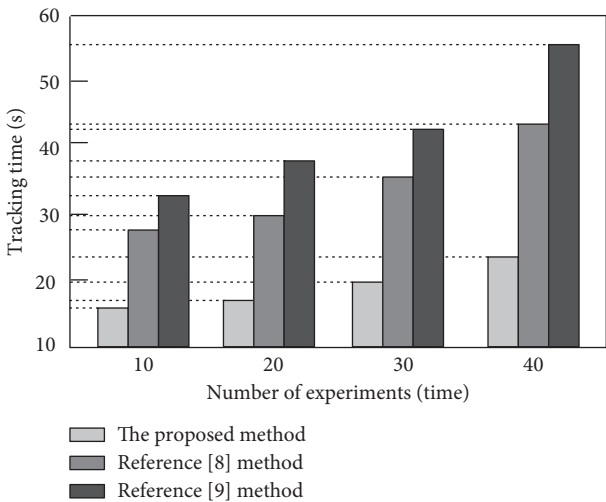
FIGURE 7: Comparison results of volleyball player's arm shot image trajectory tracking time with different methods.

5. Conclusion

With the continuous innovation of technical and tactical level of competitive volleyball and the continuous improvement of computer performance, higher requirements are put forward for badminton players. By studying the role and trajectory of volleyball players' arms in the hitting process of volleyball, on the one hand, it can provide scientific and effective exercise methods for volleyball lovers; on the other hand, it is helpful to the improvement and development of volleyball players' technical and tactical theory. Therefore, the image trajectory tracking method of volleyball player's arm hitting based on D-P algorithm is studied in this paper. The effectiveness of this method is verified by experiments; that is, it can improve the trajectory tracking accuracy of volleyball player's arm hitting image and shorten the

trajectory tracking time of volleyball player's arm hitting image.

However, the complexity of algorithm and background and the influence of noise on the tracking effect are not considered in the process of tracking the trajectory of volleyball player's arm hitting image. Therefore, in the next research, we can take Gaussian white noise as the background and expand the algorithm to further verify the trajectory tracking effect of the proposed method.

Data Availability

The raw data supporting the conclusions of this article will be made available by the corresponding author, without undue reservation.

Conflicts of Interest

The author declares that there are no conflicts of interest regarding this work.

References

[1] C. Huang and L. Jiang, "Data monitoring and sports injury prediction model based on embedded system and machine learning algorithm," *Microprocessors and Microsystems*, vol. 81, Article ID 103654, 2021.

[2] L. P. Bass, Y. A. Plastinin, and I. Y. Skryabysheva, "Machine learning in problems involved in processing satellite images," *Measurement Techniques*, vol. 63, no. 12, pp. 950–958, 2021.

[3] G. Song, Q. Dai, X. Han, and L. Guo, "Two novel ELM-based stacking deep models focused on image recognition," *Applied Intelligence*, vol. 50, no. 5, pp. 1345–1366, 2020.

[4] L. Bargsten and A. Schlaefer, "SpeckleGAN: a generative adversarial network with an adaptive speckle layer to augment limited training data for ultrasound image processing," *International Journal of Computer Assisted Radiology and Surgery*, vol. 15, no. 9, pp. 1427–1436, 2020.

[5] T. Yang, "Optimization of volleyball motion estimation algorithm based on machine vision and wearable devices," *Microprocessors and Microsystems*, vol. 81, no. 4, Article ID 103750, 2020.

[6] C. Zhang, H. Tang, and Z. Duan, "WITHDRAWN: time series analysis of volleyball spiking posture based on quality-guided cyclic neural network," *Journal of Visual Communication and Image Representation*, vol. 13, Article ID 102681, 2019.

[7] X. Jiang, "Human tracking of track and field athletes based on FPGA and computer vision," *Microprocessors and Microsystems*, vol. 83, no. 4, Article ID 104020, 2021.

[8] X. Piao and P. Kim, "An efficient trajectory tracking algorithm for the backward semi-Lagrangian method of solving the guiding center problems," *Journal of Computational Physics*, vol. 418, Article ID 109664, 2020.

[9] C. Zhang, J. Yang, Y. Yan, L. Fridman, and S. Li, "Semiglobal finite-time trajectory tracking realization for disturbed nonlinear systems via higher-order sliding modes," *IEEE Transactions on Automatic Control*, vol. 65, no. 5, pp. 2185–2191, 2020.

[10] X. Pan, Y. Wang, and K. C. Chin, "Dynamic programming algorithm-based picture fuzzy clustering approach and its application to the large-scale group decision-making problem," *Computers & Industrial Engineering*, vol. 157, Article ID 107330, 2021.

[11] L. Hutchison, "Pika parsing: reformulating packrat parsing as a dynamic programming algorithm solves the left recursion and error recovery problems," 2020, https://arxiv.org/abs/2005.06444.

[12] R. Bürgy, A. Hertz, and P. Baptiste, "An exact dynamic programming algorithm for the precedence-constrained class sequencing problem," *Computers & Operations Research*, vol. 124, Article ID 105063, 2020.

[13] F. Miretti, D. Misul, and E. Spessa, "DynaProg: deterministic Dynamic Programming solver for finite horizon multi-stage decision problems," *Software*, vol. 14, Article ID 100690, 2021.

[14] J. Hinz, T. Tarnopolskaya, and J. Yee, "Efficient algorithms of pathwise dynamic programming for decision optimization in mining operations," *Annals of Operations Research*, vol. 286, no. 1/2, pp. 583–615, 2020.

[15] L. Meng and X. Zhou, "An integrated train service plan optimization model with variable demand: a team-based scheduling approach with dual cost information in a layered network," *Transportation Research Part B: Methodological*, vol. 125, pp. 1–28, 2019.

[16] A. V. Smirnov, M. V. Klibanov, and L. H. Nguyen, "On an inverse source problem for the full radiative transfer equation with incomplete data," *Siam Journal on entific Computing*, vol. 41, no. 5, pp. B929–B952, 2019.

[17] Y. Chen, W. Zhou, and W. Zhang, "Analysis of the best driving decision of electric vehicles based on dynamic programming," *Control and Information Technology*, vol. 6, pp. 1–7, 2019.

[18] Y. Liu, H. Zhang, R. Yu, and Z. Xing, "H∞ tracking control of discrete-time system with delays via data-based adaptive dynamic programming," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 50, no. 11, pp. 4078–4085, 2020.

[19] X. He, K. Cai, W. Song, and Z. Mei, "Dynamic programming for sequential deterministic quantization of discrete Memoryless Channels," 2019, https://arxiv.org/abs/1901.01659.

[20] H. Fu, X. Chen, W. Wang, and M. Wu, "MRAC for unknown discrete-time nonlinear systems based on supervised neural dynamic programming," *Neurocomputing*, vol. 384, no. 12, pp. 130–141, 2020.

[21] H. Hu, J. Yuan, and V. Nian, "Development of a multi-objective decision-making method to evaluate correlated decarbonization measures under uncertainty - the example of international shipping," *Transport Policy*, vol. 82, pp. 148–157, 2019.

[22] S. Zhang and X. A. Sun, "Stochastic dual dynamic programming for multistage stochastic mixed-integer nonlinear optimization," 2019, https://arxiv.org/abs/1912.13278.

[23] T. Trzaskalik, "Multiobjective dynamic programming in bipolar multistage method," *Annals of Operations Research*, vol. 15, no. 2, pp. 1–12, 2021.

[24] G. Weisz, P. Amortila, B. Janzer, Y. Abbasi-Yadkori, Y. Jiang, and C. Szepesvári, "On query-efficient planning in MDPs under linear realizability of the optimal state-value function," 2021, https://arxiv.org/abs/2102.02049.

[25] J. L. Escalona, "A methodology for the measurement of track geometry based on computer vision and inertial sensors," 2020, https://arxiv.org/abs/2008.03763.

[26] A. Hvps and B. Qhm, "ViDAQ: a computer vision based remote data acquisition system for reading multi-dial gauges," *Journal of Industrial Information Integration*, vol. 15, pp. 29–41, 2019.

[27] M. A. Asghar, M. J. Khan, M. Rizwan et al., "AI inspired EEG-based spatial feature selection method using multivariate empirical mode decomposition for emotion classification," *Multimedia Systems*, vol. 21, no. 4, pp. 1–14, 2021.

[28] D. Hong, X. Wu, P. Ghamisi, J. Chanussot, N. Yokoya, and X. X. Zhu: Invariant Attribute Profiles: a Spatial-Frequency Joint Feature Extractor for Hyperspectral Image Classification.

[29] X. Zheng, Y. Y. Tang, and J. Zhou, "A framework of adaptive multiscale wavelet decomposition for signals on undirected graphs," *IEEE Transactions on Signal Processing*, vol. 67, no. 7, pp. 1696–1711, 2019.

[30] U. A. Nnolim, "Single image de-hazing via multiscale wavelet decomposition and estimation with fractional gradient-anisotropic diffusion fusion," *International Journal of Image and Graphics*, vol. 21, no. 3, Article ID 2150032, 2021.

[31] G. Xie, Y. Qian, and H. Yang, "Forecasting container throughput based on wavelet transforms within a decomposition-ensemble methodology: a case study of China," *Maritime Policy and Management*, vol. 46, no. 1-2, pp. 178–200, 2019.

[32] F. Zhang, Y. Gao, and L. Xu, "An adaptive image feature matching method using mixed Vocabulary-KD tree," *Multimedia Tools and Applications*, vol. 79, no. 2, pp. 16421–16439, 2020.

[33] S. Zheng, Y. Wang, and X. Li, "A hardware-adaptive deep feature matching pipeline for real-time 3D reconstruction," *Computer-Aided Design*, vol. 132, Article ID 102984, 2021.

[34] P. R. Vadamala and A. F. Aklak, "Adaptive patch feature matching and scale estimation for visual object tracking," *Journal of Electronic Imaging*, vol. 28, no. 3, pp. 033037.1–033037.11, 2019.

[35] C. Liu and M. Pang, "One-dimensional image surface blur algorithm based on wavelet transform and bilateral filtering," *Multimedia Tools and Applications*, pp. 1–15, 2021.

[36] F. Tian and Z. Sun, "On the significance detection method of compound wavelet transform for visual image," *Computer Simulation*, vol. 37, no. 11, pp. 427–430 + 469, 2020.

Research Article

Art Visual Image Transmission Method Based on Cartesian Genetic Programming

Jing Zhao (ID)

School of Arts and Design, NanJing Vocational University of Industry Technology, Nanjing 210046, China

Correspondence should be addressed to Jing Zhao; 2015100827@niit.edu.cn

Because most of the traditional artistic visual image communication methods use the form of modeling and calculation, there are some problems such as long image processing time, low success rate of image visual communication, and poor visual effect. An artistic visual image communication method based on Cartesian genetic programming is proposed. The visual expression sensitivity difference method is introduced to process the image data, the neural network is used to identify the characteristics of the artistic visual image, the midpoint displacement method is used to remove the folds of the artistic visual image, and the processed image is formed under the above three links. The Cartesian genetic programming algorithm is used to encode the preprocessed image, improve the fitness function, select the algorithm to improve the operation, design the image rendering platform, input the processed image to the platform, and complete the artistic visual image transmission. The analysis of the experimental results shows that the image processing time of this method is short, the success rate of visual communication is high, and the image visual effect is good, which can obtain the image processing results satisfactory to users.

1. Introduction

In recent years, the worldwide informatization process has been accelerating, and the development of all kinds of data has shown a soaring trend. As a key type of data in all kinds of information, image data can more intuitively record the state of human life. The subsequent technical problems of data analysis have gradually attracted people's attention, which has promoted the development of image visual communication research [1–3]. Most of the traditional art visual image communication methods use the form of modeling and calculation. Although this method can model quickly, the modeling effect of art visual image is not very ideal and the visual image effect is blurred, which is a kind of damage to the art image itself. Therefore, it is necessary to design an effective art visual image communication method. In order to solve this problem, relevant scholars have proposed some optimization methods [4–6].

Reference [7] proposes a plane visual communication effect optimization method based on wavelet change, which decomposes the image and reconstructs the wavelet. In the

process of reconstruction, the modulus diagram and phase angle diagram are calculated, and the edge images of each scale are extracted. Through the edge image, the corresponding edge points of the semireconstructed image are enhanced. On the above basis, the graphic beautification vector of slip model is used for operation. In order to simplify the operation, the above operation method is transformed into simple mathematical operation, and the visual communication effect is optimized through the reflected light graphic mode. The experimental results show that the proposed method effectively improves the visual communication effect of the image, makes the color of the image more vivid and rich, and reduces the optimization cost, but there is a problem of long image processing time. Reference [8] proposes a plane visual communication design method based on user experience effect. Computer vision imaging technology is used to sample the image information of plane visual communication and combined with the edge contour extraction method to detect the image contour of plane visual communication design, so as to extract the multiscale local structure feature information of plane visual

image. According to the needs of user experience effect, the boundary feature detection and multilevel structure decomposition in the process of plane visual communication design are carried out, the low-level visual structure of plane visual communication design is reconstructed by the method of adjacent pixel information fusion, and the user experience effect evaluation model of plane visual design image is established. According to the user experience effect, realize the optimization of plane visual communication design. The simulation test results show that the user experience effect of plane visual communication design using this method is good and improves the design effect of plane visual communication, but there are some problems such as blurred image edge and poor visual effect. Reference [9] designed a visual communication system design of animated character graphics and images in virtual reality environment. In the hardware design, the adaptation parameters of the renderer motherboard were designed to optimize the experience of visual communication. In the software design, by introducing Sobel edge operator, the gray function is established to solve the gradient amplitude, and the threshold is selected to compare it to complete the recognition and thinning of the edge data of animated character graphics and images. Design the motion capture module, establish the behavior control model, generate and manage the motion capture files, and complete the overall design of the system. The experimental comparison shows that the motion of the animated characters constructed by the designed system is relatively consistent, which verifies the effectiveness of the design system, but the success rate of visual communication is low.

In order to improve the visual effect of artistic visual image, improve image processing efficiency, and improve the success rate of visual communication, this paper proposes an artistic visual image communication method based on Cartesian genetic programming. Cartesian genetic programming is different from other algorithms in evolutionary art. Its works can generally be used directly and have high market value. Therefore, compared with other algorithms, Cartesian genetic programming is an art design algorithm with high artistic value and market potential, which has been successfully used in the field of image design such as wallpaper design. This paper optimizes the visual communication effect of artistic visual image with the help of this algorithm, so as to improve the user experience effect and satisfaction.

2. Art Visual Image Preprocessing

2.1. Artistic Visual Image Data Processing. Before visual communication design, the artistic visual image is preprocessed first. In this paper, the imported visual expression sensitivity difference method is used to process image data. Firstly, the data information needs to be effectively calculated. The formula is

$$F_i = (x + y + z) \times \Delta x. \tag{1}$$

In the formula, F_i represents the validity, and generally the validity is a range value; x, y, and z, respectively, represent the digitization of the length, width, and height of the image data. Through the validity calculation, data stability analysis can be performed, and the stabilized data can be converted again for easy use [10]; the formula is

$$K_i = F_i \times (v_2 - v_1) - \Delta x (a + b + c). \tag{2}$$

In the formula, a, b, and c all represent the maximum value of the effective data function. After the conversion of the formula, the matching effect can be directly calculated.

Each image data has a specific visual expression matching value. The matching value represents the corresponding attribute feature, and the attribute feature coefficient is

$$E = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{m} f_i w_j s_h. \tag{3}$$

In the formula, f_i represents the integration coefficient; w_j represents the attribute extraction coefficient; n represents the deployment function; m represents the order change function. In order to facilitate the positioning of pixels, the selected attributes need to be matched and converted:

$$h_{ij} = v_{ij} \times [\tau_n \Delta f_i (k - k_1)]. \tag{4}$$

In the formula, v_{ij} represents the coefficient value of the sensitivity and τ_n represents the matching conversion coefficient. Through the restriction of conditions, the value of each pixel with a specific attribute can be measured, and the difference calculation can be carried out after the measurement. The formula is

$$r_{ik} = \sum_{i=1}^{n} \tau_n \Delta f_{ik} \times \mu_x \mu_y. \tag{5}$$

In the formula, Δf_{ik} represents the image distortion coefficient.

The difference calculation can calculate the pixel value of each frame, and the pixel filling is performed through the difference calculation to ensure the effect of the artistic visual image [11]. The filling calculation formula is

$$f(s) = \frac{\Delta f_{ik} (x(k) - v(k))}{G_n}. \tag{6}$$

In the formula, $x(k)$ represents the regular area of the image; $v(k)$ represents the irregular area; and G_n represents the connected area of the image.

Substituting formula (4) and formula (5) into formula (6), we can get

$$f(s) = h_{ij} + r_{ik} \frac{\int E(t) + x(k) \mathrm{d}t}{H(s)}. \tag{7}$$

In the formula, $H(s)$ represents the adjustment parameter and $E(t)$ represents the similarity of the pixel structure.

The relationship between the area pixels and the actual design can be obtained by filling and sorting, and $H(s)$ in the adjustment formula (7) can be used to adjust the filling effect. In order to ensure the accuracy of the filling position, it is also necessary to establish the frame frequency position. The formula is

$$q \begin{bmatrix} \alpha \\ \beta \\ l \end{bmatrix} = \begin{bmatrix} \alpha \\ \beta \\ \chi \\ l \end{bmatrix}, \quad q = 1, 2, \ldots, m. \tag{8}$$

In the formula, α, β, and χ, respectively, represent the value of the area coordinate length, width, and height, which can be adjusted through the conversion of logical coefficients; the coefficient U is obtained by actual calculation, and the calculation formula is

$$U = \frac{v_0^d}{(u_i + u_0)^d}. \tag{9}$$

In the formula, v_0^d represents the screen filling ratio; u_i represents the uniformity of the unit pixel; and u_0 represents the primary color reuse area.

The result of formula (9) is a coordinate filling, which ensures that every coordinate can be estimated by the system. In the process of extracting features of artistic visual image data, it can effectively solve the problem of data instability or inconspicuous presentation [12].

2.2. Feature Recognition of Artistic Visual Image.

The feature recognition of artistic visual image is based on the data processing results of artistic visual image to ensure the visual communication effect and further recognize the features of artistic visual image. This paper proposes an art visual image feature recognition method based on neural network [13]. Firstly, the global color feature and LBP texture feature of the image are extracted [14], the deep network is constructed to extract the object category feature and advanced emotion feature of the image, and the four features are fused to generate a feature vector C. Secondly, the feature vector C is input into the three-layer full connection layer BP neural network classifier constructed by full connection method to realize image high-level feature recognition [15].

Connecting the c_1, c_2, c_3, and c_4 obtained through the BP neural network above to the BN layer (Batch Normalization) can not only speed up the convergence speed of the model, but also improve the classification effect. The features passing through the BN layer are called combined features, represented by vector C, and $\oplus$ represents string connection, as shown in the following formula:

$$C = BN\left(c_1 \oplus c_2 \oplus c_3 \oplus c_4\right). \tag{10}$$

By inputting the combined feature C extracted above into the emotion recognition model [16], the emotion category of the image can be obtained. While the object category c_4 in the combined feature is used as the emotion recognition feature, it also becomes the image object

category alone. The two parts of information are integrated to obtain a descriptive phrase containing image emotion and object semantic information [17].

The emotion recognition network $F(U, \eta)$ is a neural network including two hidden layers and one output layer, as shown in the following formula:

$$F(U, \eta) = f^4 + f^3 + f^2 + f^1(U). \tag{11}$$

In the formula, η represents a set of parameters, including weights and offsets, which is the final output of the network, including the probabilities of different image categories.

Specifically, for a vector $u^v = [u_1^v, u_2^v, \ldots, u_n^v]$ in layer v, where u_i^v represents neurons in this layer, assuming that e_j^{v+1} represents the value of neuron j in layer $v + 1$, it can be calculated by $f^{v+1}(u^l)$:

$$f^{v+1}\left(u^l\right) = \sum_{i=1}^{N} g_{v,c}^j + I_v^j. \tag{12}$$

In the formula, $g_{v,c}^j$ represents the weight of neuron j in the connection layer; I_v^j represents the offset of neuron j; and N represents the number of neurons. Input the value of neuron j into a nonlinear activation function ∂ [18]:

$$u_j = \partial\left(P_v^j - P_{v,\max}\right). \tag{13}$$

Use a nonlinear activation function ReLu in the hidden layer of the neural network:

$$\partial_{\text{ReLu}}(u) = \max(0, u). \tag{14}$$

The output layer of the last layer uses the softmax activation function, so that the output of the last fully connected layer can be converted into a category of probability distribution $T \in P^k$; then the sentiment classification probability of the image is

$$T_i = \sum_{i=1}^{n} \log_2\left(1 - \frac{P_d^i h_d}{T \log_2\left(1 + \gamma_d^i\right)}\right). \tag{15}$$

In the formula, γ_d^i represents the output of the fully connected layer. The loss function of the recognition probability is the multiclass cross-entropy loss function [19]:

$$\varpi_d^i = \log_2\left(1 + \gamma_d^i\right). \tag{16}$$

In the training phase, the network weights are updated by backpropagating the gradients of all layers. In order to optimize this loss function, it can be achieved by using the SGD optimizer to optimize the weight of the network. The learning rate is set to 0.001, the momentum gradient descent parameter is set to 0.9, and the training model is started until the loss value no longer decreases, thereby realizing the feature recognition of artistic visual images.

2.3. Defolding of Artistic Visual Images.

During the transformation of artistic visual image, it is possible to map multiple pixels in the reference image to the same point of the target image, which is the "wrinkle" phenomenon caused

by the change of visibility. Only by eliminating the wrinkle can the target image show the correct occlusion relationship. In order to ensure the transmission effect of artistic visual image, on the basis of identifying the image features, continue to process the image folds to improve the smoothness of the image. One method to eliminate wrinkles is to determine the scanning order of pixels in the reference art visual image to ensure that the points obscured by other points are always drawn first in the target image. This method is usually called back to front wrinkle elimination method, and occlusion compatibility algorithm is a back to front method. This article mainly uses the midpoint displacement method to achieve this goal. The following is a detailed description.

Suppose D_s and D_{s+1} are used to express the left and right end points of the initial line segment of the artistic visual image. The first iteration is achieved by random disturbance of the midpoint of the line segment. The coordinates of the midpoint of the line segment after the iteration are expressed by formula (17). It can be regarded as different control points:

$$D_{\mathrm{mid}} = \frac{D_s + D_{s+1}}{2} + e(m)\zeta_{ij}. \tag{17}$$

In the formula, $e(m)$ represents a random function; the variance and mean of the function are 1 and 0, respectively. ζ_{ij} represents the amount of disturbance. The amount of disturbance decays at a double speed. If the predetermined threshold is greater than the length of the segmentation line or attenuates to the predetermined number of division layers, the attenuation will stop.

According to the above analysis, it is learned that the control points are not changed during the recursive refinement of the difference, so significant wrinkles are prone to appear around the control points when actually simulating the vision of the artistic visual image. Therefore, in order to improve the visual effect, such wrinkles need to be removed. Based on the sampling law, increase the spatial resolution of the artistic visual image according to twice the range [20], and the original viewpoint value will be affected and changed due to the increase of high-frequency component. At this time, use formula (18) to shift the control point:

$$Q_e = \frac{\sqrt{V^2 - P_e^2 x_k^2}}{D_{\mathrm{mid}}}. \tag{18}$$

In the formula, V^2 represents the image overlay area; P_e^2 represents the edge chain width threshold; and x_k^2 represents the image edge chain length threshold.

The wrinkles existing in the artistic visual image can be removed in the form of midpoint displacement, and only a small amount of calculation content can be added on the basis of the midpoint displacement method to achieve the goal of removing wrinkles. It shows that the processing steps of this method are relatively simple, which can not only realize the effective processing of the image, but also improve the efficiency of image processing.

3. Art Visual Image Communication Method Based on Cartesian Genetic Programming

As a relatively new form of evolutionary art, Cartesian genetic programming has greatly changed its evolutionary image. It creatively uses four-digit data as a gene meaning unit, including two-input data, one-reference function data, and one-parameter data. This makes the image output by Cartesian genetic programming related to the selected function, which can create a more aesthetic image with a strong sense of art [21].

3.1. Coding. Coding is the first step of Cartesian genetic programming, which directly affects the performance of the algorithm. Binary coding is the most commonly used coding in genetic algorithm. In this coding, only 0 and 1 numbers are allowed. It adopts the minimum character coding principle, which greatly simplifies the coding and decoding operation of the algorithm [22]. However, when binary coding is used to deal with [23] multidimensional and high-precision numerical optimization problems, the error of mapping continuous functions to discrete values cannot be well eliminated. Therefore, binary coding cannot directly reflect the real situation of specific practical problems, and the accuracy of processing problems is not high. Moreover, algorithms using binary coding often have problems such as too large individual length and too much memory at runtime. In view of the shortcomings of the above binary coding, in order to overcome the shortcomings of binary coding in solving the problem of discretization of continuous functions, an improved form of binary coding, gray code, is proposed. Assuming that the binary code of an individual is $H = h_m h_{m-1}, \ldots, h_2 h_1$ and the corresponding gray code is $Q = q_m q_{m-1}, \ldots, q_2 q_1$, there is a relationship

$$\begin{cases} h_m = q_m, \\ |h_i(t) - h_i(t-1)| < k, \\ q_i(t-1) - q_i(t) \geq k. \end{cases} \tag{19}$$

Gray code not only inherits most of the advantages of binary coding, but also improves the local search ability of genetic algorithm.

3.2. Improvement of Fitness Function. In genetic algorithm, fitness is one of the main criteria for evaluating the advantages and disadvantages of individuals in the population. The algorithm carries out a series of subsequent genetic operations according to the size of individual fitness, and the main source of individual fitness is the fitness function, so the advantages and disadvantages of the fitness function are directly related to the evolutionary efficiency of the algorithm [24]. For the solution of real problems, especially for constrained optimization problems, penalty function [25] is a more used method. It can combine the objective function of the problem with constraints and finally form an objective function without constraints. This is the initial model of the fitness function of the algorithm. Attention should be paid to the following problems.

The fitness obtained from the fitness function must be greater than or equal to 0. At the same time, it is required that the optimization direction of the objective function should be consistent with the evolution direction of the fitness function used in the individual evolution algorithm. When actually using genetic algorithm to solve specific practical problems, different fitness functions have an important impact on the final convergence and convergence speed of the algorithm. The appropriate fitness function should be finally determined according to experience or algorithm analysis.

The fitness function of Cartesian genetic programming algorithm is different from that of ordinary genetic algorithm. It uses manual evaluation of each individual in the population to obtain the individual fitness, which is also one of the main characteristics of its "interaction." The manual evaluation of each individual of each generation of population is too cumbersome and complex. Ordinary users may soon feel tired and then get bored. Even professionals may cause evaluation deviation because of the fatigue generated by long-term evaluation. When solving practical problems, users may not be able to distinguish the practical significance of the fitness given by themselves. For example, it is difficult for users to recognize the actual difference between two individuals with a fitness of 80 and 85. Therefore, users may have cognitive deviation when giving individuals actual fitness, resulting in inaccurate fitness given to individuals. In the interactive Cartesian genetic algorithm, users are only allowed to select better individuals in each generation, and the selected individuals have the highest fitness in modern times. In this way, users do not have to give specific fitness to each individual, the difficulty of evaluation is greatly reduced, the fatigue of users is alleviated, and then the accuracy of user evaluation is improved.

3.3. Selection Algorithm Improvement.
The selection algorithm is to select the better individuals among the individuals who have been evaluated for fitness for subsequent genetic evolution; that is, select the individuals with high fitness as the parent generation of population evolution, which fully demonstrates the biological evolution idea of natural selection and survival of the fittest [26]. In interactive Cartesian genetic programming, the selection algorithm is based on manual evaluation of fitness, and excellent individuals have the same fitness. For this special individual fitness assignment, interactive Cartesian genetic programming adopts quasi-simulated annealing method.

The simulated annealing method records each generation of optimal individuals selected by users and establishes a dynamic database. With the increase of evolutionary algebra, the selection probability of the optimal individuals of each generation changes dynamically. The specific approach is to make the late generated population have a greater selection probability. Considering the noisy nature of manual selection, make the selection probability of each generation g equal and make the probability change in a ladder shape. The probability function used is

$$p(x) = \frac{1}{\sigma\sqrt{2g}} - \frac{\left(d_i - d_j\right)^t}{\sigma^2}. \tag{20}$$

In the formula, σ represents the use probability of the individual; t represents the number of iterations. Calculate the use probability of the optimal individual according to formula (21). The specific formula is

$$\sigma_h = \arg\max\left[\sigma\left(Z = z_1 | X = x\right)\right]. \tag{21}$$

For example, when the current algebra is 6 and 12 generations, the change trend of usage probability of each generation is shown in Figure 1 and 2.

As can be seen from Figures 1 and 2, with the increase of algebra, the selection probability increases step by step, and the probability of the last order is much greater than that of the previous orders, which greatly increases the possibility of the algorithm to select the latest generations of individuals when selecting individuals. Even if the selection probability of the previous generations decreases gradually, the algorithm also provides the possibility for its optimal individual to enter the crossover and mutation of the latest generation. Compared with interactive genetic algorithm, this algorithm increases the range of selecting individuals and expands from using only the previous generation optimal individuals for subsequent genetic evolution to using all the optimal individuals. The algorithm not only considers expanding the scope of individual use, but also solves the problem of use probability. It also introduces the artificial individual evaluation noise response mechanism, which can help users find satisfactory images faster without increasing the burden of users.

3.4. Visual Effect Rendering.
Combined with the preprocessing results of artistic visual image and Cartesian genetic programming algorithm, the effective processing of image is realized. In order to further improve the display effect of image, the image visual effect is rendered [27–29]. After the construction of Cartesian genetic programming algorithm, an image rendering platform is designed according to the actual needs to render the visual effect of artistic visual images. The following steps are the design process of image rendering platform.

The hardware composition of the image rendering platform designed in this article is shown in Figure 3. Among them, GPU is a graphics processor that supports OpenGL, responsible for rendering list calculation processing tasks; CPU is 32 bits; embedded processors such as ARM9 can be used to execute graphics applications; DMA can send the rendering list to the GPU; the VGA/LCD controller is mainly used to generate the scan timing signal of the display [30–32]. The platform is equipped with two frame buffers, one for display and the other for saving GPU drawing results. The ping-pong operation between the two can ensure the efficient operation of the entire platform.

With the support of the image rendering platform shown in Figure 3, the visual rendering process of artistic visual image is described in detail:
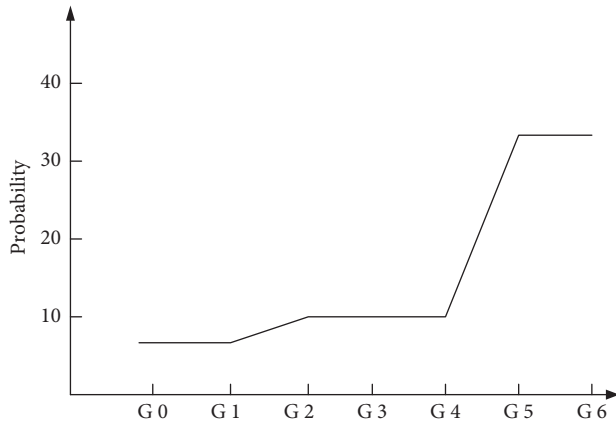
Figure 1: Change trend of generation 6 probability.
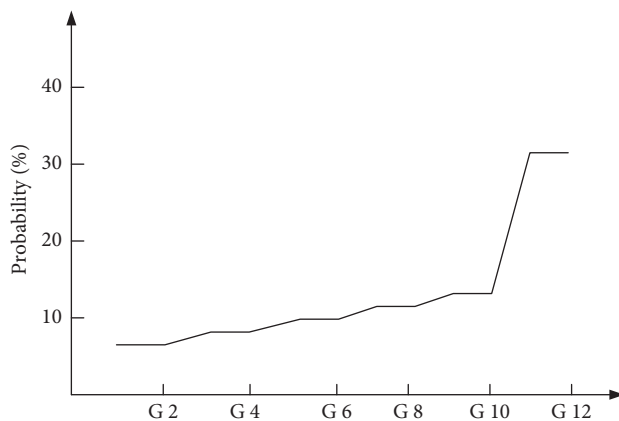
Figure 2: Change trend of generation 12 probability.
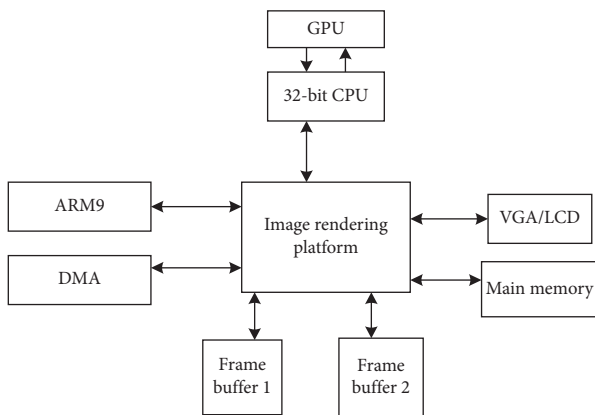
Figure 3: The hardware composition of the image rendering platform.

(1) Initialize the image rendering platform, set the environment for artistic visual image operation in advance, build a common data area, and allocate the original length memory to each list in advance, so that each parameter can be initialized [33]

(2) Definition file: read the artistic visual image file and store data in each list

(3) Configuration platform: in order to facilitate the calling of artistic visual image scene during rendering, preprocess each scene

(4) Display rendering results: obtain the state parameters generated by the platform itself and external input and control and display the objects in the scene, so as to optimize the display of artistic visual images and complete the visual communication design [34–36]

4. Simulation Experiment

In order to verify the effectiveness of the art visual image communication method based on Cartesian genetic programming designed in this paper, comparative simulation experiments are used to compare the effect with the traditional art visual image communication methods, specifically the plane visual communication effect optimization method based on wavelet change and the plane visual communication design method based on user experience effect. In order to obtain more accurate experimental results, different methods are used to process an artistic visual portrait at the same time. Set the image conversion coefficient to 0.45. In order to better carry out the experiment, the image adjustment error is controlled within 0.2. At the same time, the plane visual communication effect optimization method based on wavelet change proposed in [7] and the plane visual communication design method based on user experience effect proposed in [8] are used as traditional comparison methods to compare with this method. Before the experimental test, it is necessary to prepare to build a test environment and prepare three computers. Before the experimental test, the three methods are installed in the three computers and put into trial operation. The three computers are jointly connected to a data analyzer. The analyzer completes the recording of experimental data during the test, so as to provide an accurate data basis for experimental analysis. In order to ensure the fairness of the experimental test and improve the test difficulty of the experiment, three levels of artistic visual images with low difficulty, normal difficulty, and high difficulty are randomly selected as the experimental samples in a three-image database before the experiment. At the same time, the total calculation sends image communication samples to the three systems to start the experiment. After all computers submit the artistic visual image dynamic communication report, the experiment is ended, the experimental data and experimental site are sorted out, the experimental analysis is completed, and the experimental conclusion is drawn.

4.1. Experimental Dataset. The images used in the experiment are from the Abstract dataset. The Abstract dataset contains 228 abstract paintings painted by professional painters. The image emotion is expressed through the combination of image color and texture and does not contain any recognizable objects. The emotional labels of the dataset were selected by 14 observers through voting. In order to obtain the annotation information of the dataset,

FIGURE 4: Sample experimental image.

the image is evaluated by peer rating on the network. 280 images were classified by 230 people for image emotion evaluation, in which each image would be marked by 14 people, and finally the category with the most votes from the eight categories was selected as the sample image. After analysis, some images with unclear categories were removed from the dataset, and finally a dataset composed of 228 images was generated. Some image examples are shown in Figure 4.

4.2. Analysis of Experimental Results

4.2.1. Comparison Chart of Image Sharpness. First, take the image clarity as the experimental index, compare the traditional method with the method in this paper, randomly extract an image from the experimental image sample, and compare the image visual transmission effects of different methods. The result is shown in Figure 5.

It can be seen from the above pictures that the art visual image communication method based on Cartesian genetic programming designed in this paper can effectively solve the problem of unclear image. At the same time, it does not need to be revised in the process of visual communication and maintains the clarity of art image. As can be seen from Figure 5, in terms of image details, after using the method in this paper, the visual effect is clearer and the visual effect of the image is more realistic. After using the traditional method, the image shows blur effect, such as edge blur, high brightness, and so on. It can be seen that using this method has a better image visual communication effect.

4.2.2. Comparison of Image Visual Transmission Efficiency. Secondly, taking the image visual communication efficiency as the experimental index, different methods are compared, and the results are shown in Table 1.

By analyzing the data in Table 1, it can be seen that, with the increasing number of iterations, the image processing time of different methods gradually increases, indicating that the visual communication efficiency continues to decrease. Comparing the image processing time of this method with that of the traditional method, it can be seen that the image processing time of this method is much lower than that of the traditional method, and the minimum value of image processing time is 3.72s and the maximum value is only 4.80s; it is much lower than the traditional methods, which shows that the image visual communication efficiency of this method is higher. This is because this method uses the combination of image data and image processing platform to design and analyze the visual communication of images, accurately grasp the data status, reduce the image processing time, improve the visual communication efficiency, and obtain more accurate image data, so as to promote the analyzability of data and enhance the visual communication efficiency.

4.2.3. Success Rate of Visual Communication. Finally, taking the success rate of visual communication as the experimental index, the application effects of different methods are compared, and the results are shown in Figure 6.

According to Figure 6, compared with the traditional method, the image visual communication success rate of the method in this paper is higher, the highest value reaches

(a)   (b)   (c)
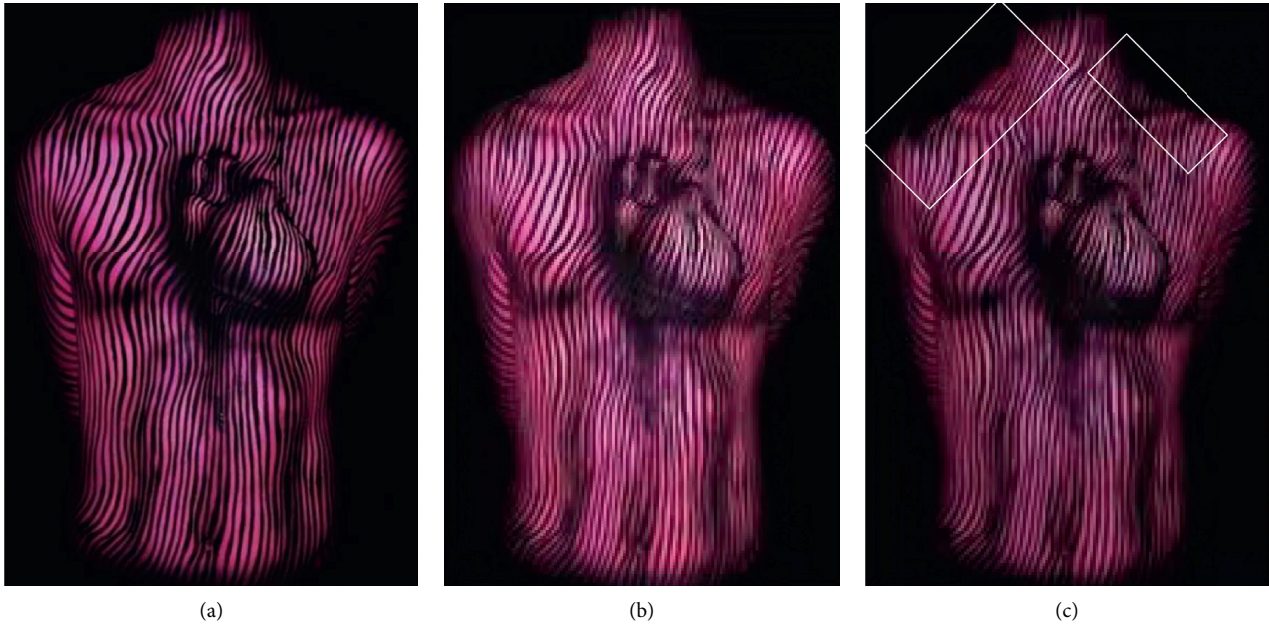
Figure 5: Visual communication effects of different methods.

Table 1: Comparison of image visual transmission efficiency.

Iterations (time)	Image processing time (s)		
	Method of this article	Image visual communication method based on wavelet transform	Image visual communication method based on user experience effect
1	3.72	7.12	5.37
2	3.82	7.36	5.62
3	3.91	7.89	5.98
4	4.05	8.12	6.37
5	4.16	8.25	6.24
6	4.20	8.56	6.70
7	4.37	8.91	7.05
8	4.59	9.13	7.19
9	4.72	9.47	7.54
10	4.80	9.84	8.90
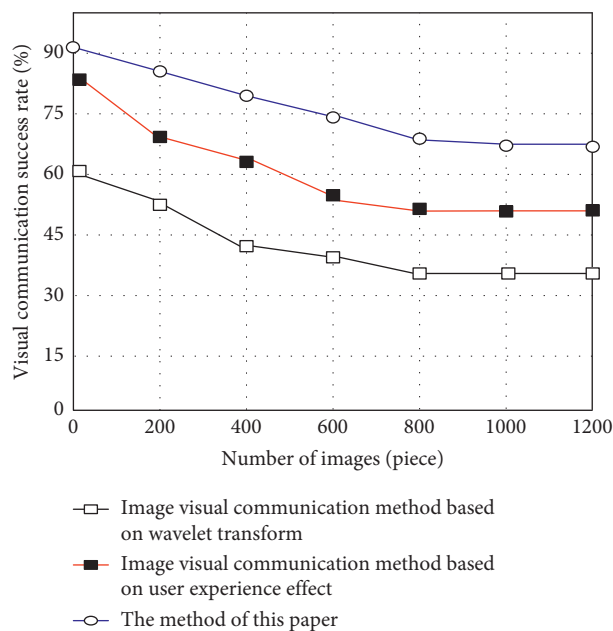
Figure 6: Comparison of the success rate of visual communication.

more than 90%, and the lowest value remains more than 65%. In contrast, the image visual communication success rate of the traditional method is lower.

To sum up, the art visual image communication method based on Cartesian genetic programming designed in this paper has a good effect, can improve the image visual communication performance to a certain extent, improve the clarity and integrity of the image, and has a good development prospect.

5. Conclusion

Aiming at the problems of long image processing time, low success rate of image visual communication, and poor visual effect in traditional methods, an art visual image communication method based on Cartesian genetic programming is proposed. The main innovations of this method are as follows:

(1) This method can not only realize the effective image processing, but also improve the efficiency of image processing.

(2) Cartesian genetic programming algorithm not only considers expanding the scope of individual use, but also solves the problem of use probability. It can help users find satisfactory images faster and improve the visual communication effect of images without increasing the burden of users.

(3) The experimental results show that this method is superior to the traditional method in image processing time, image visual communication success rate, and visual effect, which shows that this method has higher application value.

The next research work can focus on how to extend this method to the case with multiple reference images, so as to improve the processing efficiency, make the communication results more consistent with the actual results, and be more suitable for practical application.

Data Availability

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Conflicts of Interest

The authors declare that they have no conflicts of interest regarding this work.

Acknowledgments

References

[1] D. D. Rodriguez, C. Goulart, C. Valadão, D. Funayama, and T. Bastos, "Visual and thermal image processing for facial specific landmark detection to infer emotions in a child-robot interaction," *Sensors*, vol. 19, no. 13, p. 2844, 2019.

[2] E. Schwartz, R. Giryes, and A. M. Bronstein, "DeepISP: toward learning an end-to-end image processing pipeline," *IEEE Transactions on Image Processing*, vol. 28, no. 2, pp. 912–923, 2019.

[3] M. Li, J. Lv, and C. Tang, "Aesthetic assessment of paintings based on visual balance," *IET Image Processing*, vol. 13, no. 14, pp. 2821–2828, 2019.

[4] M. N. Martínez-Domingo, M. Melgosa, K. Okajima, V. J. Medina, and F. J. Collado-Montero, "Spectral image processing for museum lighting using CIE LED illuminants," *Sensors*, vol. 19, no. 24, p. 5400, 2019.

[5] A. M. Sharma, A. Dogra, B. Goyal, R. Vig, and S. Agrawal, "From pyramids to state-of-the-art: a study and comprehensive comparison of visible-infrared image fusion techniques," *IET Image Processing*, vol. 14, no. 9, pp. 1671–1689, 2020.

[6] X. Sun, L. Liu, Q. Li, J. Dong, E. Lima, and R. Yin, "Deep pixel-to-pixel network for underwater image enhancement and restoration," *IET Image Processing*, vol. 13, no. 3, pp. 469–474, 2019.

[7] M. M. Liu, "Optimization of plane visual communication effect simulation based on graphic beautification technology," *Computer Simulation*, vol. 36, no. 9, pp. 426–429, 2019.

[8] C. Li, "Research on graphic visual communication design based on user experience effect," *Modern Electronics Technique*, vol. 43, no. 11, pp. 111–114, 2020.

[9] X. Q. Zhu, "Innovative design of graphic image visual communication system for animated character in virtual reality environment," *Modern Electronics Technique*, vol. 561, no. 10, pp. 178–180, 2020.

[10] M. M. Shawky, R. A. El-Arafy, M. A. El Zalaky, and T. Elarif, "Validating (MNF) transform to determine the least inherent dimensionality of ASTER image data of some uranium localities at Central Eastern Desert, Egypt," *Journal of African Earth Sciences*, vol. 149, no. 1, pp. 441–450, 2019.

[11] G. S. Yadav and A. Ojha, "Improved security in the genetic algorithm-based image steganography scheme using Hilbert space-filling curve," *The Imaging Science Journal*, vol. 67, no. 3, pp. 148–158, 2019.

[12] M. Konnik, B. Ahmadi, N. May et al., "Training AI-based feature extraction algorithms, for micro CT images, using synthesized data," *Journal of Nondestructive Evaluation*, vol. 40, no. 1, pp. 1–13, 2021.

[13] S. Chan and A. H. Elsheikh, "Parametric generation of conditional geological realizations using generative neural networks," *Computational Geosciences*, vol. 23, no. 5, pp. 925–952, 2019.

[14] C. Yang, "Plant leaf recognition by integrating shape and texture features," *Pattern Recognition*, vol. 112, no. 1, Article ID 107809, 2021.

[15] Y. Gu, K. Vyas, J. Yang, and G.-z. Yang, "Transfer recurrent feature learning for e image recognition," *IEEE Transactions on Medical Imaging*, vol. 38, no. 3, pp. 791–801, 2019.

[16] T. Chen, S. Ju, F. Ren, M. Fan, and Y. Gu, "EEG emotion recognition model based on the LIBSVM classifier," *Measurement*, vol. 164, no. 5, Article ID 108047, 2020.

[17] W. Zhang, X. He, and W. Lu, "Exploring discriminative representations for image emotion recognition with CNNs," *IEEE Transactions on Multimedia*, vol. 22, no. 2, pp. 515–523, 2020.

[18] S. Kiliarslan and M. Celik, "RSigELU: a nonlinear activation function for deep neural networks," *Expert Systems with Applications*, vol. 174, no. 2, Article ID 114805, 2021.

[19] X. Li, L. Yu, D. Chang, Z. Ma, and J. Cao, "Dual cross-entropy loss for small-sample fine-grained vehicle classification," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 5, pp. 4204–4212, 2019.

[20] C. Geberl, K. Kugler, and L. Wiegrebe, "The spatial resolution of bat b quantified with a visual-resolution paradigm," *Current Biology*, vol. 29, no. 11, pp. 1842–1846, 2019.

[21] A. Khalilian, A. Baraani-Dastjerdi, and B. Zamani, "CGen-Prog: adaptation of cartesian genetic programming with migration and opposite guesses for automatic repair of software regression faults," *Expert Systems with Applications*, vol. 169, no. 3, Article ID 114503, 2021.

[22] M. Suganuma, M. Kobayashi, S. Shirakawa, and T. Nagao, "Evolution of deep convolutional neural networks using cartesian genetic programming," *Evolutionary Computation*, vol. 28, no. 1, pp. 141–163, 2020.

[23] C. He, Q. Zhang, T. Qu, D. Wang, and M. Liao, "Remote sensing and texture image classification network based on deep learning integrated with binary coding and s distance," *Remote Sensing*, vol. 11, no. 23, p. 2870, 2019.

[24] Y.-l. Wang, Z.-p. Wu, G. Guan, K. Li, and S.-h. Chai, "Research on intelligent design method of ship multi-deck compartment layout based on improved taboo search genetic algorithm," *Ocean Engineering*, vol. 225, no. 2, Article ID 108823, 2021.

[25] J. Wang, F. Zhang, J. Huang, W. Wang, and C. Yuan, "A nonconvex penalty function with integral convolution approximation for compressed sensing," *Signal Processing*, vol. 158, no. 5, pp. 116–128, 2019.

[26] M. D. de Lima, J. de Oliveira Roque e Lima, and R. M. Barbosa, "Medical data set classification using a new feature selection algorithm combined with twin-bounded support vector machine," *Medical, & Biological Engineering & Computing*, vol. 58, no. 3, pp. 519–528, 2020.

[27] X. Wu, C. Zhang, M. Huang, C. Yang, and G. Ding, "Quantitative atmospheric rendering for real-time infrared scene simulation," *Infrared Physics & Technology*, vol. 114, no. 3, Article ID 103610, 2020.

[28] C. Yu, "Climate environment of coastline and urban visual communication art design from the perspective of GIS," *Arabian Journal of Geosciences*, vol. 14, no. 4, p. 310, 2021.

[29] S. R. Maniyath, K. Vijayakumar, L. Singh, S. K. Sharma, and T. Olabiyisi, "Learning-based approach to underwater image dehazing using cycleGAN," *Arabian Journal of Geosciences*, vol. 14, no. 18, pp. 1–11, 2021.

[30] J. Bai, R. Chen, and M. Liu, "Feature-attention module for context-aware image-to-image translation," *The Visual Computer*, vol. 36, no. 10-12, pp. 2145–2159, 2020.

[31] T. Li, Y. Wang, R. Hong, M. Wang, and X. Wu, "pDisVPL: probabilistic discriminative visual Part Learning for image classification," *IEEE MultiMedia*, vol. 25, no. 4, pp. 34–45, 2019.

[32] H. Wu and G. Li, "Visual communication design elements of Internet of Things based on cloud computing applied in graffiti art schema," *Soft Computing*, vol. 24, no. 11, pp. 8077–8086, 2020.

[33] L. Zhang, D. Xiang, C. Jin, F. Shi, K. Yu, and X. Chen, "OIPAV: an integrated software system for ophthalmic image processing, analysis, and visualization," *Journal of Digital Imaging*, vol. 32, no. 1, pp. 183–197, 2019.

[34] J. Qin, J. Chen, X. Xiang, Y. Tan, W. Ma, and J. Wang, "A privacy-preserving image retrieval method based on deep learning and adaptive weighted fusion," *Journal of Real-Time Image Processing*, vol. 17, no. 1, pp. 161–173, 2020.

[35] W. Wang, J. Qin, Y. Zhang et al., "TNNL: a novel image dimensionality reduction method for face image recognition," *Digital Signal Processing*, vol. 115, no. 99, Article ID 103082, 2021.

[36] Y. Chen and H. Hu, "An improved method for semantic image ipi," *Neural Processing Letters*, vol. 49, no. 3, pp. 1355–1367, 2019.

Research Article

Innovative Research on the Development of Online Education Mode of Internet Thinking Based on the Discrimination of Learning Attention under the Analysis of Head Posture

Su Song ⬤ and **Fangzheng Wang** ⬤

Public Basic College, Jiangsu Vocational College of Medicine, Yancheng, Jiangsu 224000, China

Correspondence should be addressed to Fangzheng Wang; 11406@jsmc.edu.cn

With the rapid development of Internet technology and the popularity of 5G and broadband, online education in China, especially mobile online education, is in full swing. Based on the development status of online education in China, this paper analyzes the innovative application of learning attention discrimination based on head posture analysis in the development of online education mode of Internet thinking. Learning attention is an important factor of students' learning efficiency, which directly affects students' learning effect. In order to effectively monitor students' learning attention in online teaching, a method of distinguishing students' learning attention based on head posture recognition is proposed. In the tracking process, as long as the head angle of the current frame is close to the head angle of the key frame in a certain scale model, the visual angle apparent model can reduce the error accumulation in large-scale tracking. A Dynamic Bayesian Network (DBN) model is used to reason students' Learning Attention Goal (LAG), which combines the relationships among multiple LAGs, multiple students' positions, multicamera face images, and so on. We measure the head posture through the similarity vector between the face image and multiple face categories without explicitly calculating the specific head posture value. The test results show that the proposed model can effectively detect students' learning attention and has a good application prospect.

1. Introduction

Traditional education can no longer meet the needs of modern education. The cost of computer production is declining, and the modern education mode with computer as the medium is more and more widely used in education. China's online education has been on the rise, making many offline education and training institutions vigorously develop online education, and even many Internet companies that did not pay attention to the field of education began to enter the field of online education [1]. China's online education is so full of vitality that everyone applauds it, but its development cannot be smooth sailing, and it is bounded by many problems. How to solve these problems is the basic requirement to promote the further development of online education.

Since the explosive birth of information transmission technology in the 19th century, people have never stopped exploring the application of various new technologies in the field of education. Root et al. define online education management system as "a software application for classroom education and after-school training, involving educational administration, information transmission, report generation and learning effect tracking" [2]. However, Lockee et al. define online education management system as "the integration of networked tools to support online learning" [3]. Zhang J. and Zhang F. discuss the development trend and form of online higher education through the research based on relevance theory, and emphasize that Massive Open Online Courses (MOOCs) are not only an online classroom, but also the interaction between teachers and students [4]. Fehr introduces the practical application of MOOCs,

analyzes their existing problems, and holds that "There is no doubt that the development of MOOCs is of great significance to higher education all over the world [5]. It provides opportunities for the online development of traditional higher education in a flexible and convenient way, but it cannot be considered that virtual education can replace traditional education. In disciplines such as medicine and architecture, the advantages of traditional higher education model are still incomparable." By analyzing the development data of online higher education in the United States from 2002 to 2014, Mcauliffe et al. preliminarily explore the development direction and path of online higher education in China, and form its own theoretical model. In a word, scholars at home and abroad started their research on online higher education late, mostly focusing on analyzing foreign literature. Even if a few models are put forward, their feasibility and suitability are open to question [6].

Attention in visual learning is related to head posture and eye sight direction. The research shows that [7, 8], in many cases, it is enough to analyze students' Learning Attention Goal (LAG) through head posture. Because students are not used to staring at a certain goal with slanting eyes for a long time, they will turn their heads to face the goal. Therefore, this paper uses the head posture to analyze the LAG of students, so we propose a Dynamic Bayesian Network (DBN) model to deduce the LAG of students. The head pose is measured by the similarity vector between the face image and multiple face categories without explicitly calculating the specific head pose value. The observation of probability model includes face image and face position under multicameras. We collected test data in the teaching environment, and the experimental results show that our model is effective.

Structure of This Paper. The first section mainly introduces the research background, significance and main innovations of this paper. Section two presents a summary of related research. This paper introduces the research status of the key technologies (head pose estimation, face feature points, and attention recognition). In section three, the discrimination process of learning attention based on head posture analysis is mainly studied. Section four analyzes and discusses experimental results. Finally, the paper summarizes the full text, analyzes the existing problems in the current methods, and looks forward to the future research directions.

2. Related Work

In recent years, more and more researchers began to study the problem of visual LAG recognition. Bce et al. study the problem of students' LAG recognition in a small round table environment, in which an omnidirectional camera is placed on the conference table [9]. Later, they studied the method of identifying LAG in the environment of multiple remote cameras [4]. Hamrah et al. also study the problem of LAG recognition in different conference environments [10]. However, in the meeting environment, students mainly sit in fixed seats, and their bodies do not move much. Zhou et al.

also monitor students' LAG in outdoor environment, but the range of their head posture is limited to face posture close to the front. These tasks mainly deal with the analysis of multiple LAGs in fixed positions or single LAGs in multistudent positions [11]. However, our application environment includes multiple student locations and multiple LAGs. Zhao proposed a head pose recognition algorithm based on template matching technology. For each recognition object, multiple head images in different poses are extracted as sample images, and each image is marked with corresponding pose parameters [12]. Yeager et al. found that in most cases, students' attention target behavior can be obtained by analyzing the head posture angle [13]. Yfka et al. detect face feature points through random cascade regression tree, and use N-point perspective algorithm to estimate head posture, thus realizing the visualization of students' learning attention [14].

Hirata and Kusatake establish a human-computer interaction system by using the learning attention detection model, which can effectively judge the position and head posture of the target person in the multiperson environment [15]. Guo et al. study the problem of attention target recognition in different conference environments. However, in the conference environment, users mainly sit in fixed seats, and their bodies do not move much [16]. Lu and Yanmin study the problem of attention target recognition in outdoor environment. They mainly analyze whether passers-by watched posters on the wall [17]. Xiao et al. introduce the run-length matrix of binary pattern into the random feature selection of random tree, which improves the classification ability of single decision tree and achieves better recognition rate for multiclassification discrete head pose estimation [18]. Ren et al. define the distance between eyes and screen and the range of head posture by calibrating the system, and use a single light source to track the gaze direction of eyeballs [19]. In the literature [20], three-dimensional faces are recognized by combining stereo vision information such as rotation and pitch, and the learning attention direction of eye sight can be accurately tracked through the details of eye images.

Through the study of the abovementioned related literatures, it is found that most of the research methods have certain requirements for equipment. Because of the low resolution of the individual's face in the classroom scene, and the illumination change, occlusion, and large posture change in the environment, it is very difficult to learn attention recognition. Therefore, we adopt a noninvasive learning attention recognition method based on the head posture direction, and recognize the LAG of many people at the same time in the large classroom scene.

3. Research Method

3.1. Analysis of Students' Learning Attention. The main purpose of this study is to distinguish students' learning attention, and to determine the direction of students' learning attention by estimating whether students' eyes are concentrated in the blackboard area according to their head posture.

As shown in Figure 1, when students' eyes are focused on a certain point in the blackboard, such as P_1, students are considered to be focused on learning. On the contrary, when students' eyes deviate from the blackboard area for a long time, such as P_2, students are considered to be distracted in learning. Under normal circumstances, people are not used to looking at the target they pay attention to with oblique eyes. Therefore, the rotation direction of head posture can be regarded as the line of sight of students approximately to analyze students' learning attention.

According to the classroom environment, a coordinate system is established, which takes the center point on the blackboard as the coordinate origin, the horizontal right direction of the origin as the X-axis positive direction, the vertical origin direction as the Y-axis positive direction, and the vertical XY plane pointing to students as the Z-axis positive direction. According to the students' head line of sight reaching the edge of blackboard, it is regarded as the criterion of students' abnormal behavior, as shown in Figure 2.

$\alpha_1, \alpha_2, \beta_1, \beta_2$ is the threshold of abnormal head deflection of students, and α_1, α_2 is the rotation range of students in θ_{Yaw} direction; β_1, β_2 is the rotation range of students' θ_{Pitch} direction. When the rotation range of the head exceeds the threshold, it can be considered that the students' sight is outside the blackboard area, and it is judged that the learning attention is distracted.

Assume that the blackboard has a length of h, a width of d, and a head center coordinate of $F(x, y, z)$. When the students sit in the first row of the classroom and look at the left and right edges of the blackboard at points B and D shown in Figure 1, it is the maximum rotation range of the students' heads in the θ_{Yaw} direction, which is written as

$$\alpha_1 = -\arctan\left(\frac{h}{z}\right)\alpha_2$$

$$= \arctan\left(\frac{h}{z}\right). \tag{1}$$

When the student sits at point C and looks at the upper and lower edges of the blackboard, it is the maximum rotation range of the student's head in the θ_{Pitch} direction, which is written as

$$\beta_1 = -\arctan\left(\frac{y}{z}\right)\beta_2$$

$$= \arctan\left(\frac{d-y}{z}\right). \tag{2}$$

According to the actual teaching environment, assuming that the center point of the head coincides with the eyes, and the height of the adult students' eyes from the ground is 1.3 m, the head rotation range of the students is determined to be θ_{Pitch} direction [−7°, 28°], θ_{Yaw} direction [−48°, 48°].
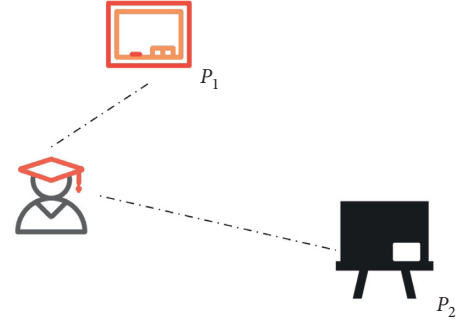
FIGURE 1: Students' learning attention.
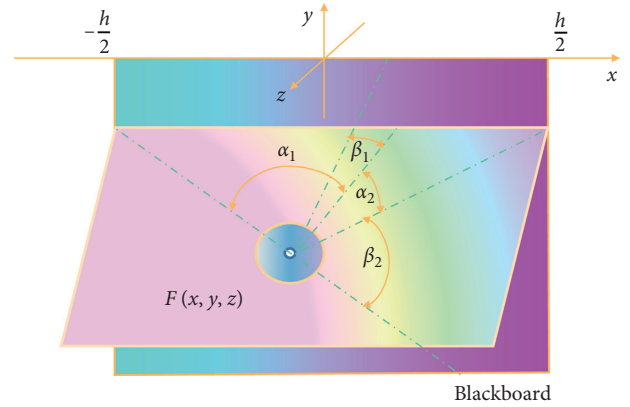
FIGURE 2: Head deflection range.

3.2. Discrimination Process of Learning Attention Based on Head Posture Analysis

3.2.1. Visualization of Students' Learning Attention Based on Head Pose Estimation. Combining the advantages of eye tracker and single camera learning attention analysis system, this paper proposes a visual analysis method of students' learning attention based on head pose estimation of single image, and constructs a corresponding visual analysis system of students' learning attention [21]. In this paper, the front camera installed in the middle position above the blackboard is used to record the students' lectures, and then the method shown in Figure 3 is used to estimate the students' head posture. Finally, the students' eyes are projected to the teacher's lecture video recorded by the rear camera by mathematical deduction.

As shown in Figure 3, this method mainly consists of the following six steps.

(1) Acquisition of data (video frame): the classroom teaching video is acquired by LifeCam camera of Microsoft 1080p, and the video frames are separated.

(2) Camera calibration: in order to improve the accuracy of head pose recognition, it is necessary to use a convenient and accurate calibration method to calibrate camera parameters.

(3) Face detection: use the disclosed face detector to detect faces from video frames [22].
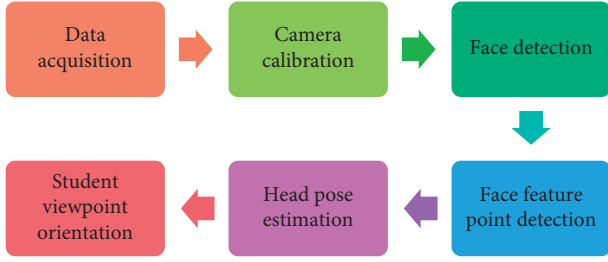
FIGURE 3: Working principle and algorithm flow of visual analysis of students' learning attention.

(4) Face feature point detection: this paper uses random cascade regression tree to obtain the coordinate information of 19 face feature points, which is used to provide two-dimensional information in solution.

Students' viewpoint positioning: according to the rotation and translation matrix information of head posture, the students' viewpoint is projected to the teacher's lecture video shot by the rear camera by using the transformation relationship of spatial coordinates, so as to realize the visual display of students' learning attention.

Combining the oculomotor and the single-camera learning attention analysis system, the random cascade regression tree is used to locate the face feature points, and a rigid model obtained by statistical measurement is introduced as the 3D face approximation. The students' eyes are projected onto the video images taught by teachers, and the visual analysis of students' learning attention is realized.

3.2.2. Visual Angle Apparent Model

(1) Key Frame Adjustment. After tracking the current frame t, in order to calculate the attitude parameters of $t + 1$ frame, the current frame t becomes a new key frame in the model, and the expectation and covariance of X should be expanded accordingly. Because x_{t+1} is unknown at this moment, the expectation and covariance of X are extended as follows:

$$E_X^{t+1} = \left[0; E_X^t\right],$$
$$\left[\left(\sigma_X^{t+1}\right)^2\right]^{-1} = \begin{bmatrix} 0 & 0 \\ 0 & \left[\left(\sigma_X^t\right)^2\right]^{-1} \end{bmatrix}. \tag{3}$$

Here, E_X^t and $(\sigma_X^t)^2$ represent the expectation and covariance of X after the tracking of the current frame t is completed, and E_X^{t+1}, $(\sigma_X^{t+1})^2$ represents the extended expectation and covariance of X.

To reduce the number of key frames in the model, only one key frame is selected from each perspective. If the attitude parameters of one key frame are very close to those of other key frames in the model, the key frame will be removed from the model [23]. Since the previous frame $t - 1$ is always used as the reference frame of the current frame t, the previous frame $t - 1$ is likely to be removed from the model after the tracking of the current frame t is completed. Let the

largest one of the three rotation angles of the previous frame be ω_{t-1}, and when the corresponding angle ω_i of a certain key frame i in the model satisfies the condition

$$\left\|\omega_{i-1} - \omega_i\right\| \leq \tau, \tag{4}$$

remove the previous frame $t - 1$ from the model, where τ is the threshold determined by the number of key frames in the model. The removal process is completed by deleting the corresponding rows and columns in E_X^t, $(\sigma_X^t)^2$.

(2) *Multiscale Visual Angle Apparent Model.* When the head movement range is small, the visual angle apparent model can effectively reduce the tracking error. Visual angle apparent model when the head moves in a small range is also called single-scale visual angle apparent model, but when the head moves in a large range, it will exceed the effective range of single-scale visual angle apparent model. There are two main reasons for this phenomenon: first, the scale transformation of the head image is large when moving in a large range; second, the apparent model parameters themselves contain large error accumulation due to long-term movement, which leads to the tracking result deviating from the true value [24]. Therefore, each visual angle apparent model is only valid within a certain range.

When the head movement range is large, multiple visual angle apparent models can be used to cooperate with each other to reduce the error accumulation caused by large-scale movement. In the specific implementation, the effective range of each visual angle apparent model is defined as a space neighborhood around its initial frame head position.

Let m represent the initial frame of the current apparent model, D_m represent the distance between the head of the initial frame and the camera, and the distance D_t between the head of the current frame and the camera satisfies the condition

$$\left\|D_t - D_m\right\| \geq \eta. \tag{5}$$

The tracking method will generate a new apparent model, where η is a predefined threshold. The current frame t will become the initial frame of the new model. Multiple visual angle apparent models which are continuous in space are called multiscale visual angle apparent models.

On the one hand, the multiscale visual angle apparent model can effectively solve the error accumulation problem of tracking large-scale motion, especially large-scale forward and backward motion. On the other hand, because each scale visual angle apparent model is only responsible for tracking motion in a small range, it is beneficial to the application of the model. For example, when the head leaves the camera view and re-enters, the head posture can be quickly recovered by using the multiscale visual angle apparent model.

3.2.3. DBN Model for Analyzing Visual LAG.

We propose a DBN model to reason students' attention goals. Reasoning students' attention goals by calculating the maximum posterior probability. The model integrates the relationships among multiattention targets, multistudent positions, and multicamera face images, and conducts joint reasoning.

(1) Overview of Models. A Dynamic Bayesian Network (DBN) model for analyzing LAG proposed in this paper [25] is shown in Figure 4.

In the model, we have the following:

Implicit variable F_t represents the LAG of students at time t, and its value is M possible lags.

Implicit variable C_t^i represents the pose category of the face image shot by camera i at time t, and its value is K face pose category.

The observation variable Z_t^i represents the face image taken by the camera i at time t.

The observation variable L_t^i represents the horizontal position of the face image shot by the camera i at time t.

The model combines multicamera information to analyze LAG more accurately. For example, when students stand in area 2 and area 3, both cameras can capture students. In some cases, the images obtained by a single camera may not accurately analyze the LAG of students. At this time, through the information of another camera, it may be easier to analyze students' goals.

(2) Description of Each Part of the Model. According to the probability model, the joint probability distribution among all variables can be written as follows:

$$
\begin{aligned}
&P\big(F_{1:\,T}, C_{1:\,T}^{1:\,R}, L_{1:\,T}^{1:\,R}, Z_{1:\,T}^{1:\,R}\big) \\
&= \prod_{t=1}^{T} P(F_t|F_{t-1}) \prod_{i=1}^{R} P\big(C_t^i|C_{t-1}^i, L_t^i, F_t\big) P\big(Z_t^i|C_t^i\big).
\end{aligned} \tag{6}
$$

Among them, $P(F_t|F_{t-1})$ stands for the purpose of transition probability matrix between different LAGs at adjacent moments, which is to enhance time smoothness.

$P(C_t^i|C_{t-1}^i, L_t^i, F_t)$ represents the probability dependence of face pose category C_t^i on LAG F_t, face position L_t^i, and face pose category C_{t-1}^i at the previous moment. This is the core of this model, which describes the probability dependence among multiple LAGs, multiple student positions, and faces obtained by multiple cameras.

$P(Z_t^i|C_t^i)$ is the likelihood of face observation. The face posture class C_t^i is known, and the likelihood represents the probability that the face observation Z_t^i is generated by the face posture class, as shown in

$$
P\big(Z_t^i|C_t^i = k\big) = \frac{1}{\Lambda} \exp\left(\frac{-d^2\big(Z_t^i, M_k\big)}{\sigma^2}\right). \tag{7}
$$

Here, Λ is the normalization factor, M_k represents the image subspace $C_t^i = k$ of the face pose category, and $d^2(Z_t^i, M_k)$ represents the distance from the face image to the image subspace, such as the reconstruction error when the image is projected to the subspace.

(3) Model Reasoning. The analysis problem of LAG is regarded as the reasoning problem of probability model. Given the observations Z and L, we hope to deduce the hidden variables F and C. That is, our objective function is to maximize the following joint probability density distribution:

$$
(\widehat{F}, \widehat{C}) = \arg\max_{F,C} p(F, C, Z, L). \tag{8}
$$

We use the approximate reasoning algorithm proposed in [26] to minimize the cost function "free energy." The free energy uses a simple probability density distribution $Q(h)$ to approximate the true posterior probability density distribution $P(h|v)$ where h and v represent hidden variable (F, C) and observed variable (Z, L), respectively. Then, $Q(h)$ is used to calculate the objective function with the following formula:

$$
Q(h) = \prod_{t=1}^{T} Q(F_t|F_{t-1}) \prod_{i=1}^{R} Q\big(C_t^i|C_{t-1}^i, F_t\big). \tag{9}
$$

According to the method of [11], free energy can be written as

$$
E = \int_h Q(h) \ln \frac{Q(h)}{P(h, v)}. \tag{10}
$$

Given the state of hidden variables F and C at time $t-1$, by minimizing E, we can get

$$
\frac{\partial E_t}{\partial Q\big(C_t^i|C_{t-1}^i, F_t\big)} = 0 \Rightarrow Q\big(C_t^i|C_{t-1}^i, F_t\big) \propto P\big(C_t^i|C_{t-1}^i, L_t^i, F_t\big) P\big(Z_t^i|C_t^i\big), \tag{11}
$$

$$
\frac{\partial E_t}{\partial Q(F_t|F_{t-1})} = 0 \Rightarrow Q(F_t|F_{t-1}) \propto P(F_t|F_{t-1}) \prod_{i=1}^{R}
$$
$$
\prod_{C_{t-1}^i=1}^{K} \prod_{C_t^i=1}^{K} \big(P\big(C_t^i|C_{t-1}^i, L_t^i, F_t\big) P\big(Z_t^i|C_t^i\big)\big)^{Q\big(C_{t-1}^i|F_{t-1}\big)}. \tag{12}
$$

In formula (12), it can be calculated as follows:

$$
Q\big(C_{t-1}^i|F_t\big) = \int_{F_{t-1}, Q_{t-1}^i} Q\big(C_t^i|C_{t-1}^i, F_t\big) Q\big(C_{t-1}^i|, F_{t-1}\big) Q(F_{t-1}). \tag{13}
$$
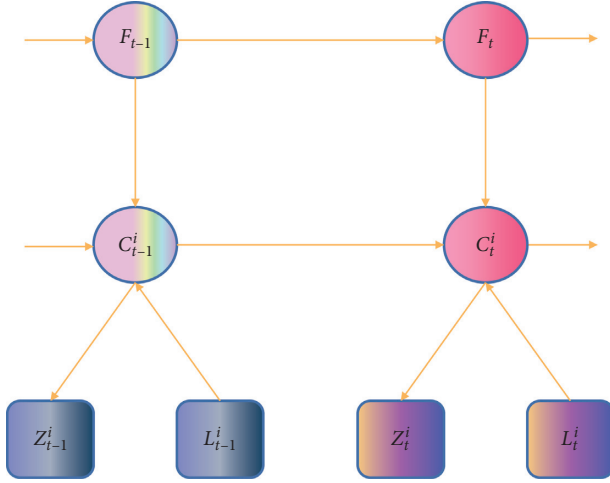
FIGURE 4: DBN model. The subscript t indicates the frame number, and the superscript i indicates the camera number.

Then, the probability density distribution can be calculated by iterative method:

$$Q(F_t) = \int_{F_{t-1}} Q(F_t|F_{t-1})Q(F_{t-1}),$$
$$Q(C_t^i) = \int_{F_t} Q(C_t^i|F_t)Q(F_t). \tag{14}$$

Finally, the LAG with the highest probability is the final reasoning result of the model.

$$\text{LAG} = \arg\max_{F_t} Q(F_t). \tag{15}$$

4. Analysis and Discussion of Experimental Results

4.1. The Efficiency Test of Visual Analysis of Learning Attention in This Paper. This paper transforms a conference room into a small classroom. The subjects sat 6 m in front of the blackboard and looked at the teachers who were writing the edition books. A front camera was installed directly above the blackboard to monitor the subjects' learning status. The rear camera is installed behind the subjects to monitor the teaching situation of teachers. In order to ensure the test accuracy, the camera was calibrated during installation.

Based on the captured video frame images, using the head pose estimation method proposed in this paper, the three-dimensional angle information and three-dimensional coordinate axis (total six-dimensional information) of the subject's head pose are calculated, in which the six-dimensional information is marked at the tip of the nose for convenience of display. Using the derived visualization method of students' learning attention based on head posture, the physical position of the gaze point of the subject can be calculated and marked on the captured video frame image.

The 1080P high-definition video taken by the front/rear camera is visually analyzed for single-person learning attention, and the video duration is 2 minutes. At the same

time, in order to test the parallel acceleration performance of the visual analysis method of learning attention in this paper, 1~4 physical threads of i5-4570 (4-core) CPU are used to process two videos in series/parallel. The experimental system uses 32 GB memory and TitanX graphics card (12 GB memory) to ensure that memory and graphics card do not become the performance bottleneck of hardware system.

Firstly, the two videos with a duration of 2 min are subjected to five steps, including frame image reading, face detection, face feature point detection, head pose estimation, and learning attention visualization, and the running time of each step and the total running time of the whole algorithm are obtained, respectively. In this paper, we first use a single thread to get the running time of serial computing. Based on this, we use 2~4 threads to get the running time of parallel computing to observe the effect of parallel acceleration (see Figure 5).

The following conclusions can be drawn from all the data listed in Figure 5.

When one thread is used for serial processing of visual analysis of students' learning attention, the processing speed of 1080P single face video is very slow, and it takes 711.58 ms to process one frame image.

Comparing the running time of each step of serial processing for visual analysis of students' learning attention, we can find that the two steps of face detection and face feature point detection are the most time consuming, which are 541.33 ms and 122.36 ms, respectively. Therefore, the key to improve the visual analysis speed of learning attention lies in how to reduce the time consumption of face detection and face feature point detection.

Using 2~4 threads for parallel computation of visual analysis of learning attention, it is found that parallel computation can effectively reduce the time-consuming of face detection, but for the other 4 steps (reading frames, face feature point detection, head pose estimation, and learning attention visualization), the acceleration effect is not obvious. In order to analyze the attention of 25–50 students in classroom teaching in real time by using 4 KB high-definition cameras in practical applications in the future, large-scale parallel processing must be carried out by computing clusters with many node machines. This method uses a single image for head pose estimation, which will be completely suitable for parallel processing of many node machines.

4.2. Head Posture Tracking Experiment Using the Visual Angle Apparent Model

4.2.1. Experiment 1. Evaluate the tracking results of visual angle apparent model when the body moves back and forth in a large range. In the experiment, a video sequence was recorded at a speed of 12 Hz by using a Digiclops stereo camera, in which the subjects moved from a position about 0.5 m away from the camera to a position about 1.5 m away from the camera along the z-axis. During the movement, the subjects constantly changed various head postures, and the rotation angle of their heads around three coordinate axes ranged from −45° to 45°. According to the effective range of
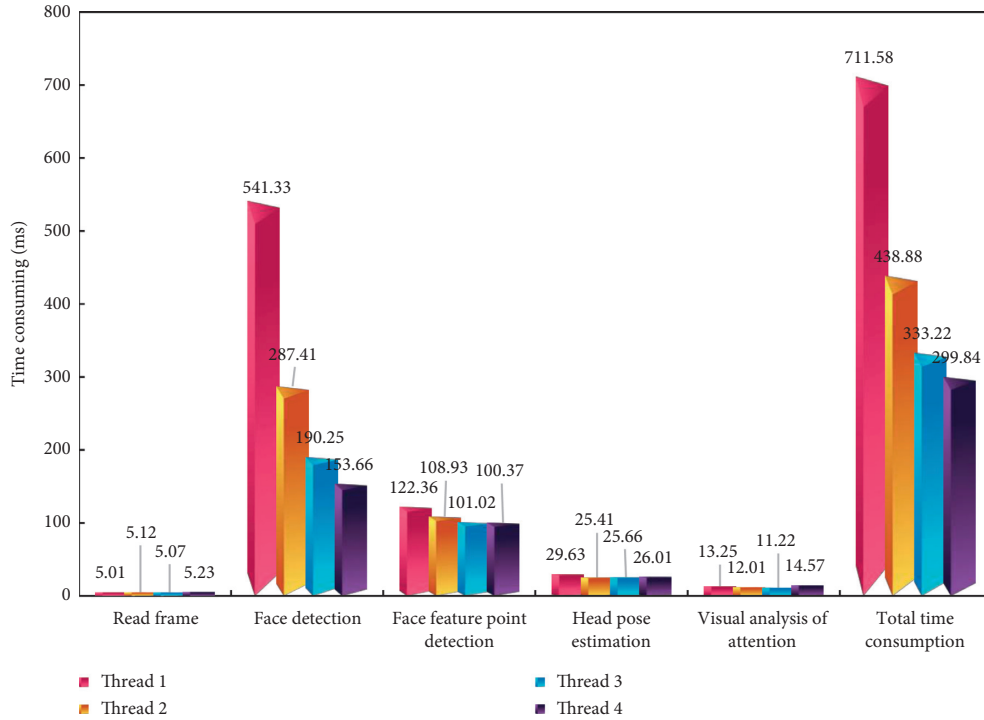
FIGURE 5: The average acceleration effect of multithreading and the average time consumption of each step of operation in this paper.

each scale visual angle apparent model, the tracking method constructs three scale visual angle apparent models, and their initial frames are about 0.5, 0.9, and 1.3 m away from the camera, respectively.

4.2.2. Experiment 2. Firstly, the tracking results of the proposed method are tested when the head leaves the camera view and then re-enters. In the experiment, another video sequence was recorded at a speed of 12 m by using the Digiclops stereo camera, in which the subjects left the camera's perspective when they were about 1.2 m away from the camera, and then re-entered the camera's perspective when they were about 0.8 m away from the camera.

4.2.3. Experiment 3. In order to further evaluate the performance of the proposed method, the tracking error of this method was measured in Experiment 3. The real head posture parameter data is obtained by using the motion sensing sensor pciBIRD. Three video sequences were recorded in the experiment, and the head pose parameters of each frame were obtained by pciBIRD. All three video sequences were recorded at a speed of 12 Hz, with an average length of 1020 frames. During recording, the motion of the tracked object was similar to that of Experiment 1, and the parameter settings were the same as those of Experiment 1.

The average tracking error (average error of three angles) using video sequence one is shown in Figure 6, in which only the tracking error of the tracked object moving from about
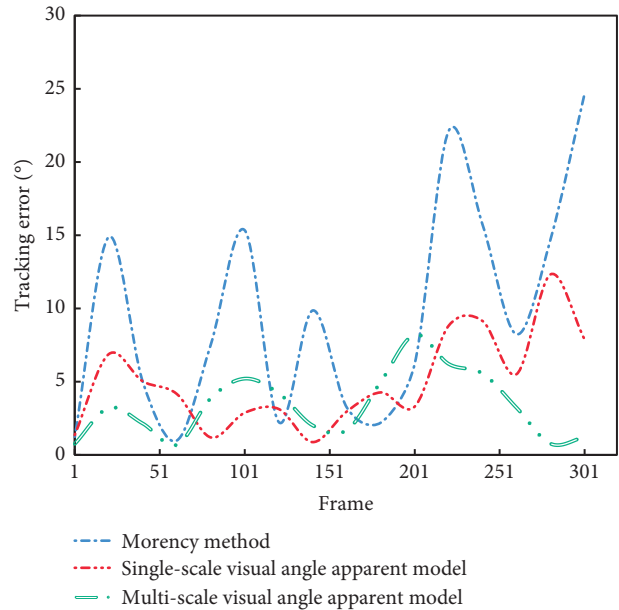
FIGURE 6: Error accumulation in tracking process.

0.9 m away from the camera to about 1.3 m away from the camera is displayed, with a total length of 320 frames.

Therefore, this method can accurately track attitude parameters (4° root mean square error), while Morency's method has a larger error (8° root mean square error), and the maximum error can reach 20°. Figure 6 also shows the error when using the single-scale visual angle apparent model. It can be seen that when the head movement exceeds the effective range of the single-scale visual angle model, the tracking error is larger.
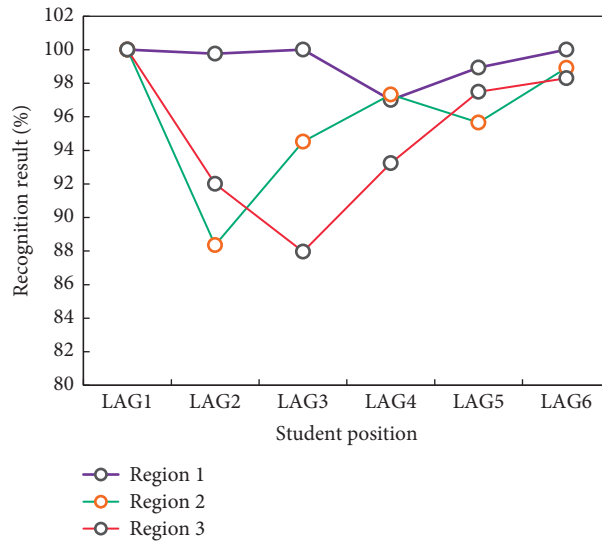
FIGURE 7: LAG recognition result: the training set and the test set are the same.

4.3. LAG Recognition.

We evaluate the validity and extensibility of the model through two sets of experiments. In the first group of experiments, we used the data of all 8 people for training and testing. In the second group of experiments, the method of row-by-row cross-validation was adopted, that is, training with the data of 7 people and testing with the data of another person. A total of 8 rounds were run, and each person's data was taken as the test set. Figures 7 and 8 are the results of two groups of experiments, respectively. The percentage in the figure is the recognition accuracy obtained by dividing the number of correctly recognized frames by the total number of frames. Accuracy evaluation is only performed on manually marked video clips. In the second group of experiments, we only consider the results of the test set video, and give the average results of 8 rounds of experiments.

As can be seen from the above figure, the result is very good. This is because the students in the training set and the test set are the same. The results of the second group of experiments are not as good as those of the first group. This is understandable, because the students in the test set have not appeared in the training set, and the appearance and illumination of different students are quite different. It can be seen that when students stand in area 1 and watch LAG4, 5, and 6, the accuracy is not very high. They often do not mistakenly identify as adjacent LAG. This is because the distance between LAG4, 5, and 6 is relatively short, and they are far away from Area 1. Therefore, when students look at these different targets, they often only turn their heads very slightly, and sometimes they only turn their eyeballs instead of their heads. If these conditions are ruled out, the experimental results are acceptable considering the difficulty of the data captured in real scenes.

4.4. Discriminant Analysis of Students' Learning Attention.

Based on the above analysis, the students' learning attention is analyzed and studied by the criterion of
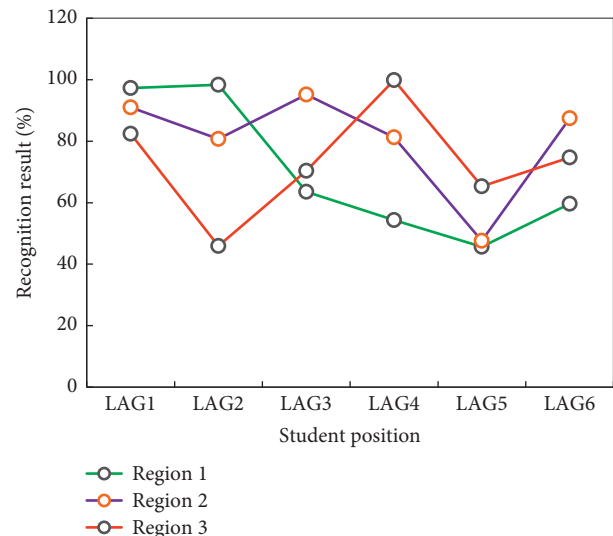
FIGURE 8: LAG recognition result: cross-validation by row one method.

students' learning attention. In order to verify the detection effect of this method, it is designed that learners imitate students' traditional classroom learning process, and test students' daily learning classroom behaviors such as irregular listening carefully, looking down at their mobile phones, and looking around. The following steps were taken: selecting a high-definition camera with 12 million pixels as an acquisition tool, fixing the camera 2 m in front of the learner, detecting the learning attention of the learner by analyzing the sampled video, and recording the learning process of the learner within 60 s under ordinary illumination.

Implementation process of the algorithm: when students sit in front of the camera, the camera will record the students' learning situation, and then each frame image of the students' learning process will be detected by the

FIGURE 9: Comparison results of pitch angles of different algorithms.

FIGURE 10: Comparison results of yaw angles of different algorithms.

FIGURE 11: Comparison results of roll angles of different algorithms.

As shown in the above figures, the displacement deviation of the head posture in $X/Y/Z$ direction by this method is within the acceptable range. The maximum error of displacement estimation in the x-axis and y-axis is less than 40 mm and 50 mm, respectively, and the maximum error of displacement estimation in z-axis is less than 230 mm. In this paper, the discriminant analysis of students' learning attention is accurate, which can basically coincide with the calibration curve, and the deviation is small.

5. Conclusion

Online education is a new trend in the development of education, which will bring profound changes in educational concepts, educational systems, teaching methods, personnel training models, etc., and play a positive role in deepening education and teaching reform, improving education quality, and promoting education equity. In this paper, a discrimination method of learning attention based on head posture analysis is proposed. By selecting key frames with different head postures and generating them online, besides attaching posture parameters to the key frames, the head region is accurately extracted from each key frame as the head perspective, and the key frames are combined into a multiscale visual angle apparent model according to the spatial distribution. A DBN model is proposed to reason students' LAG. The model integrates the relationships among multi-LAG, multistudent positions, and multicamera face images, and conducts joint reasoning. We measure the head posture by the similarity vector between the face image and multiple face categories without explicitly calculating the specific head posture value. We collected test data in the teaching environment, and the experimental results show that our model can get better results.

learning attention detection system, and the information of the students' head rotation will be recorded, in which one corner out of range will be recorded as 1, and the one without exceeding will be recorded as 0. First, if the record is 1 continuously within 2 seconds, it will be judged as learning distraction and counted. Secondly, the ratio of the sum recorded as 1 per unit time to the total time will be calculated and the students' classroom learning attention will be output. From this, we can count the number of distractions in learning and the situation of students' learning attention per unit time. Some typical behaviors are shown in Figures 9–11.

In the current experiment, when the user looks at the distant visual attention target, the attitude measurement is inaccurate due to the small difference of images. In the future, we will consider fusing motion information to detect the change of the user's visual attention target. The visual attention target in this paper is several screens on the wall. In the future, we will consider extending the visual attention target to more places, such as different areas on the workbench.

Data Availability

The dataset used in this paper are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

References

[1] G. Otte, "Online instruction as local education: cuny's online baccalaureate," *Journal of Asynchronous Learning Networks*, vol. 11, no. 1, pp. 9–14, 2019.

[2] W. B. Root and R. A. Rehfeldt, "Towards a modern-day teaching machine: the synthesis of programmed instruction and online education," *Psychological Record*, vol. 71, no. 1, pp. 85–94, 2021.

[3] B. B. Lockee, "Online education in the post-covid era," *Nature Electronics*, vol. 4, no. 1, pp. 5-6, 2021.

[4] J. Zhang and F. Zhang, "Empirical research on factors affecting users' willingness to continuously use online education platform," *Value Engineering*, vol. 38, no. 5, pp. 134–137, 2019.

[5] E. Fehr, "Ecologies of faith in a digital age: spiritual growth through online education," *Growth: The Journal of the Association for Christians in Student Development*, vol. 18, no. 18, p. 11, 2019.

[6] D. Mcauliffe, "Challenges for best practice in online social work education," *Australian Social Work*, vol. 72, no. 1, pp. 110–112, 2019.

[7] V. Kearney, J. W. Chan, T. Wang et al., "Dosegan: a generative adversarial network for synthetic dose prediction using attention-gated discrimination and generation," *Scientific Reports*, vol. 10, no. 1, Article ID 11073, 2020.

[8] H. Datta, A. Hestvik, N. Vidal et al., "Automaticity of speech processing in early bilingual adults and children," *Bilingualism: Language and Cognition*, vol. 23, no. 2, pp. 429–445, 2020.

[9] M. Bce, S. Staal, and A. Bulling, "How far are we from quantifying visual attention in mobile hci?" *IEEE Pervasive Computing*, vol. 19, no. 2, pp. 46–55, 2020.

[10] R. Hamrah, R. R. Warier, and A. K. Sanyal, "Finite-time stable estimator for attitude motion in the presence of bias in angular velocity measurements," *Automatica*, vol. 132, no. 2, Article ID 109815, 2021.

[11] C. Zhou, Q. Chen, Z. Li, B. Zhao, Y. Xu, and Y. Qin, "Aspect category detection based on attention mechanism and Bi-directional LSTM," *Xibei Gongye Daxue Xuebao/Journal of Northwestern Polytechnical University*, vol. 37, no. 3, pp. 558–564, 2019.

[12] I. Zhao, "An improved attitude estimation algorithm based on mpu9250 for quadrotor," *Mechanical Engineer*, vol. 5, pp. 36–39, 2019.

[13] M. Yeager, B. Gregory, C. Key, and M. Todd, "On using robust mahalanobis distance estimations for feature discrimination in a damage detection scenario," *Structural Health Monitoring*, vol. 18, no. 1, pp. 245–253, 2019.

[14] B. Yfka, B. Lyn, and B. Qzy, "Fluorescent probes for detection of biothiols based on "aromatic nucleophilic substitution-rearrangement" mechanism," *Chinese Chemical Letters*, vol. 30, no. 10, pp. 1791–1798, 2019.

[15] M. Hirata and N. Kusatake, "How cattle discriminate between green and dead forages accessible by head and neck movements by means of senses: reliance on vision varies with the distance to the forages," *Animal Cognition*, vol. 23, no. 2, pp. 405–414, 2020.

[16] Y. Guo, J. Zhang, and W. Lian, "Study on learning attention discrimination based on head posture," *Science Technology and Engineering*, vol. 20, no. 14, pp. 5688–5695, 2020.

[17] Y. Lu and L. Yanmin, "Design of online classroom attention evaluation system based on realsense," *china medical education technology*, vol. 34, no. 3, pp. 82–86, 2020.

[18] S. Xiao and S. Nan, "Head pose estimation of 3D point cloud based on deep learning," *Computer Applications*, vol. 40, no. 4, pp. 72–77, 2020.

[19] Y. Ren, S. Su, K. Guan, W. Fu, and X. Zhang, "Detection and fatigue analysis based on oculogram and head posture signals," *Acta changchun university of science and technology: Natural Science Edition*, vol. 1, pp. 38–44, 2020.

[20] S. Li, "An analysis of effective strategies to maintain the attention of primary school students in classroom learning," *Chinese class*, vol. 9, p. 65, 2020.

[21] L. Liang, T. Zhang, and H. Wei, "Head pose estimation based on multiscale convolution neural network," *Advances in Laser and Optoelectronics*, vol. 56, no. 13, 2019.

[22] J. Cui and J. Wang, "Facial expression recognition model based on enhanced head pose estimation," *Computer Science*, vol. 46, no. 6, pp. 328–333, 2019.

[23] X. Liu, "Analysis of children's attention in preschool education," *Encyclopedia Forum Electronic Journal*, vol. 3, pp. 662-663, 2020.

[24] Z. Zeng, X. Lu, S. Xu, and M. Chen, "False comment detection based on deep learning model of multi-layer attention mechanism," *Computer Applications and Software*, vol. 37, no. 5, pp. 183–188, 2020.

[25] T. Wu and C. Chunping, "Emotion analysis model of long-term and short-term memory based on attention crossover with position weight," *Computer Applications*, vol. 39, no. 8, pp. 2198–2203, 2019.

[26] J. Hu, W. Zhang, and S. Chen, "Research on attribution of attention deficit in college students fragmentation learning-qualitative analysis based on grounded theory," *Audio-visual Education Research*, vol. 40, no. 12, pp. 36–43, 2019.

Retraction

Retracted: Design of Online Music Teaching System Based on B/S Architecture

Scientific Programming

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope

(2) Discrepancies in the description of the research reported

(3) Discrepancies between the availability of data and the research described

(4) Inappropriate citations

(5) Incoherent, meaningless and/or irrelevant content included in the article

(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] D. D. Dai, "Design of Online Music Teaching System Based on B/S Architecture," *Scientific Programming*, vol. 2021, Article ID 1297019, 6 pages, 2021.

Research Article

Design of Online Music Teaching System Based on B/S Architecture

Dan Dan Dai

Nanchang JiaoTong Institute, Nanchang 330100, Jiangxi, China

Correspondence should be addressed to Dan Dan Dai; 06054@ncjti.edu.cn

Under the global epidemic situation, offline face-to-face teaching has not adapted to the current teaching environment. Using computer technology to carry out online assisted teaching has become an inevitable choice for music teaching in colleges and universities. Offline teaching also has many disadvantages, such as single and backward teaching methods, poor sharing of audio-visual resources, and inconsistent teaching contents. The survey found that music teachers urgently need an online teaching platform to enable teachers to obtain effective, standardized, and comprehensive teaching resources anytime and anywhere so that students can enjoy high-quality music teaching resources as much as possible. Based on these requirements, the development of online music teaching system under B/S architecture has an important practical significance.

1. Introduction

In the 1990s, China had officially put systematic music and art education on the agenda, and all regions of the country have increased music courses in senior high school, which is also inseparable from the promotion of quality education. Meanwhile, the financial investment of art colleges and universities is also gradually increasing. There is no difference in the technical level between domestic music websites and developed countries, and the content is still dominated by pop music websites. According to the survey, 63.2% of China's music websites are pop music websites and classical, ethnic, and other types of music account for a relatively small proportion, but they basically meet the needs of nonart professional audiences. However, there are four deficiencies in domestic music websites. First, the audience group is small. Because most of China's music websites focus on pop music, the audience group is mostly young people. Second, the attention of music websites is one-sided. Most music websites are driven by interests and actively cater to public tastes. Third, the role of "gatekeeper" of network media is missing. Fourth, the construction of network laws and regulations is not perfect, and piracy and infringement are common. There are few people studying the music teaching

assistant platform in China for three reasons. First, the use of the platform is small. The platform mainly serves music teachers, and music teachers are a small group. Second, the platform cannot bring economic benefits and needs the investment of colleges and universities. Third, the R&D of the platform requires comprehensive talents with certain music attainments and professional knowledge such as computer network and database. Such talents are very few. In 2008, Liaoning Normal University made an attempt to assist research in music teaching, mainly using VFP visual programming to make a teaching music database management program. Due to the lack of special personnel to maintain its database, fewer and fewer teachers are used in the program. In 2011, Shenyang Conservatory of music began to study the music teaching auxiliary platform based on C/S architecture. The teaching content of this auxiliary platform meets the needs of vocal music teachers. Each track included in the platform contains the detailed information of the author of the track and the detailed information of the representative singer, the staff, and other information. However, because the platform based on C/S architecture needs to install and configure software on each computer and the software upgrade is very troublesome, only some young teachers use the platform [1–6].

With the development and application of multimedia and Internet, music appreciation course has also produced revolutionary development. Music lessons have expanded from a single piece of chalk, a book, and a mouth to more and more colorful forms, such as the application of a variety of music scores in the teaching process, simple score and staff score, the tour broadcasting of multimedia teaching materials, and the comparison of different performance forms of the same track. Therefore, multimedia music teaching is an epoch-making milestone in improving the quality of music teaching. However, there are still single and backward ways of appreciation in the music appreciation course, such as a piece of chalk, a book, and a small speaker in the music class, so that students often only know it but do not know why, so they are more at a loss when appreciating works [7–9]. Moreover, in the current stage of music appreciation teaching, teachers often download the works to be appreciated and put them on the USB flash disk, which can be appreciated by students through audio-visual equipment. Different teachers will choose different works to enjoy. The sharing of audio-visual appreciation resources is relatively weak, the classification of appreciation works is chaotic, and the choice of appreciation works is not unified. Third, in the process of preparing lessons through the network, due to the particularity of some music, such as less creative background, music style, simplified staff, and staff resources, teachers create obstacles in the process of preparing lessons and reduce the depth of students' appreciation of music [10, 11]. In online teaching, students' autonomous learning is stronger, and the arrangement of their learning tasks can be determined according to their own needs.

In view of the current teaching difficulties, a music teaching system based on B/S structure is designed. It has the characteristics of convenience, resource sharing, openness, and professionalism. It provides more choices for the teaching of music teachers and allows students to experience a more convenient teaching environment.

2. B/S Architecture Introduction

2.1. B/S Tertiary Structure Concept.

B/S structure is an improved model of the previous traditional C/S structure after the rise of the Internet. It has no independent client, but depends on the browser. All users operate through the browser, the maintenance and use of the system become simple, and the development process can be simplified [9, 12]. The client only needs to install any type of browser, and all databases are installed on the server. This can greatly simplify the client computer review, and the system maintenance and upgrade do not need to be completed by installing the upgrade package like the C/S structure. Users no longer need to install a separate client login working interface, but log in to the browser. The detailed B/S architecture diagram is shown in Figure 1.

2.2. Characteristics of the Three-Layer Structure of the B/S Architecture.

Postmaintenance and upgrade simplification, during the use of the software, are necessary to continuously
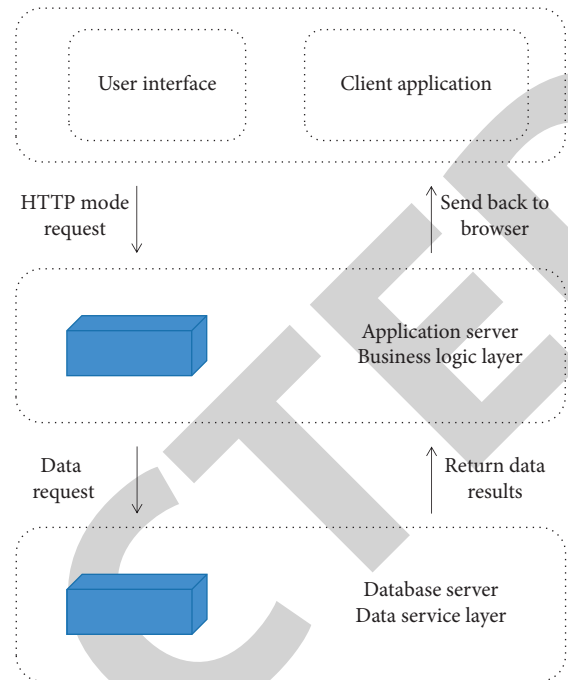
FIGURE 1: B/S architecture.

upgrade to meet the use needs of users. Compared with the C/S structure, the B/S structure shows strong convenience in upgrading and maintenance. In order to meet the functional needs, the software system needs to be continuously improved and upgraded. If the C/S architecture is adopted, hundreds of computers need to install the upgrade package, resulting in huge workload and low work efficiency. If the B/S three architecture is adopted, the software engineer can maintain the server. All users can log in to the browser after unified maintenance and enjoy the upgraded service. Meanwhile, since the upgrade and maintenance are only for server operations, the B/S architecture enables the system to support remote maintenance operations [13–16]. Therefore, the mainstream direction of the current information development is that the server is becoming more and more "fat," while the user machine is becoming more and more "thin." Under such software and hardware conditions, the system upgrade and maintenance will be easier and easier, and the user operation will develop to simplification, which can greatly save the user's human, material, and financial resources.

Cost reduction: Windows has almost become the mainstream of PC operating system, and almost every PC is equipped with browser, but Windows is not the mainstream operating system of the server. For security reasons, most B/S applications are installed on Linux servers. There are many options for the server operating system, but no matter which option, users can safely use windows as the PC operating system [17]. In addition to being free, the database of Linux operating system is also free, so the choice of Linux as a server operating system has become very popular.

The load is concentrated on the application server. The databases of this structure mode are concentrated on the server, and the logical things are reflected in the browser. It

does reduce the pressure on the users of the system, but the pressure reduced by the users is passed on to the server. In this mode, the backup management of the database must be done well [18–21]. Generally, the regular backup work is done by a separate large memory to prevent the system collapse and all data loss.

3. System Design

3.1. Overall System Architecture Design.

The schematic diagram of B/S architecture is shown in Figure 2. System users are divided into two categories: one is an administrator and the other is a teacher, who can log in to the system through the network. The responsibility of the system administrator is to input and modify the classified directory, review the teaching tracks and other teaching materials uploaded by teachers, and review the system feedback, so as to ensure the quality of documents in the system. Teachers log in to the system by entering their own account, password, and verification code. They can enjoy teaching tracks and upload teaching music and other teaching materials online during teaching. After being approved by the administrator, they will be able to automatically enter into the system to share with other teachers. The teaching tracks uploaded by each teacher are entered into the system, and teachers can freely download the tracks and other teaching materials uploaded by other teachers. Teachers can express their own unique opinions on a teaching track. Teachers can share classical and helpful tracks and other teaching materials with other music teachers. Firstly, the system needs to ensure stability and realize efficient operation on this basis. The database cluster requires two servers for cold backup and database reading. The operation mode of the whole architecture is that the operation request of the system user is sent to the reverse proxy server through the network and then sent to the database cluster through the Web cluster. The design of such architecture can not only ensure stability but also ensure efficiency [22–25].

3.2. Frame Construction.

Stability is the premise of framework implementation. This framework adopts MVC mode (action, model, and view). The most important is action. The user's request is completed through the action controller, either directly processed or forwarded. Each action of the user needs to pass through the action controller first, that is, it is the core of the whole structure. For example, after the user clicks the course material under the personal user and sends this request, the action controller first determines whether the user has this permission. If so, call the data in the database cluster, and view displays the course material interface.

3.3. System Function Module.

Teachers log in, upload and download teaching tracks and other teaching materials, enjoy teaching tracks online, and share teaching tracks and other teaching materials, and teacher comments, teacher feedback, and advanced search are all concentrated on the front desk. The foreground function of the system is mainly to serve teachers and users, and most of the key functions of
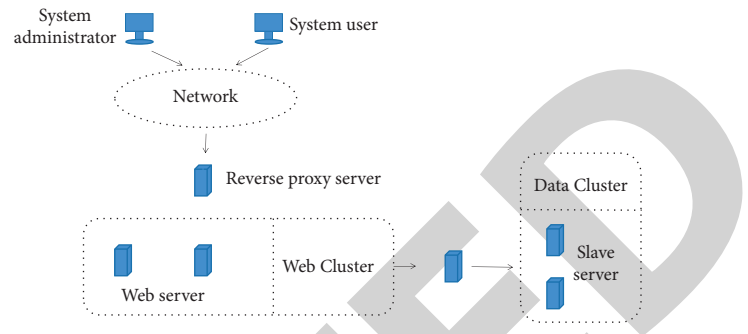
Figure 2: Schematic diagram of the B/S architecture.

the system are reflected in the foreground. After logging into the system, teachers can browse the classification of teaching music, the corresponding representative works under each classification, and authors and other teaching materials related to works. Teachers can upload tracks and other teaching materials required for teaching and can also share classic teaching tracks with other music teachers through SNS. In the teaching process, teachers have their own unique ideas about a teaching track and can comment. In the course of teaching, teachers will sort out the inconveniences of the system into feedback and brainstorm and make the system function more perfect.

The operation of the system administrator is mainly concentrated in the background. The system administrator is responsible for the creation of classification directory when the system is just launched and the maintenance of system directory in the later stage. The system administrator reviews whether the teaching tracks and other teaching materials uploaded by teachers are suitable for music teaching and whether they belong to correct classification. The system administrator reviews whether teachers' comments comply with relevant regulations. To ensure that the teaching tracks and other teaching materials in the system are classified correctly, the sound quality is clear and conducive to teaching. When teachers find the defects and deficiencies of the system in the process of teaching and using the system, they can feed back their opinions to the system, and then, the system administrator will uniformly deal with the information fed back by teachers, for example, timely reply to teachers' feedback information and opinions and feedback problems to system developers. The system background function structure is shown in Figure 3.

3.4. Main Module Design of the System.

The system mainly has the following functional modules: login management, teaching track directory management, teaching track management, teaching track sharing management, audit management, teacher comments, user feedback, teaching material upload and download, and search (advanced search).

Login management: both system administrators and ordinary teachers have their own login accounts. Only users who enter the correct user name and password can log in to the home page of the system, which is different from many application-based websites.
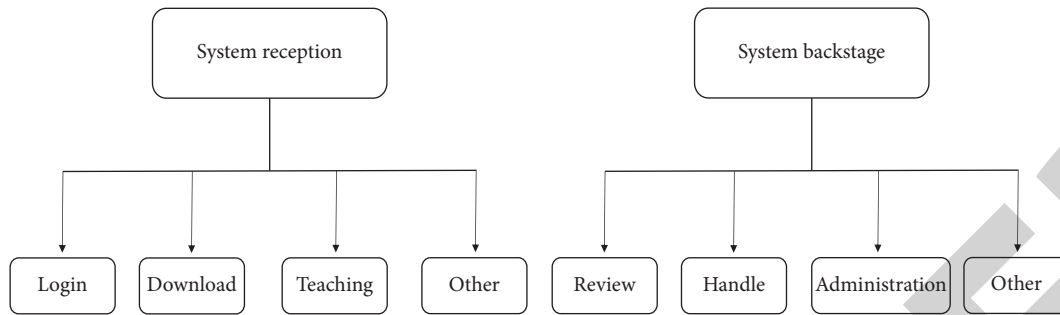
FIGURE 3: System reception and background function diagram.

Teaching track directory management: when the preliminary design of the system is successful, the system administrator is responsible for completing the system initialization. The initialization of the system is mainly to input the existing teaching music classification and corresponding representative works and representatives into the system according to the current general music syllabus. If the classification standard changes in the later stage, you can modify, delete, or add the classification directory or the corresponding content in the directory.

Management of teaching tracks: multimedia teaching should at least ensure that music teachers can play audio materials of teaching tracks in the process of teaching. With the support of this platform, music teachers can choose from two kinds of playing forms. The first is to find the teaching tracks and teaching materials needed for teaching in the system, download them to the mobile hard disk or personal computer, and play them with any playing software that can play MP3 format. The second is to find the required teaching audio materials and play them online directly.

Teaching track-sharing management: when music teachers think that a teaching track has special teaching significance or have their own unique indirect on the teaching method of a teaching track, they can share it with their colleagues so that other music teachers can discuss the teaching method of this track together. There are many ways to share. One is to share with designated teachers through sharing tools, and teachers who receive sharing materials can see the shared content. However, to share in the group designated by the music teacher, the teachers in the group can see the shared materials. Third, share in your own space. Friends on the sharing tool can see the shared content.

Audit management: during the teaching process, ordinary teachers will upload some teaching tracks according to the situation of the teaching song library in the system and their own teaching needs, and the corresponding teaching tracks need to be equipped with staff, simplified music, and other materials; At the same time, some comments and feedback will be uploaded according to the specific situation, which will be reviewed by the administrator. Review whether the uploaded tracks belong to a directory. For example, the Folk Song Directory uploaded by Pavarotti's my sun to the song cannot pass the review. If my sun is uploaded to the directory of Bel Canto under the song, it can pass the review. This function ensures the correctness of teaching tracks and other teaching materials in the system and the good use environment of the system. For music teaching tracks and other teaching materials uploaded by teachers, the comments of teachers and users can be passed only after review.

Upload and download of teaching materials: when the system was just built, there were few tracks in it. Teachers can gradually add teaching tracks under a subdirectory or upload other teaching materials corresponding to teaching tracks, such as staff, simplified music, and creative background. The teaching materials uploaded by music teachers are not included immediately after uploading, but need to be reviewed by the system administrator. After approval, it can be used and downloaded by other music teachers.

If the correctness and rationality of teaching tracks and other teaching materials pass the examination, they will be employed. If it is correct and reasonable, it will be entered into the system, and other teachers and users can enjoy or download teaching tracks and other teaching materials online. If the approval fails, the system administrator will directly delete the uploaded teaching tracks and other teaching materials and send a notice to the teacher user. The uploaded teaching tracks and other teaching materials are incorrect or unreasonable, and the upload fails.

Search: when music teachers need to view a specific teaching resource, searching one by one in the directory list is inefficient and cumbersome, so the search function is particularly important. For general teaching resources, teachers can directly enter the resource name in the upper right corner. In addition, advanced retrieval can be carried out according to the creation time, author, music category, performer, content search conditions, and other keywords of the track.

4. Feasibility Analysis

4.1. Technical Viability. The teaching auxiliary platform is developed by lamp. Lamp is a very good group of software on the production website, and the application technology is also relatively advanced. The operating system is Linux, the programming language is PHP, and the database uses MySQL management system. Lamp is the first choice for the development of music teaching system for many reasons. First, its performance can fully meet the requirements of the platform. Second, it shows great advantages in the richness of resources. Third, from the perspective of system

development cost, it is wise to adopt this software. Finally, it has unparalleled advantages in cross-platform features.

4.2. Economic Viability. The whole construction process of the platform is divided into two steps. The first step is the initial construction of the platform. At this time, teachers have not participated in it, and a software development department has completed the initial construction. At this time, a very complete database construction is not required. The second step is the construction of the platform improvement process. At this time, teachers and users need to participate in it and constantly upload teaching tracks and teaching materials in the process of use, so as to enrich the content of the database day by day. Because the first step of the construction of this platform does not need to be too complex, it does not need too large database construction. It only needs a platform with basic functions and basic classification. Therefore, the establishment cost of music teaching platform is relatively low [26, 27].

4.3. Operation Viability. In order to meet the normal operation of the foreground and background and realize the functions of users, two types of users must exist at the same time. The first is the system administrator managing the background, and the other is the teacher (whether to join the students needs further practice and research). The managers have three responsibilities: website management and information release, reviewing the materials uploaded by teachers and users, and deleting remarks that do not comply with laws and regulations. Managers realize the management of the website and the release of information through the operation interface. Managers only need to have a preliminary understanding of the classification of teaching tracks to complete the task of reviewing the data uploaded by teachers and users. The third responsibility of managers is to delete illegal remarks, which can still be operated simply. Teachers and users can upload teaching tracks and share relevant teaching materials. They only need to master simple Internet technology.

5. Conclusion

Online music teaching system based on B/S architecture has many advantages. Teachers can quickly find teaching resources through the system and share resources with students through simple operation. Multiple teachers can share resources through the system to reduce offline resource allocation. During use, teachers and students can discuss and communicate in relevant chapters of the course to reduce offline communication and notes in class and build an online teaching platform in B/S architecture to increase the diversity of teaching and bring more choices for music teaching under the epidemic situation.

Data Availability

The dataset can be obtained from the corresponding upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

Acknowledgments

References

[1] M. A. Akbar and I. Handriani, "Study and implementation information system of zakat using MVC architecture," *IOP Conference Series: Materials Science and Engineering*, vol. 453, no. 1, pp. 13–18, 2018.

[2] F. Elmaz, R. Eyckerman, W. Casteels, S. Latré, and P. Hellinckx, "CNN-LSTM architecture for predictive indoor temperature modeling," *Building and Environment*, vol. 206, Article ID 108327, 2021.

[3] G. Cheng, A. Matsune, H. Du, X. Liu, and S. Zhan, "Exploring more diverse network architectures for single image super-resolution," *Knowledge-Based Systems*, vol. 235, Article ID 107648, 2022.

[4] F. Robert Jacobs and E. Bendoly, "Enterprise resource planning: developments and directions for operations management research," *European Journal of Operational Research*, vol. 146, no. 2, pp. 14-15, 2003.

[5] C. Franke, S. Morin, A. Chebotko, J. Abraham, and P. Brazier, "Efficient processing of semantic web queries in HBase and MySQL cluster," *It Professional*, vol. 15, no. 3, pp. 36–43, 2013.

[6] X.-M. Zhang, J.-Y. Yu, Y. Yuan, C.-P. Feng, L. Jing, and S.-L. Xu, "A flipped classroom method based on a small private online course inphysiology," *Advances in Physiology Education*, vol. 3, no. 43, pp. 345–349, 2019.

[7] S. Uijl, R. Filius, and O. Ten Cate, "Student interaction in small private online courses," *Medical Science Educator*, vol. 27, no. 2, pp. 237–242, 2017.

[8] C.-H. Lai, H.-W. Lin, R.-Mu Lin, and P. D. Tho, "Effect of peer interaction among online learning community on learning engagement and achievement," *International Journal of Distance Education Technologies*, vol. 1, no. 17, pp. 66–77, 2019.

[9] M. Stoytcheva, "Collaborative distance learning: developing an online learning community," *AIP Conference Proceedings*, vol. 1910, no. 1, pp. 1–8, 2017.

[10] Winanti, F. L. Gaol, F. L. Meyliana, and H. Prabowo, "A survey positive engagement of learning community for informal education to support community," *IOP Conference Series: Materials Science and Engineering*, vol. 662, Article ID 022024, 2019.

[11] B. Hudson, "Developing an open and flexible networked learning community at doctoral level across Europe: national Teaching Fellowship update," *MSOR Connections*, vol. 5, no. 1, pp. 1–4, 2021.

[12] B. Kim, "Things in common in learning communities," *Instructional Science*, vol. 46, no. 3, pp. 627–631, 2018.

[13] J. Fleck, "Blended learning and learning communities: opportunities and challenges," *The Journal of Management Development*, vol. 31, no. 4, pp. 398–411, 2012.

[14] S. Cosmin Nistor and G. Czibula, "IntelliSwAS: optimizing deep neural network architectures using a particle swarm-

Research Article

Recognition of Power Equipment Based on Multitask Sparse Representation

Lei Lei,[1,2] **Jian Wu** ![ORCID] **,**[1,2] **Shuhai Zheng,**[3] **Xinyi Zhang,**[4] **Liang Wang,**[1,2] **Yanfei Wang,**[5] **and Hao Wan**[1,2]

[1]*State Grid Shaanxi Electric Power Science Research Institute, Xi'an, China*
[2]*State Grid (Xi'an) Environmental Technology Center Co., Ltd, Xi'an, China*
[3]*State Grid Co., Ltd. DC Construction Branch, Beijing, China*
[4]*State Grid Shaanxi Electric Power Xi'an Power Supply Company, Xi'an, China*
[5]*Sichuan Hengchuang Tiandi Automation Equipment Co., Ltd., Chengdu, China*

Correspondence should be addressed to Jian Wu; jwu.sxsp@yahoo.com

Image analysis of power equipment has important practical significance for power-line inspection and maintenance. This paper proposes an image recognition method for power equipment based on multitask sparse representation. In the feature extraction stage, based on the two-dimensional (2D) random projection algorithm, multiple projection matrices are constructed to obtain the multilevel features of the image. In the classification process, considering that the image acquisition process will inevitably be affected by factors such as light conditions and noise interference, the proposed method uses the multitask compressive sensing algorithm (MtCS) to jointly represent multiple feature vectors to improve the accuracy and robustness of reconstruction. In the experiment, the images of three types of typical power equipment of insulators, transformers, and circuit breakers are classified. The correct recognition rate of the proposed method reaches 94.32%. In addition, the proposed method can maintain strong robustness under the conditions of noise interference and partial occlusion, which further verifies its effectiveness.

1. Introduction

With the continuous increase of power equipment, traditional manual-line inspection and substation monitoring have been difficult to meet the actual requirements. In this context, a large number of power-line inspection equipment types based on helicopters, drones, and other computational platforms have been put into application [1–5]. These devices collect images of power equipment through optical and infrared sensors on themselves or nearby. Then, the image analysis and other technical means can be used to determine possible faults in the power equipment. Therefore, it is of great significance to carry out the analysis and interpretation of power equipment images. Image recognition of power equipment is one of the branches in the field of image analysis of power equipment. The basic idea is to classify the collected power equipment on the basis of the existing

database, so as to provide a prerequisite for the targeted analysis of the special type of equipment. At present, there are still few studies in this field. In fact, the power equipment image recognition problem is similar to the traditional image-based target recognition problem and basically uses two stages of feature extraction and classifiers. In the process of feature extraction, according to the characteristics of power equipment images, geometric shape features, image gray distribution features, and local texture features are developed and employed [6–12]. Afterwards, an appropriate classifier is selected to process the extracted features and determine the corresponding category. Commonly used classifiers in image recognition of power equipment include support vector machines (SVMs) and sparse representation-based classification (SRC). In recent years, the deep learning models represented by the convolutional neural network (CNN) have become a powerful tool in the field of image

processing [13–16] and have also been widely used and verified in the image recognition of power equipment [17–21].

The rapid development of modern pattern recognition technology provides a large number of tools, which can be well used for reference in the recognition of power equipment images. In recent years, the compressive sensing (CS) technology has developed rapidly and widely used in the fields of signal processing, image analysis, and other relevant applications. Specifically, in the field of image recognition, SRC is a representative classifier based on the compressed sensing theory, which has been successfully applied in face recognition and remote sensing image recognition [22–24]. This paper makes optimizations in the feature extraction and classifier stages based on the CS theory and develops an image recognition method of power equipment. In the process of feature extraction, the two-dimensional (2D) random projection is used to reduce the dimensionality of the original images [25]. Through multiple random projection matrices, the characteristics of the image can be obtained from different aspects to achieve complementary enhancement. 2D random projection is an extension of traditional one-dimensional random projection, which can effectively maintain the structural characteristics of 2D signals such as images. At the same time, the algorithm inherits the advantages of random projection, which needs no prior parameters and has high computational efficiency. Through the comprehensive operation of multiple 2D random projection matrices, it can play a complementary role, so as to provide more comprehensive feature descriptions for the input images. In the classification stage, the multitask compressed sensing (MtCS) [26] is used to jointly reconstruct and analyze the multilevel projection features. MtCS is a typical multitask joint sparse representation algorithm, which can be used to analyze the internal correlation of multiple sparse representation tasks, so as to improve the accuracy of the solution [27–31]. In the field of target recognition, based on the solved sparse representation coefficients, the test samples can be reconstructed by different classes, so as to make decisions based on the reconstruction error. MtCS combines the principles of CS and Bayesian theory to obtain the largest a posteriori solution in the theoretical sense. In particular, due to the basic principles of CS and Bayesian solution, the algorithm can effectively overcome the influence of noise and other interference factors and can better handle the changes in lighting conditions, sensor noise, and other factors that may be encountered during the image acquisition process of power equipment. In order to verify the proposed method, images of insulators, transformers, and circuit breakers are used as basic samples in the experiments. And, noise interference and partial occlusion conditions are further constructed for testing. The experimental results show the effectiveness and robustness of the proposed method.

2. 2D Random Projection for Feature Extraction

Traditionally, the projection features are extracted by first stacking the 2D image $X \in R^{n_1 \times n_2}$ into a vector as $x = \text{vec}(X)$. Such operation actually benefits the processing afterwards but inevitably corrupts the structural information of the image. In order to maintain the structural properties of images, the 2D projection algorithm can be employed like $Y = AXB^T$, in which $Y \in R^{m_1 \times m_2}$ denotes the resulted feature and $A \in R^{m_1 \times n_1}$ and $B \in R^{m_2 \times n_2}$ are the projection matrices with $m_1 \ll n_1$ and $m_2 \ll n_2$. The key in 2D projection lies on the design of the projection matrices, which influences the validity of the final features.

According to the CS theory, the sparse signal can be well reconstructed by a small number of measurements. In a similar idea, the sparse matrix can be reconstructed by the low-dimensional matrix. In this sense, the CS theory provides a simple and effective way to extract features from 2D sparse matrices [25]. With the projection features $A \in R^{m_1 \times n_1}$ and $B \in R^{m_2 \times n_2}$, the process of feature extraction can be described as $Y = AXB^T$, where $X \in R^{n_1 \times n_2}$ is the input image.

In order to guarantee the fidelity after feature extraction, the dimensions of the result Y, i.e., m_1 and m_2, should be properly chosen. According to [25], the requirements for m_1 and m_2 are as follows:

$$\|X\|_0 < \frac{\text{spark}(A)\text{spark}(B)}{4},$$

$$\left(\ell^0 \text{ norm of each column of } X\right) < \frac{\text{spark}(A)}{2}, \quad (1)$$

$$\left(\ell^0 \text{ norm of each row of } X\right) < \frac{\text{spark}(B)}{2},$$

where $\|X\|_0$ denotes the ℓ^0 norm of X and the spark of a matrix represents the minimal number of its columns which are linearly dependent. For the random projection matrices $A \in R^{m_1 \times n_1}$ and $B \in R^{m_2 \times n_2}$, they comply to $\text{spark}(A) = m_1 + 1$ and $\text{spark}(B) = m_2 + 1$.

Owing to the merits of 2D random projection, this paper employs it for feature extraction of images of power equipment. Specially, multiple random projection matrices are developed to generate multiple feature vectors, which provide complementary descriptions of the target to be analyzed and classified.

3. MtCS for Classification

3.1. Basics of MtCS. It is easy to understand that the multiple measurements from the same source are statistically related like multichannel signals and multiview signatures [27–31]. MtCS is a multitask learning algorithm based on CS and Bayesian theory, which could consider the correlations of several related tasks to achieve high reconstruction precision.

Denote the L measurements from the same source as $\{y_i\}_{i=1,\dots,L}$; they are expressed as follows:

$$y_i = A_i x_i + n_i, \quad (2)$$

where $A_i \in R^{N_i \times N}$ denotes the dictionary corresponding to ith measurement and n_i represents a zero-mean Gaussian process with the variance of ξ_0.

The likelihood function of y_i is modeled as follows:

$$p(y_i \mid x_i, \xi_0) = \left(\frac{2\pi}{\xi_0}\right)^{-N_i/2} \exp\left(-\frac{\xi_0}{2}\|y_i - A_i x_i\|_2^2\right). \quad (3)$$

The parameters $x_i\,(i = 1, 2, \ldots, L)$ are kept unchanged for all the L tasks. The L tasks are assumed to be statistically related as follows:

$$p(y_i \mid \xi, \xi_0) = \prod_{j=1}^{N} N\left(x_{i,j} \mid 0, \xi_j^{-1}\xi_0^{-1}\right), \quad (4)$$

where $x_{i,j}$ is the jth element of x_i and $\xi = [\xi_j, \ldots, \xi_N]^T$. The Gamma priors are put on the parameter ξ_0 as follows:

$$p(\xi_0 \mid a, b) = \text{Gamma}(\xi_0 \mid a, b). \quad (5)$$

With the choices of ξ and y_i, the posterior density function (PDF) of x_i can be calculated as follows:

$$p(x_i \mid y_i, \xi) = \int p(x_i \mid y_i, \xi, \xi_0) p(\xi_0 \mid a, b) d\xi_0$$

$$= \frac{\Gamma(a + N/2)\left[1 + (x_i - \mu_i)^T \Sigma_i^{-1}(x_i - \mu_i)/2b\right]^{-(a+N/2)}}{\Gamma(a)(2\pi b)^{N/2}\left|\Sigma_i^{-1}\right|^{1/2}}, \quad (6)$$

where

$$\mu_i = \Sigma_i A_i^T y_i,$$
$$\Sigma_i = \left(A_i^T A_i + \Lambda\right)^{-1}, \quad (7)$$

with $\Lambda = \text{diag}(\xi_1, \xi_2, \ldots, \xi_N)$.

The parameter ξ can be estimated by searching the maximum of the marginal likelihood as follows:

$$L(\xi) = \sum_{i=1}^{L} \log(y_i \mid \xi)$$

$$= -\frac{1}{2}\sum_{i=1}^{L}\left[(N_i + 2a)\log\left(y_i^T B_i^{-1} y_i + 2b\right) + \log|B_i|\right] + \text{const}, \quad (8)$$

where $B_i = I + A_i \Lambda^{-1} A_i^T$.

The dependence of $L(\xi)$ on ξ_j and B_i can be formulated as $B_i = B_{i,-j} + \xi_j^{-1} A_{i,j} A_{i,j}^T$ with $B_{i,-j} = I + \sum_{k \neq j} \xi_k^{-1} A_{i,k} A_{i,k}^T$. Therefore, $L(\xi)$ can be reformulated as follows:

$$L(\xi) = L\left(\xi_{-j}\right) - \frac{1}{2}\sum_{i=1}^{L}\left[\log\left(\frac{1 + s_{i,j}}{\xi_j}\right) + (N_i + 2a)\log\left(1 - \frac{q_{i,j}^2/g_{i,j}}{\xi_j + s_{i,j}}\right)\right], \quad (9)$$

where ξ_{-j} is obtained by removing the jth component in ξ.

$$s_{i,j} = A_{i,j}^T B_{i,-j}^{-1} A_{i,j}, q_{i,j}$$
$$= A_{i,j}^T B_{i,-j}^{-1} y_i \text{ and } g_{i,j} \quad (10)$$
$$= y_i^T B_{i,-j}^{-1} y_j + 2b.$$

To find the maximum of $L(\xi)$, $L(\xi)$ is differentiated with respect to ξ_j. With the assumption of $\xi_j \ll s_{i,j}$, ξ_j can be approximated as follows:

$$\xi_j \approx \begin{cases} \dfrac{L}{\sum_{i=1}^{L}\left(\left(N_i + 2a\right)q_{i,j}^2/g_{i,j} - s_{i,j}/s_{i,j}\left(s_{i,j} - q_{i,j}^2/g_{i,j}\right)\right)}, & \text{if } \sum_{i=1}^{L}\dfrac{(N_i + 2a)q_{i,j}^2/g_{i,j} - s_{i,j}}{s_{i,j}\left(s_{i,j} - q_{i,j}^2/g_{i,j}\right)} > 0, \\[20pt] \infty, & \text{otherwise.} \end{cases} \quad (11)$$

Equation (11) controls the addition and deletion of $A_{i,j}$ from the signal representation. Then, $s_{i,j}$, $q_{i,j}$, and $g_{i,j}$ can be computed as follows:

$$s_{i,j} = \frac{\xi_j S_{i,j}}{\xi_j - S_{i,j}},$$

$$q_{i,j} = \frac{\xi_j Q_{i,j}}{\xi_j - Q_{i,j}}, \quad (12)$$

$$g_{i,j} = G_i + \frac{Q_{i,j}^2}{\xi_j - S_{i,j}},$$

with

$$S_{i,j} = A_{i,j}^T A_{i,j} - A_{i,j}^T A_i \sum_i A_i^T A_{i,j},$$

$$Q_{i,j} = A_{i,j}^T y_i - A_{i,j}^T A_i \sum_i A_i^T y_i, G_i = y_i^T y_i - y_i^T A_i \sum_i A_i^T y_i + 2b. \quad (13)$$

In equation (13), A_i and Σ_i only contain the basis vectors currently included in the model. With the sedimentations of Σ_i, the sparse representation coefficients can be solved. With the solutions of sparse coefficients corresponding to different tasks, the original input can be reconstructed class by class to calculate the reconstruction errors. Finally, by comparison of the reconstruction errors from different classes, the category of the input can be decided.

3.2. Procedure of Target Recognition. Based on the above analysis, the basic process of the power equipment image recognition method designed in this paper is shown in Figure 1. First, multiple 2D random projection matrices are initialized for multilevel feature extraction. For all training samples, a corresponding dictionary is constructed through different 2D random projection matrices to form a multi-feature dictionary. The test sample uses the same random projection matrices to obtain the multilevel feature vectors. Then, based on the multifeature dictionary, MtCS is used to characterize multiple feature vectors of the test sample and the sparse representation coefficients corresponding to different features are calculated. Finally, the decision is made based on the reconstruction errors of the test sample corresponding to different training classes. The proposed method shows a certain degree of randomness in the construction of 2D random projections, so as to ensure the complementarity between them. In the classification stage, the maximum posterior strategy adopted by MtCS can well overcome the interference caused by nuisance conditions such as noise and occlusion, so as to ensure the reliability of the final decision.

4. Experiments and Discussion

4.1. Description of the Dataset. In order to test the performance of the proposed method, this paper uses three types of power equipment, i.e., insulators, transformers, and circuit breakers, for experiments, which are common in power systems. 2000 images of each of the three types of equipment are collected and used. All these images are adjusted to sizes of 400 pixels × 400 pixels by means of preprocessing. 1400 images of each of the three types of targets are randomly selected as the training samples, and the remaining 600 images are used as the test samples. As a comparison, this paper selects several types of existing relevant methods to conduct experiments at the same time, including the method based on the region moments in [11] (denoted as Region moment), the method using SRC in [5] (denoted as SRC), the method based on SVM in [8] (denoted as SVM), and the method using CNN in [21] (denoted as CNN).

In the following, the original samples are first tested as a preliminary validation. Later, the nuisance conditions, which may occur in the actual applications, are considered including noise interference and occlusion. The performance of all the methods is comprehensively investigated under the three conditions to reach the final evaluation of the proposed method.

4.2. Performance on Original Samples. At first, we use the proposed method to classify the original test samples of three types of equipment. Table 1 shows the recognition results of the three types of equipment achieved by the proposed method. It can be seen that the recognition rate of the three types of equipment has reached more than 90%, and the average recognition rate is calculated to 92.3%. This result shows the effectiveness of the proposed method for image recognition of power equipment. Table 2 compares the average recognition

rates of different methods. The performance of the proposed method is better than of others, which proves its superior performance. Specifically, compared with the SRC method, this paper extends the single sparse representation problem to a multitask one and combines multilevel 2D random projection features to provide a more adequate discrimination basis for classification decision. Therefore, the recognition result of the proposed method is greatly improved compared with the SRC method. The recognition performance of the CNN method ranks second in this case, only lower than the proposed method. For the original samples, the test samples can maintain high correlations with the training samples. At this time, the trained classification network can maintain strong adaptability to the test samples. The method based on regional moment features has poor performance under the current condition, mainly because there may be certain errors in the process of regional feature extraction, which are passed to the classification stage and cause the decrease of the recognition accuracy.

4.3. Performance on Noisy Samples. Since noise interference is inevitable in the actual image acquisition process, it is necessary to investigate the recognition performance of the proposed method under noise interference conditions. In this experiment, we first add different degrees of Gaussian white noise to the original test samples [32] and then test the recognition performance of different methods for noise samples. Figure 2 shows the average recognition rate curves of different methods as the signal-to-noise ratio (SNR) changes. It can be seen that the proposed method maintains the best recognition performance under different SNRs, indicating its robustness to noise interference. As analyzed above, both 2D random projection and MtCS are based on the basic theory of CS and have good adaptability to noise influences. At the same time, Bayesian estimation is introduced into MtCS, which can further enhance the robustness of the classification process against noise interference. Compared with the condition of the original samples, the performance of the CNN and SVM methods degrades the most significantly. Taking the CNN as an example, as the noise level of the test samples continues to increase, its similarity with the training samples decreases. As a result, the trained classification network has poor adaptability to those test samples, resulting in a decrease in the recognition accuracy. The overall robustness of the method based on the regional moment feature under noise interference is second only to the proposed method, because the regional feature is relatively insensitive to noise interference. Even under noise pollution conditions, the area characteristics of the target can generally be well maintained, so it can maintain good performance under noise interference conditions. The SRC method has some advantages over the CNN and SVM, which further reflects the robustness of the CS principle to noise interference.

4.4. Performance on Occluded Samples. In the process of collecting images of power equipment, occlusion and other situations inevitably occur, resulting in partial occlusions in the acquired images. To test the adaptability of the proposed method to occlusion conditions, this paper constructs occluded samples based on the original test set and obtains
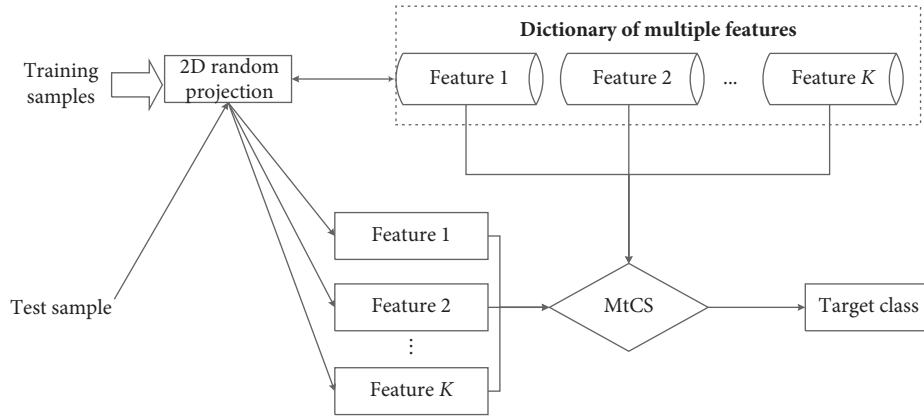
Figure 1: Procedure of recognition via electric equipment based on 2D random projection and MtCS.

Table 1: The recognition results of the three types of power equipment achieved by the proposed method.

Class	Recognition result			Recognition rate (%)
	Insulators	Transformers	Circuit breakers	
Insulators	552	21	27	92.00
Transformers	9	580	11	96.67
Circuit breakers	10	13	567	94.50
Average recognition rate (%)		94.32		

Table 2: Comparison of performance of different methods on original test samples.

Method type	Proposed	Region moment	SVM	SRC	CNN
Average recognition rate (%)	94.32	91.87	92.13	92.42	93.54
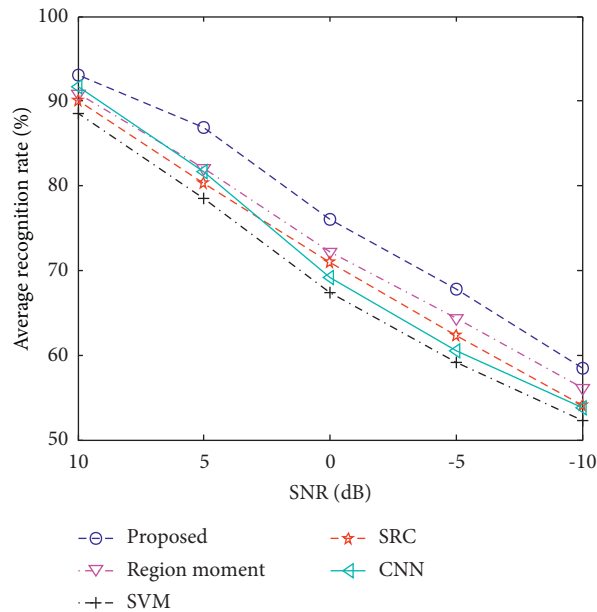
Figure 2: Comparison of performance of different methods on noisy samples.
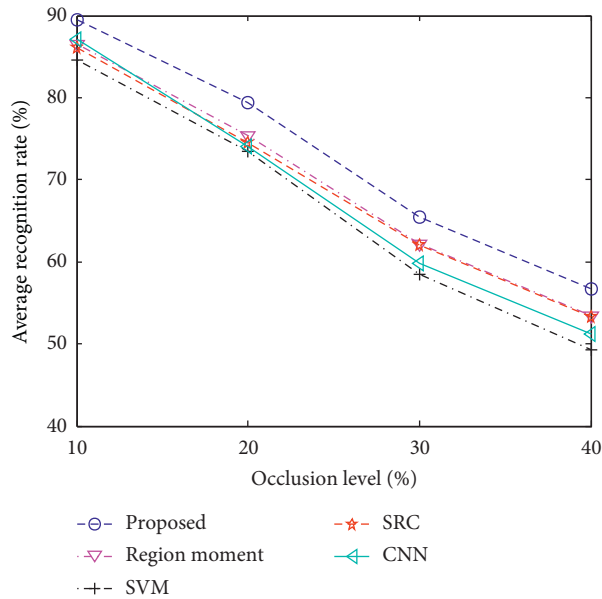
Figure 3: Comparison of performance of different methods on occluded samples.

different test sets with occlusion levels of 10%, 20%, 30%, and 40%. Afterwards, all the methods are used to classify the test sets with different occlusion conditions, and the statistical results are shown in Figure 3. It can be seen that the proposed method can achieve higher performance than the comparison methods at different occlusion levels, showing its robustness to partial occlusion. The multilevel 2D random projection features have good complementarity, so they play a positive role in the recognition problem under occlusion conditions. At the same time, the Bayesian estimation used in MtCS can be solved in a maximum posterior manner, so as to obtain a theoretical optimal decision. Similar to the case of noise interference, the performance of the SRC method is better than that of the SVM and CNN methods, which reflects that the principle of CS is also adaptable to occlusion situations. Therefore, the proposed method combines the advantages of these two tools to effectively improve the adaptability to occlusion situations.

5. Conclusion

For the problem of power equipment image recognition, this paper proposes a method based on 2D random projection and MtCS. The multilevel feature vectors of the power equipment image are obtained by 2D random projection, which have good complementarity. MtCS has good noise robustness and anti-interference performance and can robustly solve sparse representation coefficients under more complex conditions. By combining the advantages of the two algorithms, the adaptability of the recognition method to various scenarios can be improved. Experiments are carried out based on the samples of three typical power equipment: insulators, transformers, and circuit breakers. The results show that the proposed method is effective for the problem and has

stronger robustness than other methods under conditions like noise interference and partial occlusions.

Data Availability

Our experiment data are all collected from the Internet, and they are permitted for public use.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] C. Guo, M. Ren, C. Xia, M. Dong, and B. Wang, "Fault diagnosis of power equipment based on infrared image analysis," in *Proceedings of the IEEE International Conference on Advances in Electrical Engineering and Computer Applications (AEECA)*, pp. 659–663, Dalian, China, August 2020.

[2] J. Ma, K. Qian, X. Zhang, and X. Ma, "Weakly supervised instance segmentation of electrical equipment based on RGB-T automatic annotation," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 12, pp. 9720–9731, 2020.

[3] G. C. Stone, "Partial discharge diagnostics and electrical equipment insulation condition assessment," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 12, no. 5, pp. 891–903, 2005.

[4] T. Liu, "Development and application of electrical equipment intelligent detection management system," in *Proceedings of the 5th International Conference on Smart Grid and Electrical Automation (ICSGEA)*, pp. 70–73, Zhangjiajie, China, June 2020.

[5] C.-F. Lai, W.-C. Chien, L. T. Yang, and W. Qiang, "LSTM and edge computing for big data feature recognition of industrial electrical equipment," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 4, pp. 2469–2477, 2019.

[6] D. Ming, B. Wang, M. Ren et al., "Joint visualization diagnosis of outdoor insulation status with optical and acoustical detections," *IEEE Transactions on Power Delivery*, vol. 34, no. 4, pp. 1221–1229, 2018.

[7] X. Xiong, S. Xu, W. Wu, D. Tu, J. Zhang, and Z. Wei, "Identification of electrical equipment based on faster LSTM-CNN network," in *Proceedings of the IEEE International Conference on Networking, Sensing and Control (ICNSC)*, pp. 1–6, Nanjing, China, November 2020.

[8] L. Jin, J. Ai, Z. Tian, and Y. Zhang, "Detection of polluted insulators using the information fusion of multispectral images," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 24, no. 6, pp. 3530–3538, 2017.

[9] L. Li, J. Yong, and W. Xu, "On-line cable condition monitoring using natural power disturbances," *IEEE Transactions on Power Delivery*, vol. 34, no. 4, pp. 1242–1250, 2019.

[10] J. Wang, J. Wang, J. Shao, and J. Li, "Image recognition of icing thickness on power transmission lines based on a least squares hough transform," *Energies*, vol. 10, no. 4, p. 415, 2017.

[11] Y. Yin, Z. Meng, and S. Li, "Feature extraction and image recognition for the electrical symbols based on Zernike moment," in *Proceedings of the IEEE 2nd Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)*, pp. 1031–1035, Chongqing, China, March 2017.

[12] Z. Zhao, G. Xu, and Y. Qi, "Representation of binary feature pooling for detection of insulator strings in infrared images,"

IEEE Transactions on Dielectrics and Electrical Insulation, vol. 23, no. 5, pp. 2858–2866, 2016.

[13] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Proc NIPS*, vol. 2, pp. 1096–1105, 2012.

[14] C. Szegedu, W. Liu, Y. Jia et al., "Going deeper with convolutions," in *Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1–9, Boston, MA, USA, June 2015.

[15] K. He, X. Zhang, S. Ren et al., "Deep residual learning for image recognition," in *Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, Las Vegas, NV, USA, June 2016.

[16] X. X. Zhu, D. Tuia, L. Mou et al., "Deep learning in remote sensing: a comprehensive review and list of resources," *IEEE Geoscience and Remote Sensing Magazine*, vol. 5, no. 4, pp. 8–36, 2017.

[17] X. Gong, Q. Yao, M. Wang, and Y. Lin, "A deep learning approach for oriented electrical equipment detection in thermal images," *IEEE Access*, vol. 6, Article ID 41590, 2018.

[18] C. Yeung, "Image Recognition Algorithm of Electrical Engineering Equipment Based on Machine Learning Method," in *Proceedings of the IEEE International Conference on Power Electronics, Computer Applications*, pp. 711–714, Shenyang, China, January 2021.

[19] Y. Xia, J. Lu, H. Li, and H. Xu, "A Deep Learning Based Image Recognition and Processing Model for Electric Equipment Inspection," in *Proceedings of the 2018 2nd IEEE Conference on Energy Internet and Energy System Integration (EI2)*, pp. 1–6, Beijing, China, October 2018.

[20] X. Siheng, Y. Liu, R. Xu et al., "Power equipment recognition method based on mask R-CNN and bayesian context network," in *Proceedings of the IEEE Power & Energy Society General Meeting (PESGM)*, pp. 1–5, Montreal, Canada, August 2020.

[21] A. Jiang, N. Yan, B. Shen, C. Gu, H. Zhu, and H. Huang, "Research on infrared image recognition method of power equipment based on deep learning," in *Proceedings of the IEEE International Conference on High Voltage Engineering and Application (ICHVE)*, pp. 1–4, Beijing, China, September 2020.

[22] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Yi Ma, "Robust face recognition via sparse representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 2, pp. 210–227, 2009.

[23] J. J. Thiagaraianm, K. N. Ramamurthy, P. Knee et al., "Sparse representations for automatic target classification in SAR images," in *Proceedings of the 2010 4th International Symposium on Communications, Control and Signal Processing (ISCCSP)*, pp. 1–4, Limassol, Cyprus, March 2010.

[24] H. Song, K. Ji, Y. Zhang, X. Xing, and H. Zou, "Sparse representation-based SAR image target classification on the 10-class MSTAR data set," *Applied Sciences*, vol. 6, no. 1, p. 26, 2016.

[25] B. Ding, G. Wen, F. Ye, X. Huang, and X. Yang, "Feature extraction based on 2D compressive sensing for SAR automatic target recognition," in *Proceedings of the 11th European Conference on Antennas and Propagation (EUCAP)*, pp. 1219–1223, Paris, France, March 2017.

[26] S. Ji, D. Dunson, and L. Carin, "Multitask compressive sensing," *IEEE Transactions on Signal Processing*, vol. 57, no. 1, pp. 92–106, 2009.

[27] Q. Qisong Wu, Y. D. Zhang, M. G. Amin, and B. Himed, "Multi-task bayesian compressive sensing exploiting intra-task dependency," *IEEE Signal Processing Letters*, vol. 22, no. 4, pp. 430–434, 2015.

[28] J. A. Tropp, A. C. Gilbert, and M. J. Strauss, "Algorithms for simultaneous sparse approximation. Part II: c," *Signal Processing*, vol. 86, no. 3, pp. 589–602, 2006.

[29] H. Zhang, N. M. Nasrabadi, Y. Zhang, and T. S. Huang, "Multi-view automatic target recognition using joint sparse representation," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 48, no. 3, pp. 2481–2497, 2012.

[30] B. Ding and G. Wen, "Exploiting multi-view SAR images for robust target recognition," *Remote Sensing*, vol. 9, no. 11, p. 1150, 2017.

[31] S. Liu and J. Yang, "Target recognition in synthetic aperture radar images via joint multifeature decision fusion," *Journal of Applied Remote Sensing*, vol. 12, no. 1, Article ID 016012, 2018.

[32] X. Zhang, "Noise-robust target recognition of SAR images based on attribute scattering center matching," *Remote Sensing Letters*, vol. 10, no. 2, pp. 186–194, 2019.

Research Article

Research on Decision Evaluation Model of HOV Lane Setting

Ding Lv [ID],[1,2,4] Qunqi Wu,[1,4] Bo Chen,[3,4] and Yahong Jiang[3,4]

[1]*School of Economics and Management, Chang'an University, Xi'an 710064, China*
[2]*Research Center for Urban Public Security, Shaanxi Police College, Xi'an 710021, China*
[3]*College of Transportation Engineering, Chang'an University, Xi'an 710064, China*
[4]*Center of Comprehensive Transportation Economic Management, Chang'an University, Xi'an 710064, China*

Correspondence should be addressed to Ding Lv; lvding@chd.edu.cn

In order to achieve the purpose of improving the travel efficiency of commuters in the periphery of the city, expanding the beneficiary groups of urban rail transit, and alleviating urban road traffic congestion, when planning and setting up HOV in the periphery of the city, it is necessary to analyze the feasibility of HOV lane setting from both the demand conditions and the setting conditions. This paper combines machine learning to construct a decision-making evaluation model for HOV lane setting and studies the optimal layout model and algorithm of HOV lanes in service rail transit commuter chain. The setting, planning, and layout of HOV lanes are a two-way interactive process of traveler's path selection and designer's road planning. Finally, after the model is constructed, the performance of the system model is verified. The results show that the system studied in this paper can be used for traffic data and lane planning analysis. Therefore, in the process of urban operation, the HOV model constructed in this paper is mainly used to alleviate urban traffic and improve urban operation efficiency.

1. Introduction

In order to ensure the smoothness of urban traffic and the sustainable development of traffic, since the late 1990s, major cities in China have undertaken a series of actions such as road rectification, scale expansion, and construction of transportation infrastructure to improve the current state of traffic congestion. However, the construction speed is far behind the growth rate of the traffic volume in the transportation system, and the current total road scale and traffic management level are far from the actual needs of travelers in the transportation system. Moreover, the growth rate of transportation demand is far greater than the rate of new infrastructure construction. In addition, actual experience over the years has shown that, relying only on new roads, increased investment and other methods to deal with traffic congestion will not solve the problem and will even plunge the supply of road resources into a vicious circle, which will worsen the imbalance between supply and demand. In particular, when the road network structure and basic framework of medium and large cities have been

determined, the available land resources are already very limited, and blind expansion can only further squeeze the space resources of the city. However, with the increasing demand for travel, newly built roads will be filled with huge traffic, causing new traffic jams. This will further deepen the imbalance between supply and demand in the transportation system, make the contradiction between the supply and demand of road resources increasingly acute, and intensify the congestion problem. Because of the above phenomenon, a series of problems may arise, such as lower and lower travel efficiency of residents and more and more serious urban environmental pollution. Improving road utilization may become a more operational and effective method for alleviating traffic congestion in my country's big cities, and it is also a research direction with a lot of room for development in the future [1].

The setting of HOV lanes is an effective way to alleviate congestion. HOV is a dedicated lane with high loading rate. It was formally proposed in the United States as a concept in traffic demand management in the 1990s. As the name implies, the HOV dedicated lane refers to a management

lane dedicated to vehicles with high load-bearing rates. The proposal of this concept provides a reasonable new idea for alleviating urban traffic congestion [2].

For improving urban traffic conditions, commuters are the main participants in the morning and evening peaks [3]. To study the optimal layout of HOV lanes in the service rail transit commuter chain, the route with the least total travel time cost can be selected from multiple road sections in the periphery of the city as the route for setting HOV lanes. Moreover, it can guide more commuters to travel by carpooling, then transfer to rail transit, and finally connect to other modes of travel to complete the commuting process.

This paper creatively establishes the decision-making evaluation model of high-performance vehicle lane setting combined with machine learning. The optimal layout model and algorithm of high-performance vehicle lanes in service rail transit commuter chain are studied. The setting, planning, and layout of HOV lane are a two-way interactive process of traveler's path selection and designer's road planning. Finally, after establishing the model, the performance of the system model is verified. The results show that the high-performance vehicle lane path optimization model constructed in this paper has a certain effect.

This paper is divided into five parts. The first two parts introduce the theory of rail transit commuter chain and HOV lane and then determine the research method of the attraction range of rail transit stations around the city. The third part of this paper establishes the decision-making evaluation model of high-performance vehicle lane setting combined with machine learning. The fourth part studies the optimal layout model and algorithm of high-performance vehicle lane in the service-oriented rail transit commuter chain. The setting, planning, and layout of HOV lane are a two-way interactive process of traveler's path selection and designer's road planning. Finally, after establishing the model, the performance of the system model is verified.

2. Related Work

Gadawe et al. [4] comprehensively summarized the development status of existing HOV lanes abroad. This paper sorted out the whole process of HOV lane development in chronological order, introduced the concept, function, type, and existing problems of HOV lane in turn, and also briefly described the doubts raised by scholars in other fields regarding this new thing. Moreover, the paper finally boldly predicted the research direction and development trend of HOV lanes in the next few years and elaborated in detail objectively and comprehensively the irreplaceability of HOV lanes and the inevitability of putting them into practical use.

Malipatil et al. [5] used a multimethod selection logit model to discuss the construction and application of HOV lanes in urban traffic systems. Starting from the current low utilization rate of bus dedicated lines, Rehman et al. [6] proposed a targeted approach to construct HOV lanes using idle or underutilized bus dedicated lines. Moreover, they used the existing data in the actual road network to perform inference calculations and conducted reasonable analysis in both the necessity and feasibility dimensions. Hamarashid

et al. [7] selected a specific road section to analyze data and perform calculations and predicted and analyzed the final results. This type of research provides the necessary theoretical basis for introducing HOV lanes into China's actual road network for application. Matthews et al. analyzed traffic problems from the perspective of macro-coordination, and based on the comprehensive consideration of foreign transportation systems, they first proposed the concept of carpooling priority. Sivanathan et al. [8] introduced the management of HOV lanes into urban expressways and used the MNL model and analytic hierarchy process to explain in detail the necessity and possibility of prioritizing the development of HOV lanes under the current development situation. Koresh et al [9] conceived the combined construction of HOV lanes and BRT lanes, which, to some extent, broke our country's stereotype of thinking about HOV lane construction. Ning et al. [10] discussed the actual effect of HOV lanes in alleviating traffic congestion and analyzed the feasibility of setting HOV lanes in cities. This paper uses simulation and comparative analysis. Qiu et al. [11] used the traffic efficiency model as a tool to specifically analyze the functionality of HOV lanes in urban settings. Moreover, based on the current situation and forecasted traffic volume, it used the calculation and comparative analysis of traffic efficiency to analyze whether the construction of HOV lanes is feasible and proposed the decision-making basis for judgment when setting up HOV lanes. Liu et al. [12] analyzed the setting conditions of HOV lanes in actual operation, proposed HOV lane management methods that meet the actual situation, and designed four different types of cross-sections for HOV lanes according to different types. Sharif et al. [13] used VISSIM software to carry out a simulation comparison and analysis on the road conditions before and after setting HOV lanes and explained in detail the respective advantages and disadvantages and applicable conditions of HOV lanes and bus lanes. Arulselvi et al. [14] put forward the design and construction plan for the design of HOV lanes in more detail. Moreover, on the basis of on-site traffic investigation and data collection and sorting, they carefully considered the specific organic components of the road and gave a complete and specific HOV dedicated lane design plan. Finally, it used simulation software to evaluate the design. Li et al. [15] incorporated carpooling as an independent travel mode into the transportation system and introduced the two-level programming model into the problem of HOV lane construction to solve the optimal HOV lane layout. Raskar et al. [16] adjusted the relationship between the difference of the total utility and the planned HOV route from the perspective of whether the total utility of travel can be reduced and deeply studied the method of optimal design of the HOV dedicated lane. Sumi et al. [17] proposed a design method for HOV dedicated lanes that can improve road operation efficiency. For the first time in this paper, the idea of directly converting ordinary lanes into HOV lanes is proposed.

Although Adekitan et al. [18] has evaluated the operating performance of high-carrying lanes, most studies focus on the limited data obtained from one or a small number of road sections. The methods used in these studies may not be

easily applied to larger studies. Some of them neglect large-scale data processing, because it takes too much time. Moreover, other studies are only to illustrate the running performance of the short-circuit section of the high-carrying lane. For example, only the distance between vehicle detection stations (VDS) is considered in space. In addition, the conclusions drawn in the research of some specific road sections may be restricted by relevant conditions, such as the traffic demand and lane configuration of the road section. Therefore, it is difficult to extend the conclusions obtained to other suitable situations. By analyzing the five-year traffic accident data in the Traffic Accident Monitoring and Analysis System (TASAS) database, Li et al. [19] evaluated the safety performance of continuous and restricted high-carrying lanes. Among them, an interesting finding is that, in terms of accident rate and severity, restricted-access high-load lanes cannot provide better safety performance than continuous high-load lanes. Some studies compare the accident characteristics of two different types of high-carrying lanes, such as accident location, accident type, and frequency. Bellucci et al. [20] pointed out that, in terms of safety, the partition design of the space separation wall is better than the partition design of the buffer zone, at least better than the continuous access design. Wang et al. [21] pointed out that the accident rate of continuous high-carrying lanes is lower than that of restricted high-carrying lanes. However, the continuing impact after the accident has not been studied in depth. As a nonrepeatable cause of congestion, traffic accidents are an important part of the Freeway Traffic Management System (FTMS). Many accident analysis methods focus on the analysis of conventional highways. Traffic accidents on high-load lanes will affect the overall performance of the high-load lanes, and their accident research and analysis have great significance.

3. Decision Evaluation Algorithm for HOV Lane Setting

It can be found from Figure 1 that this is a process problem with many stages, and each stage has many decision-making processes. In order to solve this type of problem, a method is proposed, namely, the dynamic programming method. Therefore, the research direction of the dynamic programming method is: how to select the optimal series of decisions from a multistage and multidecision process? After the decision-making system process with time as a variable, the decision of each stage of the system is obtained according to the current state of the system. At the same time, the obtained decision immediately causes the current state to transition to the next moment state, and the next moment state is selected by the optimization principle to get the next moment decision, and at the same time, the next moment state is generated. In this way, the state of the system is constantly moving, and decisions are constantly being produced. Therefore, a series of decisions composed of one optimal decision is produced, which gives the impression that it is a process of continuous movement. Therefore, it is called a dynamic programming method.
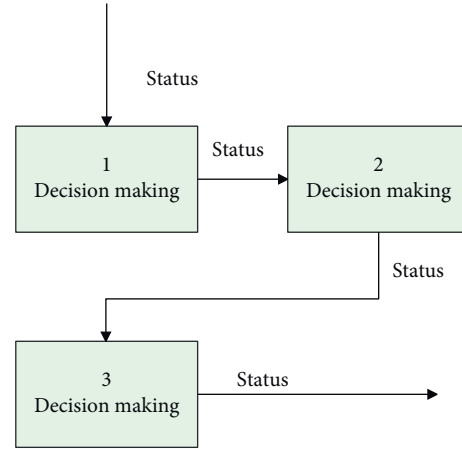
Figure 1: Schematic diagram of multistage and multidecision process.

The state that the system transitions from the current state to the next moment is related to the system function. People generally call the mathematical model of the state transition of the system as the state transition equation, which is an important part of the dynamic programming method [22]:

$$x(k+1) = F(x(k), u(k), k), \quad k = 0, 1, \ldots, N-1. \quad (1)$$

Among them, the function $F(\cdot)$ represents the function of the system, the character k represents the k-th stage in a system with N stages, the state $x(k)$ represents the state variable of the decision-making system in the k-th stage, and the decision $u(k)$ represents the action strategy taken by the system process when the state is $x(k)$.

An important member of the dynamic programming method is the efficiency function $U[x(k), u(k), k]$. The so-called efficiency function refers to the benefit value $U(\cdot)$ obtained by the system when the system state $x(k)$ is transferred to the state $x(k+1)$ at the next moment of the system process because of the decision $u(k)$ in the k-th stage of the system process.

$$J(x(i), i) = \sum_{k=1}^{N} \gamma^{k-i} U[x(k), u(k), k], \quad k = 1, 2, \ldots, N.$$
$$(2)$$

In the above formula, γ represents the discount factor, and its value range is $0 < \gamma \le 1$. $J(x(i), i)$ represents the sum of the instant benefits obtained from the initial time i and the initial state $x(i)$ when the system q executes the system decision-making process until the end of the entire system decision-making process [23].

However, in dynamic programming, $J(x(i), i)$ is not the index that people want. What people care about is its optimal index, which is the optimal performance index function:

$$J^*(x(i)) = \underset{\{u(k), \ldots, u(N)\}}{opt} \sum_{k=i}^{N} \gamma^{k-i} U[x(k), u(k), k]. \quad (3)$$

In the formula, $J^*(\cdot)$ represents the performance index function obtained when the decision of each stage of the process system is optimal. opt stands for maximum or minimum.

To understand the optimal performance index function, it is necessary for us to understand the principle of optimization. Its theory is as follows:

We assume that there is a set Ψ, which has many elements and many subsets. When Φ is a subset of Ψ, the largest element μ in set Ψ is the largest element in set Φ. When the subset Φ contains the largest element in the set Ψ, then the largest element ε in the set Φ is the largest element in the set Ψ [24].

In the 1950s, in order to solve the problem of multistage and multidecision system, Bellman proposed the core theory of dynamic programming method, namely, Bellman equation, according to the optimization principle. Its specific theory is as follows:

In a process system with N stages and N decisions, the optimal strategy sequence of the system is $u(0), \ldots, u(k), \ldots, u(N-1)$. Then, when the initial state $x(k)$ and the initial strategy $u(k)$ are used as the starting point, the system decision sequence $u(k), \ldots, u(k+1), \ldots, u(N-1)$ of the remaining $N-k$ stages of the process system must still be the optimal strategy sequence of the remaining stage.

The mathematical model is as follows:

$$J^*(x(k)) = \min_{u(k)}\{U(x(k), u(k)) + \gamma J^*(x(k+1))\}. \tag{4}$$

The nonlinear discrete system can be regarded as a multistage decision-making system, so it is suitable to use the dynamic programming method. After years of exploration and research, the basic model of using dynamic programming in discrete systems is as follows:

First, the mathematical model of the state transition equation of the nonlinear discrete system to be solved at the k-th stage is as follows:

$$x(k+1) = F(x(k), u(k), k), \quad k = 0, 1, \ldots. \tag{5}$$

The conditions of the initial state are

$$x(0) = x_0. \tag{6}$$

For $x(0)$, the corresponding system performance index function when $x(N)$ is in domain $G(x(N), N) = 0$ is

$$J = G(x(N), N) + \sum_{k=1}^{N-1} U(x(k), u(k), k). \tag{7}$$

Based on the optimality principle of Bellman equation, after step-by-step sorting and simplification, the mathematical expression of the optimal performance index function is finally obtained as

$$J^*(x(k)) = \min_{u(k)}[U(x(k), u(k)) + J^*(x(k+1))],$$
$$k = 1, 2, \ldots, N-1. \tag{8}$$

To solve this formula, we must also obtain the optimal performance index function of the last stage $x(N)$, namely $J^*(x(N))$.

$$J^*(x(N)) = \min_{u(N)}[G(x(N)) + U(x(N), u(N))]. \tag{9}$$

If the continuous time is regarded as composed of many very small time periods, then the continuous-time system becomes a discrete system. Therefore, continuous-time system models can also be solved using dynamic programming methods. The following is the mathematical model of dynamic programming method applied in continuous-time system. First, we assume that the mathematical model of the state transition equation in the continuous-time system is

$$\dot{x}(t) = f(x(t), u(t), t), \quad t_0 \le t \le t_f. \tag{10}$$

The initial state of the system is

$$x(t_0) = x_0. \tag{11}$$

For $x(t_0)$, the corresponding system performance index function when $x(t_f)$ is in domain $G(x(t_f), t_f) = 0$ is

$$J = G(x(t_f, t_f)) + \int_{t_0}^{t_f} U(x(t), u(t), t)\mathrm{d}t. \tag{12}$$

In the formula, the final moment state $x(t_f)$ of the continuous-time system is fixed, $x(t)$ and $u(t)$, respectively, represent the state vector and control vector of the continuous-time system at time t, the functions $G(\cdot)$ and $\int$ are continuously differentiable, and the function $U(\cdot)$ is the efficiency function and continuously differentiable. It represents the income of the continuous-time system at time t.

When the state vector of the system is $x(t)$, the corresponding control vector $u(t)$ can be found to make the performance index function optimal:

$$J(x(t), t) = \min_{u(t)}\left\{G[x(t_f, t_f)] + \int_{t_0}^{t_f} U(x(t), u(t), t)\mathrm{d}t\right\}. \tag{13}$$

Based on the Bellman equation, when the performance index function of the continuous-time system is taken as the minimum $u(t)$, the function equation is

$$J(x(t), t) = \min_{u(t)}\left\{\int_{t}^{t+\Delta t} U(x(t), u(t), t)\mathrm{d}t + J(x(t+\Delta t), t+\Delta t)\right\}. \tag{14}$$

After a series of derivation and simplification operations, the above equation is finally simplified to obtain the Hamilton-Jacobi equation as follows:

$$-\frac{\partial J^*(x(t),t)}{\partial t} = \min_{u(t)} \left\{ U(x(t),u(t),t) + \left(\frac{\partial J^*(x(t),t)}{\partial x(t)}\right)^T f(x(t),u(t),t) \right\}. \tag{15}$$

Its optimal performance indexes at the final moment are

$$J^*\big(x(t_f,t_f)\big) = G\big[x(t_f,t_f)\big]. \tag{16}$$

Based on the above two formulas, the solution of the continuous-time system can be solved.

Reinforcement learning is defined as a machine intelligence method that successfully combines dynamic programming and supervised learning to solve problems that neither of these two methods can solve alone. Supervised learning is a method of approaching the target through a lot of training, but its approach is not random, but determined by the tutor signal provided by the outside. That is to say, if there is no external tutor signal to supervise the learning, it is impossible to achieve success. Unfortunately, in people's real life, people have not found clear answers to many questions, so people cannot provide matching mentor signals.

In a standard reinforcement learning model, an agent (agent) interacts dynamically with its environment. These interactions are the forms that the agent uses to explore the surrounding environment and select a strategy through the results of the exploration. This strategy corresponds to an action of the agent. When the action is executed by the agent, it will have a certain impact on the environment. influences. This influence will notify the agent through some kind of enhanced signal, and the agent can judge whether the influence is good or bad based on the feedback signal. A typical reinforcement learning system consists of three basic parts: environment, reinforcement module, and value function.

(1) The environment of reinforcement learning can be understood as follows: the reinforcement learning system learns how to establish an optimal mapping to select the optimal action through error training that continuously interacts with the dynamic environment. This dynamic environment must be at least partially observable by the reinforcement learning system. It is said to be observation, but it is actually detected by various sensors and then described by some related symbols such as voltage and current, so that the agent can use it.

(2) The reinforcement module can be understood as follows: the reinforcement learning system learns a mapping from situation to action through error training that interacts with the dynamic environment. Therefore, the goal of reinforcement learning is defined by a concept called reinforcement function. The so-called reinforcement function is a function that can maximize the sum of the benefits of the agent in the future. After the reinforcement learning system performs an action according to a state, it will obtain a reinforcement signal as a reward or punishment. The purpose of the reinforcement learning system is to maximize the sum of all rewards obtained by performing actions from the beginning of the initial state to the end of the final state through learning.

One of the main tasks of reinforcement learning system designers is to define a reinforcement function, so that this function has a proper learning goal. Although the designer can freely define the reinforcement function of the reinforcement learning system, in order to achieve people's desired goals, the designer has some links when constructing the reinforcement function, for example, purely delayed rewards and punishments. When the system state is the final state, people hope that the reinforcement incentive of pure delayed reward function class is zero.

(3) Value function: how does the agent learn to choose those good actions? Or how to measure the effectiveness of an action? Therefore, the concept of value function is introduced into reinforcement learning. The so-called value function refers to the enhancement function in any state plus the maximum value of the sum of all enhancement functions from this state to the final state.

The schematic diagram of reinforcement learning is shown in Figure 2. The parameter module includes these parts: policy module P, enhancement module R, input module I, and the current state s of the agent receiving environment. When the agent (consisting of the strategy module P, the reinforcement module R, and the input module I) interacts with the dynamic environment, the agent receives the current state S of the environment and becomes i after processing I. After that, the strategy module derives strategy a based on the r at the previous moment and the current i. After that, under the action of a, the dynamic environment T changes its state from s to s' and, at the same time, sends a return r to the agent.

In fact, reinforcement learning is developed from the Markov process. In a Markov process environment, the agent selects an action a from the set of all actions, which makes the state of the environment system transfer and returns a reward and punishment value r to the agent. Its mathematical expression is as follows:
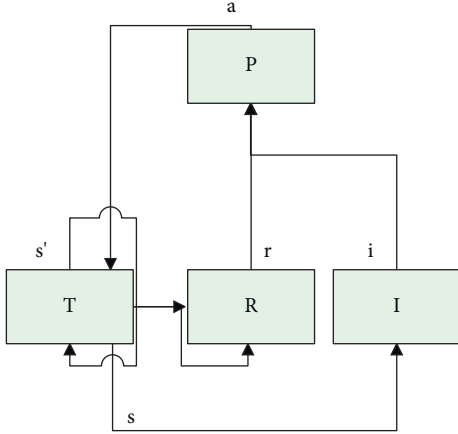
FIGURE 2: Schematic diagram of reinforcement learning.

$$\text{prob}\left(s = s' \mid s, a\right) = P\left(s, a, s'\right). \tag{17}$$

The utility function obtains the value when the state is s under the action of the strategy, and the mathematical expression form is

$$U(s) = r(s) + \gamma \max_a E\left(\sum_{s'}^{s_N} \left(T(s, a, s'), U(s')\right)\right). \tag{18}$$

In the formula, s_N is the final state of the system, and γ is the discount factor. When action a is the optimal action in the set of all actions, then $U(s)$ is also the optimal value function in state s.

From the composition of GDHP, we know that it is very complicated, and what is even more troublesome is that it must calculate the second derivative $\partial^2 J(t)/\partial R(t)\partial W_c$. Although it is much more complicated than the previous two types, the results obtained are much better than expected. In 1997, Danil and Donald gave a detailed analysis and calculation module in their paper to make the second derivative $\partial^2 J(t)/\partial R(t)\partial W_c$ easier to calculate. Its structure is shown in Figure 3. When the squared difference is the smallest, we can get the update formula for evaluating the weight of the network:

$$\Delta W_c = -\eta_h [J(t) - \gamma J(t+1) - U(t)] \frac{\partial J(t)}{\partial W_c}$$
$$- \eta_d \sum_{j=1}^{n} \frac{\partial^2 J(t)}{\partial R_j(t)\partial w_c} E_D. \tag{19}$$

Among them, η_h, η_d are the learning rates of the execution network and the evaluation network, respectively, and E_D is the squared difference of DHP.

In this figure, how does GDHP's evaluation network get $\partial^2 J(t)/\partial R(t)\partial W_c$? First of all, we must be able to work smoothly in the GDHP design structure shown in Figure 4. Here, by using the output $J(t)$ and $\partial J(t)/\partial R(t)$ of the evaluation network, the burden of calculating $\partial^2 J(t)/\partial R(t)\partial W_c$ will be minimized. In this way, the second
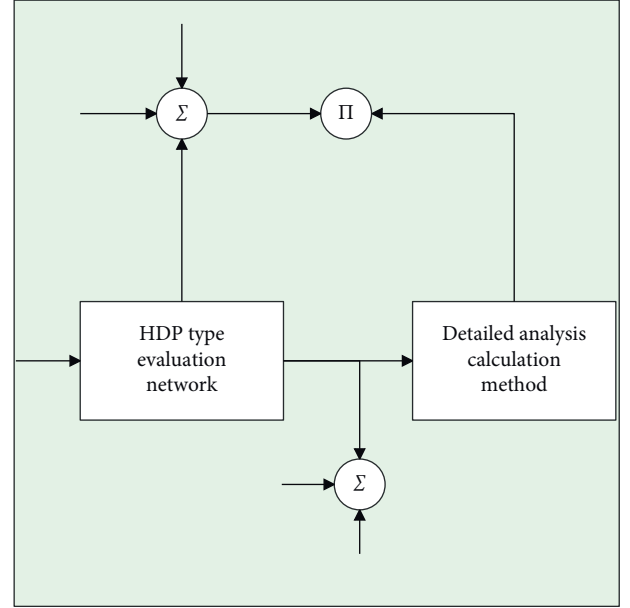
FIGURE 3: Schematic diagram of the structure of GDHP.

derivative $\partial^2 J(t)/\partial R(t)\partial W_c$ can be obtained more conveniently by backpropagation.

The output $J(t)$ of the evaluation network does not return to the input $R(t)$ through the internal path to train the execution network, because if it is executed in this way, it will return to the working mode of HDP. We already have a high-quality method to obtain $\partial J(t)/\partial R(t)$, which is DHP. Therefore, DHP is put into GDHP as a part of GDHP.

A simple evaluation network is designed as shown in Figure 5:

From Figure 3, we know that its hidden layer includes two sigmoidal nerves and a linear output $J(t)$. The following equation can be obtained from the network:

$$J(t) = f_3 w_{35} + f_4 w_{45} + R_0 w_{05}$$

$$= \frac{w_{35}}{\exp\left(-R_0(t)w_{05} - R_1(t)w_{13} - R_2(t)w_{23}\right) + 1}$$

$$+ \frac{w_{45}}{\exp\left(-R_0(t)w_{04} - R_1(t)w_{14} - R_2(t)w_{24}\right) + 1} + R_0 w_{05}. \tag{20}$$

To find the partial derivative $\partial J(t)/\partial R_j(t)$ of the input for the output, there are

$$\frac{\partial J(t)}{\partial R_j(t)} = \sum_3^4 f_i(1 - f_i)w_{i5}w_{ji} + w_{05}\delta_{j0}(t), \tag{21}$$

$$i = 3, 4; \ j = 0, 1, 2.$$

Among them, $\delta_{j0}(t)$ is the Kronecker function. Then, the partial derivative of the weight of the output layer and the weight of the hidden layer is calculated from the above formula, as shown in the following formula:
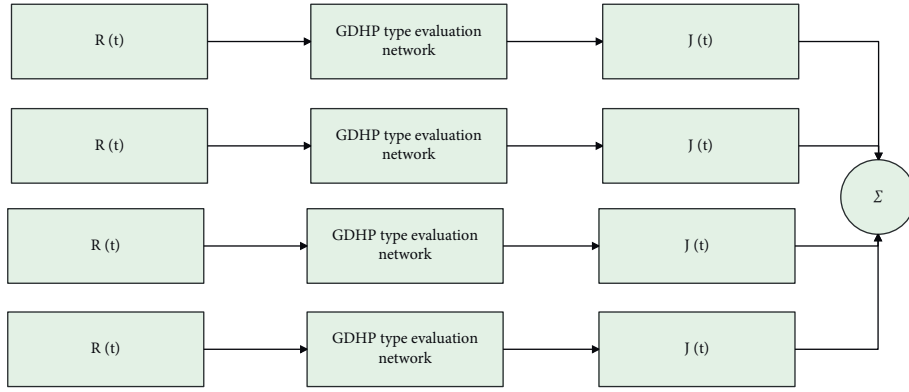
Figure 4: Evaluation network in a simple GDHP design.
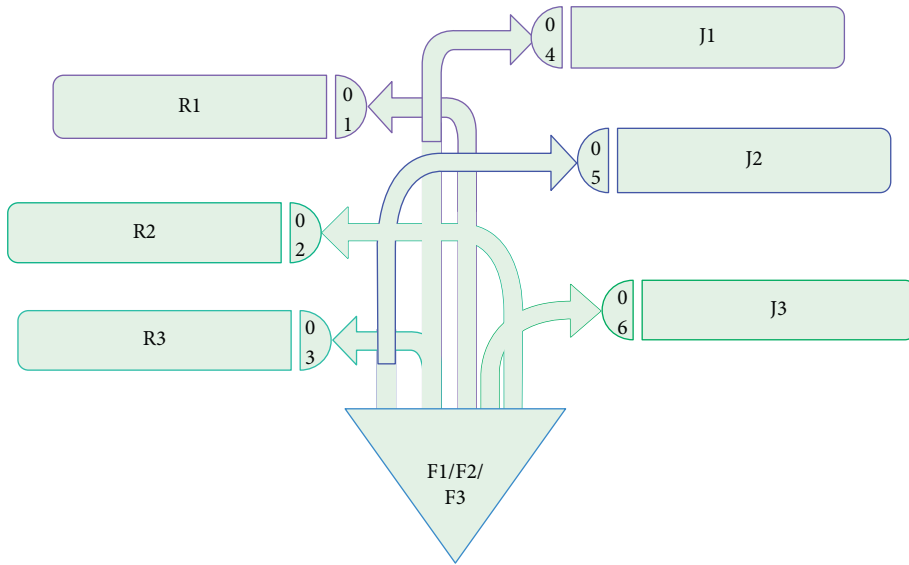
Figure 5: A simple network used to calculate an example of $\partial^2 J(t)/\partial R(t)\partial W_c$ in the GDHP design.

$$\frac{\partial^2 J(t)}{\partial R_j(t)\partial w_{i5}} = w_{ji}f_i(1-f_i), \quad i = 3,4; j = 0,1,2,$$

$$\frac{\partial^2 J(t)}{\partial R_j(t)\partial w_{05}} = \delta_{j0}(t), \quad i = 3,4; j = 0,1,2,$$

$$\frac{\partial^2 J(t)}{\partial R_j(t)\partial w_{ji}} = f_i(1-f_i)w_{i5}\left[1+(1-2f_i)R_j(t)w_{ji}\right],$$

$$i = 3,4; j = 0,1,2,$$

$$\frac{\partial^2 J(t)}{\partial R_j(t)\partial w_{ki}} = (1-2f_i)R_j(t)w_{ji}f_i(1-f_i)w_{i5,}$$

$$i = 3,4; k = 0,1,2; j = 0,1,2, k \neq j.$$

(22)

Then, we can use the basic formula to modify its form:

$$\Delta W_{ji} = -\eta_h[J(t) - \gamma J(t+1) - U(t)]\frac{\partial J(t)}{\partial W_{ji}}$$

$$- \eta_d \sum_{k=1}^{2} \frac{\partial^2 J(t)}{\partial R_k(t)\partial w_{ji}}\left[\frac{\partial J(t)}{\partial R_k(t)} - \gamma\frac{\partial J(t+1)}{\partial R_k(t)} - \frac{\partial U(t)}{\partial R_k(t)}\right].$$

(23)

This is the weight update formula of GDHP.

4. Decision Evaluation Algorithm for HOV Lane Setting

Travelers are passengers who use it in the road network, and they are more concerned about the cost of travel. Therefore, the Stackelberg game problem is the essence of the optimal design of HOV lanes. That is, the planning department of the decision-maker and the follower traveler live in two levels of decision-making process, respectively, and they are relatively independent and affect each other. In order to achieve

the purpose of maximizing their own interests, they are optimized separately, and finally the effect of the optimization of the HOV lane layout is achieved. This paper solves this problem by constructing a two-level programming model.

According to the actual operation of HOV lane, VSSM is selected as the simulation software, the micro simulation platform of HOV lane is built, and the evaluation index system of traffic flow characteristics and HOV Lane characteristics is established. In this paper, the effectiveness of the simulation platform is tested and corrected. Based on the current situation of HOV Lane setting, combined with the relevant standards and laws and regulations of foreign HOV system, this paper summarizes the setting conditions of domestic HOV lane and verifies the effectiveness of the scheme by comparing the evaluation results before and after.

The upper-level model starts from the perspective of the planner (i.e., the system). By deciding which road section to build HOV dedicated lanes, the system efficiency of the entire transportation network can be optimized. Therefore, by considering the performance indicators of the generalized cost of the selection system, the objective function of the upper model is formed. By optimizing the objective function, the optimal HOV channel set is found, and the optimization result ensures that the generalized cost of the entire road network system is minimized, and the efficiency is also optimal. The lower-level model starts from the standpoint of road network users (travelers) and considers travel users as the lower-level decision makers. By choosing the corresponding travel strategy under the road network conditions defined by the upper-level model, the lowest travel cost is achieved. The result of the traveler's choice will in turn affect the result of the HOV lane optimization design of the upper model.

The specific modeling steps and framework are shown in Figure 6. Passenger Car Flow Distribution includes all vehicles driving on regular roads.

The regional city structure diagram is shown in Figure 7.

The urban periphery studied in this paper refers to the urban fringe area, which includes inner and outer edges, and has the following characteristics: (1) it maintains close contact with the urban core area and is the main penetration zone of the core area; (2) its population grows rapidly, and its structure is complex; (3) its land properties continue to change, and the intensity of land development continues to increase; (4) it is an area with frequent commutes. Factors such as urban population density, land properties, and economic development will affect the determination of the extent of urban fringe areas. At the same time, economic, transportation, location, policy, and other factors will also affect the expansion of urban fringe areas. Therefore, according to the characteristics of different cities, the scope of the outer city should be determined specifically. The connection distance between the urban center and the periphery of the city in our country is usually 10–20 km. The division of urban structure is not static. As the city's economic strength increases, and the scope of the urban core area increases, the original urban fringe area will gradually evolve into the core area, and the affected area will evolve

into the fringe area. The development of the city is the process of continuously increasing the extent and scope of land use.

There is a dynamic relationship between transportation, economy, and land use. The convenience of transportation can drive the development of urban economy, the development of economy drives land use, and the use of land strengthens the development of transportation. The relationship between the three is shown in Figure 8.

On the basis of the definition of the urban periphery and the commuter chain, we combine the research content of this article to summarize the definition of the rail transit commuter chain. That is, commuters in the periphery of the city use urban rail transit as the main travel mode at one time and connect with other modes of transportation to form a travel chain, as shown in the schematic diagram in Figure 9. This type of travel chain includes three processes: front-end travel (the travel phase of commuters starting from the residential area to the urban rail transit station), urban rail travel, and back-end travel (the commuter travels from the urban rail transit station to the work unit stage).

The rail transit commuter chain is divided into two types: simple and complex. The simple rail transit commuter chain refers to a chain that connects to the front and back ends of urban rail transit and has only one mode of transportation except walking. The complex rail transit commuter chain refers to a chain that connects the front and back ends of urban rail transit and has two or more modes of transportation except walking. In commuting travel, complex commuter chains usually bring a psychological burden to commuters, and frequent transfers cause time waste, so simple commuter chains are the first choice for commuters. Taking into account the limitations of the research conditions, the scope of this paper is defined as the simple rail transit commuter chain of "HOV-urban rail transit-walking".

Restricted by passenger demand and construction conditions, the characteristics of urban rail transit in the central area of the city and the outer area of the city are different. The urban rail transit in the central area is networked, and the urban rail transit in the outer city is often the end of a certain urban rail transit line, so determine the type of urban rail transit station outside the city as intermediate station or terminal station.

The urban rail transit in the periphery of the city is often the end of a certain urban rail transit line. Therefore, the type of urban rail transit station in the outer periphery of the city is determined as the intermediate station or the terminal station. When the distance between the urban rail transit station and the core area of the city is different, the radiation range is different. Therefore, a specific analysis should be made for a specific station outside the city to be studied. The boundary of the coverage area is assumed to be treated as a circle, and the attraction radius is obtained through survey statistics or model calculations. Affected by many aspects such as neighboring stations, topography, and road network layout, the boundary of the attraction range cannot be processed in a perfectly circular manner, as shown in
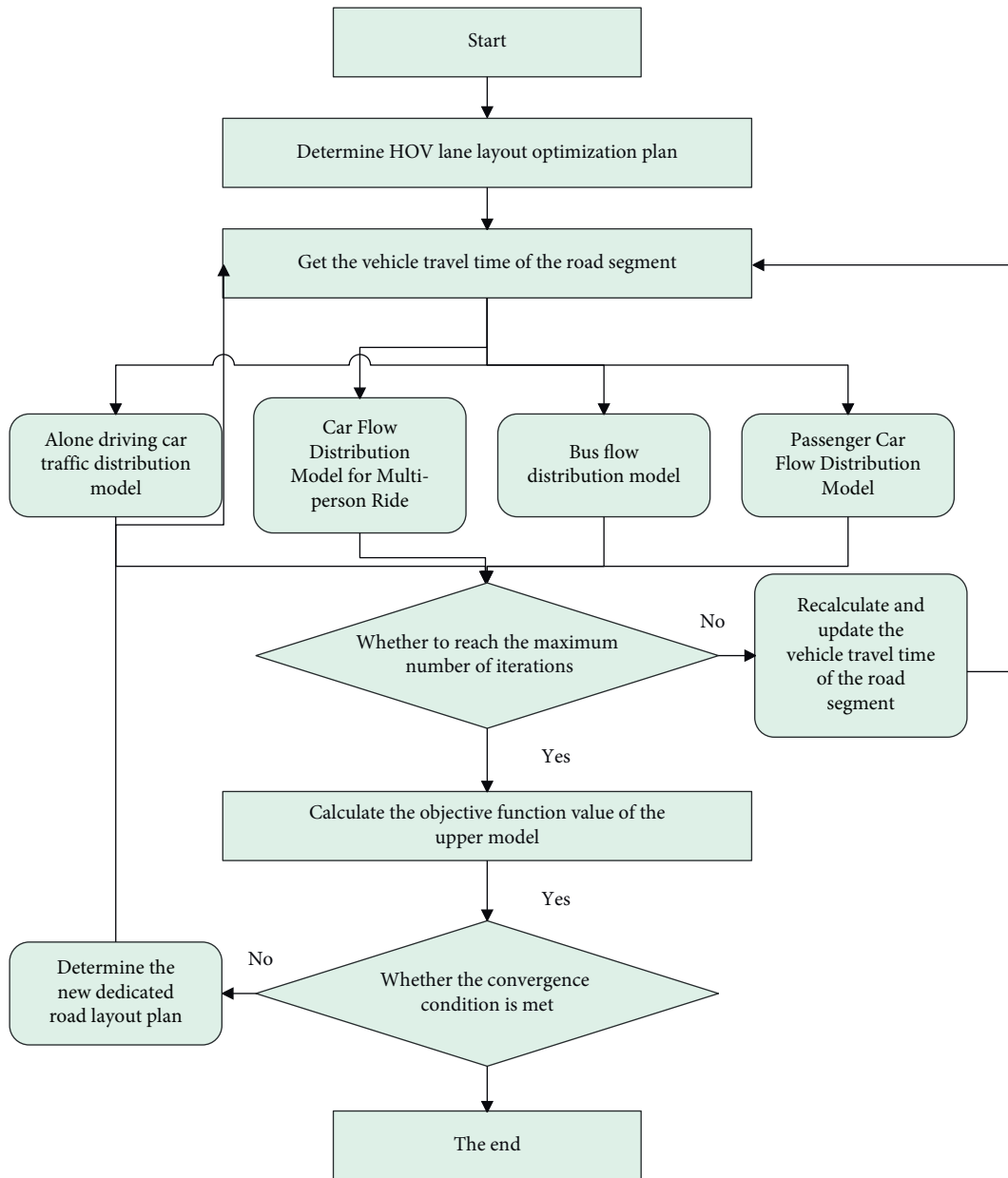
Figure 6: Model construction framework.
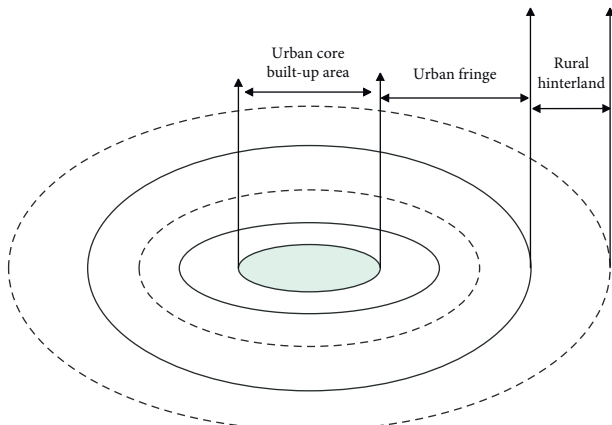
Figure 7: City structure diagram.

Figure 10. For the area where the two stations overlap, the dividing line is generally divided according to the river and road lines. This paper refers to the boundary of the attraction range of foreign "$P + R$" park and ride stations for cars and generally adopts the form of parabola for processing, as shown in Figure 11.

The same direction HOV lane with the bus lane is set in the rightmost lane, and the HOV lane is separated from the ordinary lane by marking lines. To make it easier for passengers to get on and off the bus, parking stations are arranged on the HOV lanes, speed bumps and dashed lines that can be crossed are set 500m from the front and back of the station, and solid lines that cannot be crossed are set at the rest of the road sections. The schematic diagram of road section setting is shown in Figure 12.
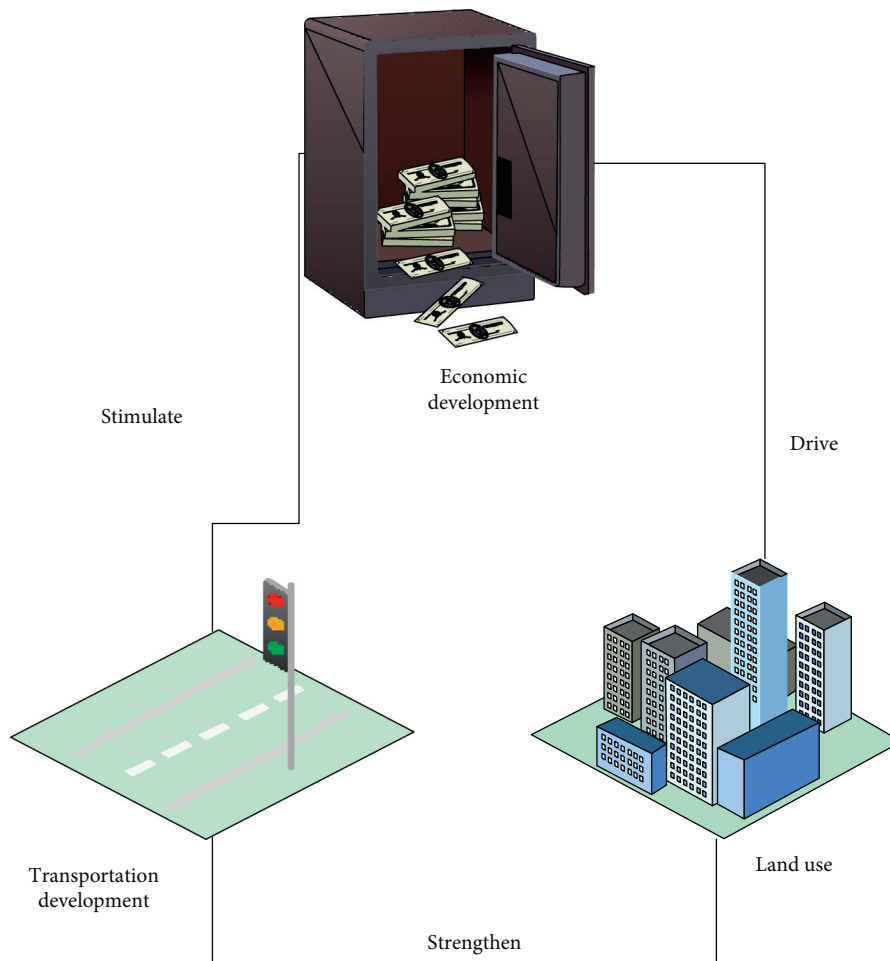
FIGURE 8: The relationship diagram between urban traffic, land use, and economic development.
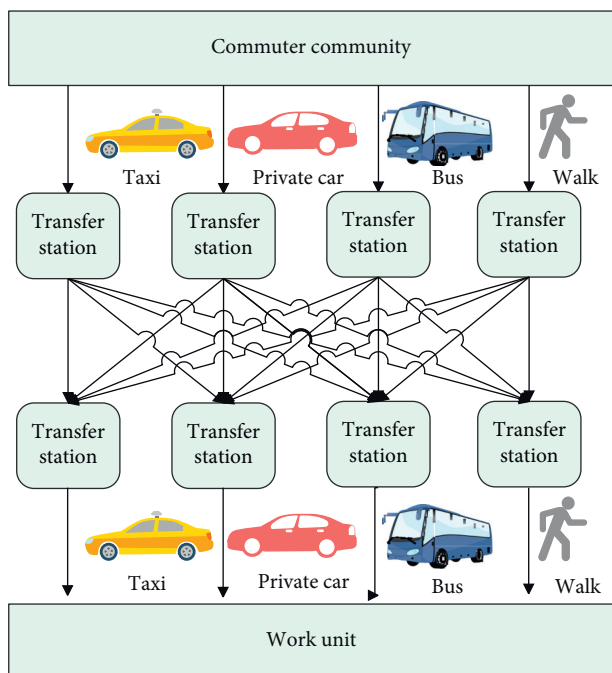
FIGURE 9: Schematic diagram of rail transit commuter chain.

5. Model Test Research

After constructing the HOV lane setting decision evaluation model, the performance of the system model is studied. The HOV model constructed in this paper is mainly used to ease urban traffic and improve urban operation efficiency. Therefore, in the process of urban operation, the system is mainly used to analyze the traffic data to determine the feasibility of the HOV lane setting, and after the feasibility is determined, the system is used to conduct the HOV lane planning analysis. In the experiment and research, this paper firstly analyzes the feasibility of HOV lane setting for the system. This paper randomly sets up 100 sets of experiments through computer simulation, collects city conditions through the network, conducts data simulation through the system, and conducts the scoring of the feasibility analysis of HOV lane setting through statistical models (quantitative analysis through subjective evaluation by experts). The results are shown in Table 1 and Figure 13. As shown in Figure 13, the abscissa is the simulation of 100 groups of HOV lanes set, and the ordinate is the feasibility score of HOV Lane set by the model.

It can be seen from the analysis results of the above charts that the system constructed in this article has a good
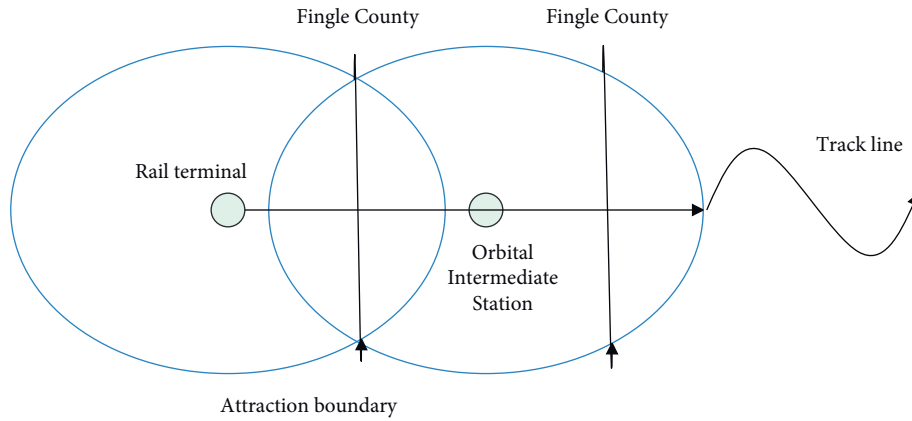
FIGURE 10: Schematic diagram of the boundary of the attraction range of a rail transit station.
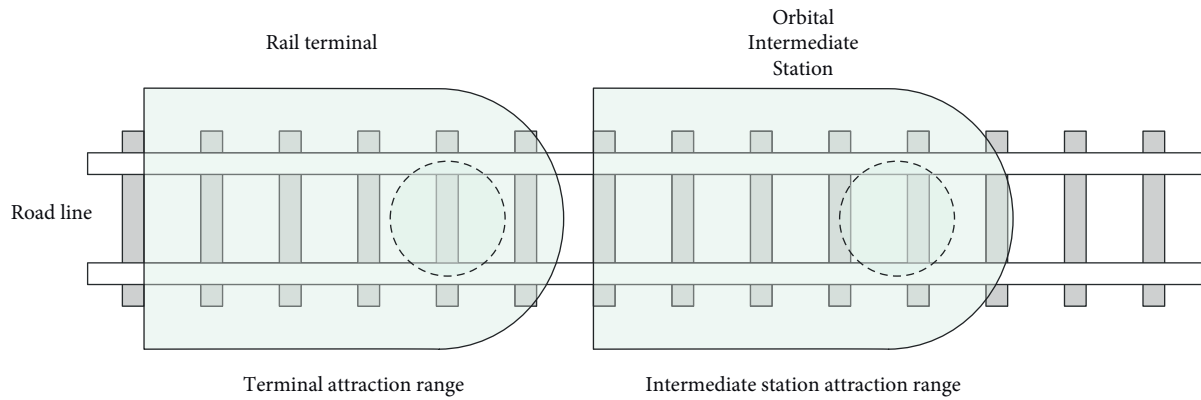
FIGURE 11: Schematic diagram of the parabolic boundary of the P&R site coverage area.
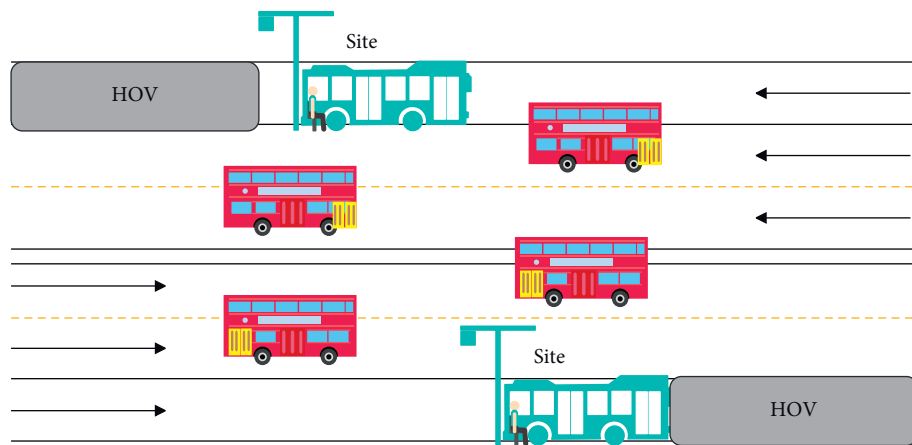
FIGURE 12: Schematic diagram of HOV lane road section.

effect in the feasibility analysis of HOV lane setting. Therefore, on the basis of the feasibility analysis, the HOV lane planning analysis is carried out, and the scoring method is used to count the test results (quantitative analysis through subjective evaluation by experts), as shown in Table 2 and Figure 14.

From the above test results, the decision evaluation model of HOV lane setting constructed in this paper

TABLE 1: Statistical table of decision scores for HOV lane setting.

Number	Scores	Number	Scores	Number	Scores
1	75.31	35	80.86	69	85.43
2	76.98	36	79.47	70	86.53
3	82.11	37	86.47	71	83.28
4	86.24	38	78.01	72	79.87
5	90.17	39	75.27	73	87.48
6	89.83	40	81.47	74	80.52
7	78.77	41	76.51	75	83.42
8	89.68	42	86.85	76	84.11
9	83.49	43	85.62	77	87.05
10	84.22	44	91.18	78	76.81
11	85.66	45	76.35	79	84.42
12	89.05	46	78.99	80	77.12
13	85.25	47	91.94	81	88.52
14	79.07	48	77.09	82	90.68
15	85.11	49	90.58	83	77.30
16	82.68	50	88.20	84	89.64
17	78.37	51	75.29	85	87.23
18	81.24	52	91.19	86	80.31
19	82.10	53	83.43	87	90.54
20	90.70	54	75.18	88	81.66
21	81.16	55	91.32	89	77.85
22	91.46	56	84.59	90	79.48
23	81.04	57	89.77	91	81.05
24	83.05	58	84.95	92	82.81
25	82.23	59	76.67	93	85.74
26	77.75	60	86.56	94	85.03
27	78.14	61	85.79	95	80.98
28	88.56	62	89.65	96	90.45
29	80.64	63	89.85	97	80.72
30	78.32	64	91.67	98	80.75
31	87.50	65	91.92	99	84.56
32	89.87	66	84.11	100	84.64
33	81.30	67	89.14		
34	83.05	68	89.43		
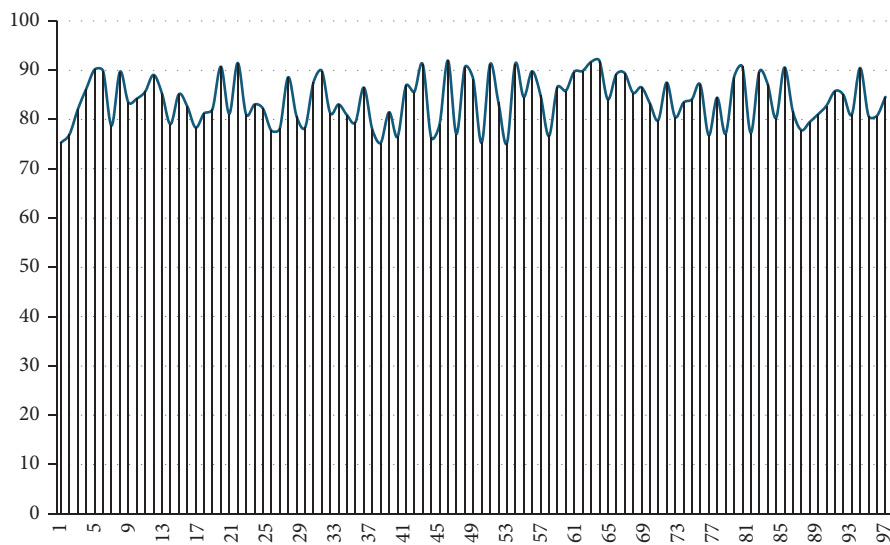
FIGURE 13: Statistical diagram of decision scores for HOV lane setting.

TABLE 2: Statistical table of scores of HOV lane planning.

Number	Scores	Number	Scores	Number	Scores
1	78.6	35	81.1	69	83.3
2	82.6	36	83.7	70	73.2
3	80.8	37	69.9	71	75.8
4	84.8	38	70.8	72	74.1
5	77.4	39	73.9	73	83.5
6	80.2	40	76.6	74	71.6
7	75.9	41	78.5	75	70.1
8	73.7	42	71.0	76	69.9
9	70.3	43	79.5	77	76.5
10	80.3	44	73.8	78	81.8
11	72.3	45	69.3	79	80.1
12	77.5	46	80.3	80	73.9
13	84.9	47	82.0	81	82.7
14	79.0	48	71.4	82	75.2
15	82.9	49	71.2	83	82.1
16	82.3	50	80.0	84	73.0
17	82.0	51	83.5	85	72.3
18	71.9	52	75.5	86	78.3
19	83.5	53	71.9	87	74.2
20	82.2	54	76.0	88	84.2
21	75.6	55	69.7	89	83.5
22	81.0	56	83.0	90	71.9
23	70.3	57	78.6	91	82.9
24	70.3	58	77.9	92	70.7
25	78.9	59	79.1	93	76.9
26	78.6	60	75.9	94	76.5
27	84.3	61	75.7	95	75.1
28	83.5	62	69.9	96	73.2
29	72.6	63	74.3	97	71.9
30	75.9	64	84.1	98	77.6
31	79.1	65	78.0	99	75.6
32	78.5	66	74.8	100	76.0
33	83.7	67	77.0		
34	81.5	68	75.6		
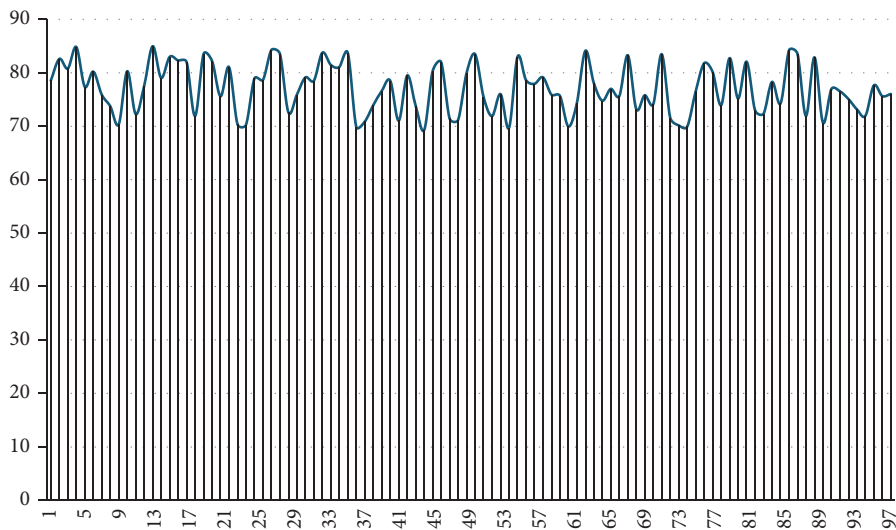
FIGURE 14: Statistical table of scores of HOV lane planning.

basically meets the expected system setting requirements and has certain practical effects.

6. Conclusion

This paper analyzes the feasibility of setting up HOV lanes in the service rail transit commuter chain. This article first gives a theoretical introduction to the rail transit commuter chain and HOV lanes and then determines the research methods for the attraction range of rail transit stations on the periphery of the city. Moreover, this paper uses two methods of theoretical analysis and numerical analysis to illustrate the travel characteristics and ride-sharing demand of commuters within the urban rail transit attraction area and provides a method for forecasting travel demand in the HOV commuter chain. In addition, this paper determines the traffic and road conditions that should be achieved when planning and setting up HOV lanes within the attraction range of urban rail transit stations from the perspective of demand and supply. This paper combines machine learning to construct a decision-making evaluation model for HOV lane setting and studies the optimal layout model and algorithm of HOV lanes in service rail transit commuter chain. The setting, planning, and layout of HOV lanes are a two-way interactive process of traveler's path selection and designer's road planning. Finally, after the model is constructed, the performance of the system model is verified. The research results show that the route optimization model of HOV lane constructed in this paper has a certain effect. This paper analyzes the opening of HOV lane, which is of great help to reduce the flow and delay, improve the average speed and service level, and increase the number of section transportations. However, in the case of limited road resources, how to reduce the waste of no-load vehicles on road resources and tilt the right of way to multioccupant and efficient transportation mode is studied. To improve the transportation efficiency of road section, further research is needed in the future [25].

Data Availability

The experimental data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest regarding this work.

Acknowledgments

References

[1] D. Nallaperuma, R. Nawaratne, T. Bandaragoda, A. Adikari, and D. Pothuhera, "Online incremental machine learning platform for big data-driven smart management," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 12, pp. 4679–4690, 2019.

[2] M. Rath and B. Pati, "Communication improvement and traffic control based on V2I in smart city framework," *International Journal of Vehicular Telematics and Infotainment Systems*, vol. 2, no. 1, pp. 18–33, 2018.

[3] W.-P. Chen, A.-H. Tsai, and C.-H. Tsai, "Smart traffic off-loading with mobile edge computing for disaster-resilient communication networks," *Journal of Network and Systems Management*, vol. 27, no. 2, pp. 463–488, 2019.

[4] "Design and implementation of smart traffic light controller using VHDL language," *International Journal of Engineering & Technology*, vol. 8, no. 4, pp. 596–602, 2019.

[5] A. R. Malipatil, K. Somasundaram, M. Ashwini et al., "Smart traffic control barricade system," *Journal of Critical Reviews*, vol. 7, no. 14, pp. 1658–1664, 2020.

[6] A. Rehman, M. Mazhar Rathore, A. Paul, F. Saeed, and R. W. Ahmad, "Vehicular traffic optimisation and even distribution using ant colony in smart city environment," *IET Intelligent Transport Systems*, vol. 12, no. 7, pp. 594–601, 2018.

[7] H. K. Hamarashid, M. H. R. Saeed, and S. Saeed, "Designing a smart traffic light algorithm (HMS) based on modified round robin algorithm," *Kurdistan Journal of Applied Research*, vol. 2, no. 1, pp. 27–30, 2017.

[8] A. Sivanathan, H. H. Gharakheili, F. Loi, A. Rad ford, and C. Wijenayake, "Classifying IoT devices in smart environments using network traffic characteristics," *IEEE Transactions on Mobile Computing*, vol. 18, no. 8, pp. 1745–1759, 2018.

[9] J. D. K. H and J. Deva, "Computer vision based traffic sign sensing for smart transport," *Journal of Innovative Image Processing*, vol. 1, no. 1, pp. 11–19, 2019.

[10] Z. Ning, J. Huang, and X. Wang, "Vehicular fog computing: enabling real-time traffic management for smart cities," *IEEE Wireless Communications*, vol. 26, no. 1, pp. 87–93, 2019.

[11] J. Qiu, L. Du, D. Zhang, S. Su, and Z. Tian, "Nei-TTE: intelligent traffic time estimation based on fine-grained time derivation of road segments for smart city," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 4, pp. 2659–2666, 2019.

[12] J. Liu, C. Zhang, and Y. Fang, "EPIC: a differential privacy framework to defend smart homes against internet traffic analysis," *IEEE Internet of Things Journal*, vol. 5, no. 2, pp. 1206–1217, 2018.

[13] A. Sharif, J. P. Li, and M. A. Saleem, "Internet of things enabled vehicular and ad hoc networks for smart city traffic monitoring and controlling: a review[J]," *International Journal of Advanced Networking and Applications*, vol. 10, no. 3, pp. 3833–3842, 2018.

[14] S. Arulselvi and M. Sundararajan, "Smart control system in traffic analysis using RTK-GPS standards," *International Journal of Pure and Applied Mathematics*, vol. 116, no. 15, pp. 349–353, 2017.

[15] Z. Li, R. Al Hassan, M. Shahidehpour, S. Bahramirad, and A. Khodaei, "A hierarchical framework for intelligent traffic management in smart cities," *IEEE Transactions on Smart Grid*, vol. 10, no. 1, pp. 691–701, 2017.

[16] C. Raskar and S. Nema, "Modified fuzzy-based smart barricade movement for traffic management system," *Wireless Personal Communications*, vol. 116, no. 4, pp. 3351–3370, 2021.

[17] L. Sumi and V. Ranga, "Intelligent traffic management system for prioritizing emergency vehicles in a smart city," *International Journal of Engineering*, vol. 31, no. 2, pp. 278–283, 2018.

[18] A. I. Adekitan, J. Abolade, and O. Shobayo, "Data mining approach for predicting the daily Internet data traffic of a smart university," *Journal of Big Data*, vol. 6, no. 1, pp. 1–23, 2019.

[19] M. Li, P. Si, and Y. Zhang, "Delay-tolerant data traffic to software-defined vehicular networks with mobile edge computing in smart city," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 10, pp. 9073–9086, 2018.

[20] P. Bellucci and E. Cipriani, "Data accuracy on automatic traffic counting: the SMART project results," *European transport research review*, vol. 2, no. 4, pp. 175–187, 2010.

[21] X. Wang and Z. Li, "Traffic and transportation smart with cloud computing on big data," *Indian Journal of Chemistry, Section A*, vol. 13, no. 1, pp. 1–16, 2016.

[22] G. Hou, S. Chen, Y. Zhou, and J. Wu, "Framework of microscopic traffic flow simulation on highway infrastructure system under hazardous driving conditions," *Sustainable and Resilient Infrastructure*, vol. 2, no. 3, pp. 136–152, 2017.

[23] Y. Xue, Y. Zhang, D. Fan, P. Zhang, and H.-d. He, "An extended macroscopic model for traffic flow on curved road and its numerical simulation," *Nonlinear Dynamics*, vol. 95, no. 4, pp. 3295–3307, 2019.

[24] G. Ghermandi, S. Fabbi, A. Bigi et al., "Impact assessment of vehicular exhaust emissions by microscale simulation using automatic traffic flow measurements," *Atmospheric Pollution Research*, vol. 10, no. 5, pp. 1473–1481, 2019.

[25] V. O. Matthews, S. I. Uzairue, and E. Noma-Osaghae, "Design and simulation of a smart traffic system in a campus community," *Journal of Emerging Technologies and Innovative Research (JETIR)*, vol. 5, no. 7, pp. 492–497, 2018.

Research Article

Digital Audio Scene Recognition Method Based on Machine Learning Technology

Sihua Sun (ID)

Anhui Art College, Hefei, Anhui Province 230011, China

Correspondence should be addressed to Sihua Sun; 114039@ahua.edu.cn

Audio scene recognition is a task that enables devices to understand their environment through digital audio analysis. It belongs to a branch of the field of computer auditory scene. At present, this technology has been widely used in intelligent wearable devices, robot sensing services, and other application scenarios. In order to explore the applicability of machine learning technology in the field of digital audio scene recognition, an audio scene recognition method based on optimized audio processing and convolutional neural network is proposed. Firstly, different from the traditional audio feature extraction method using mel-frequency cepstrum coefficient, the proposed method uses binaural representation and harmonic percussive source separation method to optimize the original audio and extract the corresponding features, so that the system can make use of the spatial features of the scene and then improve the recognition accuracy. Then, an audio scene recognition system with two-layer convolution module is designed and implemented. In terms of network structure, we try to learn from the VGGNet structure in the field of image recognition to increase the network depth and improve the system flexibility. Experimental data analysis shows that compared with traditional machine learning methods, the proposed method can greatly improve the recognition accuracy of each scene and achieve better generalization effect on different data.

1. Introduction

As an information carrier, sound is an important way for us to perceive the external environment. With the development of signal processing technology and computer science, the audio processing task of extracting information from sound assisted by machine has attracted more and more researchers' attention [1–6]. Compared with images with multimedia information, the acquisition of audio files is not limited by light environment and visual obstacles. In addition, audio occupies less capacity and faster processing speed compared with images [7–9]. Audio processing tasks include speech recognition, audio fingerprint, music mark, audio scene recognition, etc.

Audio scene recognition belongs to the subfield of computational auditory scene analysis. Its main goal is to enable devices to understand and distinguish their environment by analyzing sound. The implementation principle is that the equipment extracts different audio features through audio scene recognition technology to obtain the corresponding features and then models the audio scene according to these features, that is, constructs a classifier. After learning enough samples, the classifier will judge the audio scene category according to the extracted audio features [10–13]. The applications of audio scene recognition include context-aware services, intelligent wearable devices, robot sensing, and robot hearing. In addition, audio scene recognition also complements the research in several related fields. Among them, the detection and classification of audio events are often associated with audio scene recognition because the audio scene can be regarded as the product of the superposition of several audio events. On the other hand, audio scene recognition can improve the performance of sound event detection by providing a priori information about the probability of some events.

Previously, the implementation of audio scene recognition often applied general classifiers (such as Gaussian mixture model [14–16], support vector machine, and hidden

Markov model) to manually extracted features, such as mel-frequency cepstrum coefficient. In recent years, thanks to the improvement of computer speed and the rapid development of deep learning, people gradually realize that the characteristics of automatic feature extraction of deep learning can replace the inefficient manual extraction in the past. At the same time, more and more recording devices such as smart phones have made great contributions to the expansion of audio datasets [17]. With a large amount of audio data, it is possible to realize the deep learning method which is difficult to realize in the past.

2. Literature Review

Influenced by psychoacoustic/psychological technology, most of them emphasize the local and global features of audio scene recognition. In addition, a few researchers focus on the time-domain features of audio. Mitsukura [18] used mel-frequency cepstrum (MFCC) to describe the local spectral envelope of audio signal and Gaussian mixture model (GMM) to describe its statistical distribution. Zhao et al. [19] proposed to train the hidden Markov model by using the discriminant algorithm of the knowledge of training signal types to explain the time-domain evolution of GMM. Abrol and Sharma [20]. further improved the recognition performance by considering more features and adding a feature transformation step in the classification algorithm and obtained an average accuracy of 58% in 18 different sound fields [21].

Convolutional neural network (CNN) is a deep learning network model inspired by animal visual system. Its network composition imitates the principles of various cells in the visual system to construct the network model [22–24]. CNN was originally designed for feature extraction of two-dimensional data. It can directly establish the mapping relationship from low-level features to high-level semantic features and has achieved remarkable results in the field of two-dimensional image classification. Zhao et al. [25] proposed driver fatigue state recognition based on CNN. Cai et al. [26] proposed a CNN-based video classification method, which uses convolution filter and global average pool layer to obtain more detailed features. Tan et al. [27] discussed whether CNN in deep learning can be effectively applied to audio scene recognition.

Although applying a variety of deep learning methods to audio scene classification speeds up the research process in this field, there are still two problems worth discussing:

(1) The mainstream audio feature extraction methods are based on mel spectrum. Is there any special processing skill for mel spectrum to make it more in line with specific application scenarios, so as to improve the accuracy of audio classification?

(2) Since most of the results after audio feature extraction also belong to images, can the excellent volume of CNN architecture in the field of image recognition be applied to audio scene classification to improve the system performance?

Aiming at the above two problems, this paper proposes an audio scene recognition method based on optimized audio processing and CNN. The basic system is improved from two aspects: audio processing and network structure. In audio processing, binaural representation and harmonic impact source separation are used to process the original audio and extract the corresponding features, so that the system can make use of the spatial features of the scene, and then the classification accuracy has been significantly improved. In terms of network structure, we try to learn from the VGGNet structure in the field of image recognition, improve the system flexibility while increasing the network depth, and finally achieve better generalization effect on different data. Finally, the effectiveness of the proposed method is verified by experimental data analysis.

3. Audio Processing Optimization Method

3.1. Binaural Representation. Traditional audio processing methods always use MFCC feature in feature extraction. Although MFCC can describe features concisely with more than a dozen coefficients, MFCC is not competent for audio scene classification in scenes with obvious spatial features such as libraries.

Although it is common to use stereo equipment for recording, the signal is usually averaged to make it a mono before processing. Although using mono is easy for processing and feature extraction, it will lose a lot of spatial information. If important audio information is captured well in only one of the channels, problems may occur. Because averaging two channels to one channel will reduce the signal-to-noise ratio, the difference between the two channels is not easy to be reflected, so it is very easy to cause confusion in the classification process. Analyzing the two channels separately can alleviate this problem. In view of the excellent results obtained in previous challenges using binaural representation, this paper decides to use LR (left right) representation and MS (mid-side) representation in feature processing.

LR representation represents the left and right channels in conventional stereo recording. For example, when a car passes in front of a microphone, the sound moves from L channel to R channel or from R channel to L channel, which is only reflected in amplitude change in mono. By introducing LR representation, the movement of sound source in space can be reflected. In this section, only the left and right channels of the source audio file are separated. The MS representation emphasizes the time difference between sounds reaching each side of the stereo microphone. MS representation obtains the final result by summing and subtracting the waveforms of two stereo input channels, respectively. Mel spectrum will be extracted for the above four representations, as shown in Figure 1.
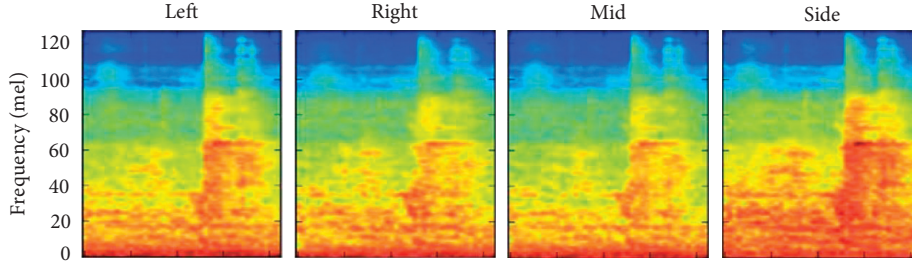
FIGURE 1: MFCC spectrum of binaural representation.

For LR spectrum and MS spectrum, they are input into the subsequent convolution neural network as one of the classification features.

3.2. Harmonic Percussive Source Separation Method.
Sound can generally be divided into two types: harmonic and impact sound. In the traditional research work, harmonic impulse source separation (HPSS) algorithm is proposed under the background of music signal processing. The goal is to decompose the input audio signal into all harmonics and signals composed of all impulse sources. In order to solve the problem of poor generalization and dependence on learning data in audio scene classification, this section draws lessons from music signal processing to try to improve the classification performance of the system. The steps of HPSS algorithm are given below.

Assume that the input discrete input audio signal is $\lambda \in \mathbb{R}$. The harmonic component signal x_h and the shock source component signal x_p shall be calculated so that $x = x_h + x_p$.

First, the short-time Fourier transform of x can be expressed as

$$X(t,k) = \sum_{n=0}^{N-1} x(n + tH)w(n)\exp\left(\frac{-2\pi ikn}{N}\right), \qquad (1)$$

where T is the number of frames, N is the frame size of Fourier transform, $w(n)$ is the window function, and H is the frame offset.

The input power spectrum Y can be calculated by the following formula:

$$Y(t,k) = |X(t,k)|^2. \qquad (2)$$

Next, the harmonic enhancement spectrum $\widetilde{Y}_h$ and the shock source enhancement spectrum $\widetilde{Y}_p$ are calculated by median filtering Y. Assuming that A is a set composed of a column of real numbers and N is the number of real numbers in the set, the median filtering of A is defined as

$$\text{median}(A) = \begin{cases} a_{N-1/2}, & N \text{ is an odd number}, \\[2mm] \dfrac{a_{N/2} + a_{N-1/2}}{2}, & N \text{ is an even number}. \end{cases}$$

$$(3)$$

Then, according to the definition of median filter, the harmonic enhancement spectrum $\widetilde{Y}_h$ and impact source enhancement spectrum $\widetilde{Y}_p$ can be obtained by performing one horizontal and one vertical median filter on Y, respectively.

$$\widetilde{Y}_h(t,k) = \text{median}(Y(t - l_n, k), \ldots, Y(t + l_n, k)),$$
$$\widetilde{Y}_p(t,k) = \text{median}(Y(t, k - l_p), \ldots, Y(t, k + l_n)), \qquad (4)$$

where l_n and l_p are filter lengths.

Then, a variable β is introduced, which is called the separation factor. The original input signal $X(t,k)$ can be intuitively judged as harmonic or impact source component. Through this rule, binary masks M_h and M_p can be defined.

$$X_h(t,k) = X(t,k) \cdot M_h(t,k),$$
$$X_p(t,k) = X(t,k) \cdot M_p(t,k). \qquad (5)$$

Finally, the required signals x_h and x_p can be calculated by transforming these spectra into time domain by using inverse short-time Fourier transform.

In the separation process, with the help of the decompose.hpss method in the Librosa library, the separation factor for the experiment is 1.05. Convert stereo audio to mono before separation. As shown in Figure 2, when the HPSS algorithm is applied to the input signal, the harmonic tends to form a horizontal structure (in the time direction) and the impact source tends to form a vertical structure (in the frequency direction) on the mel spectrum.

4. Audio Scene Recognition Method Based on Convolutional Neural Network

Convolution layer is the core module of convolutional neural network, which can complete most of the heavy computing work. Its function is to extract the local features of a region. The core operation performed by the convolution layer is called convolution. Convolution is a common operation method in analytical mathematics. It is a mathematical operator that generates the third function through two functions. In the application of machine learning, convolution is usually embodied in sliding a filter on an image or some feature and using this operation to obtain a set of new features.

In this study, an audio scene classification system based on convolutional neural network will be designed and compared with traditional machine learning methods. The classification system mainly includes training and testing, and its process design is shown in Figure 3.
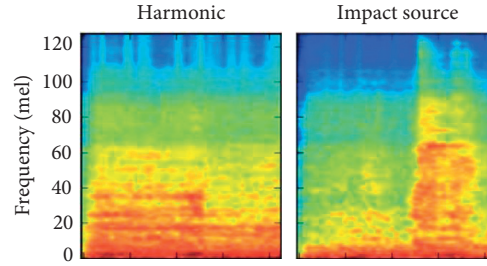
FIGURE 2: MFCC spectrum of harmonic percussive source separation method.
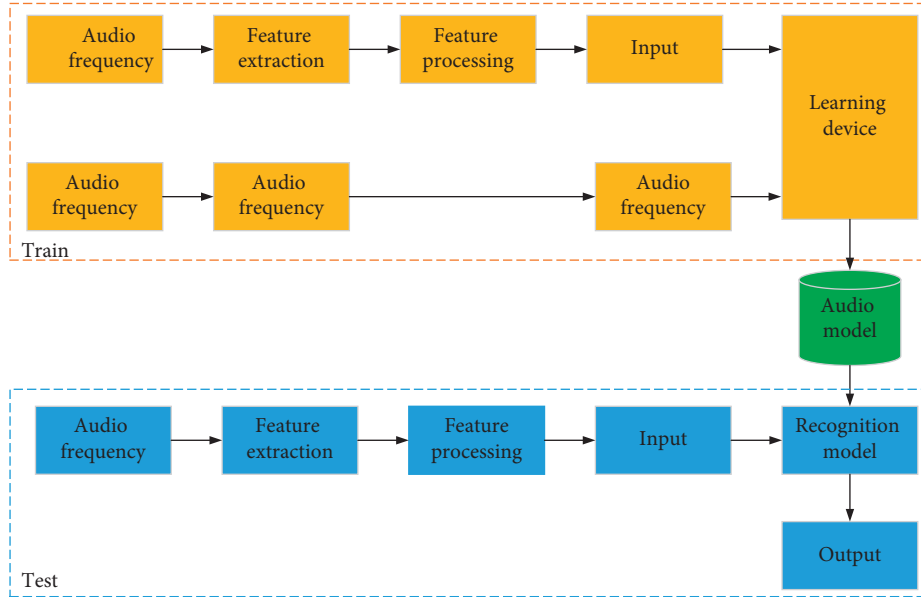
FIGURE 3: Audio scene recognition process based on convolutional neural network.

4.1. Convolutional Neural Network Architecture. Convolutional neural network structure also has the problem of relying too much on training parameters in the training process [28]. At present, there are many convolutional neural network frameworks with good performance that can be widely used. Whether its structure can be slightly modified and used in audio scene classification to improve system performance is also the focus of this section. The convolutional neural network module of the system core is shown in Figure 4.

The first layer performs convolution on the input spectrum. Since the size of convolution filter is reflected in the size of local block diagram and the size of local block diagram with different sizes determines the speed of feature extraction [29], filter sizes with different sizes will be arranged for comparison in the experiment. In addition, the number of filters will also affect the analysis angle of features, so different filter numbers will be arranged for comparison during the experiment. More filters will increase the angle of feature analysis but also increase the amount of calculation, and too many filters may lead to parameter redundancy. Generally, the number of filters is 2^n. After the convolution process is completed, the maximum pooling layer is used to subsample the obtained feature map.

The second convolution layer is basically the same as the first convolution layer, except that the second layer uses more cores (twice the first layer) to represent features at a higher level. Then, the second and last subsampling is performed for the "destruction" of the timeline. Therefore, the maximum pooling layer is still used, which operates over the entire sequence length. The activation function for the kernel in the convolution layer is ReLU.

Finally, because the classification involves 15 different classes, the last layer is the softmax layer composed of 15 fully connected neurons, which normalizes the output results of the network and makes the system output the classification results. Assuming that y_i is the output of the upper layer neuron i, y_i can be defined as

$$y_i = \text{softmax}(x_j) = \frac{\exp(x_j)}{\sum_{j=1}^{N} \exp(x_j)}, \tag{6}$$

where N is the total number of categories, x_i is the nonlinear input, and y_i is the prediction score of the input sequence belonging to class i.

4.2. Batch Standardization and Dropout Mechanism. In order to solve the problem of decreasing the network
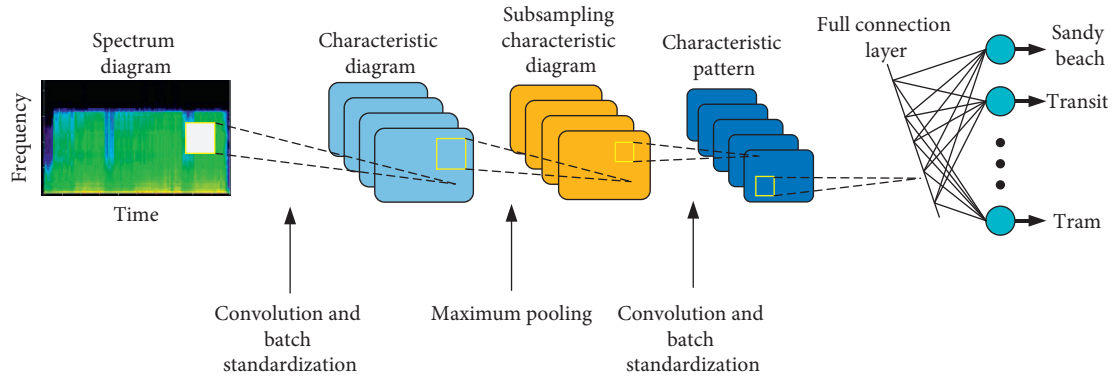
FIGURE 4: Composition of convolutional neural network module.

learning speed and convergence speed caused by internal covariate transformation, batch standardization is introduced to change the distribution of input data. The batch standardization layer applies linear transformation $BN_{\beta,\gamma}$ to its input x, as shown in the following formula:

$$BN_{\beta,\gamma} = \frac{\gamma}{\sqrt{\mathrm{Var}(x) + \varepsilon}} \cdot x + \left(\beta - \frac{\gamma \cdot E[x]}{\sqrt{\mathrm{Var}(x) + \varepsilon}}\right), \qquad (7)$$

where $E[x]$ is the average value of batch standardization layer input, $\mathrm{Var}(x)$ is the variance of batch standardization layer input, and β and γ represent the transformation parameters to be learned during training.

By using batch standardization, the input feature distributions have the same mean and variance, and the correlation between features is removed. Although it increases the complexity of the model, it slows down the transformation process of internal covariates, greatly reduces the training convergence time, and speeds up the learning progress.

In the convolutional neural network model, overfitting often occurs if there are too many parameters and too few training samples. Overfitting is embodied as follows: the loss function of the model is small in the training stage and the prediction accuracy is high, but the loss function is large and the accuracy is low in the testing stage.

In order to solve the over fitting phenomenon, the dropout mechanism is adopted in this paper. Suppose the output of the neural network is x and the output is y. After dropout is introduced, half of the hidden neurons in the network will be hidden randomly, as shown in Figure 5, while the input and output neurons remain unchanged. Then, the input x is still propagated forward through the modified network, and the loss result is propagated back through the network. After a batch of training samples, the parameters were updated by the gradient descent method on the neurons that were not deleted. Finally, continue to repeat the process.

Because half of the hidden neurons are deleted randomly, the network structure changes. The whole dropout process is equivalent to averaging multiple different neural networks. By introducing the average effect of dropout mechanism, some opposite fitting in the network is offset, so as to achieve the effect of similar model integration and reduce overfitting.

4.3. Improved Network Structure Design.

The traditional convolutional neural network contains two convolution modules and uses a single mel spectrum as the input. Due to the simple network structure, the system performance can only be improved by adjusting parameters one sidedly. However, too many parameters will make the model dependent on data, resulting in weak generalization. Therefore, changing the network structure, such as increasing the depth, to enhance the classification ability of the system for different datasets is an important means to enhance the system performance.

In recent years, deep convolution neural network has been widely used in the field of computer vision. One of the benefits brought by the increase of network depth is that the flexibility of the system is greatly enhanced. In the current widely used framework, VGGNet has been widely used because of its simple architecture and strong expansibility. VGGNet was jointly developed by the computational vision group of Oxford University and Google DeepMind. A major feature of VGGNet is that the entire neural network uses a convolution kernel size of $3 * 3$ and a maximum pool size of $2 * 2$. In addition, although VGGNet has deeper layers and more parameters than conventional neural networks, VGGNet can converge with only a few iterations because the depth of the network and the small-size filter play an implicit normalization role. On the other hand, it initializes the parameters using the data obtained by pretraining in a specific layer.

Inspired by VGGNet, the improved convolution neural network structure decides to also use the convolution kernel size of $3 * 3$ and more convolution layers to improve the classification performance. The overall architecture of the designed system is shown in Figure 6.

The whole network organization includes three layers: convolution block, mono model, and dual channel model. Among them, the convolution block is responsible for the convolution operation of the system core and includes the steps of zero filling, batch standardization, activation function, and so on. The size of each convolution kernel in the convolution block is $3 * 3$, and the zero filling size is $1 * 1$. The mono channel model consists of convolution blocks, maximum pooling, and overall average pooling steps, which is responsible for processing one of the input
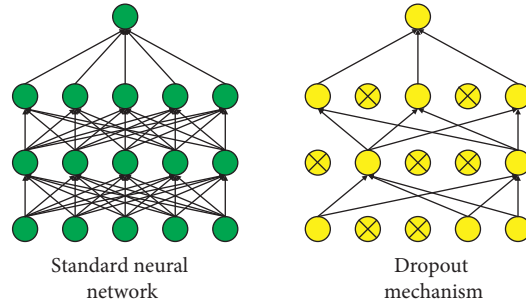
FIGURE 5: Dropout mechanism.

channels, and each channel is provided with four layers of convolution blocks. Similar to the design of VGGNet, the number of filters in each convolution block is doubled to 32, 64, 128, and 256.

5. Experimental Results and Analysis

5.1. Experimental Environment and Dataset. This experiment is implemented under the Ubuntu system with Linux as the kernel. The processor used in the system is an 8-core Intel E3-1270@3.8 GHz CPU, with 32 GB memory. The experiment uses Python language and introduces the external library Librosa for feature extraction. In the training part, sklearn library is introduced for unsupervised learning using the GMM model. The audio feature part includes 60-dimensional MFCC feature vector, including 20 MFCC static coefficients. For each audio file, log mel energy is extracted in 40 frequency bands using 40 ms analysis frame and 50% frame shift window. The improved convolutional neural network structure was introduced in Section 4.3. Its input feature is a single channel learner that only extracts mel spectrum. In addition, there is a dropout layer behind each CNN layer, and the dropout ratio is 30%. The parameters of convolutional neural network are shown in Table 1.

The TUT Acoustic Scenes 2017 dataset was used in this experiment, and the team responsible for collection is the audio research group of Tampere University of Technology. The dataset consists of 15 audio scenes with different labels: beach, bus, cafe/restaurant, car, city center, forest path, grocery store, family, library, subway station, office, park, community, train, and tram. All audio files are cut into segments with a length of 30 seconds, and the audio file format is wav. The dataset used in this paper is divided into development set and verification set. Among them, the development set contains 4680 audio files, and the number of files in each type of scene is 312. About 70% of the data are used to train the audio scene classification model, and the remaining 30% are used for testing. In the verification set, there are 1620 audio files, including 108 audio files of each type. The length of each audio segment is 10 seconds, and the audio of each scene is 18 minutes in total.

5.2. Identification of Performance Evaluation Indicators. The evaluation standard adopted is accuracy (ACC), that is, the ratio between the predicted correct number of samples and the total number of samples, which is calculated as follows:

$$ACC = \frac{N_{\text{true}}}{N_{\text{total}}}, \tag{8}$$

where N_{true} represents the number of correctly predicted samples and N_{total} represents the total number of samples.

5.3. Result Analysis. The classification accuracy of each scene in the development set and verification set of the proposed method is shown in Table 2.

By observing Table 2, we can find that although the average classification accuracy of the verification set is significantly lower than that of the development set, it still achieves a good result of 81.5%. It can be seen that the generalization ability of the improved system has been significantly improved, especially in the scenes of subway station, forest path, car, and home, and the classification accuracy is more than 90%, but there is still room for improvement in parks, libraries, and other scenes.

In addition, the proposed method is compared with GMM and traditional CNN in the development and verification of two datasets in order to analyze the specific performance of the improved method. The comparison between GMM and the proposed method is shown in Figure 7.

As can be seen from Figure 7, the classification accuracy of the proposed method is significantly ahead of GMM in most scenes, especially in beach, bus, library, park, and other scenes, both in the development set and verification set. Analyzing the possible reasons, on the one hand, the convolutional neural network itself has stronger learning ability for data; on the other hand, the sound field space of the above scene is large and has fixed bottom noise, which just fits the role of audio processing on the audio environment of fixed scene.

The comparison between CNN and the proposed method is shown in Figure 8.

It can be seen from the analysis of Figure 8 that although the proposed method does not achieve overall advantages in the development set, the overall classification accuracy of the proposed method in the verification set is about 19% higher than that of traditional CNN. This shows that the generalization performance of the network structure has been significantly enhanced after the introduction of VGGNet. It also shows that the flexibility performance of the system has been greatly improved by increasing the network depth and simplifying the network parameters.
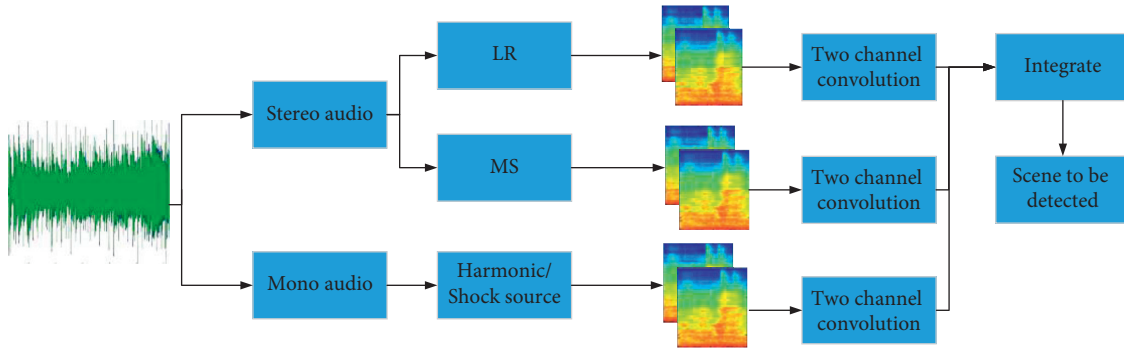
FIGURE 6: Improved convolution network structure design.

TABLE 1: Parameters of convolutional neural network.

Layer (type)	Output size	Parameter
Conv_1	(40, 500, 128)	6400
Batch standardization_1	(40, 500, 128)	160
Activation function_1	(40, 500, 128)	0
Maximum pooling_1	(8, 100, 128)	0
Dropout_1	(8, 100, 128)	0
Conv_2	(8, 100, 256)	1605888
Batch standardization_2	(8, 100, 256)	32
Activation function_ 1	(8, 100, 256)	0
Maximum pooling_1	(2, 1, 256)	0
Dropout_1	(2, 1, 256)	0
Dense (full connection layer)_1	100	51300
Dropout_3		0
Dense (full connection layer)_2	100	1515

TABLE 2: Scene classification results of the proposed method (%).

Audio scene	Development set	Validation set
Sandy beach	89.6	76.3
Transit	98.2	71.9
Coffee/restaurant	88.3	81.2
Automobile	99.0	92.4
Center	89.7	88.7
Forest giants	99.8	95.5
Grocery store	93.6	75.8
Home	84.1	93.6
Library	88.5	62.1
Metro station	99.2	98.1
Office	98.9	83.4
Park	80.3	66.8
Neighborhood	86.8	74.3
Train	91.4	90.6
Tram	92.7	71.0
Average	92.0	81.4
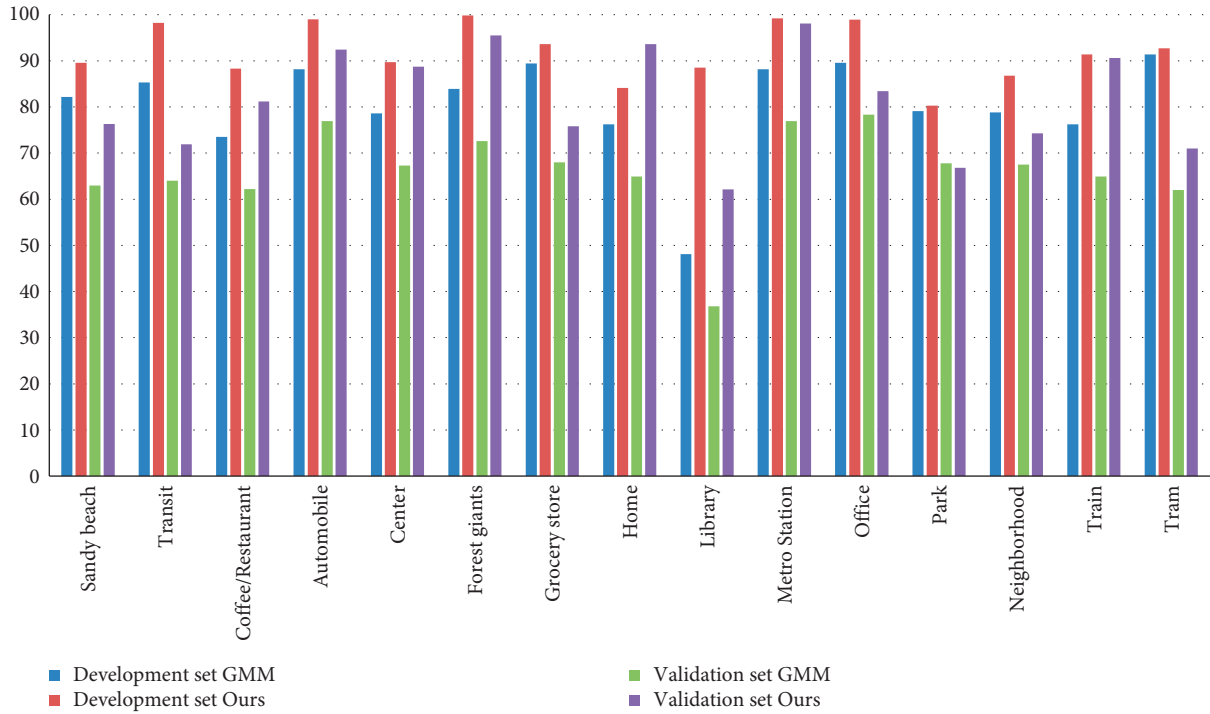
Figure 7: Comparison between GMM and the proposed method.
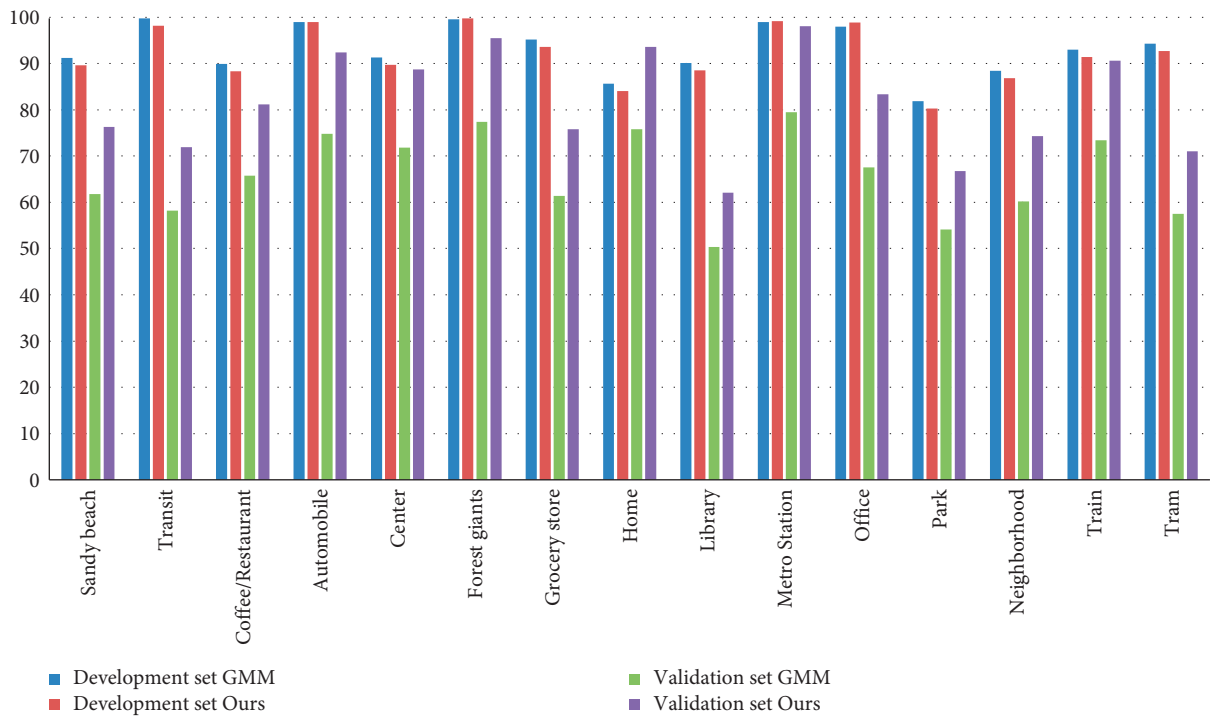
Figure 8: Comparison between CNN and the proposed method.

6. Conclusions

This paper presents an audio scene recognition method based on optimized audio processing and convolutional neural network. In audio processing, binaural representation and harmonic impact source separation are used. In terms of network structure, VGGNet-like structure is introduced to improve the flexibility of network structure. Experimental results show that compared with other existing methods, the proposed method can have better generalization performance while maintaining high recognition accuracy.

Data Availability

The experimental data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The author declares that there are no conflicts of interest.

References

[1] K. Nogueira, O. Penatti, and J. Santos, "Towards better exploiting convolutional neural networks for remote sensing scene classification," *Pattern Recognition*, vol. 61, pp. 539–556, 2016.

[2] X. Qi, C.-G. Li, G. Zhao, X. Hong, and M. Pietikäinen, "Dynamic texture and scene classification by transferring deep image features," *Neurocomputing*, vol. 171, no. 6, pp. 1230–1241, 2016.

[3] C. Gong, Z. Li, X. Yao, L. Guo, and Z. Wei, "Remote sensing image scene classification using bag of convolutional features," *IEEE Geoscience & Remote Sensing Letters*, vol. 14, no. 10, pp. 1735–1739, 2017.

[4] V. Bisot, R. Serizel, S. Essid, and G. Richard, "Feature learning with matrix factorization applied to acoustic scene classification," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 25, no. 6, pp. 1216–1229, 2017.

[5] D. Liang and H. Ling, "Dynamic scene classification using redundant spatial scenelets[J]," *IEEE Transactions on Cybernetics*, vol. 46, no. 9, pp. 2156–2165, 2016.

[6] J. Zou, W. Li, C. Chen, and Q. Du, "Scene classification using local and global features with collaborative representation fusion," *Information Sciences*, vol. 348, pp. 209–226, 2016.

[7] M. Hayat, S. H. Khan, M. Bennamoun, and S. An, "A spatial layout and scale invariant feature representation for indoor scene classification," *IEEE Transactions on Image Processing*, vol. 25, no. 10, pp. 4829–4841, 2016.

[8] H. Phan, L. Hertel, M. Maass, P. Koch, R. Mazur, and A. Mertins, "Improved audio scene classification based on label-tree embeddings and convolutional neural networks," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 25, no. 6, pp. 1278–1290, 2017.

[9] A. Rakotomamonjy, "Supervised representation learning for audio scene classification," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 25, no. 6, pp. 1253–1265, 2017.

[10] M. Raissi and G. E. Karniadakis, "Hidden physics models: machine learning of nonlinear partial differential equations," *Journal of Computational Physics*, vol. 357, pp. 125–141, 2018.

[11] C. Voyant, G. Notton, S. Kalogirou et al., "Machine learning methods for solar radiation forecasting: a review," *Renewable Energy*, vol. 105, no. 5, pp. 569–582, 2017.

[12] M. P. Pound, J. A. Atkinson, A. J. Townsend et al., "Deep machine learning provides state-of-the-art performance in image-based plant phenotyping," *Gigascience*, vol. 6, no. 10, pp. 1–10, 2018.

[13] I. Kavakiotis, O. Tsave, A. Salifoglou, N. Maglaveras, I. Vlahavas, and I. Chouvarda, "Machine learning and data mining methods in diabetes research," *Computational and Structural Biotechnology Journal*, vol. 15, no. C, pp. 104–116, 2017.

[14] W. Yang and S. Krishnan, "Combining temporal features by local binary pattern for acoustic scene classification," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 25, no. 6, pp. 1315–1321, 2017.

[15] M. A. Alamir, "A novel acoustic scene classification model using the late fusion of convolutional neural networks and different ensemble classifiers," *Applied Acoustics*, vol. 172, no. 3, pp. 112–122, 2020.

[16] X. Lin, X. Wang, and L. Li, "Intelligent detection of edge inconsistency for mechanical workpiece by machine vision with deep learning and variable geometry model," *Applied Intelligence*, vol. 50, no. 7, pp. 2105–2119, 2020.

[17] T. Kim, I. Y. Jung, and Y. C. Hu, "Automatic, location-privacy preserving dashcam video sharing using blockchain and deep learning," *Human-centric Computing and Information Sciences*, vol. 10, no. 1, pp. 1–23, 2020.

[18] Y. Mitsukura, "EEG signal processing for real applications," *Journal of Signal Processing*, vol. 20, no. 1, pp. 1–7, 2016.

[19] F. a. Zhao, X. Zhang, X. Mu, Z. Yi, and Z. Yang, "Learning multi-modality features for scene classification of high-resolution remote sensing images," *Journal of Computer and Communications*, vol. 06, no. 11, pp. 185–193, 2018.

[20] V. Abrol and P. Sharma, "Learning hierarchy aware embedding from raw audio for acoustic scene classification," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 157, no. C, pp. 552–559, 2020.

[21] S. Waldekar and G. Saha, "Two-level fusion-based acoustic scene classification," *Applied Acoustics*, vol. 170, no. 5, Article ID 107502, 2020.

[22] B. S. Chandra, C. S. Sastry, and S. Jana, "Robust heartbeat detection from multimodal data via CNN-based generalizable information fusion," *IEEE Transactions on Biomedical Engineering*, vol. 66, no. 3, pp. 710–717, 2019.

[23] S. Ravikumar and D. Kavitha, "CNN-oppositional-based Henry gas solubility optimization model for autonomous vehicle control system," *Journal of Field Robotics*, vol. 38, no. 4, pp. 1–30, 2021.

[24] R. Yang, X. Zha, K. Liu, and S. Xu, "A CNN model embedded with local feature knowledge and its application to time-varying signal classification," *Neural Networks*, vol. 142, no. 1, pp. 564–572, 2021.

[25] L. Zhao, Z. Wang, G. Zhang, and H. Gao, "Driver drowsiness recognition via transferred deep 3D convolutional network and state probability vector," *Multimedia Tools and Applications*, vol. 11, no. 2, pp. 1–19, 2020.

[26] J. Cai, J. Hu, S. Li, J. Lin, and J. Wang, "Combination of temporal-channels correlation information and bilinear feature for action recognition," *IET Computer Vision*, vol. 14, no. 8, pp. 634–641, 2020.

[27] D. Tan, H. Nguyen, D. T. Ngo, L. Pham, and H. H. Kha, "Acoustic scene classification using a deeper training method for convolution neural network," in *Proceedings of the 2019 International Symposium on Electrical and Electronics Engineering (ISEE)*, pp. 131–136, Ho Chi Minh City, Vietnam, October 2019.

[28] S. Bao, S. Ma, and C. Yang, "Multi-scale retinex-based contrast enhancement method for preserving the naturalness of color image," *Optical Review*, vol. 27, no. 6, pp. 475–485, 2020.

[29] G. Antipov, M. Baccouche, S.-A. Berrani, and J.-L. Dugelay, "Effective training of convolutional neural networks for face-based gender and age prediction," *Pattern Recognition*, vol. 72, pp. 15–26, 2017.

Research Article

Image Recognition Technology with Its Application in Defect Detection and Diagnosis Analysis of Substation Equipment

Long Luo ⓘ,[1] **Rukuo Ma,**[1] **Yuan Li,**[1] **Fangnan Yang,**[2] **and Zhanfei Qiu**[1]

[1]State Grid Qinghai Maintenance Company, Xining 810000, China
[2]State Grid Qinghai Electric Power Company, Xining 810000, China

Correspondence should be addressed to Long Luo; lluo.sgqh@yahoo.com

Detection of substation equipment can promptly and effectively discover equipment overheating defects and prevent equipment failures. Traditional manual diagnosis methods are difficult to deal with the massive infrared images generated by the autonomous inspection of substation robots and drones. At present, most of the infrared image defect recognition is based on traditional machine learning algorithms, with low recognition accuracy and poor generalization capability. Therefore, this paper develops a method for identifying infrared defects of substation equipment based on the improvement of traditional ones. First, based on the Faster RCNN, target detection is performed on 6 types of substation equipment including bushings, insulators, wires, voltage transformers, lightning rods, and circuit breakers to achieve precise positioning of the equipment. Afterwards, different classes are identified based on the sparse representation-based classification (SRC), so the actual label of the input sample can be obtained. Finally, based on the temperature threshold discriminant algorithm, defects are identified in the equipment area. The measured infrared images are used for experiments. The average detection accuracy achieved by the proposed method for the 6 types of equipment reaches 92.34%. The recognition rate of different types of equipment is 98.57%, and the defect recognition accuracy reaches 88.75%. The experimental results show the effectiveness and accuracy of the proposed method.

1. Introduction

The power station is an important node in the power grid that is responsible for converting voltage and distributing electric energy. Its safety and reliability are directly related to the safety and stability of the power system [1–5]. According to the statistics, about half of the power equipment failures have abnormal temperature in the early stage. Infrared detection of power equipment has the advantages of non-power failure, noncontact, high sensitivity, and mature technology [6–8]. It can detect equipment overheating defects in a timely and effective manner and prevent equipment failures. It has been widely used in the field of power equipment thermal fault diagnosis. In recent years, infrared thermal imagers have been equipped with inspection platforms such as substation inspection robots and drones to conduct intelligent inspections on substation equipment,

reducing the workload of operation and maintenance personnel to collect infrared images of equipment [9–14]. However, a large number of inspections or online monitoring still require manual analysis, which is time-consuming and inefficient, and also difficult to cope with the massive infrared images generated by inspections. Therefore, the researchers have carried out a series of works on infrared defect recognition. These methods used different types of features or classification models to analyze and process images and then determined and recognized defects. Most of the existing research studies on infrared image defect recognition are based on traditional machine learning algorithms, which have disadvantages such as low accuracy, poor generalization ability, and rigid model. With the development of artificial intelligence technology, image detection algorithms based on deep learning have been widely used in the field of visible light object detection [13–16], but there are

few applications in infrared image recognition and detection. So, it is urgent to carry out infrared image-based fault identification algorithms of substation equipment [17–19].

Based on the improvement of traditional image analysis technology, this paper proposes an infrared defect recognition method for substation equipment. First, based on the Faster RCNN algorithm [20–23], target detection is performed on 6 types of substation equipment, including bushings, insulators, wires, voltage transformers, lightning rods, and circuit breakers, to achieve precise positioning of the equipment. Next, for the detected image area of the substation equipment, fine classification is performed based on the sparse representation-based classification (SRC) [24–27]. And, the specific substation equipment label to which it belongs is determined based on the training samples. Finally, based on the threshold on the temperature, defects are identified in the equipment area to determine whether the equipment has defects and the relevant severity. Compared with the traditional image defect detection algorithms of substation equipment, this paper further introduces SRC to realize the accurate confirmation of the category of the input substation equipment. Such operation has important auxiliary significance for subsequent targeted defect detection and failure analysis. In the experiment, the proposed method is tested based on the collecting infrared images from actual measurement to verify its feasibility and accuracy.

2. Equipment Detection Based on Faster RCNN

This paper first builds an infrared image defect detection model for substation equipment based on Faster RCNN. The specific frame structure is shown in Figure 1. First, the Faster RCNN detection model is trained based on the infrared image training dataset of the substation equipment to generate a defect detection network. Then, the test infrared image is input into the defect detection network to verify the detection performance. As shown in Figure 1, the Faster RCNN detection network is composed of two parts: a region proposal network (RPN) and a region convolutional neural network (RCNN). The Faster RCNN model is an integrated network that can effectively realize object detection in input images. It mainly completes region proposal and CNN detection, which are completed by RPN and RCNN, respectively. RPN is a fully convolutional neural network that can effectively predict object boundaries and object scores at various positions in the input image. Object boundaries and object scores are important indicators for evaluating whether an area is proposed. The image input proposal given by RPN will be input into RCNN to realize the classification of the proposed area and calculate the score of each proposed area. The Faster RCNN model integrates the RPN network and RCNN into one network by sharing convolutional features and provides a joint training method. Compared with a single network structure, the Faster RCNN model deepens the network depth and improves the image detection performance. There have been many research results based on Faster RCNN's target detection and power equipment detection [20–23]. This paper mainly uses the

direct application shown in Figure 1 and will not introduce it in detail here.

3. Equipment Recognition and Defect Diagnosis

3.1. SRC. The basic principle of SRC [24–27] is to use training samples to linearly fit the test samples of unknown categories and constrain the high sparsity of linear representation coefficients. On this basis, the reconstruction error of each training class for the test sample is calculated separately to determine the target label of the test sample. Assume that the training samples of the M-class targets construct a global dictionary $A = [A^1, A^2, \ldots, A^M] \in R^{d \times N}$, where $A^i \in R^{d \times N_i} (i = 1, 2, \ldots, M)$ are all the training samples of the ith class. For the test sample y with an unknown target label, the sparse representation process is as follows:

$$
\begin{aligned}
\widehat{x} &= \arg\min \|x\|_0, \\
\text{s.t. } y &= Ax.
\end{aligned}
\tag{1}
$$

In equation (1), x is the linear coefficient vector to be solved. Under the sparsity constraint, only a small number of elements in the coefficients are nonzero. With the optimal estimated $\widehat{x}$ to be obtained, equation (3) is used to calculate the reconstruction error of each training class for the test sample, and then, the decision is made according to the minimum error criterion:

$$
r(i) = \left\| y - A\delta_i(\widehat{x}) \right\|_2^2, \quad i = 1, 2, \ldots, M,
\tag{2}
$$

$$
\text{identity}(y) = \arg\min_i (r(i)),
\tag{3}
$$

where $\delta_i(\widehat{x})$ represents the sparse representation coefficient vector corresponding to the ith class and $r(i)(i = 1, 2, \ldots, M)$ is the reconstruction error of the ith class.

It can be seen that the key problem in SRC is the solution of the sparse coefficients. With the accurate and reliable solutions, the target label of the test sample can be accurately obtained. Since the problem of minimizing the ℓ_0-norm in equation (1) is an NP-hard problem, it is difficult to solve it directly. Researchers found that the ℓ_1-norm minimization optimization and the ℓ_0-norm minimization optimization are equivalent under certain conditions [24], so equation (1) can be transformed into

$$
\begin{aligned}
\widehat{x} &= \arg\min \|x\|_1, \\
\text{s.t. } y &= Ax.
\end{aligned}
\tag{4}
$$

The introduction of the ℓ_1 norm can turn the original problem into a convex optimization problem which is easy to solve [5]. So, the linear programming method can be used to solve the sparse representation coefficients. In the actual process, due to the influence of noise and the unstable error between the training and the test samples, it is difficult to achieve a completely accurate reconstruction. Therefore, the sparse coefficient vector is solved as follows under the given reconstruction error limit:
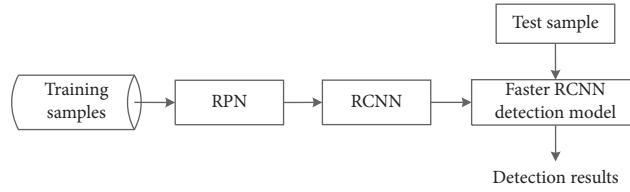
FIGURE 1: Detection process of substation equipment based on Faster RCNN.

$$\widehat{x} = \arg \min \|x\|_1,$$
$$\text{s.t. } \|Ax - y\|_2 < \varepsilon, \quad (5)$$

where ε is the allowable reconstruction error.

Based on the detection results of substation equipment by Faster RCNN, this paper uses SRC for precise identification for each region of the detected results. A global dictionary is constructed based on the training samples with known target labels. Then, the reconstruction error of the input sample in any class is calculated according to the abovementioned sparse representation and reconstruction process. According to the reconstruction errors, the substation equipment label is determined precisely.

3.2. Defect Diagnosis. Abnormal heating of substation equipment is an important basis for judging the defects. Frequent defects in substation equipment include defects, loose parts, loose wires, corrosion or oxidation of equipment contact surfaces, and excessive current-carrying. The equipment's own defects cause the equipment to fail to operate normally and heat up; then, the equipment must be replaced. Loose parts and loose wires are common failures of the capacitor interface accessories of the power transmission and transformation equipment. This type of failure is caused by the change in the area. Corrosion or oxidation of the equipment contact surface is a common fault in electrical contact parts such as disconnecting switches of power equipment. This type of fault increases current-carrying resistance and causes the fault point to heat up. When the current-carrying is too large, the equipment is under high load. After running for a certain period of time, abnormal heating will also occur.

Electricity can divide equipment thermal defects into current heating and voltage heating according to different heating properties. Current heating defects refer to the heating caused by the resistance of the current flowing through the conductive loop, which is mainly concentrated at the equipment connectors. And, the temperature rise is more obvious and easier to be found, accounting for about 90% of the total number of faults [12]. Voltage heating defect refers to the heating of the device caused by the action of the electric field when the operating voltage is applied to the insulating medium of the device. Only a small part of the heat of this type of fault can be conducted to the surface of the equipment shell, and the temperature rise that can be measured is very small, which is difficult to be detected. This paper proposes an infrared defect recognition method for current-induced defects. Current heating-type defects include joint heating, disconnection switch knife edge and switching hair heating, circuit breaker contact heating, transformer internal connection heating, and bushing string heating. The surface temperature discrimination method is the simplest and most effective method for judging current-heated defects. According to previous works, three temperature threshold values of 50°C, 80°C, and 105°C are selected to identify defects in the equipment. Urgent defects represent defects that have a serious threat and need to be addressed immediately. A serious defect indicates a fault that may threaten to be dealt with as soon as possible. General deficiencies indicate that there is little threat to the safety of equipment operation that requires enhanced surveillance. On the basis of completing the detection and identification of substation equipment, the process of infrared image defect identification in this paper is shown in Figure 2.

4. Experiments and Analysis

4.1. Preparation. This paper selects infrared images of 6 types of substation equipment including bushings, insulators, wires, voltage transformers, lightning rods, and circuit breakers for defect detection. These images mainly come from the grid company reserves and collected data in actual scenes. The dataset includes a total of 800 defective images and 1900 normal images. In order to avoid additional deviations introduced in the data division process and affect the final result, the training set and the test set should be randomly selected from the original ones. Finally, this paper randomly chooses 1700 infrared images of substation equipment from the total sample set as the training set and 600 infrared images as the test set.

4.2. Results and Discussion. For Faster RCNN-based substation equipment detection, in the experiment, the model training learning rate is set to 0.001, mini-batch size 256, momentum 0.9, weight attenuation 0.0001, and total number of iterations 20000. The 1700 training samples are trained to obtain substation equipment detection models. The detection accuracy of 6 types of substation equipment achieved by the proposed method is shown in Table 1, and the average accuracy is 92.34%. This result verifies the effectiveness of Faster RCNN for image detection of substation equipment.

The training samples are also used to train the SRC to obtain the corresponding classification model. According to the Faster RCNN results of the test samples, the concrete label of the substation equipment to which the test samples belong is confirmed. Table 2 shows the recognition rates of different types of substation equipment based on SRC, and the average recognition rate reaches 98.57%. This result proves the effectiveness of SRC for the identification of substation equipment.
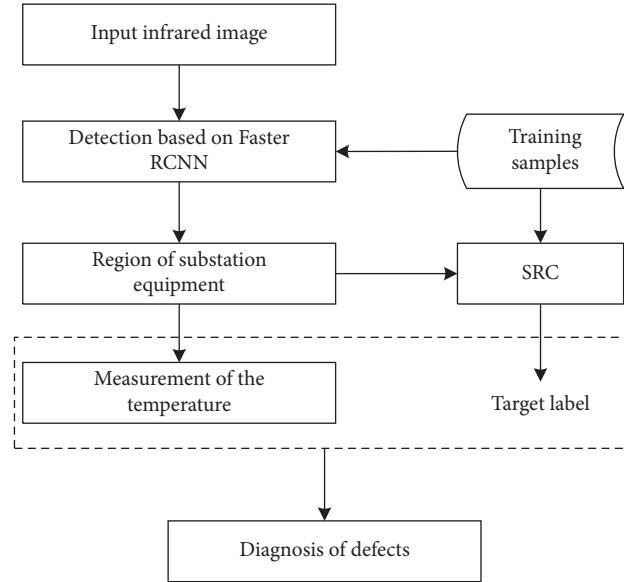
Figure 2: Procedure of defect diagnosis based on temperature.

Table 1: Analysis of detection performance based on Faster RCNN.

Equipment	Detection accuracy (%)
Bushing	92.32
Insulator	93.46
Wire	91.23
Voltage transformer	93.56
Lightning rod	89.24
Circuit breaker	90.03

Table 2: Analysis of recognition performance based on SRC.

Equipment	Recognition rate (%)
Bushing	98.56
Insulator	98.92
Wire	97.31
Voltage transformer	97.85
Lightning rod	99.02
Circuit breaker	98.34

On the basis of the abovementioned detection and identification process, the defect detection of substation equipment is carried out based on the idea of temperature thresholds. According to statistics, there are a total of 160 defective devices in the test set of 600 infrared images, and 142 devices are correctly judged as defective based on the proposed algorithm, with an accuracy rate of 88.75%. Figure 3 shows examples of the defect detection results of the proposed method for some substation equipment. The region marked in the dashed box is the abnormal part where the temperature exceeds the set threshold, that is, the local area where defects may exist. The results are consistent with the actual situation. The experimental results show the effectiveness and accuracy of the method in this paper. In the specific defect processing process, because this article not only carried out the inspection of

(a)   (b)
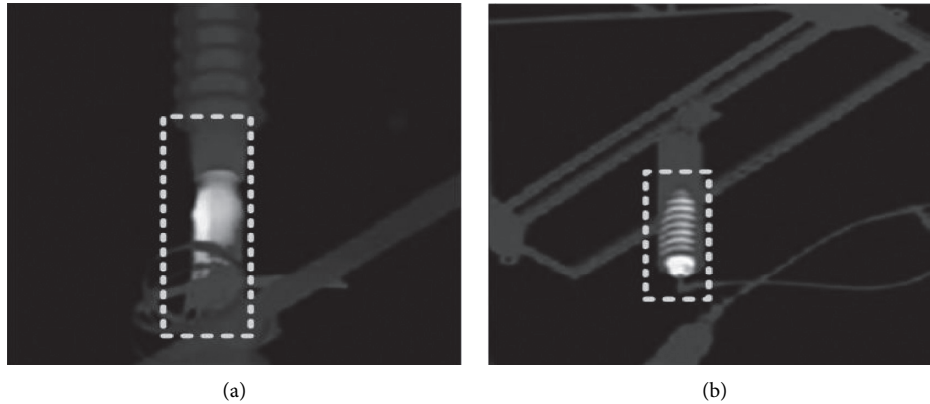
FIGURE 3: Examples of defect detection results.

the equipment but also confirmed its category, it can assist the staff to formulate more effective treatment measures.

5. Conclusion

This paper applies image recognition technology to the detection and diagnosis of defects in substation equipment. First, for an input infrared image of substation equipment, Faster RCNN is used to detect the key areas to obtain the substation equipment of interest. On this basis, relying on the training samples that have been classified, the detected area is confirmed based on the SRC, and the specific type of the substation equipment is determined. Finally, determine whether it is in a normal state according to the temperature of the device reflected in the current infrared image. Compared with traditional substation equipment defect analysis algorithms, this paper further confirms the category of input equipment by introducing SRC, so the analysis results are more informative, which is conducive to the development of targeted diagnosis and repair. Experiments are carried out on the infrared image set of actual measurement substation equipment, and the results show the effectiveness of the proposed method for defect detection and identification of substation equipment.

Data Availability

The data used in this study are available from the authors upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

References

[1] G. C. Stone, "Partial discharge diagnostics and electrical equipment insulation condition assessment," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 12, no. 5, pp. 891–903, 2005.

[2] T. Liu, "Development and application of electrical equipment intelligent detection management system," in *Proceedings of the 5th International Conference on Smart Grid and Electrical Automation (ICSGEA)*, pp. 70–73, June 2020.

[3] X. Tian, Z. Lan, F. Huo et al., "Calculation of electric field on substation equipment considering AC ion flow field," *Mathematical Problems in Engineering*, vol. 2020, Article ID 3914872, 2020.

[4] C. Shi, J. Fei, X. Zhang, Q. Yao, and J. Fan, "Continuous trust evaluation of power equipment and users based on risk measurement," *Scientific Programming*, vol. 2020, Article ID 8895804, 2020.

[5] X. Wang, Y. Chen, W. Ruan, Q. Gao, G. Ying, and Li Dong, "Intelligent detection and recovery of missing electric load data based on cascaded convolutional autoencoders," *Scientific Programming*, vol. 2020, Article ID 8828745, 2020.

[6] C. Guo, M. Ren, C. Xia, M. Dong, and B. Wang, "Fault diagnosis of power equipment based on infrared image analysis," in *Proceedings of the IEEE International Conference on Advances in Electrical Engineering and Computer Applications (AEECA)*, pp. 659–663, Dalian, Chinapp, Aug 2020.

[7] N. K. Dhote and J. B. Helonde, "Fuzzy algorithm for power transformer diagnostics," *Advances in Fuzzy Systems*, vol. 2013, Article ID 421621, 7 pages, 2013.

[8] Q. Wang, S. Wang, R. Shi, and Y. Li, "A power transformer fault diagnosis method based on random vector functional-link neural network," *Mathematical Problems in Engineering*, vol. 2021, Article ID 6656061, 2021.

[9] K. Liu, S. Wu, Z. Luo et al., "An intelligent fault diagnosis method for transformer based on ipso-gcforest," *Mathematical Problems in Engineering*, vol. 2021, Article ID 6610338, 2021.

[10] J. Yang, B. Liu, C. Liu, F. Lu, and W. Rong, "The application of evidence theory in the field of equipment fault diagnosis," in *Proceedings of the 2006 6th World Congress on Intelligent Control and Automation*, pp. 5824–5827, Dalian, China, June 2006.

[11] G. Wang, C. Feng, X. Mu, and H. Liu, "Intelligent diagnosis of power transformer working fault," in *Proceedings of the 2019 Chinese Control And Decision Conference (CCDC)*, pp. 383–387, Nanchang, China, June 2019.

[12] Y. Wang, F. Lu, and H. Li, "The fault diagnosis method for electrical equipment based on Bayesian network," in *Proceedings of the2005 International Conference on Electrical Machines and Systems*, pp. 2259–2261, September 2005.

[13] Y. Lin, W. Zhang, H. Zhang, D. Bai, J. Li, and R. Xu, "An intelligent infrared image fault diagnosis for electrical equipment," in *Proceedings of the 2020 5th Asia Conference on*

Power and Electrical Engineering (ACPEE), pp. 1829–1833, Chengdu, China, June 2020.

[14] W. Yongqiang, L. Fangcheng, and L. Heming, "The fault diagnosis method for electrical equipment using bayesian network," in *Proceedings of the2009 First International Workshop on Education Technology and Computer Science*, pp. 563–565, Wuhan, China, March 2009.

[15] H. Zou and F. Huang, "A novel intelligent fault diagnosis method for electrical equipment using infrared thermography," *Infrared Physics & Technology*, vol. 73, pp. 29–35, 2015.

[16] L. Yifei, Z. Hangyan, and Y. Yujie, "Fault diagnosis algorithm of Electrical system based on Deep Learning," in *Proceedings of the2020 2nd International Conference on Information Technology and Computer Application (ITCA)*, pp. 226–228, Guangzhou, China, December 2020.

[17] A. Jiang, N. Yan, B. Shen, C. Gu, H. Zhu, and H. Huang, "Research on infrared image recognition method of power equipment based on deep learning," in *Proceedings of the2020 IEEE International Conference on High Voltage Engineering and Application (ICHVE)*, pp. 1–4, September 2020.

[18] X. Gong, Q. Yao, M. Wang, and Y. Lin, "A deep learning approach for oriented electrical equipment detection in thermal images," *IEEE Access*, vol. 6, pp. 41590–41597, 2018.

[19] C. Wei, F. Tao, Y. Lin et al., "Substation equipment thermal fault diagnosis model based on ResNet and improved bayesian optimization," in *Proceedings of the 2019 9th International Conference on Power and Energy Systems (ICPES)*, pp. 1–5, Perth, December 2019.

[20] S.M. Suh, Y. Park, K.M. Ko et al., "Weighted mask R-CNN for improving adjacent boundary segmentation," *Journal of Sensors*, vol. 2021, Article ID 8872947, 8 pages, 2021.

[21] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: towards real-time object detection with region proposal networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, 2017.

[22] Y. Cheng, L. Xia, B. Yan, J. Chen, D. Hu, and L. Zhu, "A defect detection method based on faster RCNN for power equipment," *Journal of Physics: Conf. Ser,* vol. 1754, Article ID 012025, 2021.

[23] X. Xu, L. Yang, and F. Yang, "Railway subgrade defect automatic recognition method based on improved faster R-CNN," *Scientific Programming*, vol. 2018, Article ID 4832972, 2018.

[24] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 2, pp. 210–227, 2009.

[25] J. J. Thiagarajan, K. N. Ramamurthy, P. Knee, A. Spanias, and V. Berisha, "Sparse representations for automatic target classification in SAR images," in *Proceedings of the 4th Int. Symp. Commun., Control Signal Process.*, pp. 1–4, Limassol, Cyprus, March 2010.

[26] L. Zhang, Z. Tao, and B. Wang, "SAR image target recognition using kernel sparse representation based on reconstruction coefficient energy maximization rule," in *Proceedings of the 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 2369–2373, Shanghai, China, March 2016.

[27] X. Xing, K. Ji, H. Zou, and J. Sun, "Sparse representation based SAR vehicle recognition along with aspect angle," *The Scientific World Journal*, vol. 2014, Article ID 834140, 2014.

Research Article

Intelligent Detection Method of English Text in Natural Scenes in Video

Liqin Dai [1] **and ChunHua Chen**[2]

[1]*Jiangxi University of Chinese Medicine, JiangXi 330022, NanChang, China*
[2]*Huai Yin Institute of Technology, JiangSu 223001, HuaiAn, China*

Correspondence should be addressed to Liqin Dai; 20030748@jxutcm.edu.cn

With the rapid development of Internet technology, breakthroughs have been made in all branches of computer vision. Especially in image detection and target tracking, deep learning techniques such as convolutional neural networks have achieved excellent results. In order to explore the applicability of machine learning technology in the field of video text recognition and extraction, a YOLOv3 network based on multiscale feature transformation and migration fusion is proposed to improve the accuracy of english text detection in natural scenes in video. Firstly, aiming at the problem of multiscale target detection in video key frames, based on the YOLOv3 network, the scale conversion module of STDN algorithm is used to reduce the low-level feature map, and a backbone network with feature reuse is constructed to extract features. Then, the scale conversion module is used to enlarge the high-level feature map, and a feature pyramid network (FPN) is built to predict the target. Finally, the improved YOLOv3 network is verified to extract key text from images. The experimental results show that the improved YOLOv3 network can effectively improve the false detection and missed detection caused by occlusion and small target, and the accuracy of English text extraction is obviously improved.

1. Introduction

As the mainstream part of today's media industry, images and videos are rich in information and easy to understand, which makes them an indispensable part of life. Computer vision analysis is also the key development direction of Internet communication industry at present. For example, character recognition has great application value in many scenes, such as vehicle license plate detection, image-text conversion, image content translation, and image search. However, because the precision of text recognition technology is not ideal, its application scenarios are relatively simple, such as content search in images [1–6].

In image content search scene, the background is simple, the font is single, and some physical features of the image are used to assist, but the error rate is still not ideal. Therefore, most of the current research focuses on text detection and then will consider recognition. Especially in the application of natural scenes, because the background is extremely complex,

the font size varies and is affected by special circumstances such as overlapping and pollution of various perspectives, and the difficulty of text detection in natural scenes in videos is much greater than that in ordinary scenes. Image content can be divided into two parts: perceptual content and semantic content [7–9]. The visual part of the image, including the direct visual impression of color, shape, and texture, is the perceptual content. Indirect understanding parts in images, such as objects, words, and events contained in images, are all semantic contents. Among them, words are an important tool for information understanding and communication. Compared with other semantic contents, words are the main content of expressing information and information interaction [10–12]. In addition, the characters are easy to extract and have strong descriptive ability, so how to understand the semantic information of characters in images is an urgent problem to be solved.

This paper focuses on the extraction of English text from natural scenes in video. The meaning of video text

recognition is to extract the text content in the video and then recognize it by the recognition system and finally get the text content. Video is essentially a sequence stream composed of a series of images, and the words in the frame images can express the contents in a short time. One kind of text is the text in the natural scene of the image [13–15], such as the license plate number and bus stop sign text in the image [16], and the other kind of text is artificially added, such as movie subtitles, advertising information, and medical image analysis text [17]. Therefore, all the words should be extracted except the repeated words within a short time delay. The final result of the recognition system can only be determined if the images and characters have good detection performance, so the text detection is the key research content of this paper.

2. Literature Review

In the aspect of text detection, traditional methods are mainly based on the characteristics of text coherence and single color. For example, Minetto et al. [18] proposed an improved image text detection algorithm. Firstly, this algorithm extracts three edge images with different colors by using edge detection operators obtained from three directions and then applies common operations in morphology to these three edge images in turn to obtain different connected domain images. Finally, the three connected graphs are AND-operated, and the noise is filtered, so as to obtain the text region. Yin et al. [19] also adopted a similar method, combined with the morphological method, and used the OSTU algorithm to obtain adaptive threshold, so as to obtain a clean and clear binary image. According to the characteristics of different gray trends of characters and backgrounds, Risnumawan et al. [20] proposed a video character location method based on gradient discrete cosine transform algorithm. In this method, each frame is divided into $n \times n$ blocks, and the discrete cosine transform coefficients of each block are calculated. The amplitude obtained by the gradient operator is used as the block strength for smooth filtering and morphological processing. Finally, the image is projected horizontally and vertically, and the candidate text regions are extracted by wavelet transform and unsupervised clustering. Zhuge and Lu [21] used the model based on level set function to realize text segmentation with small color difference between target and background and large text groove and solved the problem of difficult parameter selection in variational model through optimization calculation.

With the popularization of computers and the great improvement of computer technology, especially the powerful parallel computing capability of GPU and the Big Data resources in the era of mobile Internet, CNN continues to develop. Text detection also follows this trend, turning to the method based on CNN technology, and the effect is greatly improved, which is a step closer to the application of real scenes. Alqhtani et al. [22] trained CNN to detect characters in text, calculated the confidence of pixel blocks, and adopted the nonmaximum suppression method when locating text lines. Then, a similar transcription process is

adopted, the classification scores of each character are inquired, and the best words are selected for matching. Tong et al. [23] integrated text detection and recognition into an end-to-end network. RPN (ridge polynomial network) is used for detection, and bilinear sampling is used to unify the text regions into highly consistent variable-length feature sequences, and then recurrent neural network (RNN) is used for recognition. You Only Look Once (YOLO) is a target recognition and location algorithm based on the deep convolution neural network [24]. Its most obvious advantage is its fast detection speed [25], which is especially suitable for the real-time detection system, which is the fundamental reason why YOLOv3 network is selected in this paper. YOLO has been continuously improved on the basis of the original version and has developed to v3 version [26]. However, conventional object detection methods in the visual field (SSD, YOLO, faster-RCNN, etc.) are not ideal when directly applied to English text detection tasks. The main reasons are as follows: compared with conventional objects, the length of text lines and the ratio of length to width vary widely. Therefore, after analyzing YOLO series networks, we propose a new multiscale feature fusion method, which improves the performance of YOLOv3 networks.

In order to construct features with multiscale characteristics and rich expressive ability, we introduce a feature scale transformation and migration fusion method to improve the traditional YOLOv3 network. Different from the existing multiscale feature fusion model by scaling single-channel features, we achieve the purpose of feature scale reduction and enlargement by splitting and combining multichannel feature maps. At the same time, we migrate and fuse the converted features in the backbone network and construct the backbone network with reduced feature scale and FPN prediction network with enlarged feature scale, which achieves better detection results than YOLOv3 network when detecting occluded and smaller targets.

3. English Text Detection Based on Improved YOLOv3 Network

3.1. Convolutional Neural Network. At present, the principles of target detection algorithms are generally divided into two methods: region division and position regression. For example, fast-RCNN method can obtain high accuracy, but its running speed is slow. The latter, for example, SSD and YOLO pursue the real-time performance of the algorithm but can also obtain acceptable detection results. Among these methods, YOLO has become a widely used and efficient algorithm because of its fast speed and high precision. YOLO is a target recognition and location algorithm based on the deep convolution neural network, and its most obvious advantage is its fast detection speed, which is especially suitable for the real-time detection system, which is the fundamental reason why YOLOv3 network is selected in this paper.

Convolution neural network extracts features by convolution operation on local "receptive field" [27], and it is mainly used in image processing-related problems.

CNN is a kind of feedforward neural network with deep structure. Firstly, the image is input at the input layer and then calculated by the convolution layer, pooling layer, and nonlinear activation function, and the semantic information of high-level abstraction is gradually extracted from the image. This is the "feedforward operation" of the convolutional neural network. Finally, for the fully connected layer, all the features extracted from the previous network are connected for prediction, and the difference between the detected value and the true labeled value of the network is calculated. The loss is propagated back to the first convolution layer from the fully connected layer by the gradient descent method, so that all the parameters of the network are updated, and the whole network model converges after several rounds of training. The shallow features extracted by convolution network are shown in Figure 1.

3.2. Feature Fusion.

Fusion of convolution features of different scales is an important means to improve the performance of target detection. The low-level features have higher resolution and contain more position and detail information, but because there are few convolution layers, they have less semantic information and more noise. High-level features have stronger semantic information, but their resolution is very small, and their ability to perceive details is poor. How to fuse them efficiently? This section introduces the feature fusion methods of upsampling, deconvolution, and scale transformation in detail.

3.2.1. Upsampling.

The upsampling method is to directly interpolate the original feature map [28] to enlarge the feature map, and interpolation is the most common and practical method. On the basis of the original feature map, interpolation algorithm is used to insert new pixel values between the original pixel positions, thus enlarging the scale of the feature map smoothly.

Bilinear Interpolation is an interpolation method for two-dimensional features, which is an extension of linear interpolation algorithm and is widely used in the field of image processing. The purpose of bilinear interpolation method is to calculate the position element value by using the existing values of the target point in two vertical directions in the original feature map, and its main purpose is to jointly determine a linear interpolation in two directions. Schematic diagram of bilinear interpolation is shown in Figure 2.

In Figure 2, the data points with known values are marked in red and represented by the letter Q; mark the data points to be interpolated as green and use the letter P to represent them; mark the data points in the middle transition in blue, which is indicated by the letter R. The values of R_1 and R_2 can be obtained by formulas (1) and (2), respectively.

$$f(R_1) = \frac{x_2 - x}{x_2 - x_1} f(Q_{11}) + \frac{x - x_1}{x_2 - x_1} f(Q_{21}), \quad (1)$$

$$f(R_2) = \frac{x_2 - x}{x_2 - x_1} f(Q_{12}) + \frac{x - x_1}{x_2 - x_1} f(Q_{22}). \quad (2)$$

The value of the target point P is inserted by linear interpolation in the y direction.

$$f(P) = \frac{y_2 - y}{y_2 - y_1} f(R_1) + \frac{y - y_1}{y_2 - y_1} f(R_2). \quad (3)$$

Bilinear interpolation method uses four points in two vertical directions in the original image to calculate the target pixel value for interpolation.

3.2.2. Deconvolution.

The formula for calculating a single deconvolution layer is as follows:

$$\sum_{k=1}^{K_1} z_k^i \oplus f_{k,c} = y_c^i. \quad (4)$$

In this layer, an image y^i composed of feature images $y_1^i, \ldots, y_{K_0}^i$ of K_0 color channels is used as input. Each channel c of the image can be expressed as the linear sum of K_1 potential feature maps and convolution kernel.

The deconvolution layer makes the potential feature graph z_k^i sparse by introducing regularization terms. The total loss function of is composed as follows:

$$C_1(y^i) = \frac{\lambda}{2} \sum_{c=1}^{K_0} \left\| \sum_{k=1}^{K_1} z_k^i \oplus f_{k,c} - y_c^i \right\|_2^2 + \sum_{k=1}^{K_1} \left| z_k^i \right|^p, \quad (5)$$

where p is sparse norm and λ is constant.

The implementation process of deconvolution is shown in Figure 3.

3.2.3. Scale Conversion.

Scale problem is the core problem of target detection. In order to obtain feature maps with different resolutions with strong semantic information, we use the scale conversion method of feature map instead of the upsampling method in the original YOLOv3 network. Scale conversion is very efficient and can be directly embedded into dense blocks of Darknet. Assuming that the size of the input tensor of scale conversion is $H \times W \times (C \cdot r^2)$, where H and W are the length and width of the feature graph, $C \cdot r^2$ is the number of channels of the input feature graph, and r is the upsampling factor, this paper sets $r = 2$. Scale enlargement of feature map is shown in Figure 4.

It can be seen that reducing and enlarging the width and height of the transport layer is achieved by increasing and decreasing the number of channels, and the scale conversion module is an operation of periodic rearrangement of elements.

$$I_{x,y,c}^{SR} = I_{x/r, y/r, r \bmod(y,r) + \bmod(x,r) + cr^2}^{LR}. \quad (6)$$

(a)

(b)

(c)
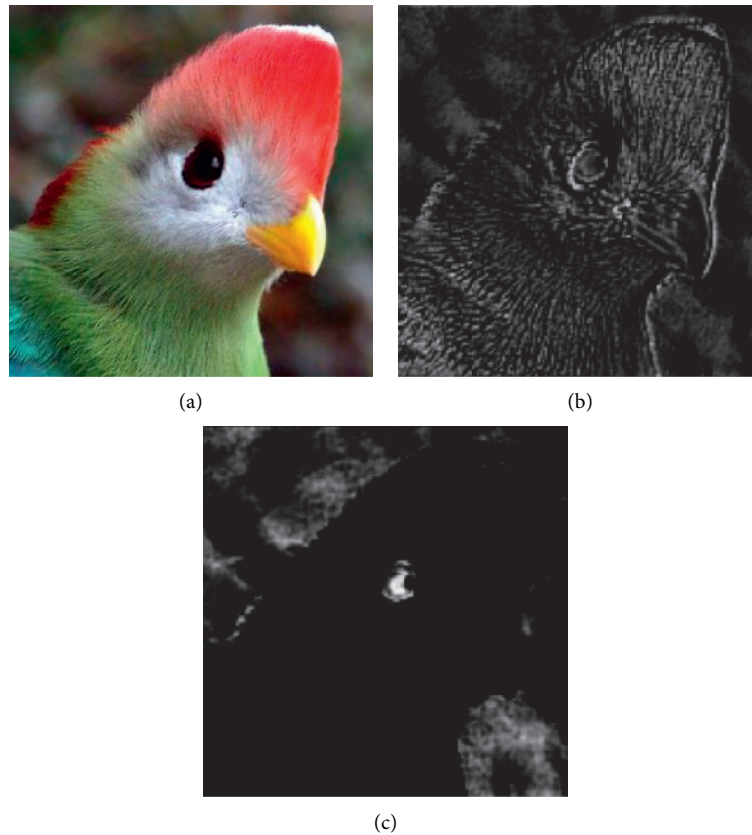
FIGURE 1: Shallow features extracted by the convolution network. (a) Original picture; (b) texture information; (c) shape information.
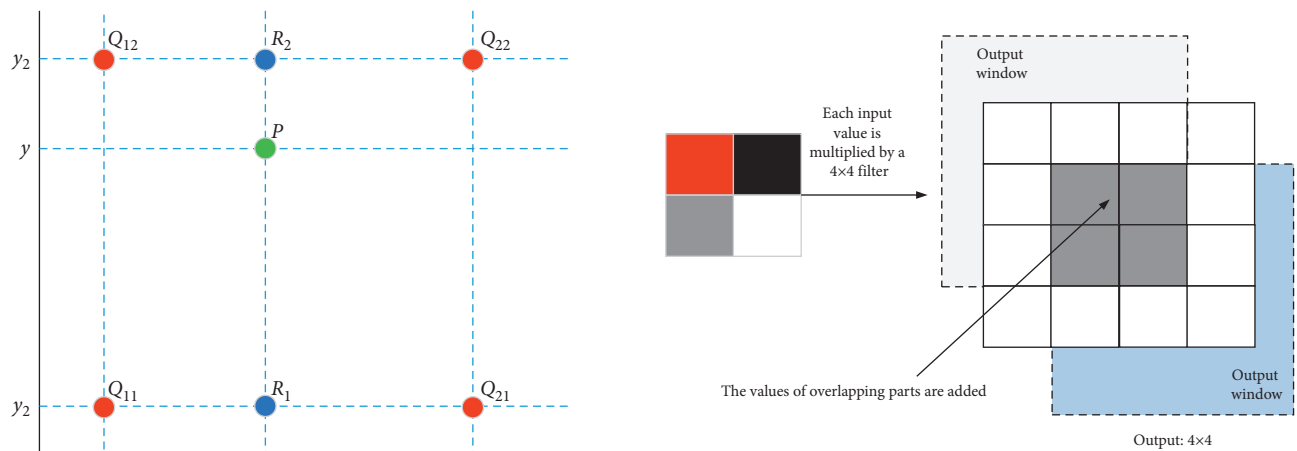
FIGURE 2: Schematic diagram of bilinear interpolation.

FIGURE 3: Implementation process of deconvolution.

where I^{SR} is a high-resolution feature map and I^{LR} is a low-resolution feature map. Scale transformation and use deconvolution layer must fill in zeros in the amplification step before convolution operation, without extra parameters and calculation overhead.

3.3. YOLOv3 Network Based on Feature Transformation, Migration, and Fusion. YOLO took the lead in innovatively combining the tasks of candidate selection stage and target recognition stage into one, and only one feature extraction can detect how many target objects and their positions [29]. Each grid in YOLO predicts two Bounding Box (BBox) in target detection. In YOLO's network structure, the task of extracting candidate regions is removed, and only the task of suggestion box loss regression exists. Therefore, the network structure is very simple, with only convolution and pooling operations, and finally, it is predicted by two fully connected layers, as shown in Figure 5.
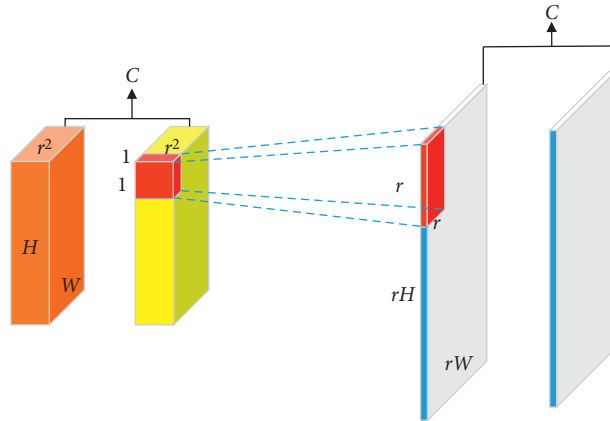
FIGURE 4: Scale enlargement of feature map.

convolution
residual
Characteristic graph

Regional recommendation
Up-sampling
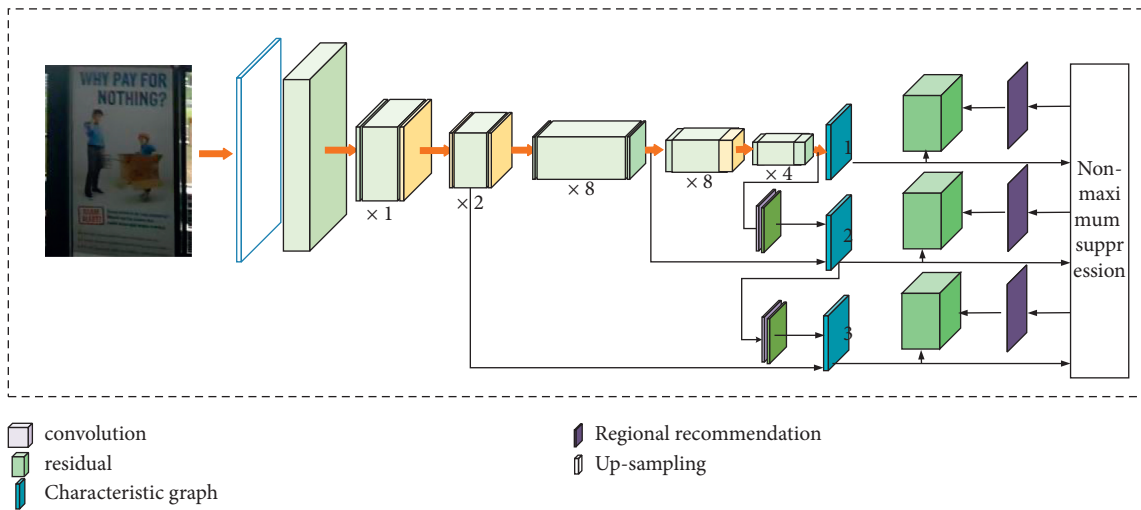
FIGURE 5: YOLO network structure.

It can be seen from Figure 5 that the first 24 convolution layers of the network are the backbone network for extracting image features, and finally, the extracted features are predicted through two fully connected layers. All-connection layer requires input, which is characterized by fixed dimensions. Each grid is predefined with 30-dimensional vector information corresponding to two suggestion boxes, as shown in Figure 6, which includes the positions of two BBox, the confidence of two BBox, and the classification probabilities of 20 objects.

During target detection, the specific category confidence score of each BBox is as follows:

$$P_{BBox} = P_r\left(\mathrm{Class}_i\right) * IOU_{\mathrm{pred}}^{\mathrm{truth}}, \tag{7}$$

where $P_r\left(\mathrm{Class}_i\right)$ is the probability of target occurrence and $IOU_{\mathrm{pred}}^{\mathrm{truth}}$ is the intersection over union (IOU) value of prediction frame and real frame.

In order to solve the gradient divergence problem caused by deepening the network model, YOLO3 borrowed the method of residual network and added the method of shortcut connections between some layers. The residual

Two BBox positions (one BBox needs four values to indicate its position, and two BBox have eight values in total)

Confidence of twoBBox

Probability of 20 objects
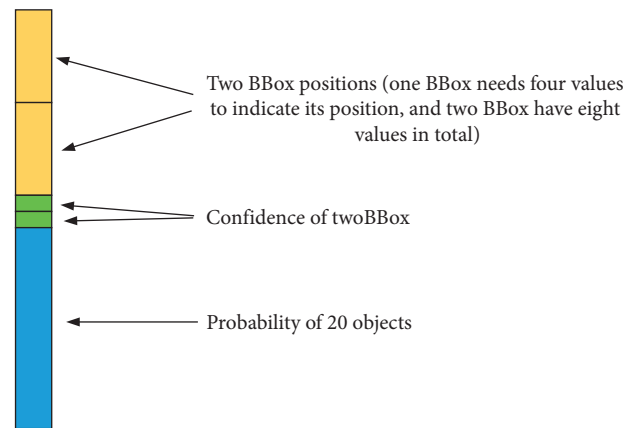
FIGURE 6: 30-dimensional output vector.

component of the specific direct connection method is shown in Figure 7 by directly transmitting the input x to the output, the output result is $f(x) + x$, and when $f(x) = 0$, then $H(x) = x$, the residual result approaches 0, and the training
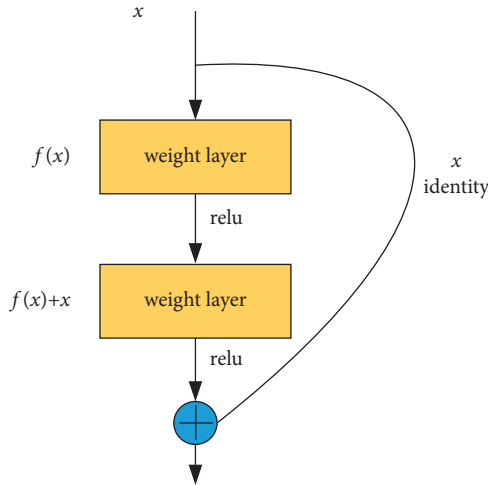
FIGURE 7: Residual components.
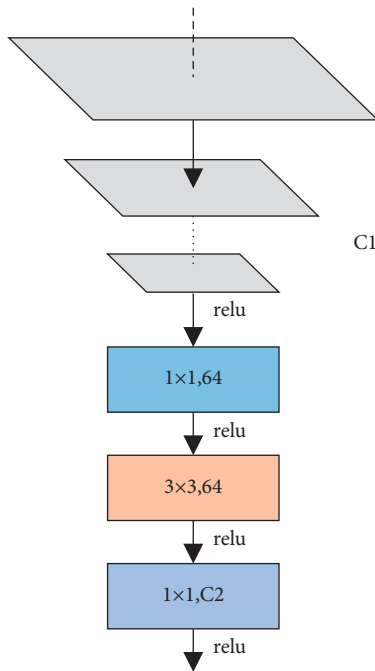
FIGURE 8: Scale-down migration fusion.

model converges, so that the accuracy will not decrease as the network deepens.

According to the above feature scale conversion operation, the scale size of the feature map is converted without destroying the original data of the feature. By using this feature scale transformation method, feature maps are reduced, migrated, and fused in the backbone network, and features are reused, so as to improve the expressive ability of features extracted by convolutional neural networks. The network layer with the same scale of input and output features is called the same level feature, and the last level feature of each level is selected as the reference feature because it has the strongest abstract expression ability after many convolution calculations. As shown in Figure 8, based on these reference features,

this paper first reduces the feature scale of low-level features, sets the downsampling factor r to 2, performs convolution dimension reduction operation through 64 1×1 convolution kernels, then extracts features through convolution operation of 3×3 convolution kernels, then selects 1×1 convolution kernels matching the number of fusion layers to perform convolution dimension enhancement operation, and finally adds them to the fusion layers as the input of the subsequent network to continue extracting features. On the basis of the original YOLOv3 backbone network Darknet-53, this scale reduction migration fusion method is added to the feature layers with reference feature scales of 128×128, 64×64, 32×32, and 16×16.

At the same time, according to the abovementioned feature scale conversion operation, the feature scale amplification migration fusion method is adopted in the feature layers with reference feature scales of 8×8 and 16×16 in the trunk network Darknet-53, instead of the interpolation upsampling method which destroys the original data and has a huge amount of computation in the original FPN network. The specific implementation of feature scale amplification migration fusion is shown in Figure 9. First, the advanced feature map is subjected to feature scale amplification operation, the upsampling factor r is set to 2, and 64 1×1 convolution kernels are used for convolution dimension reduction operation; then, features are extracted by convolution operation of 3×3 convolution kernels, and then 1×1 convolution kernels matched with the number of fusion layers are selected for convolution dimension enhancement operation and finally added with the fusion layers as prediction features.

4. Experimental Results and Analysis

4.1. Experimental Environment and Data Set. The related software and hardware platforms are as follows: Intel Core i7 processor, 2.93 GHz computer with 8G memory, Ubuntu 16.04 LTS operating system, NVIDIA GPU with 24G memory, CUDA8.0, OPENCV3.2.0, and Darknet as the deep learning framework. The data set is ICDAR2015 data set, but the format of the data set needs to be changed into a trainable format accordingly, that is, image2voc, voc2 label files can convert data files into trainable data sets. Configure the internal classification standard of the model as a kind of TEXT, train according to the training set of the existing data set, and save the final model.

4.2. Determine Loss Function. As the loss function is the "baton" of the whole network learning and plays a very important role in the quality of the network model, it is necessary to design and optimize the loss function before network training. Therefore, during model training, the sum loss of square error is adopted for the coordinates, height and width of BBox, and the cross entropy loss is adopted for the classification scoring of BBox. The joint loss of multiple parts is as follows:
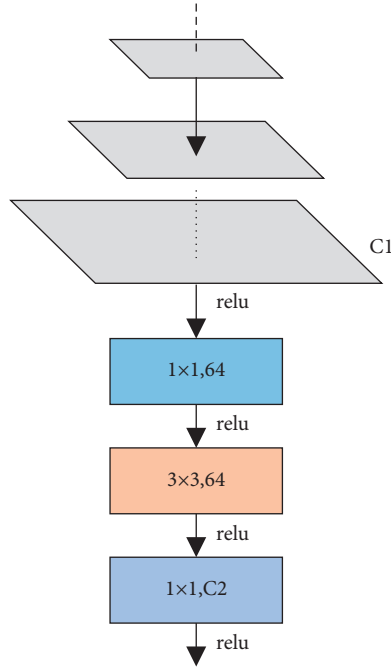
FIGURE 9: Scale-up migration fusion.

$$L = \sigma \sum_{i=1}^{N} \left[(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 + \left(\sqrt{w_i} - \sqrt{\hat{w}_i} \right)^2 + \left(\sqrt{h_i} - \sqrt{\hat{h}_i} \right)^2 \right] + \lambda \sum_{i=1}^{N} \sum_{c \in C} -\hat{p}_i(c) \log(p_i(c)), \tag{8}$$

where N is the number of times when the IOU value of prior frame and real frame is greater than the threshold; (x_i, y_i), w_i and h_i are the center point coordinates, width and height of the i-th prediction frame; $(\hat{x}_i, \hat{y}_i)$; $\hat{w}_i$ and $\hat{h}_i$ are the center point coordinates, width and height of the real frame matched with the i-th prediction frame; $p_i(c)$ indicates the confidence score that the i-th prior frame belongs to category c; $\hat{p}_i(c)$ indicates the confidence score that the i-th prior frame matches the real frame belongs to category c; and σ and λ are the loss weights of BBox location and classification, respectively.

4.3. Set Network Parameters and Train. The training round of the proposed network on ICDAR2015 data set is 135. In the whole training process, the number of images in each batch is 64, the weight attenuation is 0.0005, and the momentum is 0.9. All additional network layer parameters are initialized with Xavier 8. The setting of learning rate is as follows: because the model tends to converge and be stable in the training and learning process, if the learning rate is kept high, the model will usually diverge due to unstable gradient. Therefore, the learning rate of the first 75 training rounds is set to 10^{-2}, the learning rate of the middle 30 training rounds is set to 10^{-3}, and the learning rate of the last 30 training rounds is set to 10^{-4}. After the network design is completed and the network parameters are set, the random gradient descent algorithm (SGD) is used to update the network parameters under the guidance of the loss function.

4.4. Result Analysis. In natural scenes, the text detection effect of traditional YOLOv3 network is not good, which may be due to the error in prediction when predicting the text pixels. In contrast, this text can be partially recognized in the improved YOLOv3 network, which improves the recall rate of locating targets. Text detection examples of traditional YOLOv3 network and improved YOLOv3 network are shown in Figures 10 and 11, respectively.

According to the effect comparison between Figures 10 and 11, the detection effect of the improved YOLOv3 network characters is better. Comparing the accuracy and detection speed of the proposed improved YOLOv3 network with the traditional YOLOv3 network, the results are shown in Table 1.

It can be seen from Table 1 that the MAP (mean average precision) of the proposed improved YOLOv3 network is increased by 9.5%, while the single frame detection time on a Tian X GPU is only increased to 27 ms from the original 22 ms. This shows that this chapter improves the algorithm of

(a)

(b)

(c)

(d)

FIGURE 10: Text detection example of the traditional YOLOv3 network. (a) Subway advertisement. (b) Notice board. (c) Wall sign. (d) Signboard.

(a)

(b)

(c)

(d)

FIGURE 11: Text detection example of the improved YOLOv3 network. (a) Subway advertisement. (b) Notice board. (c) Wall sign. (d) Signboard.

TABLE 1: Comparison of precision and detection speed

Network structure	mAP	Time (ms)
YOLOv3	67.2	**22**
Improved YOLOv3	**73.6**	27

Figure 12: PR curve.

Figure 13: AP curve.

5. Conclusions

In this paper, through the introduction of feature scale transformation method, a feature fusion method of feature scale enlargement migration fusion and feature scale reduction migration fusion is proposed. The original YOLOv3 network is modified by adopting the backbone network of feature scale reduction migration fusion to extract features and the FPN prediction of feature scale enlargement migration fusion. Compared with traditional YOLOv3, the improved YOLOv3 network through multiscale feature transformation, migration, and fusion provides more robust feature expression for the two tasks of object detection, frame regression, and category recognition. From the experimental results, it can be seen that the proposed method can effectively avoid the phenomenon of missing detection and wrong detection caused by occlusion or small target and obviously improve the detection accuracy of English text in natural scenes in videos.

Data Availability

The experimental data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] B. Demir and C. Bruzzone, "A novel active learning method in relevance feedback for content-based remote sensing image retrieval," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 539–556, 2015.

[2] L. Zhu, J. Shen, H. Jin, R. Zheng, and L. Xie, "Content-based visual landmark search via multimodal hypergraph learning," *IEEE Transactions on Cybernetics*, vol. 45, no. 12, pp. 2756–2769, 2015.

[3] S. R. Dubey, S. K. Singh, and R. K. Singh, "Multichannel decoded local binary patterns for content-based image retrieval," *IEEE Transactions on Image Processing*, vol. 25, no. 9, pp. 4018–4032, 2016.

[4] V. Bisot, R. Serizel, S. Essid, and G. Richard, "Feature learning with matrix factorization applied to acoustic scene classification," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 25, no. 6, pp. 1216–1229, 2017.

[5] X. Wang, "Application of network protocol improvement and image content search in mathematical calculus 3D modeling video analysis[J]," *AEJ - Alexandria Engineering Journal*, vol. 60, no. 5, pp. 4473–4482, 2021.

[6] J. Zou, W. Li, C. Chen, and Q. Du, "Scene classification using local and global features with collaborative representation fusion," *Information Sciences*, vol. 348, pp. 209–226, 2016.

[7] M. Hayat, S. H. Khan, M. Bennamoun, and S. An, "A spatial layout and scale invariant feature representation for indoor

YOLOv3 network and does not bring too many parameters to slow down the detection speed of the model when constructing the backbone network of feature scale reduction migration fusion and FPN of feature scale enlargement migration fusion, thus improving the detection accuracy of the algorithm without affecting the excellent real-time detection performance of the original YOLOv3.

In addition, using AP (average precision) and PR (precision recall) as evaluation indexes, the performance of the improved YOLOv3 network is compared with that of the traditional YOLOv3 network, SSD512, and faster-RCNN. The PR curve and AP curve of each model on ICDAR2015 data set are shown in Figures 12 and 13, respectively.

scene classification," *IEEE Transactions on Image Processing*, vol. 25, no. 10, pp. 4829–4841, 2016.

[8] D. A. Chandy, A. H. Christinal, A. J. Theodore, and S. E. Selvan, "Neighbourhood search feature selection method for content-based mammogram retrieval[J]," *Medical, & Biological Engineering & Computing*, vol. 55, no. 3, pp. 1–13, 2017.

[9] J. S. Deville, D. Kihara, and A. Sit, "2DKD: a toolkit for content-based local image search," *Source Code for Biology and Medicine*, vol. 15, no. 1, pp. 125–141, 2020.

[10] M. Moirangthem and T. R. Singh, "Brain tumor detection through content-based medical image retrieval using roi segmentation with Harmony search optimization," *Journal of Green Engineering*, vol. 10, no. 10, pp. 8939–8969, 2020.

[11] Z. Mehmood, M. Rashid, A. Rehman, T. Saba, H. Dawood, and H. Dawood, "Effect of complementary visual words versus complementary features on clustering for effective content-based image search," *Journal of Intelligent and Fuzzy Systems*, vol. 35, no. 5, pp. 5421–5434, 2018.

[12] M. Kurkure, A. Thakare, and S. Gudadhe, "Genetic candidate group search approach for post clustering content based image retrieval," *International Journal of Computer Application*, vol. 132, no. 16, pp. 6–9, 2015.

[13] B. Gábor, H. Szcs, and S. Dávid, "Content-based image retrieval for multiple objects search," *Cybernetics and Information Technologies*, vol. 17, no. 2, pp. 104–116, 2017.

[14] Z. Lan, T. Jung, K. Liu et al., "PIC: enable large-scale privacy preserving content-based image search on cloud," *IEEE Transactions on Parallel and Distributed Systems*, vol. 28, no. 11, pp. 3258–3271, 2017.

[15] M. E. Bates, "Picture this: challenges of video content and searchability," *Online*, vol. 41, no. 3, p. 72, 2017.

[16] X. Lin, X. Wang, and L. Li, "Intelligent detection of edge inconsistency for mechanical workpiece by machine vision with deep learning and variable geometry model," *Applied Intelligence*, vol. 50, no. 7, pp. 2105–2119, 2020.

[17] T. Kim, I. Y. Jung, and Y. C. Hu, "Automatic, location-privacy preserving dashcam video sharing using blockchain and deep learning," *Human-centric Computing and Information Sciences*, vol. 10, no. 1, pp. 1–23, 2020.

[18] R. Minetto, N. Thome, M. Cord, N. J. Leite, and J. Stolfi, "SnooperText: a text detection system for automatic indexing of urban scenes," *Computer Vision and Image Understanding*, vol. 122, no. 5, pp. 92–104, 2014.

[19] X.-C. Yin, W.-Y. Pei, J. Zhang, and H.-W. Hao, "Multi-orientation scene text detection with adaptive clustering," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 9, pp. 1930–1937, 2015.

[20] A. Risnumawan, P. Shivakumara, C. S. Chan, and C. L. Tan, "A robust arbitrary text detection system for natural scene images," *Expert Systems with Applications*, vol. 41, no. 18, pp. 8027–8048, 2014.

[21] Y.-Z. Zhuge and H.-C. Lu, "Robust video text detection with morphological filtering enhanced MSER," *Journal of Computer Science and Technology*, vol. 30, no. 2, pp. 353–363, 2015.

[22] S. M. Alqhtani, S. Luo, and B. Regan, "Fusing text and image for event detection in twitter," *The International journal of Multimedia & Its Applications*, vol. 7, no. 1, pp. 710–717, 2015.

[23] L. Tong, S. Palaiahnakote, C. L. Tan et al., "Video text detection systems," *Advances in Computer Vision & Pattern Recognition*, vol. 12, no. 4, pp. 379–388, 2014.

[24] R. Yang, X. Zha, K. Liu, and S. Xu, "A CNN model embedded with local feature knowledge and its application to time-varying signal classification," *Neural Networks*, vol. 142, no. 1, pp. 564–572, 2021.

[25] D. Wu, Q. Wu, X. Yin et al., "Lameness detection of dairy cows based on the YOLOv3 deep learning algorithm and a relative step size characteristic vector," *Biosystems Engineering*, vol. 189, pp. 150–163, 2020.

[26] Y. Li, Z. Zhao, Y. Luo, and Z. Qiu, "Real-time pattern-recognition of GPR images with YOLO v3 implemented by tensorflow," *Sensors*, vol. 20, no. 22, p. 6476, 2020.

[27] F. Montalbo, "A computer-aided diagnosis of brain tumors using a fine-tuned YOLO-based model with transfer learning [J]," *KSII Transactions on Internet and Information Systems*, vol. 14, no. 12, pp. 4816–4834, 2021.

[28] A. Shukla, I. Garkoti, A. M. B. Choudhary, and P. V. Dhaka, "Social distancing detection using open CV and yolo object detector[J]," *International Journal for Modern Trends in Science and Technology*, vol. 7, no. 1, pp. 93–95, 2021.

[29] L. Cao, H. Li, R. Xie, and J. Zhu, "A text detection algorithm for image of student exercises based on CTPN and enhanced YOLOv3," *IEEE Access*, vol. 8, pp. 176924–176934, 2020.

Retraction

Retracted: Research on Intelligent English Translation Method Based on the Improved Attention Mechanism Model

Scientific Programming

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope

(2) Discrepancies in the description of the research reported

(3) Discrepancies between the availability of data and the research described

(4) Inappropriate citations

(5) Incoherent, meaningless and/or irrelevant content included in the article

(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] R. Wang, "Research on Intelligent English Translation Method Based on the Improved Attention Mechanism Model," *Scientific Programming*, vol. 2021, Article ID 9667255, 8 pages, 2021.

Research Article

Research on Intelligent English Translation Method Based on the Improved Attention Mechanism Model

Rong Wang (ID)

School of Foreign Languages, Shaanxi Xueqian Normal University, Xi'an 710100, Shaanxi, China

Correspondence should be addressed to Rong Wang; rongw@mail.bnu.edu.cn

The use of neural machine algorithms for English translation is a hot topic in the current research. English translation using the traditional sequential neural framework, which is too poor at capturing long-distance information, has its own major limitations. However, the current improved frameworks, such as recurrent neural network translation, are not satisfactory either. In this paper, we establish an attention coding and decoding model to address the shortcomings of traditional machine translation algorithms, combine the attention mechanism with a neural network framework, and implement the whole English translation system based on TensorFlow, thus improving the translation accuracy. The experimental test results show that the BLUE values of the algorithm model built in this paper are improved to different degrees compared with the traditional machine learning algorithms, which proves that the performance of the proposed algorithm model is significantly improved compared with the traditional model.

1. Introduction

Natural language is an important vehicle for knowledge and information dissemination, as well as an outward expression of human civilization and wisdom [1]. NLP, a cross-cutting field spanning computer linguistics and artificial intelligence, explores how computers can understand and process complex human language and, on this basis, achieves a true sense of human-computer interaction [2]. NLP covers a wide range of research areas, including machine translation, information extraction, text summary generation, question and answer systems, sentiment analysis, reading comprehension, opinion analysis, data mining, and many other areas that are currently very compelling [3]. As the Internet continues to evolve and the era of artificial intelligence approaches, the amount of data generated on the Internet is growing exponentially and how to efficiently process and extract useful data has become an urgent problem for major companies [4].

Language is the main tool for cultural exchange, but there is a huge language gap between the mother tongues of different countries, which undoubtedly brings many obstacles to the cultural exchange of people in the world [5, 6]. Therefore, the demand for human translators is increasing, and this has led to the high price of human translators, which is difficult for the general public, to afford such an expensive price. The automated nature of machine translation makes translation between languages easy and efficient, which can undoubtedly contribute to a wide range of communication between countries around the world. According to investor opinion's forecast, with the increasing globalization of the economy, machine translation technology will play an increasingly important role in it and the market size of machine translation will reach \$1.5 billion in 2024 [7, 8].

Deep learning, a new machine learning method, is capable of automatically learning abstract features and establishing mapping relationships between input and output signals. Due to its powerful feature learning and representation capabilities, the application of deep neural networks to both speech recognition and image processing has yielded results that far exceed traditional methods, which have brought more dynamism to these fields [9]. With the recent success of deep learning in signal variable processing, more and more

researchers have started to work on applying neural networks to symbolic variables and have made similar progress, a classic example being the application of deep learning to reading comprehension [10], which likewise provides a new way of thinking for machine translation research.

Neural machine translation (NMT), based on deep learning, is capable of automatically learning abstract features and establishing relationships between source and target utterances and has recently obtained far better performance than SMT in various tasks of machine translation [11–14]. The encoder-decoder model is currently one of the most widely used models in NMT [10, 15], which first maps the source utterance into a distributed representation at the encoder side (source side) and then uses an attention mechanism at the decoder side (target side) to generate a target word cyclically [12]. Most encoders in previous studies have used recurrent neural networks (RNNs) to sequentially to encode the source-side utterances [13], but this has some shortcomings. Most notably, although RNNs can learn relationships between words, they still cannot solve the long-term dependency problem and are very slow to train due to the model structure. Self-attention networks (SANs) can not only learn relationships between words like RNNs [14] but also capture relationships between words by explicitly focusing on all words regardless of the distance between them and solve the problem of slow computation of RNNs.

Therefore, machine translation models based on SANs are currently doing the most excellent performance in various machine translation tasks. In this paper, we propose an English-Chinese translation model with improved attention mechanism to address the shortcomings of the traditional decoder-translator algorithm framework. The core idea of the algorithm is to combine the attention mechanism with neural networks and use deep learning methods to train the local attention of the translation model. Therefore, it can improve the translation system's ability to connect the context and then effectively improve the translation quality.

2. Related Work

Banchs and Costa-jussà [15] spent decades perfecting his invention; the machine they constructed not only failed to achieve the desired results in practice but also translated sentences with very poor readability, which made many scientists at the time to consider the invention useless. The conjecture of machine translation was therefore shelved until the emergence of electronic computers, and British and American engineers Booth and Weaver thought that automatic translation could be carried out with the help of computers [16].

With the introduction of statistical machine translation in 1990, the research on machine translation gradually began to go into a boom period and thus also entered a period of rapid development [17]. Unlike the rule-based or instance-based machine translation in the past, statistical machine translation has a very good mathematical foundation and can effectively utilize large-scale corpus, which also results in significant improvement in the performance of translation. This time also saw the emergence of many studies related to

statistical machine translation, such as word-based SMT, phrase-based SMT, and grammar-based SMT [18].

And, with the proposal of neural machine translation in 2013 [19], the research of machine translation has thus entered a brand new era. Taking neural network as the basic unit to construct the corresponding machine translation model, using the ability that neural network can automatically learn abstract features and establish the mapping relationship between input signal and output signal, with the support of large-scale corpus, the performance of neural machine translation has reached an unprecedented height and the performance on various machine translation tasks is far more than that of statistical machine translation [20], which becomes the dominant research approach. With the introduction of Transformer model by Google in 2017 [21], it has pushed the quality of machine translation to another peak, the accuracy rate has reached an astonishing 85%, and most sentences can be translated more accurately, which can basically meet people's daily needs, and the field of machine translation has thus shown a prosperous scene.

The Institute of Linguistics, Chinese Academy of Social Sciences, Tsinghua University, Harbin Institute of Technology, Northeastern University, Hong Kong University of Science and Technology, and University of Macau are all conducting research on machine translation [22, 23]. In the past few years, we have obtained a series of self-developed machine translation systems, such as "THUMT" of Tsinghua University, "Multi-language Machine Translation System" of Chinese Academy of Sciences, and "Chinese-Portuguese-English Translation System" of the University of Macau [24, 25], and a series of excellent achievements. In addition, Internet companies such as Baidu, Netease, Tencent, KDXunfei, and Sogou are also increasing their research on machine translation, which has given birth to translation software such as Baidu Translator, Sogou Translator, and Youdao Translator and real-time voice translation tools such as Xunfei Translator [26, 27].

3. Basic Modeling Research

3.1. Encoder-Decoder Architecture Model.
The encoder encodes the input data information of the neural network into a fixed-length piece of data. The decoder takes the data encoded by the encoder and decodes it in reverse and then outputs it as a translated sentence. This is the idea underlying the sequence model [28], as shown in Figure 1.

3.2. Recurrent Neural Networks.
The encoder-decoder framework is part of a neural network, and running the framework requires building a proper neural network model. Among the many neural network models, recurrent neural networks are the most widely used. RNNs are deformed models of feedforward neural networks, whose main feature is that they can handle data sequences of different lengths. Figure 2 shows the structure of recurrent neural network algorithm, which has the recursive property that the state of each time has a greater relationship with the previous activation state [29].
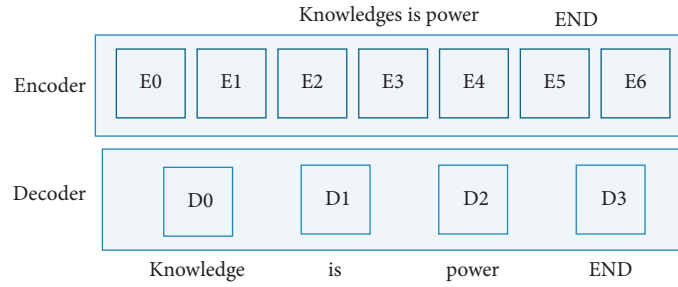
FIGURE 1: Encoder-decoder translation flowchart.
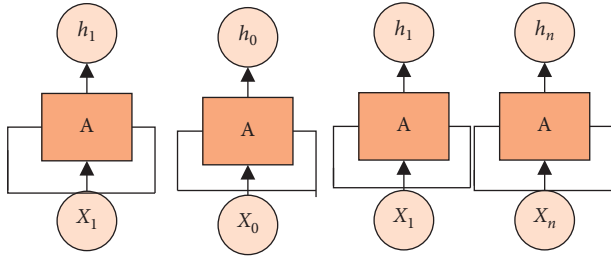
FIGURE 2: Recurrent neural network algorithm structure.

Although the input data of the theoretical recurrent neural network can be of infinite length, in the experimental state, if the data length is infinite, it will cause the problem of excessive gradient of the neural network, so the neural network may not be able to capture the correlation between the chapter contexts, resulting in the degradation of translation quality [30].

3.3. Recurrent Neural Networks with Attention Mechanism. Due to the various drawbacks of recurrent neural networks, the current mainstream approach is to combine recurrent neural networks with encoder-decoders, which can reduce the gradient problem caused by long data sequences [31].

Since the recurrent neural network operation mechanism is a left-right sequential operation process, this will significantly limit the parallel operation capability of the model itself and the sequential operation of data will also cause the problem of data module loss. The above problem will be improved by using the attention mechanism, which can change the data distance to 1 at any position in the translated data, so that it does not depend on the effect of the previous sequential operation on the current operation and the system will have better parallelism [32].

The mathematical principle of the attention mechanism is that the input data of the neural network is first weighted and imported into the encoder, which then imports the data to the decoder and the decoder queries the data weights in the decoding process as the inverse input data and then realizes the weighted average of the data of each state. The simplified implementation flow of the attention mechanism is shown in Figure 3 [11].

3.4. Model Construction. To better highlight the advantages of the attention mechanism, the model framework in this paper is based on a recurrent neural network model incorporating the attention mechanism for building an encoder-decoder framework to implement the translation task. The neuronal connectivity part of the model is implemented using the attention mechanism, which can bring out the advantages of the attention mechanism better.

3.5. Model Framework. Figure 4 shows the attention mechanism model constructed in this paper, the overall structure of which consists of an encoder and a decoder. Among them, the encoder consists of a header attention single-layer structure and a preceding network single-layer structure and the number of blocks of the whole encoder is N_c. The decoder structure is similar to the encoder and also consists of N_c block structures, except that there is no head attention layer.

This neural network uses a differential network connection approach, the distinctive feature of which is that the network enters the normalization hierarchy (Add and Norm) for data processing [33].

3.6. Construction of the Attention Mechanism Module. The attention mechanism module is mainly divided into the encoder module and the decoder module. The input part of the encoder module is the whole data sequence, and the input matrices are set to 3, Q, K, and V. The attention mechanism function can be regarded as a mapping relationship, and the function equation is [15]

$$\text{attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_k}}\right)V. \tag{1}$$

Another important structure in the encoder module is the head attention single-layer structure, the model of which is shown in Figure 5.

4. Experiment and Analysis

4.1. Extraction of Feature Parameters. In order to further improve the system computing efficiency and reduce the interference of data unrelated to the speech signal, the relevant information data should be unified and thus find the parameter features and then realize the subsequent calculation. Figure 6 shows the structure for extracting feature parameters.
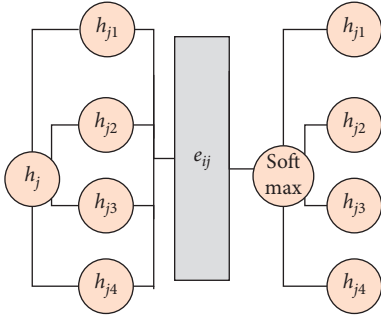
FIGURE 3: Sketch of attention mechanism.
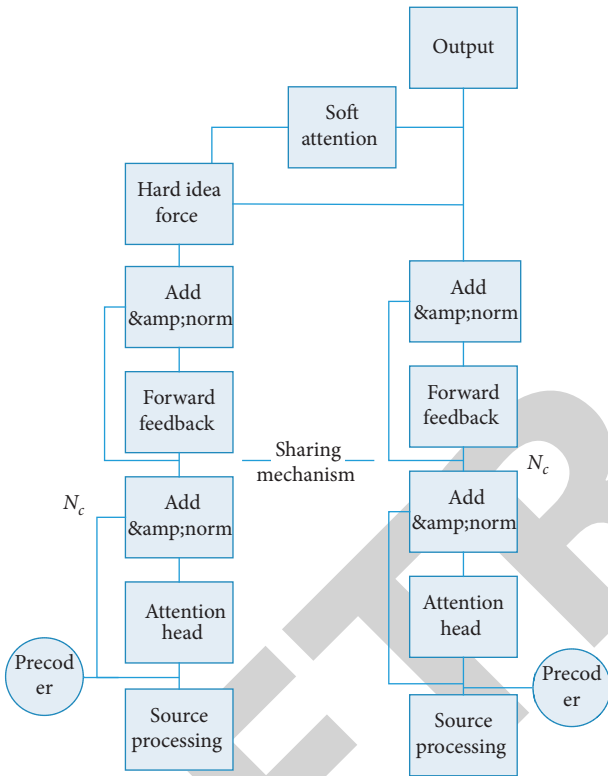
FIGURE 4: Model framework structure.
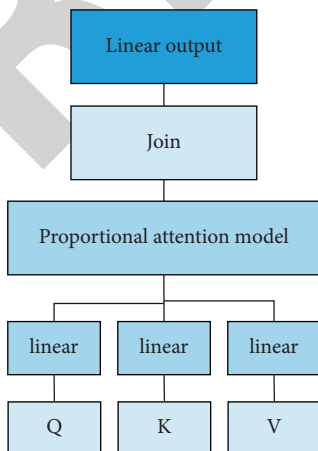
FIGURE 5: Structure schematic.

Nonperiodic continuous time signals use the Fourier transform to calculate the continuous map of the signal, but the actual control system is used to derive continuous signal discrete sampling values, so as to calculate the signal map by discrete sampling values. Fast Fourier transform [2] is as follows:

$$X[K] = \sum_{n=0}^{N-1} x[n]e^{-j(2\pi/N)nk}, \quad k = 0, 1, 2, \ldots, N, \quad (2)$$

where $x[n]$ is the discrete speech sequence obtained by sampling and $X[K]$ is the k-point reset sequence. The discrete speech sequences are transformed into Mel frequency scale:

$$\text{Mel}(f) = 2579\lg\left(1 + \frac{f}{700}\right), \quad (3)$$

where $\text{Mel}(f)$ is the Mel frequency and f is the actual frequency.

The discrete cosine transform DTC is carried out on the filtered output to derive the feature parameter extraction result P of the speech signal $w(n)$, which is calculated as

$$P = \overline{Z}_{n=1}^{N} F(l)w(n)\cos(\pi n(M + 0.5)). \quad (4)$$

After a speech signal generates a spectrogram to be weighted, windowed, and framed, each short-time analysis window is able to obtain spectral information by fast Fourier transform and then the MFCC two-dimensional map is derived using Mel filtering [14, 34].

4.2. Experimental Environment Setup. The experimental dataset used in this paper is the IWSLT2018 corpus data collection, which has a small amount of data. The experimental environment uses Python to program the attention mechanism neural network, and the parameters of the experimental setup environment are shown in Table 1.

4.3. Experimental Tests and Result Analysis. The experiments were performed as follows:

(1) Processing the corpus and cutting long sentences into words

(2) Numbering the words and storing them as files, and then, storing the files to the PC

(3) Normalizing the text, completing the sentences that are not long enough, and intercepting the sentences that are too long

(4) Training the processed sentences, and then, evaluating the BLUE values [12, 13]

In the experimental tests, the RNN (recurrent neural network) translation model, LSTM translation model, and neural network translation model incorporating the attention mechanism was evaluated using the comparison test method. The test results are shown in Table 2.

Table 2 shows the test results of the three models with the default corpus. It can be seen that the basic RNN model has
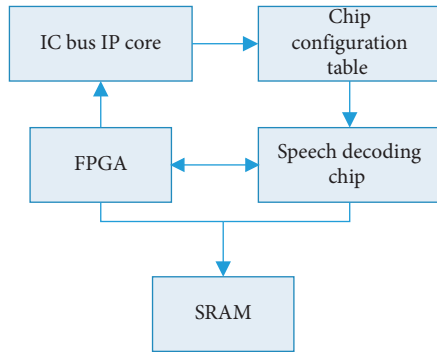
Figure 6: Structure for extracting feature parameters.

Table 1: Environmental parameters of the experimental setup.

Parameters	Content
Corpus data size	25000
Word dimension	512
Layers of neural network	2
Deep learning framework	TensorFlow
Hardware configuration	i7-9700K, 128 GB

Table 2: Test results of the 3 models.

Model	BLUE value
RNN	15.88
LSTM	21.03
ATT + RNN	24.33

the lowest BLUE value and the model built in this paper has the highest BLUE value, which also indicates that the proposed model does have some improvement on translation quality.

In the second comparison experiment, the sentences of the corpus are clustered according to their lengths and then the translation models are tested with sentences of different lengths to verify the long sentence translation ability of the translation models; the experimental results of different models are shown in Table 3.

Table 3 shows the performance comparison of machine translation models with different network structures under the English-German machine translation task. The first column of the table indicates the different kinds of translation models, which include RNN and its variant models such as LSTM state and GRU [27]. All three different RNN structures were modeled as well as experimented, and the one with the highest BELU value among the three was taken for presentation in the table. "Attn" indicates the attention mechanism, while "+Attn" indicates the addition of the attention mechanism to the model. As for "+IntHeads (8)," it indicates replacing the multiheaded attention network in the Transformer model with an interactive multiheaded attention network with 8 attention heads, while "+IntHeads (16)" replaces the multiheaded attention network in the Transformer model with an interactive multiheaded attention network with 16 heads, noting that the number of heads increases, but the parameters of the model hardly increase.

In order to compare the performance of these translation models more visually, the data in Table 3 are visualized as a histogram, as shown in Figure 7.

As can be seen from Figure 7, the Transformer model has the highest score both in the development set and in the test set, and as a machine translation model built using only the self-attention mechanism, its performance does outperform the machine translation models based on RNN and CNN. Comparing the first to the third columns of Table 3, we can find that the performance of the translation model based on RNN or its variants is average without using the attention mechanism, while the performance of the model achieves a more obvious improvement after adding the attention mechanism. The fourth row of the table shows the translation models based on CNN and attention mechanism. Due to the characteristics of CNN, its ability to obtain local information is greatly enhanced, so its performance is greatly improved compared with the first three models, which also confirm the importance of local information to the performance of machine translation.

In addition, by comparing the last three histograms in each dataset, it can be found that the performance of the model is significantly improved after replacing the multiheaded attention network in the Transformer model with interactive multiheaded attention network, except for the test set Test1. This also validates the effectiveness of the interactive multiheaded attention network, which enables the attention heads to share the learned feature representations by improving the performance of the model. It is also found that the more the number of attention heads, the better the interaction between attention heads in the interactive multiheaded attention network and thus the better the performance of the model.

For example, compared with CNN-based machine translation models, the self-attention mechanism focuses on all words in a sentence at the same time, which makes the Transformer model unable to learn local information well, so if certain measures can be taken to enhance the ability of the Transformer model, the performance of the Transformer model will be further improved to acquire local information.

In order to fully demonstrate the advantages of the designed model, the syntax-based and phrase-based intelligent recognition models are used in the experimental process to realize the comparison experiment. The number of control points in each system is recorded, and the distribution of control points is analyzed. The distribution of nodal points can describe the semantic and contextual relevance of English translation, and the dense distribution of nodal points indicates that the system has a high accuracy of English translation recognition. Figures 8 and 9 both show the distribution of the system's recognition nodes. The compact distribution of nodes in Figure 8 indicates that the system has higher recognition performance and more accurate calibration results, and the problem of incoherence in English translation has been solved; the loose distribution of nodes in the syntactic and phrase-based recognition system in Figure 9 and the compact distribution of nodes in the 1st, 4th, and 5th experiments indicate that the system has higher calibration accuracy. But the translation result warning

TABLE 3: Translation model performance comparison.

Translation model	Dev	Test1	Test2	Test3	Average
RNN	16.6	15.9	17.2	16.3	16.5
RNN + Attn	19.0	19.5	20.9	19.4	19.8
CNN + Attn	28.1	26.1	29.1	25.6	27.0
Transformer	31.8	30.8	33.6	29.3	31.2
Transformer IntHeads (8)	32.7	30.7	34.8	30.7	32.1
Transformer IntHeads (16)	33.0	30.9	35.2	31.1	32.4
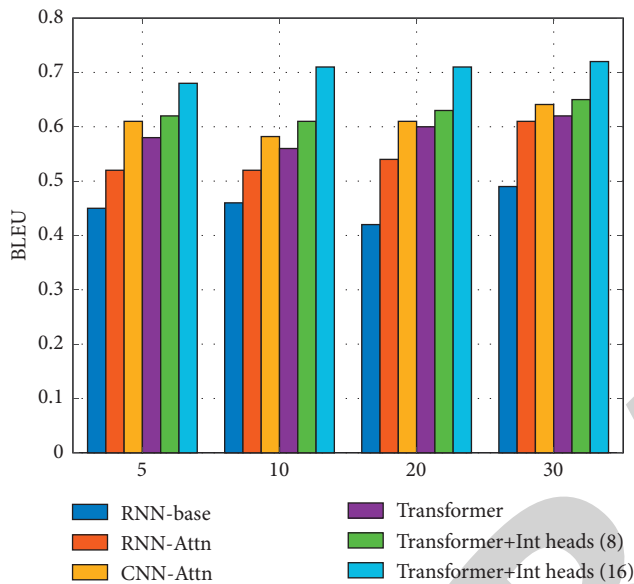
FIGURE 7: Performance of different machine translation models.
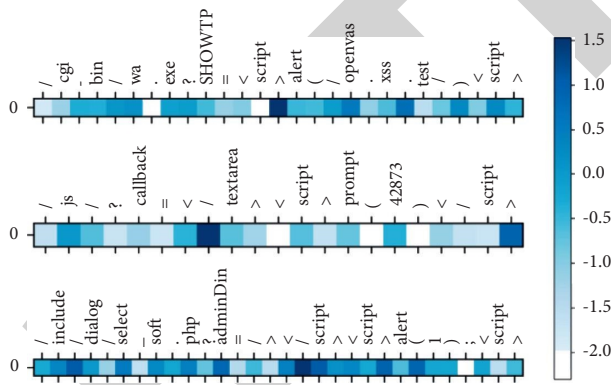
FIGURE 8: Attention weight distribution in word translation.

coherence is relatively poor. Secondly, this system alternates between loose and compact distribution of nodal control points, which indicates that it is not stable [15].

The above description shows that the intelligent recognition model for English translation designed in the paper has high proofreading accuracy and can identify the problem of contextual incoherence in English translation results, giving translation results that are consistent with contextual coherence and rationality.
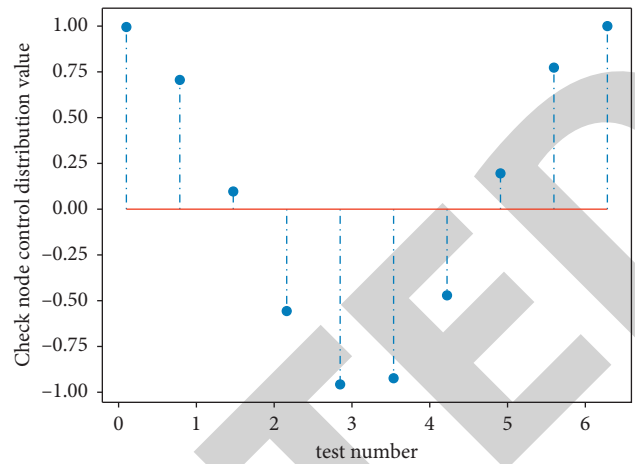
FIGURE 9: Intelligent proofreading control point distribution.

5. Conclusions

To address the shortcomings of the traditional encoding-decoding algorithm translation model, which suffers from inaccurate translation and semantic mutilation, this paper proposes to combine the attention mechanism and recurrent neural network model to build an English-Chinese translation model with improved attention mechanism and use TensorFlow to train the translation model at the same time.

Data Availability

The datasets used in this paper are available from the corresponding author upon request.

Conflicts of Interest

The author declares no conflicts of interest regarding this work.

Acknowledgments

References

[1] W. Gong, H. Chen, and Z. Zhang, "A Novel deep learning method for intelligent fault diagnosis of rotating machinery based on improved CNN-svm and multichannel data fusion," *Sensors*, vol. 19, no. 7, Article ID 1693, 2019.

[2] T. Nishihara, A. Tamura, T. Ninomiya, Y. Omote, and H. Nakayama, "Supervised visual attention for multimodal neural machine translation," in *Proceedings of the 28th International Conference on Computational Linguistics*, pp. 4304–4314, Barcelona, Spain, December 2020.

[3] S. Feng, S. Liu, N. Yang, M. Li, M. Zhou, and K. Q. Zhu, "Improving attention modeling with implicit distortion and fertility for machine translation," in *Proceedings of the COLING 2016, the 26th International Conference on Computational Linguistics*, pp. 3082–3092Technical Papers, Osaka, Japan, December 2016.

[4] P. Koehn, *Statistical Machine translation*, Cambridge University Press, Cambridge, UK, 2009.

[5] J. Bao, D. Tang, N. Duan et al., "Table-to-text: describing table region with natural language," in *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence*, New Orleans, LA, USA, February 2018.

[6] X. Wang, Z. Tu, and M. Zhang, "Incorporating statistical machine translation word knowledge into neural machine translation," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 26, no. 12, pp. 2255–2266, 2018.

[7] R. L. Smith and Z. Fang, "Techniques, applications and future prospects of diamond anvil cells for studying supercritical water systems," *The Journal of Supercritical Fluids*, vol. 47, no. 3, pp. 431–446, 2009.

[8] X. M. Yu, W. Z. Feng, H. Wang, Q. Chu, and Q. Chen, "An attention mechanism and multi-granularity-based Bi-LSTM model for Chinese Q&A system," *Soft Computing*, vol. 24, no. 8, pp. 5831–5845, 2020.

[9] H. Ryo, N. Takahiro, and M. Toshiyuki, "Translation and inclination control for intelligent tension Pole based on mode decoupling method," *Ieej Journal of Industry Applications*, vol. 7, no. 2, pp. 158–165, 2018.

[10] G. R. Mirams, Y. Cui, A. Sher et al., "Simulation of multiple ion channel block provides improved early prediction of compounds' clinical torsadogenic risk," *Cardiovascular Research*, vol. 91, no. 1, pp. 53–61, 2011.

[11] C. Shi, D. Lin, and T. Ishida, "Agent metaphor for machine translation mediated communication," in *Proceedings of the 2013 international conference on Intelligent user interfaces*, South California, U.S.A, March 19-22 2013.

[12] T. Xie, C. Zhang, Z. Zhang, and K. Yang, "Utilizing active sensor nodes in smart environments for optimal communication coverage," *IEEE Access*, vol. 7, pp. 11338–11348, 2018.

[13] S. Chhaniyara, C. Brunskill, B. Yeomans et al., "Terrain trafficability analysis and soil mechanical property identification for planetary rovers: a survey," *Journal of Terramechanics*, vol. 49, no. 2, pp. 115–128, 2012.

[14] Y. Gao, Y. Li, R. Y. Lau, Y. Xu, and M. Bashar, "Finding semantically valid and relevant topics by association-based topic selection model," *ACM Transactions on Information Systems*, vol. 9, no. 1, pp. 3.1–3.22, 2018.

[15] R. E. Banchs and M. R. Costa-jussà, "A vector-space dynamic feature for phrase-based statistical machine translation," *Journal of Intelligent Information Systems*, vol. 37, no. 2, 2010.

[16] C. Karthik, K. Suresh, K. Valarmathi, and R. Jacob Rajesh, "Model-based control for moisture in paper making process," *Advances in Intelligent Systems and Computing*, vol. 324, pp. 257–264, 2015.

[17] Z. Zhang, C. Zhang, M. Li, and T. Xie, "Target positioning based on particle centroid drift in large-scale WSNs," *IEEE Access*, vol. 8, pp. 127709–127719, 2020.

[18] L. Jin, "Research on pronunciation accuracy detection of English Chinese consecutive interpretation in English intelligent speech translation terminal," *International Journal of Speech Technology*, no. 5, pp. 1–8, 2021.

[19] T. B. Adji, B. Baharudin, and N. B. Zamin, "Annotated disjunct in link grammar for machine translation," in *Proceedings of the International Conference on Intelligent &*

[20] X. Li, L. Liu, Z. Tu, S. Shi, and M. Meng, "Target foresight based attention for neural machine translation," in *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, vol. 1, pp. 1380–1390(Long Papers), New Orleans, Louisiana, June 2018.

[21] Z. Zhang, W. Li, and Q. Su, "Automatic translating between ancient Chinese and contemporary Chinese with limited aligned corpora," in *Proceedings of the CCF International Conference on NLP and Chinese Computing*, pp. 157–167, Springer, Dunhuang, China, October 2019.

[22] Y. Yin, J. Su, H. Wen, J. Zeng, Y. Liu, and Y. Chen, "POS tag-enhanced coarse-to-fine attention for neural machine translation," *ACM Transactions on Asian and Low-Resource Language Information Processing*, vol. 18, no. 4, pp. 1–14, 2019.

[23] Y. X. Li, Q. Wang, and Y. J. Zhang, "A Novel information technology based college physical education model[J]," *Advances in Intelligent Systems and Computing*, vol. 191, pp. 53–56, 2013.

[24] A.. Gelbukh, "Lecture Notes in Computer Science] Computational Linguistics and Intelligent Text Processing," *English-Arabic Statistical Machine Translation: State of the Art*, vol. 9041, pp. 520–533, 2015, (Chapter 39).

[25] C. Zhang, T. Xie, K. Yang et al., "Positioning optimisation based on particle quality prediction in wireless sensor networks," *IET Networks*, vol. 8, no. 2, pp. 107–113, 2019.

[26] Z. Shi, X. Zhang, J. Tu, and Z. Yang, "An efficient and improved particle swarm optimization algorithm for swarm robots system," *Proceedings of The Eighth International Conference on Bio-Inspired Computing: Theories and Applications (BIC-TA), 2013*, vol. 212, no. 212, pp. 329–337, 2013.

[27] J. Mizera-Pietraszko and P. Pichappan, "Study of the Establishment of a Reliable English-Chinese Machine Translation System Based on Artificial Intelligence," *Advances in Intelligent Systems and Computing*, vol. 613, pp. 13–23, 2018, (Chapter 2).

[28] A. Vaswani, N. Shazeer, N. Parmar et al., "Attention is all you need," in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, pp. 5998–6008, Long Beach, CA, USA, December 2017.

[29] D. Wu, C. Zhang, L. Ji, R. Ran, H. Wu, and Y. Xu, "Forest fire recognition based on feature extraction from multi-view images," *Traitement du Signal*, vol. 38, no. 3, pp. 775–783, 2021.

[30] W. Qiu, Y. J. Shu, and Y. J. Xu, "Research on Chinese multi-documents automatic summarizations method based on improved TextRank algorithm and seq2seq," in *Proceedings of the 2021 International Conference on Bioinformatics and Intelligent Computing*, pp. 196–201, Harbin, China, January 2021.

[31] S. Zhao and Z. Zhang, "Attention-via-attention neural machine translation," in *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence*, New Orleans, LA, USA, February 2018.

[32] S. R. Indurthi, I. Chung, and S. Kim, "Look harder: a neural machine translation model with hard attention," in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pp. 3037–3043, Florence, Italy, July 2019.

[33] H. Choi, K. Cho, and Y. Bengio, "Fine-grained attention mechanism for neural machine translation," *Neurocomputing*, vol. 284, pp. 171–176, 2018.

Advanced Systems, pp. 205–208, IEEE, Kuala Lumpur, Malaysia, November 2007.

Research Article

Intelligent Retrieval Method of Approximate Painting in Digital Art Field

Jixin Wan (ID) and Yu Xiaobo

Xiamen Academy of Arts and Design, Fuzhou University, Fujian, Xiamen 361021, China

Correspondence should be addressed to Jixin Wan; t04303@fzu.edu.cn

With the rapid development of Internet technology and the wide application of image acquisition equipment, the number of digital artwork images is exploding. The retrieval of near-similar artwork images has a wide application prospect for copyright infringement, trademark registration, and other scenes. However, compared with traditional images, these artwork images have the characteristics of high similarity and complexity, which lead to the retrieval accuracy not meeting the demand. To solve the above problems, an intelligent retrieval method of artwork image based on wavelet transform and dual propagation neural network (WTCPN) is proposed. Firstly, the original artwork image is replaced by the low-frequency subimage after wavelet transform, which not only removes redundant information and reduces the dimension of data but also suppresses random noise. Secondly, in order to make the network assign different competition winning units to different types of modes, the dual propagation neural network is improved by setting the maximum number of times of winning neurons. Experimental results show that the proposed method can improve the accuracy of image retrieval, and the recognition accuracy of verification set can reach over 91%.

1. Introduction

With the advent of the 21st century, computer technology has been rapidly improved, and the wide popularization of image technology has led to an explosive growth in the number of digital artwork images. There are a large number of near-similar images in a large number of picture resources, and the retrieval of near-similar images has broad application prospects, such as detecting copyright infringement of works of art on the Internet and detecting whether new trademarks are similar to existing trademarks [1–6]. In recent years, near-similar image retrieval has gradually become an important branch of image retrieval and attracted more and more attention.

Nowadays, image retrieval technology is the fastest-growing technology in all information retrieval applications, and the number of image retrieval in several major retrieval engines all over the world is increasing very fast. This also makes the near-similar image retrieval technology become the focus of extensive attention at home and abroad and also

becomes a key research issue in the development of information industry and digital media technology. Near-similar image detection uses an image as a query keyword [7–10] and retrieves all database images that are all similar or partially similar to the query image. There are many methods for near-similar image detection, among which feature-based detection, index-based detection, and content-based detection are the most commonly used. At present, content-based image retrieval technology has entered a new research stage. Content retrieval based on image visual features begins to use machine learning to extract image features [11–13]. Artificial neural network (ANN), which belongs to machine learning technology [14, 15], is a widely used network model in the development of machine learning and can extract higher-level features of pictures. Artificial neural network is a neural network constructed artificially and capable of realizing a certain function. It is a theoretical mathematical model of human brain neural network based on the understanding of human brain neural network. It is an information processing system established by imitating

the structure and function of brain neural network. Artificial neural network is a brand-new system with logical thinking, fuzzy processing, and accurate calculation. It is a complex network formed by connecting a large number of simple neurons.

The traditional neural network is based on BP algorithm, but the BP algorithm has been studied with it, and some problems have been found [16–18]. (1) The signal of error correction will become smaller and smaller from top to bottom, and gradient dispersion will occur. (2) It can only converge to the local minimum. However, in practice, the use of tags is usually presented in the form of no data, while BP algorithm can only use tagged data for training. Therefore, it is difficult for BP algorithm to be widely used in practical environment. Counter propagation network (CPN) is an advanced artificial neural network proposed by Robert Hecht-Nielsen, an American scholar, in 1987.

To solve the problem of poor retrieval accuracy of approximate artwork images, an intelligent retrieval method of artwork images based on wavelet transform and dual propagation neural network (WTCPN) is proposed. The main contributions of this paper are as follows:

(1) Replacing the original artwork image with the low-frequency subimage after wavelet transform not only removes redundant information but also reduces the dimension of data and the irrelevant factors that interfere with the recognition performance of CPN

(2) The CPN is improved by setting the maximum times of winning neurons, which not only effectively overcomes the instability of the competition layer of CPN network but also avoids the local minimum points in the training process

2. Literature Review

Content-based image retrieval technology is the research direction. Dubey et al. [19] realized the local binary mode of multichannel decoding through the adder and the decoder, which can be effectively used for content-based image retrieval and is superior to other multichannel-based methods in average retrieval accuracy and average retrieval rate. Alshehri [20] proposed an image retrieval method based on BP neural network prediction technology, which classified and predicted the retrieved data by fuzzy inference of neural network. However, according to the above analysis, the research and development of traditional neural networks are limited by the problems of BP algorithm. Recently, the traditional neural network has been replaced by new network models, such as Boltzmann machine, convolutional neural network (CNN), and residual neural network (RNN), which have promoted the further development of this field.

Filip et al. [21] proposed a CNN image retrieval method without manual annotation, which expanded the maximum average pool through the trainable generalized average pool layer and showed better image retrieval performance on Oxford Buildings and Holiday Datasets. Liu et al. [22] aimed at the retrieval of large-scale network image resources and realized a simple and effective image indexing framework by

combining the pretrained large-scale convolutional neural network with the structured support vector machine. However, because this method is aimed at large-scale image retrieval, the retrieval speed has been greatly improved, but the retrieval accuracy is not high. Rajkumar and Sudhamani [23] proposed an image retrieval system based on residual neural network and used Euclidean distance measure to measure similarity. The retrieval test is carried out on a dataset with 50,000 network images in 250 categories. Experimental results show that compared with Google's random query image retrieval system, the performance of the proposed system is improved by 15%. Wang et al. [24] proposed a large-scale similar image retrieval method based on deep neural network, which mainly used deep frame learning multilevel nonlinear transformation to obtain advanced image features and achieved good retrieval results.

Usually, the dimensions of digital artwork images are very high, which make the algorithm need a long time and a large amount of computation. If the distribution of artwork image points is not very compact, it is not conducive to feature classification because artwork images usually have high-dimensional spatial points. To solve the above problems, this paper uses wavelet transform, which can not only reduce the dimension of artwork image but also filter the high-frequency interference information, highlight the main features of the image, and obtain a low-dimensional image suitable for neural network recognition. In addition, by combining the main features of wavelet transform and CPN network, a higher retrieval recognition rate is achieved.

3. Image Retrieval Based on Wavelet Transform and Improved CPN Network

3.1. Operation Principle and Learning Algorithm of CPN. CPN has three layers of standard structure, and neurons in each layer are all connected with each other. Figure 1 shows the topology of CPN network, which is composed of the input layer, the competition layer, and the output layer. Among them, the input layer and the competition layer form a feature mapping network. The competition layer and the output layer constitute a basic competitive network.

In each layer of CPN, the input vector is represented by X:

$$X = (x_1, x_2, \ldots, x_n)^T. \tag{1}$$

After the competition, the output of the competition layer is expressed by Y:

$$Y = (y_1, y_2, \ldots, y_m)^T, \quad y_i \in \{0, 1\}, \ i = 1, 2, \ldots, m. \tag{2}$$

The output of the network is represented by O:

$$O = (o_1, o_2, \ldots, o_l)^T. \tag{3}$$

The expected output of the network is denoted by d:

$$d = (d_1, d_2, \ldots, d_l)^T. \tag{4}$$

The weight matrix between the input layer and the competition layer is expressed by V:
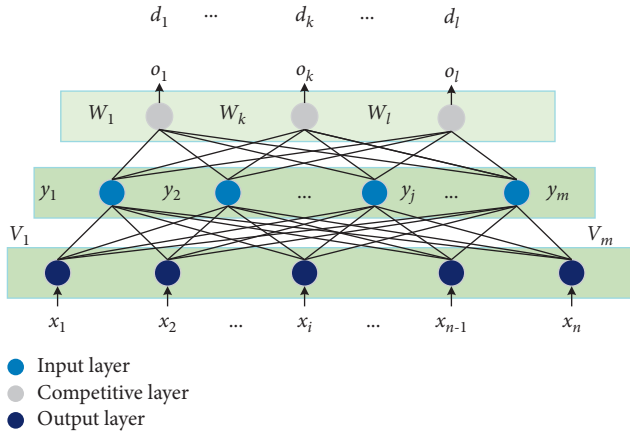
Figure 1: Topological structure of CPN.

$$V = \left(v_1, v_2, \ldots, v_j, \ldots, v_m\right), \tag{5}$$

where the column vector v_j is the inner star weight vector corresponding to the jth neuron in the competition layer.

The weight matrix between the competition layer and the output layer is represented by W:

$$W = \left(w_1, w_2, \ldots, w_k, \ldots, w_l\right), \tag{6}$$

where the column vector w_k is the weight vector corresponding to the kth neuron in the output layer.

It can be seen from Figure 2(a) that, after each layer of the network is trained according to the learning rules, it sends input vectors to the network in the running stage, and then, the competition layer performs competition calculation on them. When the net input value of a neuron is the maximum, it wins the competition, becomes the representative of the current input mode class, and becomes the active neuron shown in the figure at the same time, with an output value of L. While the rest neurons are inactive, and the output value is 0.

It can be seen from Figure 2(b) that, after the competing neurons win the competition, the neurons in the output layer are excited to produce the output pattern shown in the figure. The output value of failed neurons is 0, and the neurons in the output layer do not contribute to the net input and do not affect the output value, so the output is determined by the alien vector corresponding to the neurons that compete for victory.

The learning rules of CPN are composed of unsupervised learning and supervised learning, so the input vector and the expected output vector in the training sample set should be paired. Training is divided into two stages, and each stage adopts a learning rule. In the first stage, the competitive learning algorithm is used to train the inner star weight vector from the input layer to the competitive layer, and the steps are as follows:

(1) All inner star weights are randomly assigned with initial values between 0 and 1 and normalized to unit length, and all input modes in the training set are also normalized.

(2) Enter a pattern X^p, $p = 1, 2, \ldots, P$, where P is the total number of patterns in the training set.

(3) Determine the competition winning neuron. The competition algorithm of CPN has no winning neighborhood, so only the inner star weight vector of the winning neuron is adjusted. The adjustment rule is as follows:

$$V(t+1) = \widehat{V}(t) + \eta(t)[\widehat{X} - \widehat{V}(t)], \tag{7}$$

where $\eta(t)$ is the learning rate which is an annealing function that decreases with time.

(4) Repeat steps (2) to (3) until it drops to 0. It should be noted that the weight vector must be normalized again after adjustment.

In the second stage, an alien learning algorithm is adopted to train the alien weight vector from the competition layer to the output layer, and the steps are as follows:

(1) Input a mode pair X^p and d^p, wherein the weight matrix from the input layer to the competition layer keeps the training result of the first stage.

(2) Determine the neuron $\text{net}_{j^*} = \max_j\{\text{net}_j\}$ that wins the competition, and satisfy

$$y_j = \begin{cases} 0, & j \neq j^*, \\ 1, & j = j^*. \end{cases} \tag{8}$$

(3) Adjust the alien weight vector from the competition layer to the output layer, and the adjustment rule is

$$W_{j^*}(t+1) = \widehat{W}_{j^*}(t) + \beta(t)[d - O(t)], \tag{9}$$

where $\beta(t)$ is the learning rate of alien rules, and it is also an annealing function that decreases with time; $O = (o_1, o_2, \ldots, o_l)$ is the output value of the neurons out of the layer, which is calculated by the following formula:

$$o_k(t) = \sum_{k=1}^{l} w_{jk} y_j, \quad k = 1, 2, \ldots, l,$$
$$O_k(t) = w_{j^*k} y_{j^*} = w_{j^*k}. \tag{10}$$

The alien weight vector adjustment rules are as follows:

$$W_{j^*}(t+1) = W_{j^*}(t) + \beta(t)[d - W_{j^*}(t)]. \tag{11}$$

(4) Repeat steps (1) to (3) until $\beta(t)$ drops to 0.

3.2. Principle of Wavelet Transform. Firstly, the wavelet transform filters the signal [25], using a group of high-pass and low-pass filters with different scales. Then, we analyze and process the signals decomposed into different frequency bands, which are decomposed from the high-frequency and low-frequency components in the original signal. Finally, in order to reach the preset threshold, the above filtering process should be repeated.
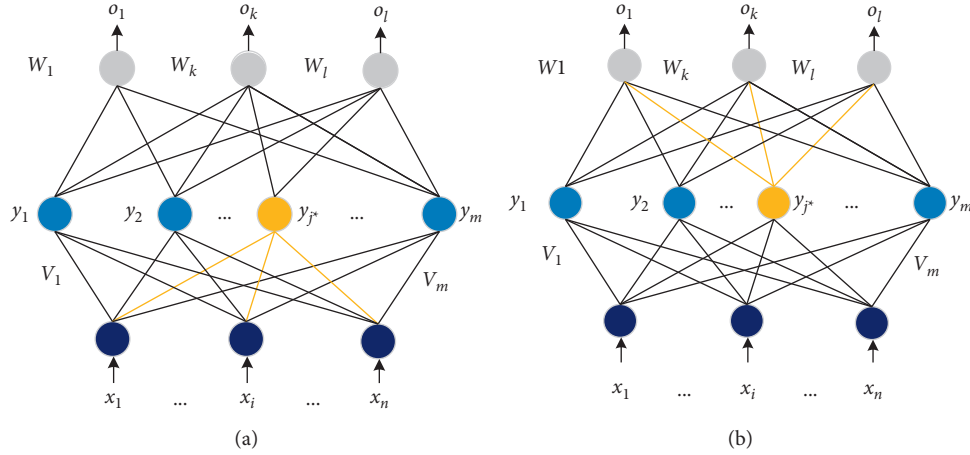
FIGURE 2: Operation process of CPN network. (a) Competition generates winning nodes; (b) win node's alien vector decision output.

At different scales α, the function $\psi(t)$ of a basic wavelet (mother wavelet) and the signal $x(t)$ to be analyzed carry out the inner product, which constitutes the wavelet basis function:

$$WT_x(\alpha, \tau) = \frac{1}{\sqrt{\alpha}} \int_{-\infty}^{+\infty} x(t)\psi^*\left(\frac{t - \tau}{\alpha}\right)dt, \quad \alpha > 0. \quad (12)$$

The equivalent frequency domain is expressed as follows:

$$WT_x(\alpha, \tau) = \frac{\sqrt{\alpha}}{2\pi} \int_{-\infty}^{+\infty} x(\omega)\psi^*(\alpha\omega)e^{j\omega\pi}d\omega. \quad (13)$$

Images in real life are generally two-dimensional signals. Therefore, extending wavelet from one dimension to two dimensions is the basic idea of applying two-dimensional wavelet transform to image processing. Two-dimensional wavelet transform is used to decompose the artwork image in frequency domain, and four regions can be obtained: low-frequency region is different from high-frequency region and the former LL is an approximate component, while the latter LH, HL, and HH are the horizontal component, the vertical component, and the diagonal component, respectively. Among them, the transformed low-frequency region LL can also be subjected to wavelet transform again.

Figure 3 is a schematic diagram of primary and secondary wavelet decomposition, respectively, in which LL1 is the low-frequency subimage of the original image, LH1 and HL1 are horizontal and vertical subimages, and HH1 is the high-frequency subimage of the image.

3.3. Proposed WTCPN.
The proposed WTCPN method is divided into two processes: training process and recognition process. In the training process, firstly, the training sample of the image is read, and the original training sample set is X, and the artwork image is decomposed by wavelet transform. The wavelet basis function selected here is Daubechies [26], and the wavelet coefficients obtained are independent of each other. Finally, the transformed LL part is selected as the approximation of the original image.

Then, the improved CPN algorithm is used to classify the images. In the process of identification, firstly, the test

samples are subjected to wavelet transform, and their low-frequency component subgraphs are selected as the approximation of the test samples, and then, the test sample identification space is constructed, which is classified and identified according to the improved CPN algorithm.

Let $\mathbf{U}_k = [u_1^k, u_2^k, \ldots, u_n^k]^T$ be the input mode of CPN network and $\mathbf{V}_k = [v_1^k, v_2^k, \ldots, v_m^k]^T$ be the output of the competition layer, but the actual output of the output layer is $\mathbf{C}_k = [c_1^k, c_2^k, \ldots, c_z^k]^T$ and the desired output of the output layer is $\mathbf{Y}_k = [y_1^k, y_2^k, \ldots, y_z^k]^T$. The number of neurons in the input layer, the competition layer, and the output layer is n, m, and z, respectively. P is the number of input modes. $\mathbf{W}_j = [w_{j1}, w_{j2}, \ldots, w_{jn}]^T$ is the connection weight vector from the input layer to the competition layer. $\mathbf{Q}_l = [q_{l1}, q_{l2}, \ldots, q_{lm}]^T$ is the connection weight vector from the competition layer to the output layer.

The improved CPN learning algorithm includes the following steps:

(1) Assign each component of $\mathbf{W}_j$, $(j = 1, 2, \ldots, m)$ and $\mathbf{Q}_l (l = 1, 2, \ldots, z)$ to a random value in the [0, 1] interval for initialization. At the same time, a variable t (with an initial value of 0) is added to each neuron in the competition layer to record the number of times the neuron won. The maximum number of neuron wins is set as T, and the error tolerance is specified as e.

(2) The kth input pattern U_k is provided to the network input layer.

(3) Carry out normalization processing on the connection weight vector $\mathbf{W}_j (j = 1, 2, \ldots, m)$.

(4) Find the input activation value of neurons in the competition layer:

$$S_j = \sum_{i=1}^{n} w_{ji}u_i^k, \quad (j = 1, 2, \ldots, m). \quad (14)$$

(5) Find out the maximum activation value S_α from the calculated S_j. If t of S_α is less than T, $t = t + 1$, and the neuron corresponding to S_α is taken as the winning neuron g in the competition layer. Otherwise, if $t \geq T$,
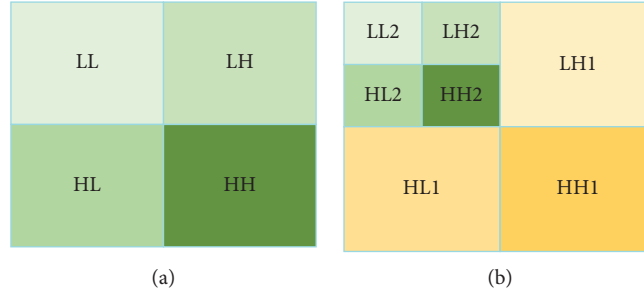
FIGURE 3: Schematic diagram of wavelet decomposition. (a) Primary wavelet decomposition; (b) secondary wavelet decomposition.

select the maximum activation value S_b except S_α. If t of S_b is less than T, $t = t + 1$, and the neuron corresponding to S_b is regarded as the winning neuron g in the competition layer. Otherwise, search in S_j in order of activation value from large to small. Set the output of the winning neuron g in the competition layer as 1, and the rest as 0, and its corresponding connection weight is W_g.

(6) Adjust W_g in the following ways:

$$w_{gi}(t + 1) = w_{gi}(t) + \alpha\left(u_i^k - w_{gi}(t)\right)$$
$$(i = 1, 2, \ldots, n),$$

(15)

where α is the learning rate.

(7) Adjust the connection weight vector $\mathbf{Q}_g = [q_{1g}, q_{2g}, \ldots, q_{lg}]$ from the winning neuron g in the competition layer to the neurons in the output layer, while other connection weights remain unchanged as follows:

$$q_{lg}(t + 1) = q_{lg}(t) + \beta v_j\left(y_1^k - c_l^k\right), \quad l = 1, 2, \ldots, z.$$

(16)

(8) Calculate the weighted sum of the comprehensive input signals of each neuron in the output layer and take it as the actual output of the output neuron.

(9) Calculate the error between the actual output $\mathbf{C}_k$ of the network and the desired output $\mathbf{Y}_k$:

$$\text{err} = \sqrt{\sum_{i=1}^{z}\left(c_i^k - y_i^k\right)^2}.$$

(17)

(10) Judging whether the error calculated in step (9) is less than the error tolerance, and if so, continue step (11) to learn the next mode. If it is greater than the error tolerance, return to step (3) to continue learning.

(11) Return to step (2) until all the p input modes are provided to the network.

4. Experimental Results and Analysis

4.1. Common Datasets and Performance Evaluation Methods.
The effectiveness of the WTCPN method is analyzed experimentally on two public datasets, specifically Oxford [27]

and Holiday [28]. Oxford architecture dataset contains 55 query images corresponding to 11 different buildings. Each query image has a rectangular area to define the building. Holiday dataset includes 1491 amplitude false pictures, which are divided into 500 groups, and each group has a different scene or object. The error tolerance $e = 0.01$, learning rate $\alpha = 0.4$, and learning rate $\beta = 0.5$ in WTCPN algorithm. The number of neuron nodes in the input layer is 4096, the number of neuron nodes in the output layer is 21, and the number of neuron nodes in the competition layer is 30. The experimental environment is carried out on our own server, the operating system is Ubuntu14.04, the CPU is I5, and the graphics card is GTX1060. Divide the training dataset and verification dataset according to the ratio of $4:1$.

The average correct rate is used to evaluate the performance of image retrieval. The average accuracy is calculated by using the area of the P-R curve, where P represents accuracy, which is the ratio of the number of retrieved positive samples to the number of all retrieved images, and R represents the recovery rate, which is the ratio of the number of retrieved positive samples to the number of all positive samples in the dataset:

$$\text{Average accuracy rate} = \frac{1}{Q}\sum_{i=1}^{Q}\frac{\sum_{k=1}^{n} P_i(k) * rel_i(k)}{R_i},$$

(18)

where Q represents the number of query images, R_i represents the number of dataset images belonging to the same group as the ith query image, $P_i(k)$ represents the accuracy of k and ith query images, $rel_i(k)$ is an indicator, and its value is 1 when the query result of the ith query image belongs to the same group as the kth image, otherwise it is 0, and n represents the total number of all images.

4.2. Experimental Results of Public Datasets.
Because the convolution kernel size and activation function have great influence on the performance of CPN, two different control experiments are conducted to summarize the CPN architecture with the best recognition effect. In the first group, the convolution kernel sizes were set to 3×3, 5×5, 7×7, and 9×9, respectively. There were four groups, and the unified activation function was ReLU. All experimental results are shown in Table 1.

Through the analysis of the experimental results shown in Table 1, it is found that the 3×3 small convolution kernel has a better effect, but with the increase of the convolution

TABLE 1: Network training results for different convolution kernel sizes.

Kernel layer	Average accuracy rate (%)
3×3	94.73
5×5	93.87
7×7	91.44
9×9	89.56

TABLE 2: Network training results for different activation functions.

Activate function layer	Average accuracy rate (%)
Sigmoid	93.09
ReLU	94.73
Tanh	88.67

kernel size, the experimental results are getting worse and worse because the large-sized convolution kernel is obviously too "rough" for the small-sized input image. This leads to the unsatisfactory experimental results of large-scale convolution kernels.

In the second group of experiments, the convolution kernel size was set to 3×3, and the activation function was changed. The network training results of different activation functions are shown in Table 2.

Through the experimental results, it is easy to find that convolution neural network with ReLU activation function has better effect. Figure 4 shows the results retrieved from two datasets by the WTCPN method. The first and second rows represent the images in Holiday dataset, and the third and fourth rows represent the images in Oxford dataset. The word Query below each image represents a query image, while TP represents a related image and FP represents an error image.

4.3. Retrieval Performance of Art Datasets.
Some highly similar artwork images were obtained from open source websites, with a total of more than 3,600 images, which were divided into six categories, with 600 images in each category. These six categories are national clothing (clothing), saddle (saddle), leather hip flask (pnjh), national craft ornaments (gongyi), high hat (gdm), and Ma Touqin (mtq). Divide into training dataset and verification dataset according to the ratio of 4:1. The convolution kernel size is 3×3, and the activation function is ReLU. The retrieval result of the WTCPN method on artwork image dataset is shown in Figure 5.

In order to show the advantages of WTCPN, WTCPN is compared with BP network and CNN network which are widely used. These three network models have basically the same topological structure and all adopt three-tier structure, thus ensuring that the three methods are under the same conditions and are comparable. When the error capacity during training is the same, the number of neurons in the input layer and the output layer of both networks is the same. Accuracy indicates the accuracy in the verification process, while loss indicates that the loss value in the verification process is obvious. Generally, the higher the accuracy value and the lower the loss value, the better the trained neural network is. Comparison of retrieval performance of three network models on artwork image dataset is shown in Figures 6 and 7, respectively.

It can be seen from Figure 6 that the accuracy value of WTCPN is higher than that of BP network and CNN network, which is 91.83%. It can be seen from Figure 7 that
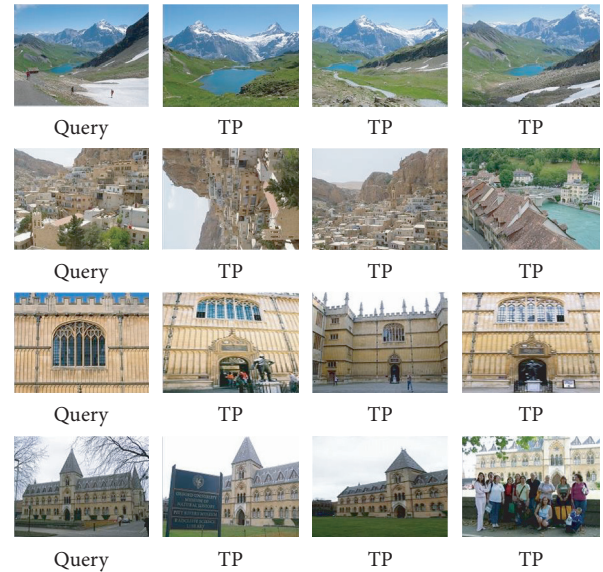
FIGURE 4: Results retrieved by the WTCPN method on two datasets.

FIGURE 5: Retrieval results of the WTCPN method on artwork image dataset.

the loss value of WTCPN is the lowest among the three network models, which is about 0.2. Therefore, it is obvious that the training results of BP network and CNN network are not as good as those of WTCPN, which means that WTCPN can get the best retrieval accuracy.
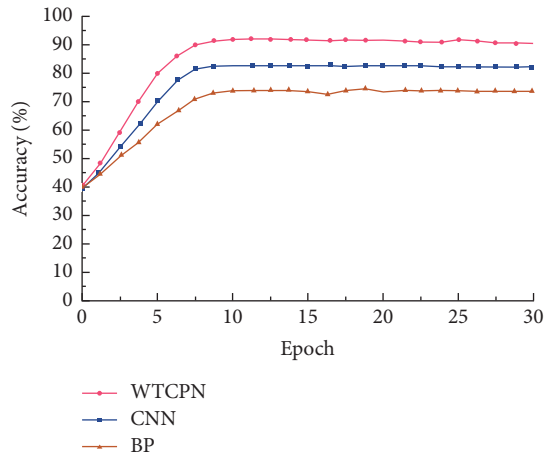
FIGURE 6: Accuracy results.
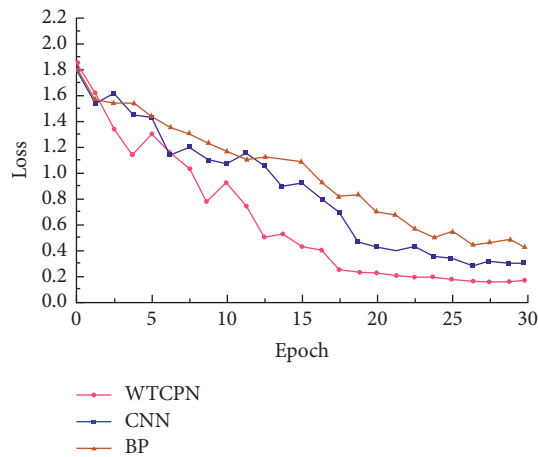
FIGURE 7: Loss results.

5. Conclusion

In this paper, an intelligent retrieval method of artwork image based on wavelet transform and dual propagation neural network (WTCPN) is proposed. By combining the main features of wavelet transform and CPN, the retrieval recognition rate is high. Test results on Oxford and Holiday datasets show that the optimal convolution kernel size of the WTCPN method is 3×3, and the optimal activation function is ReLU. Test results on art datasets show that the WTCPN method has the best retrieval accuracy compared with BP network and CNN network, with an accuracy of 91.83% and a loss value of about 0.2. In the future, aiming at the image retrieval in concurrent environment, based on the WTCPN method, we can increase message queue (MQ), MapReduce, and cache strategy, to improve the access and computing ability in concurrent environment and further improve the art image retrieval system.

Data Availability

The datasets used in the findings of the study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] G. Amato, F. Fa Lchi, and L. Vadicamo, "Image retrieval," *Multimedia Tools and Applications*, vol. 77, no. 5, pp. 5385–5415, 2016.

[2] Z. Xia, X. Wang, L. Zhang, Z. Qin, X. M. Sun, and R. Kui, "A privacy-preserving and copy-deterrence content-based image retrieval scheme in cloud computing," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 11, pp. 2594–2608, 2017.

[3] X.-S. Wei, J.-H. Luo, J. Wu, and Z.-H. Zhou, "Selective convolutional descriptor aggregation for fine-grained image retrieval," *IEEE Transactions on Image Processing*, vol. 26, no. 6, pp. 2868–2881, 2017.

[4] Z. Xia, N. N. Xiong, A. V. Vasilakos, and X. Sun, "EPCBIR: an efficient and privacy-preserving content-based image retrieval scheme in cloud computing," *Information Sciences*, vol. 387, pp. 195–204, 2017.

[5] M. H. Memon, J. P. Li, I. Memon, and Q. Ali, "GEO matching regions: multiple regions of interests using content based image retrieval based on relative locations," *Multimedia Tools and Applications*, vol. 76, no. 14, pp. 1–35, 2017.

[6] Z. Zeng, "A novel local structure descriptor for color image retrieval," *Information*, vol. 7, no. 1, p. 9, 2016.

[7] W. Zhou, S. Newsam, C. Li, and Z. Shao, "Learning low dimensional convolutional neural networks for high-resolution remote sensing image retrieval," *Remote Sensing*, vol. 9, no. 5, p. 489, 2016.

[8] M. Paulin, J. Mairal, M. Douze, Z. Harchaoui, F. Perronnin, and C. Schmid, "Convolutional patch representations for image retrieval: an unsupervised approach," *International Journal of Computer Vision*, vol. 121, no. 1, pp. 1–20, 2016.

[9] Y. Wang, L. Zhang, X. Tong et al., "A three-layered graph-based learning approach for remote sensing image retrieval," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 10, pp. 1–15, 2016.

[10] S. R. Dubey, S. K. Singh, and R. K. Singh, "Local bit-plane decoded pattern: a novel feature descriptor for biomedical image retrieval," *IEEE Journal of Biomedical and Health Informatics*, vol. 20, no. 4, pp. 1139–1147, 2016.

[11] X. Wang, X. Duan, and X. Bai, "Deep sketch feature for cross-domain image retrieval," *Neurocomputing*, vol. 207, pp. 387–397, 2016.

[12] A. K. Dhara, S. Mukhopadhyay, A. Dutta, M. Garg, and N. Khandelwal, "Content-based image retrieval system for pulmonary nodules: assisting radiologists in self-learning and diagnosis of lung cancer," *Journal of Digital Imaging*, vol. 30, no. 1, pp. 63–77, 2017.

[13] L. Zhu, J. Shen, L. Xie, and Z. Cheng, "Unsupervised topic hypergraph hashing for efficient mobile image retrieval," *IEEE Transactions on Cybernetics*, vol. 47, pp. 1–14, 2016.

[14] L. Xie, L. Zhu, P. Pan, and Y. Lu, "Cross-Modal Self-Taught Hashing for large-scale image retrieval," *Signal Processing*, vol. 124, no. 7, pp. 81–92, 2016.

[15] A. Kumar, S. Dyer, J. Kim et al., "Adapting content-based image retrieval techniques for the semantic annotation of medical images," *Computerized Medical Imaging and Graphics*, vol. 49, pp. 37–45, 2016.

[16] Z. Gao, J. Xue, W. Zhou, S. Pang, and Q. Tian, "Democratic diffusion aggregation for image retrieval," *IEEE Transactions on Multimedia*, vol. 18, no. 8, pp. 1661–1674, 2016.

[17] R. Hong, L. Li, J. Cai, D. Tao, M. Wang, and Q. Tian, "Coherent semantic-visual indexing for large-scale image retrieval in the cloud," *IEEE Transactions on Image Processing*, vol. 26, no. 9, pp. 4128–4138, 2017.

[18] X. Li, "Tag relevance fusion for social image retrieval," *Multimedia Systems*, vol. 23, no. 1, pp. 29–40, 2017.

[19] S. R. Dubey, S. K. Singh, and R. K. Singh, "Multichannel decoded local binary patterns for content-based image retrieval," *IEEE Transactions on Image Processing*, vol. 25, no. 9, pp. 4018–4032, 2016.

[20] M. Alshehri, "A content-based image retrieval method using neural network-based prediction technique," *Arabian Journal for Science and Engineering*, vol. 45, no. 4, pp. 2957–2973, 2020.

[21] R. Filip, T. Giorgos, and C. Ondrej, "Fine-tuning CNN image retrieval with no human annotation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, pp. 1655–1668, 2017.

[22] R. Liu, Y. Zhao, S. Wei, and Y. Yang, "Indexing of CNN features for large scale image search," *Pattern Recognition*, vol. 48, no. 10, pp. 2983–2992, 2018.

[23] R. Rajkumar and M. V. Sudhamani, "Image retrieval system using residual neural network in a distributed environment," *International Journal of Recent Technology and Engineering*, vol. 8, no. 6, pp. 2277–3878, 2020.

[24] X. Wang, F. Lee, and Q. Chen, "Similarity-preserving hashing based on deep neural networks for large-scale image retrieval," *Journal of Visual Communication and Image Representation*, vol. 61, no. 5, pp. 260–271, 2019.

[25] M. Verma and B. Raman, "Local tri-directional patterns: a new texture feature descriptor for image retrieval," *Digital Signal Processing*, vol. 51, pp. 62–72, 2016.

[26] Y. Uchida, S. Sakazawa, and S. Satoh, "Image retrieval with Fisher vectors of binary features," *Ite Transactions on Media Technology & Applications*, vol. 4, no. 4, pp. 23–28, 2016.

[27] X. Li, Y. Ye, and M. K. Ng, "MultiVCRank with applications to image retrieval," *IEEE Transactions on Image Processing*, vol. 25, no. 3, pp. 1396–1409, 2016.

[28] F. Zhang, Y. Song, W. Cai et al., "Dictionary pruning with visual word significance for medical image retrieval," *Neurocomputing*, vol. 177, pp. 75–88, 2016.

Research Article

Research on Optimal Matching Scheme of Public Resource Management Based on the Computational Intelligence Model

Linna Li[1] and Renjun Liu ⓘ [2,3]

[1]*School of Design, The University of Melbourne, Melbourne 3052, Australia*
[2]*Wuhan Zhenghua Architectural Design Co., Ltd., Wuhan 430010, China*
[3]*Graduate School of Architecture, Planning, and Preservation, Columbia University, New York, NY 10027, USA*

Correspondence should be addressed to Renjun Liu; rl2783@columbia.edu

The management of public resources means that people's governments at all levels and other public administrative subjects should use certain means and methods, follow certain principles, rationally allocate and utilize public resources, and maximize their functions and benefits. Under the background of limited human resources, this study adheres to the principle of maximizing the benefits of human resources and rationally allocates the use of human resources. In this study, this kind of resource allocation problem is regarded as a linear programming problem by specifying the benefit function, and then, genetic algorithm, ant colony algorithm, and hybrid genetic-ant colony algorithm are used to solve the problem; the cost and time consumption of different algorithms under different scales are evaluated. Finally, it is found that genetic algorithm is superior to ant colony algorithm when the task scale is small and the effect of genetic algorithm is lower than ant colony algorithm with the expansion of task scale, whereas the improved hybrid genetic-ant colony algorithm is better than ordinary algorithm in general.

1. Introduction

With the development of the times and the progress of society, the current situation of unreasonable resource allocation has become more and more serious. The distribution of public resources has also become an issue that cannot be ignored in today's society. From the perspective of computational intelligence, resource allocation is, in the final analysis, a linear programming problem, which generally refers to the allocation of a limited amount of resources to all types of activities. Wu et al. [1] used an adaptive ant colony algorithm to solve the optimization problem of multiresource allocation. After a series of verification experiments, the final results show that the ant colony algorithm has the ability to solve large-scale and multiresource problems, and the comparison results show that the adaptive ant colony algorithm has more advantages than the genetic algorithm. Xu et al. [2] aimed at specific medical resource allocation problems, combined with computational

intelligence to establish a customized medical resource allocation model, and combined the model with genetic algorithm to solve the model. The final experimental results showed that the model has excellent practical value. In order to solve the UAV assignment problem [3], it is transformed into an optimization problem under multiple constraints, and then, genetic algorithm and several other computational intelligence algorithms are used. Because the actual size of UAVs is large, it is a big problem in assigning tasks. The simulation experiment on the scale of the assigned task in this document shows that the genetic algorithm has better performance in smaller-scale problems, but when the problem scale becomes larger, its average calculation time far surpasses other algorithms. Nezhad and Shahbazian [4] converted the orthogonal frequency division multiple access technology (OFGMA) resource allocation problem into a function optimization problem and obtained the optimal solution to the objective function after mixing the traditional genetic algorithm and the particle swarm algorithm. The

research results showed that the hybrid algorithm has an effective improvement in system fairness and throughput compared with ordinary algorithms. Zhang and Chen [5] resourced the logistics personnel in the medical system for allocating various tasks in the hospital, in order to obtain the highest operating efficiency of the hospital and used genetic algorithms to build an automatic task allocation system. Among them, the selection operator adopts the championship mode, and some parent genes with low fitness are also involved in the parent selection to make the offspring more diversified while ensuring the speed of its operation. According to the comparison with the simulated annealing algorithm, its performance is better than that of the simulated annealing algorithm, but it performs poorly in terms of overall computing speed. Some researchers [6, 7] studied the ant colony algorithm to solve the multiobjective resource allocation problem. The literature optimized the pheromone update system and the probability selection form of the ant colony algorithm and finally compared it with the genetic algorithm in solving the employee assignment problem. The results show that the optimized ant colony algorithm is better than the genetic algorithm. Bublitz et al. [8] compared the four important branches of neural network, evolutionary computing (genetic algorithm), swarm intelligence (particle swarm algorithm), and artificial immune system through the research of the current branch algorithms of computational intelligence. Although the genetic algorithm can achieve the global optimal effect, it cannot effectively use the feed forward information, which results in the low efficiency of the algorithm. The particle swarm algorithm is also easy to fall into the local optimum. Li et al. [9] solved the problem of TSP. When optimizing the problem, the genetic algorithm is compared with the ant colony algorithm. The results show that the genetic algorithm is better than the ant colony algorithm when the problem is small, but as the scale increases, the effect is lower than the ant colony algorithm. Forootani et al. [10] studied the use of optimal strategies to solve the problem of resource allocation, combined with the company's dynamic allocation of project funds and processing resources under actual conditions, using the principle of optimization and multistage decision making while combining the chart method to maximize the company's profits. The study shows that the planning problem has practical feasibility in dynamic resource allocation. Tan et al. [11] cited genetic algorithm in the financial field and transformed the problem of project fund allocation into a planning problem. The article mentioned that the income function can be obtained by fitting actual data by neural network. Finally, combined with examples, it proves that genetic algorithm is solving the problem. When solving the problem optimally, it is possible to achieve the global optimum. Ikeda et al. [12] proposed a genetic algorithm based on fuzzy logic in order to optimize the multimedia communication problem. The simulation results show that the proposed framework has good performance and is a promising decision-making method. Lin and Lin [13] uses genetic algorithm for multiobjective optimization and solved some sample optimization problems involving two and three objective functions. It is suitable for both the

operation of existing factories and the design of new factories. A set of nondominated Pareto solutions is obtained for the research problem. Marler and Arora [14] introduced the current investigation of continuous, nonlinear, multiobjective, optimization concepts and methods. It integrated and correlated with seemingly different terms and methods. The characteristics of these methods are summarized. The conclusion reached depends on the type of information provided in the question, user preferences, solution requirements, and software availability. Vrugt et al. [15] introduced a hydrological model. Through the practical experience of research and calibration, an efficient Markov chain Monte Carlo sampler called a multiobjective, shuffled, complex, evolution urban algorithm was proposed, which can solve the problem of hydrological models. Hai [16] mentioned that the current cloud computing used in various fields also requires optimization of resource scheduling and allocation. The ant colony algorithm is introduced to apply to the resource scheduling problem of cloud computing. In order to make the ant colony algorithm better, it is mentioned that constraints can be added. The conditions also give a series of starting points for changing and adding constraints, and it also mentions the use of other algorithms combined with ant colony algorithm to achieve complementarity. Bai et al. [17] analyzed the current research status of cloud computing resource scheduling algorithm, genetic algorithm, and ant colony algorithm. Genetic algorithm has high search efficiency in the early stage, and it is easy to produce local optimal solutions in the later stage. However, the search efficiency of ant colony algorithm is low in the early stage and the later stage. Because of its positive feedback and other characteristics, it is easy to obtain the optimal solution. Through experiments, the improved genetic algorithm and the improved ant colony algorithm are compared, and the effectiveness of the algorithm in the cloud computing resource scheduling process is proved. Through the research on the management of public resources in China, it shows that there are still a series of major problems in the management of public resources [18], including the problem of resource allocation, and the unreasonable resource allocation directly leads to the excess of resources and the lack of resources, thus making society inefficient.

This study first uses genetic algorithm to solve the problem in the second part and uses the ant colony algorithm to solve the problem in the third part. After that, the two algorithms are merged, combining the advantages and disadvantages of the two algorithms to solve the problem again. The final result shows that the genetic-ant colony hybrid algorithm has the ability to solve large-scale and multiresource problems, and the comparison result shows that the adaptive ant colony algorithm has more advantages than the genetic algorithm.

2. Overview of the Problem

The solution of resource allocation problem mainly includes two aspects, namely, problem modeling and solving.

2.1. Modeling of Resource Allocation Problems. From the perspective of the problem itself, as one of the common types of linear programming problems, resource allocation problems are widely used in real life. Therefore, a large number of computational intelligence methods are also widely used in the analysis and solution of such problems, such as genetic algorithm, ant colony optimization algorithm [19], particle optimization algorithm [20], dynamic

programming [21], branch and bound, and decision tree methods.

In real-life public resource management problems, there are many types of resource allocation problems, and the constraints are also different, but the ideas for solving the problems are roughly the same. For the allocation of multiple public resources to different activities, under the premise of limited resources, seek the optimal allocation plan with the greatest benefit. Given decision variables,

$$X_{ij} \begin{cases} = 1, & \text{if } j \text{ quantity of resources is allocated to project } i \\ = 0, & \text{else}, \end{cases}$$

$$\max \sum_{i=1}^{N} \sum_{j=0}^{M} e_{ij} X_{ij}, \left\{ \min \quad \sum_{i=1}^{N} \sum_{j=0}^{M} c_{ij} X_{ij} \text{s.t.} \sum_{i=1}^{N} \sum_{j=0}^{M} X_{ij}, \quad j \le M, \sum_{j=1}^{M} X_{ij} = 1, \quad \forall i, X_{ij} = 0 \text{ or } 1 \forall i, j, \right. \tag{1}$$

where $i (i \in \{1, 2, 3, \ldots, n\})$ represents the task index, $j (j \in \{1, 2, 3, \ldots, n\})$ represents the number of resources, n represents the total number of tasks, m represents the total number of resources, cij represents the cost of resource j occupied by task i, and e_{ij} represents the benefit generated by resource j occupied by task i. The objective function (2) and the objective function (3) indicate that the maximum benefit should be generated, and the least resources should be consumed. Constraint $\sum_{i=1}^{N} \sum_{j=0}^{M} X_{ij}, j \le M$, ensures that the total number of resources does not overflow; constraint $\sum_{j=1}^{M} X_{ij} = 1$ ensures that each task i can only get one resource allocation. Pareto optimal solution is usually the solution of multiobjective programming problem. e_{ij} represents the cost required to allocate resource j to task i, whereas c_{ij} represents the benefit generated by allocating resource j to task i. The purpose of this constraint condition is to maximize the benefit under the minimum cost and conditions. Although there are some changes in e_{ij} and c_{ij} under different problems, their ideas are still practical.

2.2. Solving the Problem of Resource Allocation. In the current background of rapid development of artificial intelligence [22], more and more intelligent algorithms are used to solve such problems. For example, the common dynamic programming, branch and bound, and decision tree methods are used in the study of this article. However, when such algorithms deal with np-hard problems, such as resource allocation, as the scale of the problem becomes larger, the computational difficulty and computational cost also increase exponentially.

On this basis, this article calculates some control methods in the process of artificial intelligence immune system, such as genetic algorithm, ant colony control algorithm, artificial intelligence immune control algorithm, and so on. Such algorithms do not have the characteristics of polynomial time when solving resource allocation problems, so they are often widely used in the process of solving such problems. But for specific problems, it is necessary to

separately describe and model the problem of resource allocation.

3. Method

3.1. Genetic Algorithm. In the genetic algorithm, a population of size n is initialized, and the parental fitness value of each individual in the population is adjusted. Larger individuals are more likely to choose fitness values as individuals with parents. Through crossbreeding, n parents will produce n offspring, the offspring of n parents will mutate with a certain probability, and they will all survive and form a new next generation. In the genetic algorithm, the crossover operator is considered to be a major genetic operator, and the individual is mutated with a low probability.

Adaptive genetic algorithm [23] can be used as a kind of genetic algorithm, and the general way of realization is similar to that of genetic algorithm. First, analyze various variables for initial and evolution and then determine the scale of occurrence of each biological population, the probability of crossover, the probability of occurrence of mutation, and a series of parameters. Then, generate the initial population, take the initial population as the parent, and take the individuals with higher fitness to perform cross-mutation and other operations. After the offspring are produced, the qualified offspring are screened out, and the accompanying individuals with higher fitness are used to replace the fitness of the offspring. Then, it is judged whether there is an individual in the offspring that meets the termination condition of the algorithm. If it cannot be terminated, the offspring will be used as the new parent to repeat the above steps until the termination condition is met, and the optimal solution is found.

3.2. Implementation Steps of Genetic Algorithm. The optimization model that defines its two-dimensional resource allocation problem is as follows:

$$\max b = F(y, x) \begin{cases} \sum_{i=0}^{n} y_i = P_1, & \sum_{i=0}^{n} x_i = P_2, \\ y_i \geq 0, & x_i \geq 0. \end{cases} \quad (2)$$

The resources A and B are allocated to N activities, and the allocation schemes are denoted by y_i and x_i, respectively. The total resource of A is P_1, and the total resource of B is P_2. Then, stipulate the benefit function $F(y, x)$ according to the actual situation. For the multidimensional resource allocation problem, it can also be solved in the above-mentioned way, with only a little improvement. The meaning expressed in formula (2) is to ensure that the number of allocated resources does not exceed the total number of resources, and x_{ij} is a decision variable. If resource j is allocated to task i, it means that the value of resource j in formula (2) is 1 and the sum of allocated resources cannot be greater than the total number M. The two resource allocations are independent and the total amount is fixed.

Step 1: generate the initial population

Use binary encoding to generate a set of chromosomes G_k ($k = 1, 2, \ldots, N$), where N is the size of the group. Because this article is mainly based on the two-dimensional resource allocation problem as an example, the length of the variable code string is $2N$, so the length value of one of the chromosomes is $2N * l$, and the code string length for the variable is l digits, and each l digit. The length of the code string is equivalent to a decimal integer b_i ($i = 1, 2, 3, n$). What is different is that G_k is the first-generation population.

Step 2: feasibility process

The chromosome code is mapped to the position to meet the feasibility of the problem. Normalize b_i to b'_i, let $x_i = ab'_i$, at this time all the vectors Z_p ($p = 1, 2, 3 \ldots, n$) composed of x_i can satisfy the constraints. That is a vector corresponding to the same chromosome produces a feasible solution.

Step 3: calculate fitness

Take a new objective fitness function as an optimal fitness function. For each chromosome in different population pedigrees of the same generation of chromosome G_k, combine all the corresponding optimal feasible solutions Z_k that we require above substituting an objective fitness function, you can directly and accurately find their optimal fitness function value.

$$f_x = \sum_{i=1}^{n} g_i(x_i). \quad (3)$$

The higher the fitness value, which shows that their corresponding optimal feasible solutions are closer to the optimal solutions.

Step 4: chromosome selection

Copy the chromosomes with the highest fitness in each generation population to the next generation population. For the cumulative population up to the next generation, first, the system will automatically generate a random probability number between 0 and f'_k, and count each random value, starting from the first chromosome G_1 for random accumulation. When the value of accumulation and random adaptability is greater than the random probability number generated by the second-generation accumulation, the system will automatically suspend the accumulation and finally add the accumulated chromosomes that have been replaced. This may also be each chromosome that needs to be randomly selected as described in this article. Finally, the random probability of each cumulative chromosome being selected and the cumulative fitness value are kept in direct proportion, and then, the unselected chromosomes are maintained. Each chromosome is eliminated.

Step 5: chromosome hybridization

After the new population reproduces, select "individual pairs" according to the probabilities P_C of various crosses and perform a single-point cross (a total of $n/2$ times). In this study, P_C is 0.7. Single-point cross hybridization is adopted, and the method of cross-exchange points is randomly selected, and the feasibility optimization of all the obtained chromosome progeny data can effectively ensure that the chromosomes after cross hybridization in the population correspond to each other and the feasibility optimization. Calculate the fitness value that the offspring needs and then enter the new population by replacing other fathers and sons.

Step 6: chromosome variation

In order to effectively maintain the evolutionary diversity of various groups on biological chromosomes and maintain the "evolution" of populations on biological chromosomes, this article randomly selected individual populations on some biological chromosomes for adaptive mutation, considering that these are some of the biological variant chromosomes, i.e., the basic feature, the probability of mutation is not very large.

$$P_m = \frac{N_m}{N}. \quad (4)$$

This article first uses random $P_m = 0.05$ and randomly selects some sites that adapt to mutation as d ($1 \leq d \leq 1 * n$). The adaptive mutation and infeasibility of these mutated biological chromosome populations can be effectively done to ensure that the solutions corresponding to these mutated chromosomes are feasible solutions and then sequentially calculate the values, and values that are adapted on these chromosomes may enter the varieties.

Step 7: judging the conditions for stopping evolution

If during the evolution process from $m - s$ generation to $m + s$ generation, the highest fitness value in the chromosomes of each generation remains unchanged, you can stop its evolution; otherwise, return to Step 4.

3.3. Adaptive Ant Colony Algorithm. The resource allocation problem can be regarded as a boxing problem. The public resource v to be allocated is regarded as an item, and each activity p is regarded as a box. This article will be analogous to the idea of solving the boxing problem to deal with the optimization and matching problem of public resource management. Ant colony algorithm has the characteristics of distributed calculation and positive feedback of information and heuristic search, so it has become one of the most classic algorithms for solving box packing problems. Based on the above characteristics, compared with other algorithms, it can more effectively converge to the optimal solution when solving the packing problem. Therefore, this article will select the ant colony algorithm to deal with the optimization matching problem of public resource management. This algorithm is completed by simulating the entire process of ant foraging. First, the ant-bee randomly selects a feasible path. When the ant-bee reaches the expected planned path, it stops and starts to calculate the adaptability of this path. The ant-bee establishes a corresponding information on the path according to its own adaptability. Finally, in order to concentrate the entire ant clustering on a higher fitness approach and find the best solution as quickly as possible, it is necessary to update the pheromone and select actions. For example, the use of honeycomb swarm algorithm for task scheduling. First, the initialized pheromone and heuristic information and other related parameters. Second, each ant receives the assigned task, and when the assigned task is successfully assigned to each ant's resource, it will be recorded in its own taboo list. Third, for the following tasks that have not been accessed, repeat the above steps until the task is no longer added to table, and all tasks are considered to be completely scheduled. The goal of adaptive ant colony algorithm is to find the optimal solution of optimization problem. Ants often choose a path with a large amount of information in the process of traveling, However, when many ants choose the same path, the amount of information in the path will suddenly increase, which makes many ants concentrate on a certain path, resulting in a blockage and stagnation phenomenon, which is easy to lead to prematurity and local convergence when using ant colony algorithm to solve problems. This is the root cause of the shortcomings of ant colony algorithm. Therefore, we modify the selection strategy, adopt the combination of deterministic selection and random selection, and dynamically adjust the probability of deterministic selection in the search process. When the evolution direction has been basically determined after a certain algebra, the information amount on the path is dynamically adjusted. The shortcoming of the basic ant colony algorithm can be effectively overcome by narrowing the information gap between the best path and the worst path and appropriately increasing the probability of random selection so as to search more completely less than one pair of solution spaces.

3.4. Pheromone Function Definition. The whole process of simulating a path in which ants form a path is to allocate resources to each task reasonably. Therefore, the selection of each task and the update of pheromone in the heuristic information and algorithms are greatly interconnected. $t(T_i, R_j)$, which means that a task is allocated to the information element required by a certain resource. In the initialization phase, the pheromone is calculated by the following formula:

$$t_0 = \frac{1}{[n * (p'(S_0) + w(s_0))]}, \tag{5}$$

where n represents the number of activities, the solution S_0 obtained using the descending first adaptation algorithm to solve the boxing problem is marked as S_0, and $W(S_0)$ the amount of waste of resources generated by this solution of S_0. $p'(S_0)$ is the result of normalizing the energy consumption of this solution of S_0. It can be calculated by the following formula:

$$p'(S_0) = \sum_{j=1}^{m} \left(\frac{p_j}{p_j^{\max}} \right), \tag{6}$$

where $p_j^{\max}$ represents the amount of resources consumed by task j.

3.5. Transition Probability of Behavioral Choice. The following methods are used to select its behavior so that we can obtain the optimal solution: A hypothesis is that in the process of the k-th replacement, the resource set marked by the deadline and budget constraints required by this task will be able to be met into $g_k(T_i, R_j)$. Therefore, in the k-th iteration, the resource R_j that task T_i needs to select is the probability of completing the scheduling work, which is given by the following formula:

$$P_{k(T_i, R_j)} = \begin{cases} \dfrac{[t(T_i, R_j)]^a \cdot [\eta(T_i, R_j)]^\beta}{\sum_{h \in g_k(T_i, R_j)} [t(T_i, R_j)]^a \cdot [\eta(T_i, R_j)]^\beta} \end{cases} \tag{7}$$

In equation (7), $t(T_i, R_j)$ indicates that the task T_i is allocated to a pheromone of the resource R_j in a certain path and $\eta(T_i, R_j)$ indicates the heuristic information corresponding to the path. It can be set as the reciprocal of the starting time of a task assigned to the resource. The parameters α and β are, respectively, regarded as the weighting factors occupied by heuristic information and other pheromones. They both represent the relative importance of heuristic information and other pheromones, and they play an important regulatory role.

3.6. Fitness Function. For example, when an ant passes through all active tasks, a path has been formed, and this path is a feasible solution to the problem. In order to effectively avoid falling into a local optimal state, to ensure the accuracy and quality of the obtained solution, and to ensure that the obtained solution is the global optimal solution as much as possible and in order to accurately evaluate the accuracy and pros and cons of the obtained solution, we need to choose a function of fitness. This article mainly adopts the research method of defining the fitness function

based on the theoretical basis of the optimization model problem. Taking the scheduling optimization model established in the article as the theoretical basis, two scheduling goals are analyzed and designed, and the cost is reduced to a minimum. Therefore, the fitness function of the following formula is the evaluation value.

$$\text{Fit}(x) = \gamma e^{(-F(x))} + \delta e^{(-B(x))}, \tag{8}$$

where γ and δ in represent the two weighting factors of operating cost and resource waste, respectively. The two conditional formulas are $\gamma < \delta$, γ and $\delta \in (0, 1)$, and $F(x)$ and $B(x)$ are the objective functions used to represent operating cost and resource waste, respectively. It can be concluded from the above summary that the lower the value of the fitness function, the less waste caused by operating costs and human resources.

3.7. Pheromone Update. Ants cruise the path according to the pheromone concentration on the path. If a path is highly adaptable, the pheromone concentration of this path should be strengthened so that more ants can find this path. Therefore, it is necessary to update the pheromone at each point on the path.

The update rule is as follows:

$$t(T_i, R_j) = (1 - \rho)\Delta t(T_i, R_j) + \Delta t(T_i, R_j). \tag{9}$$

In formula (9), ρ represents the speed and degree of pheromone disappearing in the air, that is, the evaporation factor, and $t(T_i, R_j)$ represents the number and increment of pheromone in the air. The higher the adaptability of the path, the more the corresponding increase in diffusion, which is a positive feedback adjustment mechanism.

$$\Delta t(T_i, R_j) = Q \cdot \left(\gamma e^{(-F(x))} + \delta e^{(-B(x))}\right), t(T_i, R_j) \in \text{path}. \tag{10}$$

Q in equation (10) is a constant, and its initial value is 10. Among them, the lower the value of $f(x)$ and $b(x)$, the higher the value increment of the pheromone. Based on the positive feedback mechanism, the better solution in the path may gradually increase with the further update of the pheromone, whereas the worse solution may also gradually weaken with the further update of the pheromone and finally be discarded. After several consecutive years of iteration, more and more ants will move towards an optimal path. In order to prevent the solution, we can only be optimized in one part; the evaporative factor of pheromone can play an important regulatory role.

3.8. Ant Colony Algorithm Process. The flow chart of the ant colony algorithm is shown in Figure 1.

(1) Parameter initialization: the main functions include parameter initialization of application pheromone types, initialization of heuristic application information, and optimization of parameter initialization value information of various factors such as
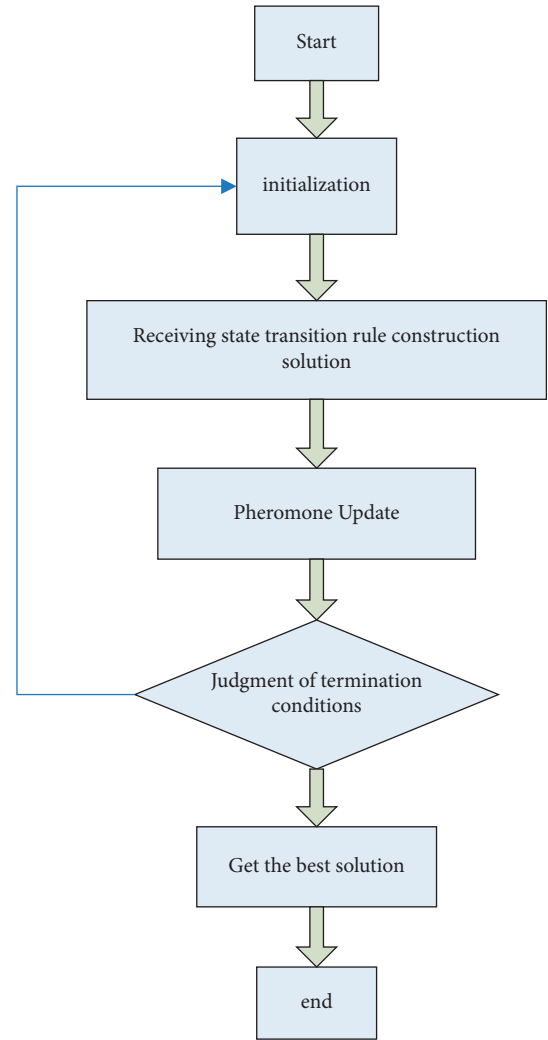
FIGURE 1: Ant colony algorithm flow chart.

information population distribution and pheromone biological volatilization rate.

(2) The ant starts the first cycle: $k = k + 1$.

(3) Randomly place several ants at different starting points and establish a search space.

(4) The probability of each ant moving to the next node is calculated separately. The ant can move on its own according to the results of these numerical calculations and finally reach its corresponding new node.

(5) During this period, when an ant moves to the path to a new node, the information elements in all the paths it passes are updated, and the content in the taboo table will be adjusted accordingly.

(6) Repeat steps 3 to 5, until each individual in the entire ant colony has found a feasible path.

(7) For an output result, output and optimize it according to requirements, including related indicators, such as running time, convergence, and the number of iterations.

4. Improved Genetic-Ant Colony Algorithm Fusion

At present, most scholars [24] optimize the ant colony algorithm mainly in two aspects: one is to add constraints to the ant colony algorithm to overcome the shortcomings of the original algorithm; secondly, to integrate other algorithms with the ant colony algorithm. Complementary, use the advantages of other algorithms to improve the shortcomings of the ant colony algorithm. In order to effectively avoid the shortcomings of the two algorithms, this study adopts the second method of algorithm optimization, which is to combine the two algorithms to form complementary advantages. Hybrid algorithm not only synthesizes the advantages of the two algorithms but also inherits the shortcomings of the two algorithms to a certain extent, so we still need to adapt to local conditions and choose the appropriate algorithm in the actual production process. When the amount of data is small, a single genetic algorithm can be directly used to solve it.

4.1. Improved Dynamic Fusion Strategy of Genetic-Ant Colony Algorithm. In this study, the optimal solution obtained by the genetic algorithm is transformed into the initial value allocation strategy of the ant colony algorithm, and the specific operations are as follows:

(1) First, use the characteristics of genetic algorithm to generate a series of optimized solutions for resource scheduling problems. In this evolution process, the evolution speed of each generation of population in the genetic algorithm is calculated. Within the range of satisfying the maximum number of runs, if the evolution speed of the population for n consecutive generations is less than the preset minimum value, the genetic algorithm is terminated.

(2) Enter the ant colony algorithm after meeting the fusion conditions. First, the first 10% of the optimized solution of genetic algorithm is converted into the pheromone value of the initial path of the ant colony algorithm, and then, the optimal scheduling plan is calculated according to the characteristics of the ant colony algorithm. According to the above operation process, it can be seen that the main difficulty of fusion is to determine the conditions for terminating the genetic algorithm. After reading a large number of documents, the evolution speed of the two algorithms over time is summarized. The evolution speed of genetic-ant colony algorithm changes with time as shown in Figure 2.

It can be seen from the figure that the evolution rate of the genetic algorithm in the initial stage, that is, from the beginning to the t_b stage, is very high, but after the time point t_b, its evolution rate decreases with the increase in time, until the time point t_c drops to a certain level degree, which is relatively stable. The evolution rate of the ant colony algorithm from the beginning to the t_b stage is very low, but after the time point t_b, its evolution rate increases with the
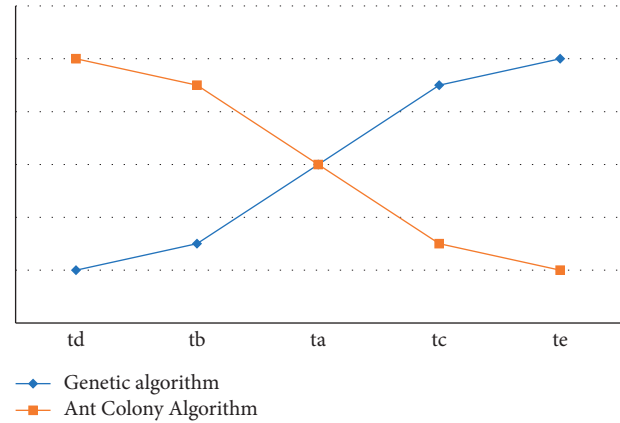
FIGURE 2: Evolutionary speed of genetic-ant colony algorithm over time.

increase in time, until the time point t_c, where the evolution rate has risen to a certain degree and is relatively stable. In the evolution rate-time curve, the most special time point is t_a. Before this time point, the evolution rate of genetic algorithm is higher than that of ant colony algorithm. At this time, the evolution rate of the two algorithms is the same. After that, the evolution rate of ant colony algorithm is higher than that of genetic algorithm. Therefore, the dynamic fusion of genetic algorithm and ant colony algorithm at time t_a is the most ideal, that is, the best results can be obtained in all aspects of time and efficiency. But in the execution process, it is difficult to accurately determine the time point t_a. Therefore, when we are improving the algorithm, as long as we determine that the evolution rate is switched during the $t_b \sim t_c$ time period, the algorithm has been improved to a large extent.

4.2. Improved Genetic-Ant Colony Algorithm Fusion Parameter Setting. It can be seen from the above that the evolutionary speed of genetic algorithm and ant colony algorithm change over time completely opposite. Therefore, the combination of these two algorithms does not jump to the ant colony algorithm after the genetic algorithm is fully executed but enters the ant colony algorithm when the combination conditions are met. In order to switch the evolution rate of the algorithm within $t_b \sim t_c$, the parameters of the two algorithms need to be set as follows:

Step 1: suppose that the maximum number of iterations of the genetic algorithm is g_{max}, the current number of iterations is g_{num}, the minimum number of iterations is g_{min}, the minimum evolution rate is g_{rate}, and the population of offspring evolves rate is g_{pop}.

Step 2: if g_{num} is less than g_{max}, then compare g_{pop} and g_{rate}; if the evolution rate of successive generations and the evolution rate of the progeny population g_{pop} is less than g_{rate}, the genetic algorithm is ended, and the ant colony algorithm is run.

Step 3: as a key parameter of algorithm fusion, the main significance of evolution rate is to obtain the current

evolution degree of genetic algorithm. Therefore, evolution rate is the comparison between the fitness value of the latest individual and the previous evolution value, as shown in the following equation:

$$R1 = \left(\frac{(f(1) - f(i-1))}{(f(i-1))} \right). \tag{11}$$

The optimal solution of the genetic algorithm is transformed into the path pheromone value t_G of the ant colony algorithm, so that the pheromone constant is t_C. In this article, the constant t_G is represented by the path value in MMAS as $t_{\min}$, and t_G is the path value passed plus 2, then the initial pheromone value of the ant colony algorithm is given as

$$t_s = t_c + t_G. \tag{12}$$

Step 4: the method of conversion between algorithms is as follows: First, each individual in the top 10% of the genetic algorithm can be represented as a task T_i executed on the resource node P_j, which is equivalent to selecting P_j for the task T_i of the ant and then changing the resource P_j. The pheromone value can save 70% of the time wasted by the ant colony algorithm at the beginning due to the lack of pheromone.

4.3. Implementation Steps of the Improved Genetic-Ant Colony Algorithm. Genetic algorithm stage:

Step 1: corresponding parameter settings are minimum number of iterations $g_{\min}$, maximum number of iterations $g_{\max}$, minimum evolution rate g_{gate}, progeny population evolution rate g_{pop}, and population iteration number g_{num}. The initial population size is the initial population G of Popsize.

Step 2: at the beginning of the algorithm, the chromosome is encoded with decimal real numbers.

Step 3: the fitness function is defined as the average completion time of the task and the average load weighted sum function $f(j)$ of the virtual machine, and the genetic operation is started.

Step 4: use the fitness value ratio selection method (the selection strategy of the selection operator $p(i)$ to filter out the high fitness value.

The parent individual waits to perform the crossover operation.

Step 5: calculate the adaptive crossover probability P_c, select the parental individuals for single-point crosses to produce a given number of new individuals.

Step 6: calculate the adaptive mutation probability P_c and use the method of randomly generating mutation positions to perform mutation operations on the new individuals obtained. The group $G(g)$ is genetically operated to obtain the next generation group $G(g+1)$.

Step 7: compare the new generation with the parents and select outstanding individuals as the final offspring.

Step 8:

$$g_{num} = g_{num} + 1 \text{ if (calculate the evolution rate } R(I)). \tag{13}$$

Step 9: determine whether the fusion condition is met: If $g_{num} > g_{max}$, or the evolution rate of consecutive g_{die} generations is less than g_{rate}, then end the genetic algorithm, run the ant colony algorithm, and go to Step 10. Otherwise, go to Step 3.

Ant colony algorithm stage:

Step 10: convert the top 10% individuals with the best optimization solution in the genetic algorithm into the path value t_G of the ant colony algorithm, get the value of t_S, set the ant colony algorithm parameters, and set the number of ants to the number of tasks m and the current iteration number NC = 0, the maximum number of iterations NC_{max}, and the taboo table $tabu_k$ is empty.

Step 11: m ants are randomly placed in n resource nodes. For each ant, use the probability $P_{ij}^k(t)$ to select the resource node to perform the next task.

Step12: when an ant completes this task, add the node to $tabu_k$, m minus 1, and record the current score.

Configure the strategy and update the pheromone value of the resource node locally.

Step 13: when all the ants complete the task, the number of iterations NC increases by 1, the optimal allocation strategy is screened out, and the pheromone value of the resource node is updated globally.

Step 14: $(NC > NC_{max})$\{break the cycle\}else go to Step 11.

5. Comparative Analysis of Simulation Experiments

After understanding the specific implementation steps of the three algorithms, this article applies the three algorithms to the actual problem solving and uses experiments to verify the efficiency of the three algorithms in solving the resource allocation problem.

5.1. Problem Description. This example is used to deal with the problem of hospital medical resource allocation. For example, what kind of medical activities should be allocated, what kind of medical resources, and what kind of deployment methods can maximize the efficiency of medical activities. The medical resources here mainly refer to medical personnel. After instantiating the problem, this article assumes that there are ten medical personnel and four medical activities at the same time. Aiming at these four medical activities, this article needs to obtain the optimal medical personnel deployment plan.

A large number of experimental research results show that different parameter values have a great impact on the accuracy of the algorithm. In this study, a large number of experimental results are used to select the more accurate parameter value configuration as follows:
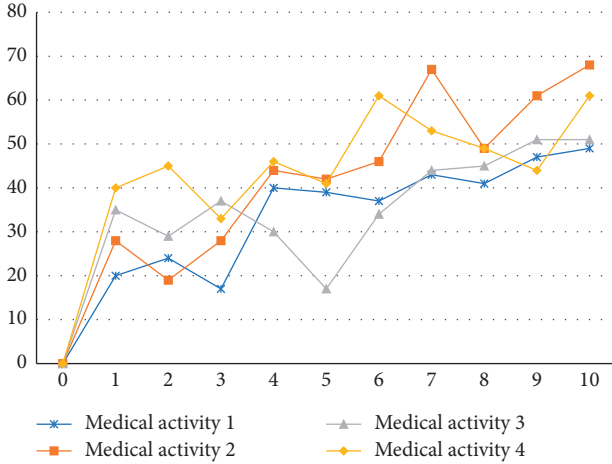
FIGURE 3: Working cost of medical staff.
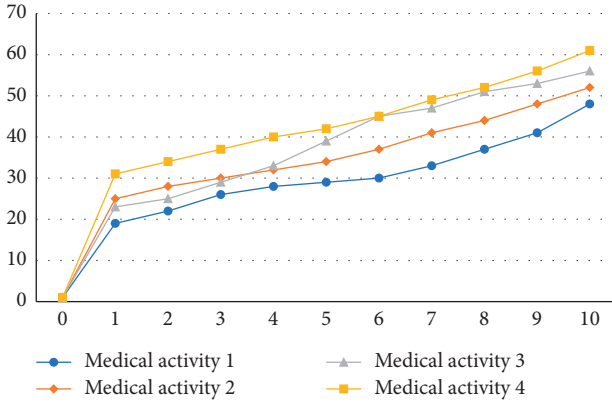
FIGURE 4: Work efficiency of medical staff.

TABLE 1: Nondominated solutions of the genetic algorithm.

Solution	1	2	3	4	Cost	Efficiency
1	3	2	1	4	101	115
2	0	2	6	2	99	106
3	3	2	5	0	87	92
4	3	1	6	0	83	88
5	1	1	6	2	107	121
6	0	1	6	3	96	105

Genetic algorithm: $P_C = 0.7$, $P_m = 0.05$.

Ant colony algorithm: $\alpha = 0.5$, $\rho = 0.3$, $t_0 = 0.01$

Hybrid algorithm: $g_{num} > 1$, $g_{num} > g_{max}$, $NC > NC_{max}$.

Figures 3 and 4 show the work cost and work efficiency of four different medical activities corresponding to different numbers of medical personnel. The cost and efficiency here are an estimate.

5.2. Comparison of Experimental Results. Tables 1–3 show the solutions of the three algorithms to solve the problem, respectively. Figures 5 and 6 show the comparison of the cost and time consumed by the three algorithms to solve the problem when the computational scale increases.

TABLE 2: Nondominated solution of ant colony algorithm.

Solution	1	2	3	4	Cost	Efficiency
1	3	2	1	3	88	112
2	1	2	6	0	77	91
3	3	2	0	0	76	53
4	1	1	6	1	102	118

TABLE 3: Nondominated solutions of the hybrid algorithm.

Solution	1	2	3	4	Cost	Efficiency
1	1	3	2	3	74	98
2	1	6	0	2	63	77
3	3	0	0	2	62	39
4	1	6	1	1	88	104
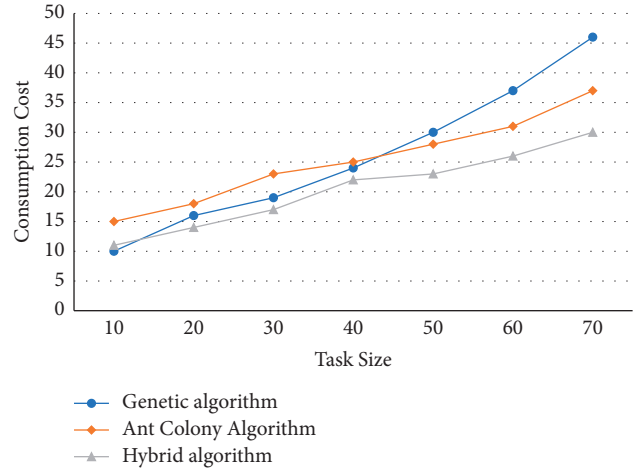
FIGURE 5: Diagram of the cost consumption of various algorithms.
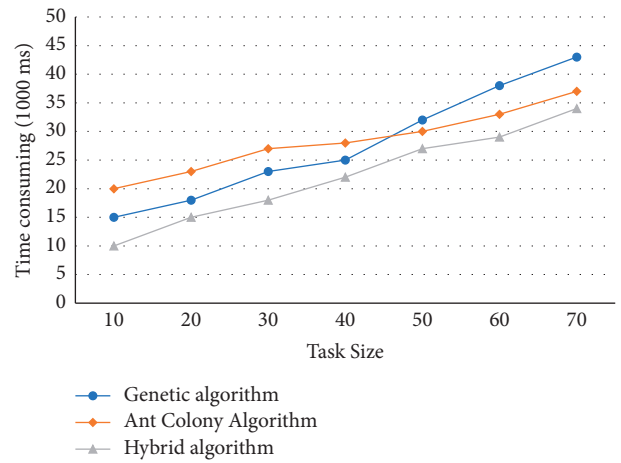
FIGURE 6: Diagram of the time consumption of various algorithms.

6. Conclusion

Both genetic algorithm and ant colony algorithm are more efficient optimization search algorithms. In solving resource

allocation problems, they have their own strengths and their own shortcomings. In this deployment problem, when the task size is below 45, the genetic algorithm can have better computing power; when the task size gradually becomes larger, the genetic algorithm appears to be a little bit powerless. At this time, the ant colony algorithm needs to be used to solve the problem. As the task size increases, the genetic algorithm is more affected than the ant colony algorithm. The hybrid algorithm can take into account the advantages of both and achieve better results overall.

Data Availability

The experimental data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest regarding this work.

Acknowledgments

References

[1] Z. Wu, L. Zhang, Y. Wang, and K. Wang, "Optimization for multi-resource allocation and leveling based on a self-adaptive ant colony algorithm," in *Proceedings of the International Conference on Computational Intelligence and Security*, IEEE Computer Society, Suzhou, China, December 2008.

[2] Y. Xu, S. Liu, and B. Wang, "Research on computational intelligence in medical resource allocation based on mass customization," *Journal of Universal Computer Science*, vol. 24, no. 6, pp. 753–774, 2018.

[3] H. A. Le Thi, D. M. Nguyen, and T. Pham Dinh, "Globally solving a nonlinear UAV task assignment problem by stochastic and deterministic optimization approaches," *Optimization Letters*, vol. 6, no. 2, pp. 315–329, 2012.

[4] Y. M. Nezhad and M. Shahbazian, "Hybrid predictive control design based on particle swarm optimization and genetic algorithm," in *Proceedings of the 2011 3rd International Conference on Computer Research and Development*, pp. 129–134, Shanghai, China, March 2011.

[5] X. Zhang and X. Chen, "UAV task allocation based on clone selection algorithm," *Wireless Communications and Mobile Computing*, vol. 2021, Article ID 5518927, 9 pages, 2021.

[6] M. Vaibhav, S. Lakshminarayanan, R. Gade Pandu, and L. Titus, "A novel optimal experiment design technique based on multi-objective optimization and its application for toxin kinetics model of hemodialysis patients," *Computer Aided Chemical Engineering*, vol. 30, pp. 1362–1366, 2012.

[7] S. K. Chaharsooghi and A. H. Meimand Kermani, "An effective ant colony optimization algorithm (ACO) for multi-objective resource allocation problem (MORAP)," *Applied Mathematics and Computation*, vol. 200, no. 1, pp. 167–177, 2008.

[8] F. M. Bublitz, A. Oetomo, K. S. Sahu et al., "Disruptive technologies for environment and health research: an overview of artificial intelligence, blockchain, and internet of things," *International Journal of Environmental Research and Public Health*, vol. 16, no. 20, pp. 1–24, 2019.

[9] K. Li, L. Kang, W. Zhang, and B. Li, "Comparative analysis of genetic algorithm and ant colony algorithm on solving traveling salesman problem," in *Proceedings of the IEEE International Workshop on Semantic Computing and Systems*, pp. 72–75, IEEE, Huangshan, China, July 2008.

[10] A. Forootani, R. Iervolino, M. Tipaldi, and J. Neilson, "Approximate dynamic programming for stochastic resource allocation problems," *IEEE/CAA Journal of Automatica Sinica*, vol. 7, no. 4, pp. 975–990, 2020.

[11] B. Tan, H. Ma, and Y. Mei, "A group genetic algorithm for resource allocation in container-based clouds," in *Evolutionary Computation in Combinatorial Optimization. EvoCOP.*, L. Paquete and C. Zarges, Eds., Springer, New York, NY, USA, 2020.

[12] M. Ikeda, L. Barolli, A. Koyama, A. Durresi, G. De Marco, and J. Iwashige, "Performance evaluation of an intelligent CAC and routing framework for multimedia applications in broadband networks," *Journal of Computer and System Sciences*, vol. 72, no. 7, pp. 1183–1200, 2006.

[13] C.-H. Lin and P.-L. Lin, "Improving the non-dominated sorting genetic algorithm using a gene-therapy method for multi-objective optimization," *Journal of Computational Science*, vol. 5, no. 2, pp. 170–183, 2014.

[14] R. T. Marler and J. S. Arora, "Survey of multi-objective optimization methods for engineering," *Structural and Multidisciplinary Optimization*, vol. 26, no. 6, pp. 369–395, 2004.

[15] J. A. Vrugt, H. V. Gupta, L. A. Bastidas, W. Bouten, and S. Sorooshian, "Effective and efficient algorithm for multi-objective optimization of hydrologic models," *Water Resources Research*, vol. 39, no. 8, 2003.

[16] Y. Hai, "Improved ant colony algorithm based on PSO and its application on cloud computing resource scheduling," *Advanced Materials Research*, vol. 989-994, pp. 2192–2195, 2014.

[17] X. Bai, H. W. Kang, Q. Y. Chen, X. P. Sun, Y. Shen, and H. Zhou, "Research of the emergence of virtual organizations based on cloud computing environment," *Applied Mechanics and Materials*, vol. 644-650, pp. 2290–2294, 2014.

[18] W. Wang, J. Zhao, and J. Huang, "Improved ant colony genetic algorithm for solving traveling salesman problem," *Journal of Physics: Conference Series*, vol. 1693, no. 1, pp. 1–7, 2020.

[19] M. Vafaei, A. Khademzadeh, and M. A. Pourmina, "A new QoS adaptive multi-path routing for video streaming in urban VANETs integrating ant colony optimization algorithm and fuzzy logic," *Wireless Personal Communications*, vol. 118, no. 6, pp. 1–34, 2021.

[20] G. Chen, L. Wang, M. Alam, and M. Elhoseny, "Intelligent group prediction algorithm of GPS trajectory based on vehicle communication," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 7, pp. 3987–3996, 2020.

[21] D. A. Zhuang, C. L. Xiaoyue, C. Xi, and M. Xiaolei, "Joint optimization of scheduling and capacity for mixed traffic with autonomous and human-driven buses: a dynamic programming approach," *Transportation Research Part C: Emerging Technologies*, vol. 114, pp. 598–619, 2020.

[22] J. Li and J.-S. Huang, "Dimensions of artificial intelligence anxiety based on the integrated fear acquisition theory," *Technology in Society*, vol. 63, Article ID 101410, 2020.

[23] H. Chen, S. Chen, M. Li, and J. Chen, "Optimal operation of integrated energy system based on exergy analysis and adaptive genetic algorithm," *IEEE Access*, vol. 8, p. 1, 2020.

[24] G. Chen and S. Li, "Network on chip for enterprise information management and integration in intelligent physical systems," *Enterprise Information Systems*, vol. 15, no. 7, pp. 935–950, 2021.

Retraction

Retracted: Application Analysis of 3D Printing Technology in Design Field: Taking Shoe Design as an Example

Scientific Programming

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope

(2) Discrepancies in the description of the research reported

(3) Discrepancies between the availability of data and the research described

(4) Inappropriate citations

(5) Incoherent, meaningless and/or irrelevant content included in the article

(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] T. Gong and L. Kang, "Application Analysis of 3D Printing Technology in Design Field: Taking Shoe Design as an Example," *Scientific Programming*, vol. 2021, Article ID 5662460, 8 pages, 2021.

Research Article

Application Analysis of 3D Printing Technology in Design Field: Taking Shoe Design as an Example

Taisheng Gong[1,2] and Luping Kang [1]

[1]College of Art and Design, Shaanxi University of Science and Technology, Xi'an 710021, Shaanxi, China
[2]National Demonstration Center for Experimental Light Chemistry Engineering Education,
 Shaanxi University of Science and Technology, Xi'an 710021, Shaanxi, China

Correspondence should be addressed to Luping Kang; 1910033@sust.edu.cn

The development of 3D technology has brought opportunities and challenges to the footwear industry because people's living standards have been improving due to economic development, and people have higher requirements for the design of fashion shoes and boots. The use of 3D printing technology in the design of fashion shoes and boots can enable faster molding of footwear products, enrich the shape of footwear, and meet people's aesthetic needs for fashion shoes. In this paper, we firstly describe the advantages of 3D printing molding shoe models and then use 3D laser foot scanning and measuring instrument to scan and obtain the cloud map of shoe lasts and foot-related data. Secondly, we realize the digital management of shoe lasts by establishing the database of solid models. On this basis, we apply the technology of least squares support vector machine improvement algorithm to make partial modifications to the lasts according to the need to leave the appropriate helper and foot lining degrees. Finally, based on this technology, we apply the least squares support vector machine improvement algorithm technology to make partial modifications to the last shape according to the need for appropriate helper and foot lining degrees to realize the process of shoe last redesign. The last model and role model can also be produced by 3D printing technology, which can be used as a mold to facilitate the processing of the cut last 2D unfolding material for shoe and boot production later. Therefore, the article studies and analyzes the design and manufacturing process of digital shoe lasts based on individual foot shape and uses CAD/CAM technology to realize the digitalization of shoe last design.

1. Introduction

With the continuous improvement of people's quality of life, people's requirements for shoes and boots are getting higher and higher, and people are also putting forward higher requirements for the design of fashionable shoes [1]. With the development of 3D printing technology, people recognize the importance of 3D printing technology, and the level of recognition is also increasing. 3D printing technology can play an important role in the field of footwear [2].

At present, the shoe lasts made in large quantities by shoe-making enterprises and factories are designed and made according to the relevant shoe last models and design standards. The relevant data given in these last design standards are mainly formulated according to statistical laws

and experience [3, 4, 5]. This results in the same model of shoes and boots made from the same model of shoe lasts. Some people wear them to fit while others wear them not to fit. Long-term wearing of shoes that do not fit the physiological condition of the human foot will not only produce a sense of discomfort but also seriously affect the development and formation of the bones of the human foot, and for people with foot diseases, unsuitable shoes and boots may even worsen their symptoms and bring serious consequences [6]. For athletes and other special groups, unsuitable shoes may significantly affect athletic performance. In addition, shoes and boots are gradually becoming a fashion trend, which is the presenting body and carrier of beauty. When designing shoe and boot styles, it is necessary to meet not only their comfort but also the demand for aesthetic

presentation when they are produced as daily-use products [7, 8], but there is still a lack of ways to quantify the beauty in the current shoe last production, which makes it objectively difficult to achieve the ideal design. With the gradual improvement of people's living standard, there are correspondingly higher requirements for shoes and boots. Fashion and personalization have become a development trend, and the reasonable use of 3D technology can make them more perfect.

3D printing technology originated in the United States in the 1980s. At the beginning of this century, with the development of mainstream technologies such as stereo light-curing molding, selective laser sintering, and fused filament deposition modeling, 3D printing technology gradually matured [9, 10]. In 2007, Objet Geometries launched the Connex500TM system that supports the simultaneous printing of multiple model materials. Objet Geometries launched the Connex500TM system, which supports the printing of multiple model materials at the same time. Currently, 3DSystems, Stratasys, HP Inc., MakerBot, EOS, SLM Solutions in Germany, and Objet in Israel are at the forefront of this field [11, 12]. In developed countries such as Europe and the United States, 3D printing technology has been widely applied in various fields; with the development of 3D printing technology in China, it has been applied in various fields, such as biology, medicine, national defense, aerospace, education, culture and creativity, architecture, clothing, food, and has shown superior performance and good market prospects [13].

Compared with traditional production methods, 3D printing can significantly shorten the production cycle, which is conducive to the convenient and efficient development of products. However, on the other side of the coin, we should also see that after the rapid development of 3D printing in the past few years and market pursuit, a series of problems have emerged, such as high costs, bottlenecks in mass production, restrictive technology and materials, and the proliferation of capital and market hype, all of which have yet to return to a rational and pragmatic attitude.

2. Related Progress

The use of 3D printing technology on the upper of sports shoes is still very limited, but this technology is an urgent challenge in the field of sports shoe production, and Nike is at the forefront for the time being. As the first sports brand to introduce 3D printing technology, Nike developed the first generation of 3D printed soccer shoes, "new Nike Vapor LaserTalon" and rugby shoes "Vapor Laser Talon" in 2013. After that, Nike launched the "Vapor Carbon 2014" elite running shoe and the "VaporHyperAgility" soccer shoe [14, 15]. It provided the famous American sprinter Allyson Felix with a 3D printer soccer shoe. Before the Rio Olympics, Nike developed an exclusive pair of high-performance spikes, "Zoom Superfly Flyknit" for the famous American sprinter Allyson Felix to meet her competition need, which eventually helped the "female flyer" win the Olympic championship [16]. In 2016, Nike announced a strategic partnership with printing giant Hewlett-Packard to use the HPJet Fusion3D printer to improve the efficiency of shoe prototypes and reduce costs.

In 2018, Nike even broke new ground with the launch of the Flyprint running shoe, which differs from its own and other brands' previous 3D printed sneakers in that it uses solid deposition modeling (SDM) technology to create the upper [17].

Adidas released its 3D printed running shoe "Futurecraft 3D" in 2015, and its limited edition "3D Runner" went on sale at the end of 2016. This running shoe can be tailored to the user's foot structure and personal sports habits [18, 19]. The shoe can be tailored to the user's foot structure and personal sports habits, with a "Primeknit" upper woven technology and an outsole and hollow midsole made of 3D printing technology.

Following the wave of 3D printing technology, Andromeda also launched a limited-edition trainer, "UA Architech," in March 2016 and a pair of 3D printed sneakers for American swimmer Michael Phelps in the same year [20]. In 2017, the company launched a new generation of 3D printed sneakers, "The ArchiTech Futurist," which features a 3D printed midsole with an interlocking grid structure that provides a "dynamic stability platform" for better stability and cushioning and a one-piece, seamless upper for a superior fit. Reebok uses 3D drawing technology to cleanly and accurately "draw" shoe components at the 3D level. This proprietary layering technology provides for the efficient, fast, and straightforward creation of personalized footwear without the need for molds. The sneaker features a high-rebound outsole with 3D printed laces that integrate the outsole with the upper for all-around foot comfort and responsive energy feedback [21].

New Balance improved its 3D printed concept sneaker "Fresh Foam Zante" in April 2016, rebranding it as the "Zante Generate" for a limited public release.

In early 2018, Anta also introduced Uniontech printing equipment, mainly used for research and development of footwear products. In June 2018, Li Ning released the latest version of its "Reignited" series at Paris Fashion Week, with its skeleton sole made of 3D printing technology and transparent plastic for the outer layer of the upper. The use of both materials is just right, creating the impression of fashion, technology, and youthfulness [22, 23].

In addition to major sports shoe brands at home and abroad, there are many related organizations at home and abroad conducting similar research, including many companies and brands that develop niche professional sports products are also trying their best to use 3D printing technology in the field of footwear design and production as early as possible.

2.1. Advantages and Limitations of 3D Printing Technology in Athletic Shoe Design and Manufacturing. 3D printing technology is a typical "additive manufacturing." It can be highly efficient, energy-saving, distributed, personalized, on-demand production mode manufacturing products, changing the previous large-scale and centralized industrial pattern, which is conducive to reduce the labor cost of low-

end manufacturing industry and the rapid development of innovative small enterprises [24]. At the same time, 3D printing technology circumvents the complex and time-consuming mold development and manufacturing process. With the further development of this technology, the printing efficiency and quality are bound to improve, so the manufacturing cycle of sports shoes will be further shortened, which also means that the speed of product iteration will be further enhanced.

3D printing technology has been a hot topic in design and manufacturing in recent years, bringing together many specialized research institutions and talents. The technology associated with it is constantly evolving. For example, before Nike's Flyprint running shoes, almost all 3D printing was focused on athletic shoe soles. However, they used solid deposition modeling manufacturing to create the idea of translating athletes' foot shape and movement data into textile geometry and collecting data through computational design tools to achieve the ideal material ratios [25]. The data was also collected through computational design tools to achieve the ideal material ratios. The performance of the shoe can also be adjusted at a later stage by adding or subtracting "warp and weft yarn" as needed, and the printing process is about 16 times more efficient than previous manufacturing processes.

With in-depth research of 3D printing technology, it is believed that, with its new technology and techniques, some difficulties in the manufacturing of sports shoes can be solved and optimized to help achieve better product performance and appearance design effects [26].

3. 3D Advantages of Printed Modeling Technology for Shoe Modeling

In the production of shoes and boots, the traditional production process is mainly design, pattern making, cutting, sewing, and molding. The entire book production process takes much time and is very complicated. Combining 3D modeling technology and 3D printing technology can quickly realize the initial production of molds and verify the effect, thus reducing the time spent on shoe production, improving the competitiveness of shoe and boot enterprises in the fierce market competition, and promoting the further development of related enterprises [1, 9, 27].

(2) Advantages of shoe modeling have great advantages in terms of freedom. The use of 3D software and technology for shoe modeling allows creating and changing the shape of the shoe or boot in accord with the design requirements. Some products have complex mold shapes before production, and 3D modeling technology can be used to better design them. The 3D modeling technology is needed to integrate fashion elements faster and better [13, 28].

When making fashion shoes and boots, it is difficult to combine some traditional shapes by traditional hand-design methods, and polygonal cross-sectional structures cannot be used as models for making shoes. In addition, shoes are made from open panels, which are prone to the problem of unevenly combined shapes and, therefore, poor integration of geometry [3]. Thus, processing technology seems to play a

great role in the production of shoes in modern society and is a realistic and important driving force in its development and advancement, which can effectively drive the industry forward.

3.1. Least Squares Support Vector Machine Improvement Algorithm. SVM is essentially a machine learning-based classification model [7], which is widely used in the field of statistical classification and regression analysis [29]. It achieves the classification of a dataset by finding a hyperplane that satisfies the classification requirements so that the different types of points of the selected training set are as far away as possible relative to the classification plan (see Figure 1).

The least-squares support vector machine (LS-SVM) algorithm is a variation of the standard SVM, which converts the SVM to solve a linear system of equations, avoiding the use of insensitive loss functions and greatly reducing the computational complexity.

The specific derivation of the LS-SVM algorithm is as follows.

Given N training samples, where x_i is the n-dimensional training sample input and y_i is the training sample output, in this paper, $(x_{i1}, x_{i2}, x_{i3}, y_i)$, where x_{i1} denotes the first variable factor in the i-th training sample, x_{i2} denotes the second variable factor in the i-th training sample, and x_{i3} denotes the second variable factor in the i-th training sample.

The objective optimization function of the LS-SVM algorithm is

$$J_1(w, e) = \frac{\mu}{2} w^T w + \frac{1}{2} \gamma \sum_{i=1}^{N} e_i^2,$$

$$\text{s.t. } y_i = w^T \phi(x_i) + b + e_i; \quad i = 1, \ldots, N, \tag{1}$$

where $\varphi( )$ is the kernel space mapping function, w is the weight vector, e_i is the error variable, b is the bias, and μ and γ are adjustable parameters. To solve for the minimum of the function, construct the Lagrange function:

$$L = J_1(w, e) - \sum_{i=1}^{N} \alpha_i \{ w^T \phi(x_i) + b + e_i - y_i \}, \tag{2}$$

where α_i is the Lagrangian multiplier.

The partial derivative of equation (3) yields

$$\frac{\partial L}{\partial w} = 0 \longrightarrow w = \sum_{i=1}^{N} \alpha_i \phi(x_i) \frac{\partial L}{\partial b} = 0 \longrightarrow \sum_{i=1}^{N} \alpha_i = 0,$$

$$i = 1, \ldots, N \frac{\partial L}{\partial e_i} = 0 \longrightarrow \alpha_i = \gamma e_i, \tag{3}$$

$$\frac{\partial L}{\partial \alpha_i} = 0 \longrightarrow w^T \varphi(x_i) + b + e_i - y_i.$$

By eliminating w and e, the solved optimization problem is transformed into solving the linear equation:
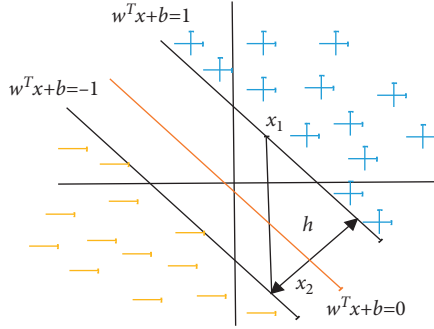
Figure 1: Fundamentals of support vector machines.

$$\begin{bmatrix} 0 & I_v^T \\ I_v & \Omega + \dfrac{1}{\gamma}I_N \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ y \end{bmatrix}, \tag{4}$$

where

$$\begin{aligned} y &= [y_1; \ldots; y_N], \\ I_v &= [1; \ldots; 1], \\ \alpha &= [\alpha_1, \ldots, \alpha_N], \\ \Omega &= \phi(x_i)^T, \\ \varphi(x_i) &= K(x_i, x_l)i, \\ l &= 1, \ldots, N. \end{aligned} \tag{5}$$

The LS-SVM for function estimation is obtained by solving equation (5) and then

$$y(x) = \sum_{i=1}^{N} \alpha_i K(x, x_i) + b, \tag{6}$$

where $k(x, x_i)$ is the kernel function. The commonly used kernel functions are mainly polynomial, RBF (radial basis), Sigmoid, and so on. The radial basis kernel function is generally used:

$$K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right) \sigma > 0. \tag{7}$$

The LS-SVM based on the radial basis kernel function needs to determine two parameters, penalty factor γ and kernel parameter σ [8]. The kernel parameter σ is a self-contained parameter of the radial basis kernel function, which mainly determines the actual size of the dimensionality of the resulting effect after mapping. The larger σ is, the faster the weight of the higher-level features decay, which is equivalent to mapping to a low-dimensional subspace; conversely, the smaller σ is, the more linearly divisible the results of arbitrary data mapping will be. The choice of γ, mainly for the segmentation process, may appear in the singularity of the introduction of soft interval hyperplane and determine the degree of tolerance for the accuracy of the segmentation surface. γ is larger, indicating that the segmentation process to select the support vector error

tolerance increases, resulting in inaccurate segmentation; otherwise, it may cause the segmentation not to be completed.

For the determination of γ and σ, the traditional algorithm generally uses the grid search method. Herein, we improve the genetic algorithm to optimize the selection of parameters γ and σ and use the 3-fold cross-validation classification accuracy as the fitness function. The value range of γ is (0.01 to 100), and the value range of σ is (0.01 to 100).

The genetic algorithm (GAA) is a heuristic algorithm that draws on natural selection and natural genetic mechanisms in biology and is more convenient, more robust, and easier to process in parallel than traditional methods such as a grid search. When dealing with the two model parameters γ and σ of the support vector machine, we first initialize the model parameters, set the binary code, randomize the initial population of the model parameters, and train the support vector machine model. The genetic algorithm is the maximization of the fitness function for the optimization, while the support vector machine model parameter selection is a minimization optimization problem, so the following conversion is made:

$$\text{Fit} = \begin{cases} C_{\max} - f(x), \\ 0. \end{cases} \tag{8}$$

After the fitness function is calculated, the global optimal solution is judged, and if the condition is satisfied, the determined parameters γ and σ are incorporated into the support vector machine model training. Otherwise, population regeneration, selection, crossover, and variation are performed iteratively until the termination condition is satisfied, and the computational effects are compared in Table 1.

3.2. Digital Shoe Last Design and Manufacturing Process. The shoe last is complex. Multidirectional irregularly twisted free-closed surface with long edge lines, many feature areas, and drastic curvature changes [2] causes difficulties in its design process, including (1) poor control of the external shape and poor density and (2) being not conducive to the rapid local modification of the last and not convenient to design personalized lasts according to the special foot shape. Therefore, the computer-aided design (CAD) and computer-aided manufacturing (CAM) design of shoe lasts have become a problem and a hot issue for discussion in recent years. The design and manufacturing process of shoe lasts is synthesized from the characteristics of shoe lasts and the experience of the traditional shoe manufacturing process. This paper summarizes the design and manufacturing process of shoe lasts into the following steps:

(1) Take human foot as the design prototype: first use 3D laser scanning measuring instrument to measure the human foot shape data, and after data processing, form the shoe last cloud model to complete the digitalization of human foot.

TABLE 1: Comparison of classification results.

Prediction algorithm	$y. \sigma$ method	Parameter value	Correct rate (%)
Least squares support	Grid search method	1,78	85.2941
Vector machine	Genetic algorithm	1.2,76	92.6471
BP neural network			60.2941

(2) After modifying the human foot model according to the demand, the data will be transferred to CAM to generate the CNC program and transferred to the cutting machine control computer. The vibrating cutting machine will be used for cutting, and the 2D unfolded shoe last section can be processed.

(3) Various file formats (e.g., stud, its file formats) are generated by CAD. The shoe lasts can be cut out by using advanced manufacturing technology CAM, or the shoe last models can be made by 3D printers, according to which the corresponding shoe last molds are made and the molds are used to reform the shoe lasts.

The specific design steps are shown in Figure 2.
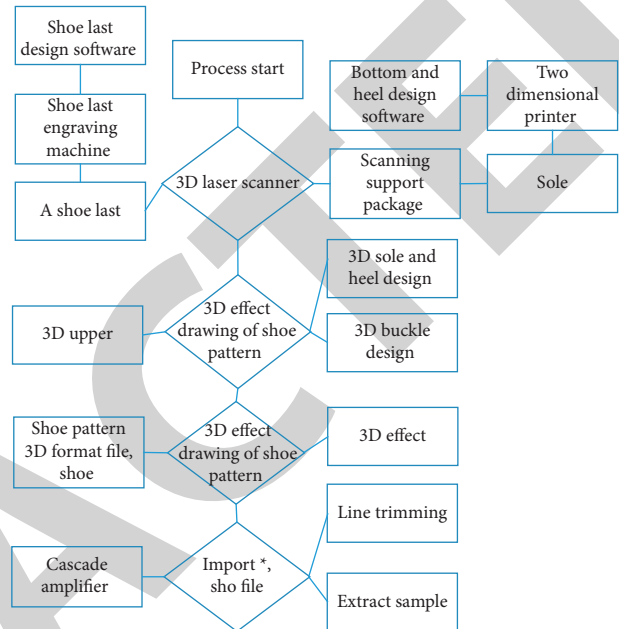
3.3. Measurement Methods and Data Processing.

The process from the human foot to the shoe last belongs to a kind of reverse engineering; that is, the human foot is used as the design object, and the last is designed and manufactured in turn to obtain the shoe last [4]. In this process, there are two key technologies: (1) the method of obtaining data related to the surface of a human foot; (2) the technology of constructing the surface of the shoe last based on the human foot data, and the need to solve the technical problem of transferring the raw data obtained from the measurement to the CAD/CAM system. At present, there are two measurement methods for foot geometry related data: contact probe measurement (such as mechanical measuring instrument CMM) and noncontact probe measurement (such as photoelectric scanner and laser scanner) [5, 6]; considering the peculiarities of human foot muscles, it is more appropriate to use laser scanner measurement, which is characterized by fast scanning speed and convenient measurement of foot size and can carry out intensive scanning measurement of human foot surface. The laser scanner is characterized by fast scanning speed, convenient foot size measurement, and intensive scanning measurement of the human foot surface, thus obtaining many "point cloud" raw data value points. In data processing, firstly, the useless points and bad points in the original data are judged and eliminated according to experience; secondly, the data point files are programmed to extract the useful data required for CAD modeling and handed over to the CAD system for curve, surface, and solid modeling; and finally, the CAD digital model of the prototype is obtained, as shown in Figures 3 and 4.

3.4. Shoe Last Modeling Design Analysis.

Analyzing the structure of the human foot, we can see that its palm surface is flat, but not completely flat, and the surface of the human foot is uneven and completely irregular. Then, as a model of the human foot, the surfaces on the last form are different

FIGURE 2: Digital shoe last design process.
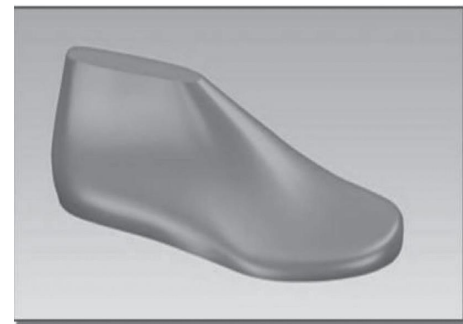
FIGURE 3: Model of shoe last after data processing.

FIGURE 4: Adjustment of heel height and last centerline.

everywhere, and the curvature changes greatly. There are both smooth surface parts and sharp corners and edges, so it is a typical nonexpandable free-form surface. For such a complex form, obviously, if the geometric modeling is completely in accordance with the shape of the human foot [7], it is technically quite difficult and practically unnecessary, and the process can be appropriately simplified. In general, the length and girth of the human foot (generally the top to girth and the front tarsal girth) are the two main design control parameters of the shoe last. On this basis, for a definite human foot type, the changes of the last required by various shoe types are mainly based on the changes of last and heel height, as shown in Figure 4. These changes will cause the last surface to change, and the point line surface describing the last surface will change accordingly. Thus, toe type and heel height are determined as the main parameter variables. In modeling, the last surface can be divided into different surface pieces, such as the last sole surface, the front surface piece, and the back surface piece. After modeling them separately, each surface piece will be put together to form the last shape. In the design process, considering that the shoes will scrape the ankles, in order to produce shoes and boots later on the physical ankle without the friction caused by the heel of the shoe, the curvature and the helper's foot degree are designed for the heel of the last, and the heel curvature and the helper's foot line are modified according to the characteristics and needs of different people's feet, so that the appropriate curvature and helper's foot angle can be reserved [8], as shown in Figure 5. The 3D effect of the shoes obtained through the software is shown in Figure 5.

4. 3D Printing Performance

As shown in Figure 6, from the perspective of 3D printed footwear appearance and design, the postprocessing and beautification of the finished print effect is still relatively limited, generally using sanding and color spraying to postprocess the product, but in the color and texture, texture effect processing is still poor, and richness is very insufficient. Although the current multicolor multimaterial printing can be achieved, the number of colors and material types of consumables is still relatively limited, often high color saturation, completely unable to meet the performance of the various subtle intercolor, multicolor, spot color, and popular colors required for sports shoes products. Of course, complex color effects and pattern effects are more difficult to achieve. In addition, due to the limited materials and melting layered way, resulting in 3D printed products texture, the performance is also relatively monotonous, unable to simulate the texture of various shoe materials, such as hardware accessories, textiles or leather, and other materials of various decorative processes, the lack of texture changes. Therefore, how to print products on the exquisite postprocessing so that shoes present richer and more diverse visual effects to meet the diverse needs of the consumer market is also one of the problems to be solved.

As shown in Figure 7, 3D-aided design and printing technology can maximize individual requirements in terms of foot size, posture, and sport. As a result, athletic shoe

Figure 5: 3D software to create shoes effect.
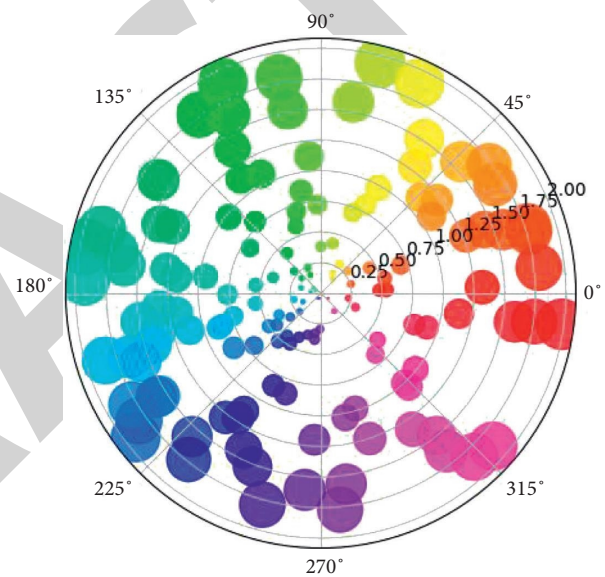
Figure 6: Printed polarization angle of 3D.
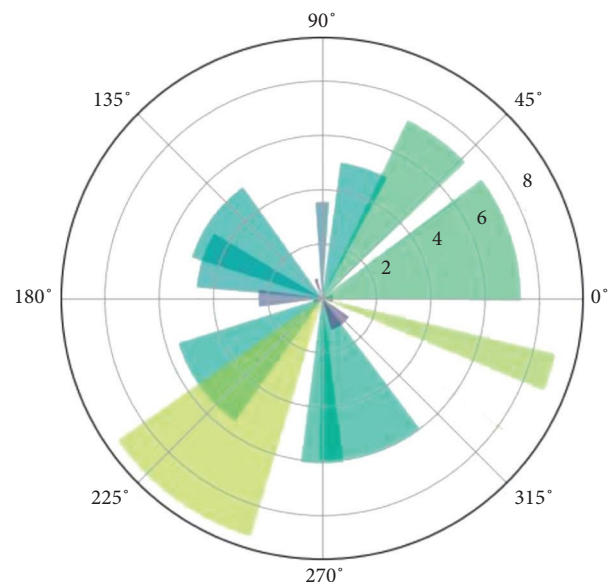
Figure 7: 3D printing sector.

companies, which have always pursued the use of technology, are scrambling to introduce 3D printing technology to improve footwear performance and enhance athletic performance. However, in general, some challenges still need to be overcome.

5. Conclusions

The new manufacturing technology represented by 3D printing is making a far-reaching change to the traditional manufacturing industry model. It can help footwear designers more conveniently transform design ideas, greatly improve the efficiency of product design, greatly expand the creative space of footwear design, and quickly create product prototypes. The forms and structures that are difficult to manufacture by traditional forming methods are produced in an efficient and low consumption way to improve the original and innovative design, processing, and manufacturing mode and the ecological environment of the manufacturing industry.

Under the baptism of 3D printing technology, the sports footwear industry will also undergo significant changes in the original design, development, and production methods, which will be beneficial in the development of sports footwear products, product innovation, appearance design, and function research and development, as well as the development of customized services. It will help enterprises better adapt to the current consumption patterns and market trends of fast fashion, personalization, small batch, and diversification; promote the footwear enterprises in China to adapt to the design and manufacturing mode under the Internet + and big data model; meet the Internet marketing and the possible distributed production pattern in the future; and gradually cultivate and develop and grow their own brands.

With the in-depth development of 3D printing technology, inevitably, the categories of printing consumables, performance, print quality, process level, appearance effects, and other aspects will continue to improve, providing more possibilities for the development of various types of sports shoes products; better meet the performance requirements of shoes on printing materials; achieve more technological product performance and richer product effects.

Data Availability

The dataset used in this paper are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest regarding this work.

References

[1] T. E. Clarke, E. C. Frederick, and C. L. Hamill, "The effects of shoe design parameters on rearfoot control in running," *Medicine & Science in Sports & Exercise*, vol. 15, no. 5, pp. 376–381, 1983.

[2] R. Vanderlinde and J. van Braak, "The e-capacity of primary schools: development of a conceptual model and scale construction from a school improvement perspective," *Computers & Education*, vol. 55, no. 2, pp. 541–553, 2010.

[3] C. Zhang, T. Xie, K. Yang et al., "Positioning optimisation based on particle quality prediction in wireless sensor networks," *IET Networks*, vol. 8, no. 2, pp. 107–113, 2019.

[4] M. A. Azad, D. Olawuni, G. Kimbell, A. Z. M. Badruddoza, M. S. Hossain, and T. Sultana, "Polymers for extrusion-based 3D printing of pharmaceuticals: a holistic materials-process perspective," *Pharmaceutics*, vol. 12, no. 2, p. 124, 2020.

[5] E. Karayel and Y. Bozkurt, "Additive manufacturing method and different welding applications," *Journal of Materials Research and Technology*, vol. 9, no. 5, pp. 11424–11438, 2020.

[6] N. Mika, "The socio-cultural construction of "theory of mind": a critical evaluation of cognitive science from phenomenological perspectives," *Japanese Journal of Developmental Psychology*, vol. 27, no. 4, pp. 288–298, 2016.

[7] H. Li, D. Zeng, L. Chen, Q. Chen, M. Wang, and C. Zhang, "Immune multipath reliable transmission with fault tolerance in wireless sensor networks," in *Proceedings of the International Conference on Bio-Inspired Computing: Theories and Applications*, pp. 513–517, Springer, Xian, China, October 2016.

[8] F. Hitt and C. Kieran, "Constructing knowledge via a peer interaction in a CAS environment with tasks designed from a task-technique-theory perspective," *International Journal of Computers for Mathematical Learning*, vol. 14, no. 2, pp. 121–152, 2009.

[9] E. H. Backes, L. D. N. Pires, C. A. G. Beatrice, L. C. Costa, F. R. Passador, and L. A. Pessan, "Fabrication of biocompatible composites of poly (lactic acid)/hydroxyapatite envisioning medical applications," *Polymer Engineering & Science*, vol. 60, no. 3, pp. 636–644, 2020.

[10] A. Hreljac, "Etiology, prevention, and early intervention of overuse injuries in runners: a biomechanical perspective," *Physical Medicine and Rehabilitation Clinics of North America*, vol. 16, no. 3, pp. 651–667, 2005.

[11] L. A. Dellinger, "From dollars to pesos: a comparison of the U.S. And Colombian anti-money laundering initiatives from an international perspective," *California Western International Law Journal*, vol. 38, no. 1, pp. 419–454, 2008, https://dialnet.unirioja.es/ejemplar/199438.

[12] C. A. Oyelami and J. L. Van Rooy, "A review of the use of lateritic soils in the construction/development of sustainable housing in Africa: a geological perspective," *Journal of African Earth Sciences*, vol. 119, pp. 226–237, 2016.

[13] A. Daneshmand and B. Hosseini Hashemi, "Performance of intermediate and long links in eccentrically braced frames," *Journal of Constructional Steel Research*, vol. 70, pp. 167–176, 2012.

[14] D. Wu, C. Zhang, L. Ji, R. Ran, H. Wu, and Y. Xu, "Forest fire recognition based on feature extraction from multi-view images," *Traitement du Signal*, vol. 38, no. 3, pp. 775–783, 2021.

[15] F. Pahlevanzadeh, R. Emadi, A. Valiani et al., "Three-dimensional printing constructs based on the chitosan for tissue regeneration: state of the art, developing directions and prospect trends," *Materials*, vol. 13, no. 11, p. 2663, 2020.

[16] T. Nachtigall, O. Tomico, and R. Wakkary, "Encoding materials and data for iterative personalization," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pp. 1–12, Glasgow, Scotland, May 2019.

[17] D. J. N. Amorim, T. Nachtigall, and M. B. Alonso, "Exploring mechanical meta-material structures through personalised

Retraction

Retracted: Research on the Application of Environmental Art Design Based on the Combination of VR and Panoramic Video Technology

Scientific Programming

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope

(2) Discrepancies in the description of the research reported

(3) Discrepancies between the availability of data and the research described

(4) Inappropriate citations

(5) Incoherent, meaningless and/or irrelevant content included in the article

(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] Y. Kong, "Research on the Application of Environmental Art Design Based on the Combination of VR and Panoramic Video Technology," *Scientific Programming*, vol. 2021, Article ID 9820550, 11 pages, 2021.

Research Article

Research on the Application of Environmental Art Design Based on the Combination of VR and Panoramic Video Technology

Yingqi Kong ⓘ

Wenzhou Business College, WenZhou 325006, ZheJiang, China

Correspondence should be addressed to Yingqi Kong; 00204434@wzbc.edu.cn

The panoramic video technology is introduced to collect multiangle data of design objects, draw a 3D spatial model with the collected data, solve the first-order differential equation for the 3D spatial model, obtain the spatial positioning extremes of the object scales, and realize the alignment and fusion of panoramic video images according to the positioning extremes above and below the scale space. Then, the panoramic video is generated and displayed by computer processing so that the tourist can watch the scene with virtual information added to the panoramic video by wearing the display device elsewhere. It solves the technical difficulties of the high complexity of the algorithm in the system of panoramic video stitching and the existence of stitching cracks and the "GHOST" phenomenon in the stitched video, as well as the technical difficulties that the 3D registration is easily affected by the time-consuming environment and target tracking detection algorithm. The simulation results show that the panoramic video stitching method performs well in real time and effectively suppresses stitching cracks and the "GHOST" phenomenon, and the augmented reality 3D registration method performs well for the local enhancement of the panoramic video.

1. Introduction

Augmented reality (AR) is an enhancement of the real world. In some real scenes (such as cultural relics and museums) that are not convenient for observers to enter, a panoramic video capture device is placed in the scene to capture the landscape of the real scene, and then the panoramic video is obtained and displayed through computer stitching processing; augmented reality technology is then used to add virtual information to the panoramic video and establish an augmented reality system based on panoramic video imaging [1].

Augmented reality systems are now widely studied at home and abroad and have a wide range of uses. Xue [2] studied scene recognition and tracking registration techniques for mobile augmented reality; Zhao [3] proposed an augmented reality system based on Hanzi markers; Bo et al. [4] proposed a new tracking registration method; Multisilta [5] proposed a tracking registration method based on TLD [6]; and Yu and Wang [7] studied augmented reality system modeling based on visual markers and alignment error

problems. However, there are few studies on augmented reality systems based on panoramic video imaging, and the establishment of a new AR system can help integrate knowledge in various fields and is very innovative.

Panoramic video technology allows videos to be viewed in any direction, free from the constraints of the original viewing angle and providing an authentic sensory experience [8–10]. Today, panoramic video is a popular and widely used form of media that has greatly improved people's lives. Environmental art design is an important part of digital architectural design, which is the construction and creation of artistic landscapes in a specific context [11]. With the gradual improvement of people's living standards, the requirements for living and leisure environments have increased. Traditional videos can only provide certain perspectives and cannot present a comprehensive view of the environment, making it difficult for users to grasp the details of the environment, wasting time and costs, and no longer meeting people's needs [12, 13].

In this paper, we explore and study the application of panoramic video technology in environmental art design by

linking panoramic video technology and environmental art design to the above problems. The application of panoramic video technology provides a solid foundation for a good living environment and creates very convenient living conditions. The panoramic video augmented reality system is proposed, and the key panoramic video stitching technology and augmented reality 3D registration technology to realize the system are proposed based on improved fast ORB (oriented FAST (features from accelerated segment test)) [14] and rotated BRIEF (binary robust independent elementary features). Binary robust independent elementary features [15], feature detection algorithm [16], and improved fast KCF (improved kernelized correlation filters, I-KCF) [17] augmented the reality 3D registration method for target tracking.

1.1. System Principle and Key Technology. The panoramic video augmented reality system can be regarded as composed of two parts: the real scene is a panoramic video camera placed on the circumference of a circle with center O and radius R, and then video images are collected, and the collected videos are stitched together according to the sequential relationship and displayed as the panoramic video background; in the virtual world part, the computer-generated virtual information is placed somewhere between the head display device and the background. If the virtual object is viewed through the head display device, the real background part behind the virtual object will be blocked by the virtual object, and the virtual object is superimposed with the real background through 3D registration technology, so as to realize the purpose of combining reality and imagination [18, 19]. The principle of the AR system with circular panoramic video imaging is shown in Figure 1.

The key technology of the panoramic video augmented reality system is mainly the panoramic video generation technology and 3D registration technology [20]. The specific steps of the system are shown in Figure 2.

1.2. Improved ORB Panoramic Video Generation. The key technology of panoramic video generation is image stitching, and image matching is the core technology of image stitching. The ORB algorithm introduces FAST [1] feature direction to make it rotation invariant because FAST feature points do not have to scale invariant information, so the ORB algorithm does not have to scale invariance [21]. The multiscale space theory is used to extract the stable feature points, and the number and spacing of feature points are parameterized during feature detection to make the ORB algorithm scale invariant and uniformly distributed; then, the Hamming distance is used for feature matching, and the RANSAC algorithm [22] is used to remove the mismatching points; finally, the best stitching line is found by using a dynamic programming algorithm with low complexity, and the stitched image is found to be the best stitching line. Finally, the optimal suture line is found by using a dynamic planning algorithm with low complexity, and the stitched image is smoothed by Poisson fusion. The specific process is shown in Figure 3.
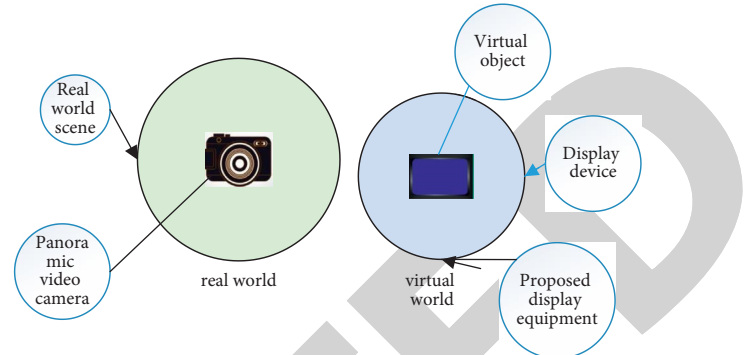
Figure 1: The principle diagram of the circular panoramic video AR system.
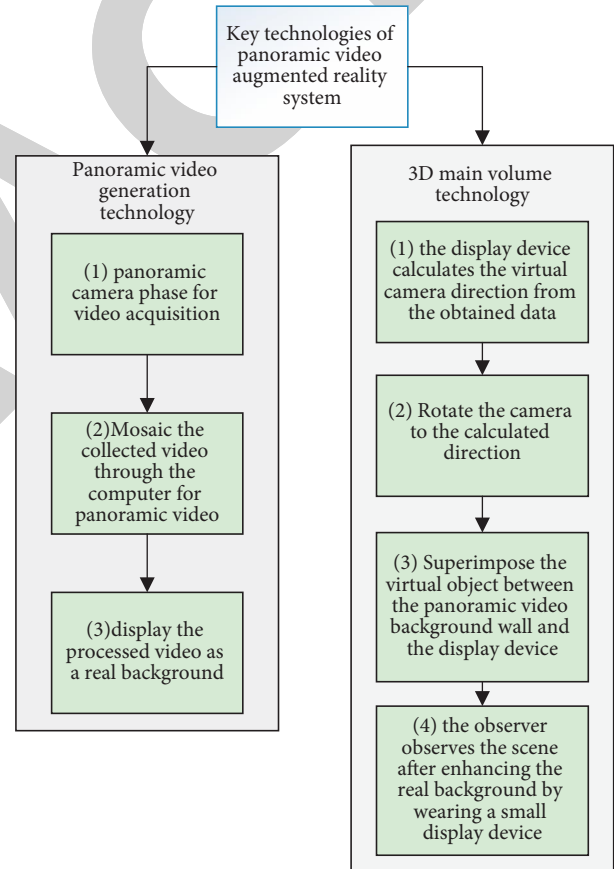
Figure 2: The key technologies of the panoramic video augmented reality system.

2. Improvement of the ORB Algorithm

2.1. Build the Scale Space and Find the Extreme Value Point. Under certain restricted conditions, the Gauss function is the only kernel smoothing function in the scale space [23]. The scale of the image $F(x, y)$ is defined as the function $L(x, y, \sigma)$, where σ is the scale factor, which is obtained by composing the image $F(x, y)$ with the Gauss function $G(x, y, \sigma)$ [24].
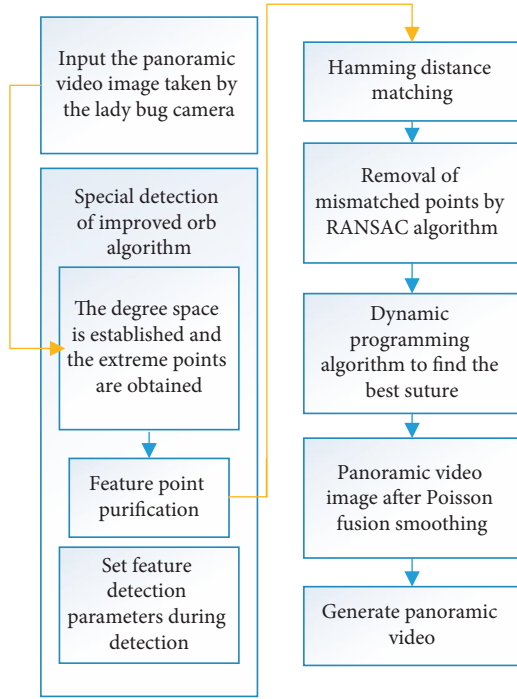
Figure 3: Flow of the improved ORB panorama video generation method.

$$L(x, y, \sigma) = G(x, y, \sigma)F(x, y),$$

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2}\exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right). \tag{1}$$

In order to obtain stable feature points in the scale space, the polar points are searched in the space $D(x, y, \sigma)$ obtained by image convolution with the Gauss difference function [25], and their local polar points are considered as the candidate feature points in the scale space with the relationship, as shown in the following equation:

$$D(x, y, \sigma) = [G(x, y, k\sigma) - G(x, y, \sigma)]F(x, y)$$
$$= L(x, y, k\sigma) - L(x, y, \sigma), \tag{2}$$

where the constant k separates the two adjacent scales.

2.2. Characteristic Point Purification.

After the detection of the extreme points, some unstable extreme points are removed to make the feature matching more stable. The positions of the extreme points of the original image are calculated by using a 3-dimensional second-order function at a certain scale, and the low-contrast extreme points are removed. First, a Taylor expansion of $D(x, y, \sigma)$ is performed at an extreme point:

$$D(X) = D + \frac{\partial D^T}{\partial X}X + \frac{1}{2}X^T\frac{\partial^2 D}{\partial X^2}X. \tag{3}$$

By finding the partial derivative of $\dot{X}$ from equation (3) and making it zero, we can obtain the extreme value point $\dot{X}$ with the following equation:

$$\dot{X} = -\frac{\partial^2 D^{-1}}{\partial X^2}\frac{\partial D}{\partial X}. \tag{4}$$

Substituting equation (3) into equation (4), we have

$$D(\dot{X}) = D + \frac{1}{2}\frac{\partial D^T}{\partial X}\dot{X}. \tag{5}$$

If $\dot{X}$ is present, the extreme points corresponding to it are removed, and the extreme points with low contrast are filtered out. The polar points on the edges are removed by calculating the ratio of the principal curvature to obtain stable polar points [26]. The Hessian matrix of the extreme points to be detected is calculated as follows :

$$H = \begin{pmatrix} D_{xx} & D_{xy} \\ D_{xy} & D_{yy} \end{pmatrix}. \tag{6}$$

Let the maximum and minimum eigenvalues of H be α and β, respectively:

$$\begin{cases} \text{Trace}(H) = D_{xx} + D_{yy} = \alpha + \beta, \\ \text{Det}(H) = D_{xx}D_{yy} - (D_{xy}^2) = \alpha\beta. \end{cases} \tag{7}$$

From equation (7), the principal curvature of the extreme point D is proportional to the eigenvalue corresponding to H. Let $\gamma = \alpha/\beta$; then, we have $(\text{Trace}(H)^2/\text{Det}(H)) = ((r + 1)^2/r)$, and then as γ increases, $(r + 1)^2/r$ also increases. Then, the ratio of principal curvature is checked by equation (9) to see if it is less than a certain threshold value γ:

$$\frac{\text{Trace}(H)^2}{\text{Det}(H)} = \frac{(r + 1)^2}{r}. \tag{8}$$

Usually, the threshold $\gamma = 8$; i.e., the extreme points of a principal curvature ratio not greater than 8 are retained, and the other extreme points are removed.

2.3. Setting of Feature Detection Parameters.

During feature detection, the number of extracted feature points is controlled by setting an integer N and setting the minimum spacing D between feature points so that the set of feature points is evenly distributed. If N is too small, it affects the uniform distribution of feature points. If N is too large for feature matching efficiency, N value can be obtained from the size of the reference image; if D is too large for feature matching success rate, D is too small for dense patches of feature distribution; D value can be set according to the demand of feature density [27]. In this paper, we set $N = 2,300$ and $D = 6$ px, where the values N and D are obtained experimentally.

Therefore, while retaining the advantages of fast operation and rotation invariance of the ORB algorithm, the ORB algorithm is made scale invariant, and the number of feature points and the spacing of feature points are parameterized in the feature extraction to make the feature distribution uniform [28].

2.4. Finding the Best Suture Line and Image Fusion. If the stitching is done directly after feature extraction and matching, it may make the panoramic video less effective due to the change of external ambient light. Unlike the traditional method of using motion estimation [29], the stitching crack "GHOST" phenomenon is suppressed by finding an optimal stitching line at the time of stitching and taking the content of one frame on both sides of this line to fill. A dynamic programming algorithm is a substratum of an optimal policy that must be optimal for its initial and final states. In applying the dynamic programming algorithm, suppose that there are n stages, $r_1(S_1, X_1)$ is the quantity indicator of the decision in stage i, S_i is the starting point of stage i, and X_i is the starting point of stage i and $i + 1$; then, the dynamic programming is to solve the value E:

$$E = \text{opt}[r_1(S_1, X_1) * \cdots * r_n(S_n, X_n)], \quad (9)$$

where "$*$" is the operation symbol and opt is max or min. In solving the shortest path, opt is taken as min and "$*$" is taken as "+" in order to minimize the summation of the stages. Drawing on the idea of dynamic programming, the above equation is used as the criterion equation for solving the value of the strategy index. First, initialize; then, expand down the row of calculated suture strength until the last row; and finally, select the best suture with the smallest strength value from the resulting set of all sutures [17, 27].

At this point, the search for the best stitching line is completed, and then the Poisson fusion algorithm is used to smooth the stitched image so that the stitched video image can reach the requirement of seamless and eliminate the "GHOST" phenomenon.

3. Panoramic Video Image Alignment and Fusion

3.1. Projection Plane Alignment. The process of acquiring spatial extremes of the scale ensures that the panoramic video is output in real time with standard video positioning, and then an algorithm is used to obtain the relevant parameters for the optimal projection matrix to enable interframe alignment of the video. The implementation process is simplified by using an efficient method to project and fuse all video frames from different cameras [25]. The implementation process is based on the alignment parameters obtained from the primitive process, and the alignment is completed by projecting all frames from different cameras into the same plane, which is represented as follows:

$$\begin{cases} x' = \dfrac{u_0 x + u_1 y + u_2}{u_6 x + u_7 y + 1}, \\[2mm] y' = \dfrac{u_3 x + u_4 y + u_5}{u_6 x + u_7 y + 1}, \end{cases} \quad (10)$$

where the 8 parameters of the optimal projection matrix obtained during the priming process are described by $u_i (i = 0, 1, 2, \ldots, 7)$, the left side of the pixel point in the projection frame is described by (x, y), and the coordinates of the pixel point projected into the reference frame are described by $(x\prime, y\prime)$.

In order to improve the image effect after alignment, we need to set up a large background image space and use this background image to align all frames. The reference coordinates used in the alignment process are the frame coordinates of the middle camera, and the background image is shaped according to these coordinates, i.e., the frames of the middle camera are filled into the middle of the background image, and the frames of other neighboring cameras are projected and transformed $W1$ and $W2$, respectively, and then aligned with the reference frames in the background image, and at the same time, the blending range between different frames is fused using bilinear fusion and nonlinear fusion techniques.

4. Hybrid Range Fusion

4.1. Bilinear Weighted Fusion. For example, if there are two frames in the blending range Q_1 and Q_2 and the weighting function is $a(x, y)$, the pixel value $E_{\text{hybrid}}(x, y)$ of the blending range of the fused frame can be expressed as

$$E_{\text{hybrid}}(x, y) = a(x, y) * Q_1(x, y) + (1 - a(x, y)) * Q_2(x, y). \quad (11)$$

Mixed range pixel values between adjacent frames are operated by bilinear weighted fusion [19]. The weighting function is shaped by the distance between the target pixel coordinates and the original frame boundary. Set the projected two images to be g and P, respectively, and a point in the mixing range of the two images to be s; then, the pixel values of the point in the two images are S_G and S_P, respectively, the minimum distances of the operation point s from the four sides of the two images are J_G and J_P, and the pixel value S_{Blend} of the point s in the fusion image is obtained as follows:

$$S_{\text{Blend}} = \frac{J_G}{J_G + J_P} \cdot S_G + \left(1 - \frac{J_G}{J_G + J_P}\right) S_P. \quad (12)$$

Fused pixel values of equation (5) fully analyze the value of adjacent frames, and the fusion of a single pixel range to the mixed range can be accomplished under normal conditions, but when the image has a parallax problem, two frames of the mixed range need to be fused, and the difference in content is high, resulting in ghosting and blurring in the mixed range.

4.1.1. Nonlinear Fusion. If the observation point fluctuates, the relative distance between the same object and nearby objects fluctuates significantly, resulting in parallax. Especially, when the distance between the object and the camera is particularly small, it will form a serious parallax problem. In this case, a nonlinear template fusion method is used to

solve the ghosting problem by constructing a risk-weighted template function:

$$b(x, y) = \begin{cases} 1, & \min(x, y, |x - R|, |y - Z|) > M, \\ \dfrac{\{\sin(\pi \cdot (\min(x, y, |x - R|, |y - Z|)/M - 0.5)) + 1\}}{2}, & \text{otherwise,} \end{cases} \tag{13}$$

where R and Z denote the width and height of the original frame, respectively, and M denotes the width of the non-linear transition range. The value of a in the center of the frame remains stable, and when it is close to a certain range of the boundary, it enters the transition range of width M, which decreases rapidly in a nonlinear manner, and the rate of decrease is regulated by the M value. The M value is used to adjust the respective weight and weighted areas of neighboring frames in the fusion process, and the higher the M value, the wider the transition range between neighboring frames and the smoother the transition, and vice versa. This nonlinear fusion process can adjust the transition rate and range at the boundary of the blending range in real time to solve the ghosting problem and enhance the fusion effect of the image [30–32].

5. Experimental Results and Analysis

5.1. Experimental Software and Hardware Platform. The video sequence-based environmental art design method was selected for comparison experiments. The experiment is divided into two parts: the simulation part compares the size of the shadow area visually through data software, and the actual test part compares the fusion effect to judge the advantages and disadvantages of the two methods. A special camera was used for data acquisition during the experiment, as shown in Figure 4, and relevant experimental parameters were given as shown in Table 1.

6. Simulation Experiment Results

The simulation results of the two environmental art design methods are shown in Figure 5. Among them, Figure 5(a) shows the deviation in positioning magnitude of the comparison method, and Figure 5(b) shows the deviation in positioning magnitude of this method. Since the experiments have eliminated all the influencing factors in the design process to the greatest extent, the analysis of Figure 5 shows that the common shadows in Figure 5(a) are larger and those in Figure 5(b) are smaller, which indicates that the comparison method fails to locate the object itself, resulting in the phenomenon of shadow superposition, while the design method in this paper is better.

6.1. Algorithm Performance. A panoramic video image of the Jinchengguan Intangible Cultural Heritage Exhibition Hall in Lanzhou city taken by a ladybug is simulated. The initial two frames shown in Figure 6 are compared with the ORB

algorithm for feature detection, and the results of the ORB algorithm are shown in Figure 7, and the results of the improved ORB algorithm are shown in Figure 8. The results of the ORB algorithm and the improved ORB algorithm are shown in Figure 7, respectively, after using Hamming distance to match the detected feature points and the RANSAC algorithm to remove the wrong matching points, and finally, the results of the improved ORB algorithm are shown in Figure 9, respectively, after using weighted average fusion and Poisson fusion smoothing.

In Figure 6, the gymnast identification and tracking system is mainly composed of two parts: hardware for image information data acquisition and software for machine vision identification and tracking.

The correlation R, entropy, spatial frequency, and average gradient indexes in [18] are used to quantitatively evaluate the fusion algorithm. R indicates the correlation between the stitched image and the original image, and ideally, R is 1; entropy is a measure of the average amount of information, and the larger the entropy value, the better the fusion effect; spatial frequency reflects the overall active level of the image, and the larger the value, the clearer the image; the larger the average gradient, the clearer the image. The results shown in Figure 10 are obtained statistically. It can be seen that the Poisson fusion algorithm has the highest R value and has advantages in entropy, spatial frequency, and average gradient.

The I-KCF and KCF target tracking algorithms are quantitatively analyzed in terms of success rate, which is evaluated in terms of the overlap rate of the border frames. Assuming that the tracked bounding boxes τ_t and τ_a are accurate bounding boxes, the repetition rate is defined as $s = (|\tau_t \cap \tau_a|/|\tau_t \cup \tau_a|)$, $|.|$ denoting the number of pixels in the region; to evaluate the performance of the tracking algorithm in the video sequence, the number of successful frames with $S > t_0$ is calculated, and the results are shown in Figure 11.

The simulation results show that the panoramic video generation algorithm proposed in this paper has a good stitching effect and effectively suppresses the effects of stitching cracks and the "GHOST" phenomenon. The local enhancement effect of the I-KCF-based augmented reality 3D registration algorithm is good.

6.2. Superparameter Value of Our Network Loss Function. In order to test and verify the segmentation effect of the WBCE Tversky loss function used to train the main network

Figure 4: Experimental camera set.

Table 1: Experimental camera parameters.

Attribute	Value
Color (rgb)	24
Capture rate (cif)	30
Focal length	Manual balance
White balance	Brake balance
Exposure form	Braking exposure
Gain range (cm)	3
Signal-to-noise ratio	>48 dB
Camera lens	F1.8/F7.85
Focus range	>80 mm
Apparent elevation angle	50°
Current (mA)	200
Lighting (lux)	2.5

(a)

(b)
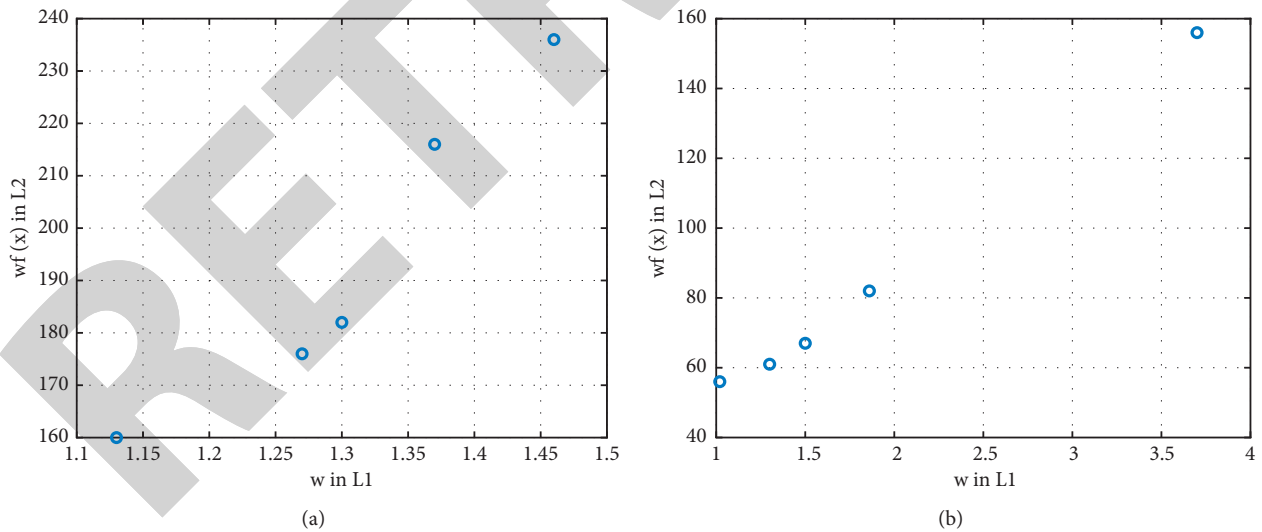
Figure 5: The average distribution of the weight.

of our network and to determine the superparameters β according to the range from 0.5 to 1.0, the value is compared in steps of 0.05 and compared with WBCE and Tversky loss function alone to prove the improvement of its performance. Based on GAN and Unet, the above different loss functions and parameter values are used for comparative experiments, and the experimental results are based on β. The values obtained by different methods are drawn into a line graph as shown in Figure 12. The WBCE loss function has no superparameters β. The experimental results are as follows: DSC is 48.31%, F2 is 47.45%, PR is 65.17%, and re is 46.66%. Comparing the WBCE experimental results with Figures 12(a) and 12(b), it can be seen that β, the DSC, and F2 of the WBCE Tversky loss function are higher than

FIGURE 6: The initial image.

FIGURE 7: The ORB algorithm matching effect.
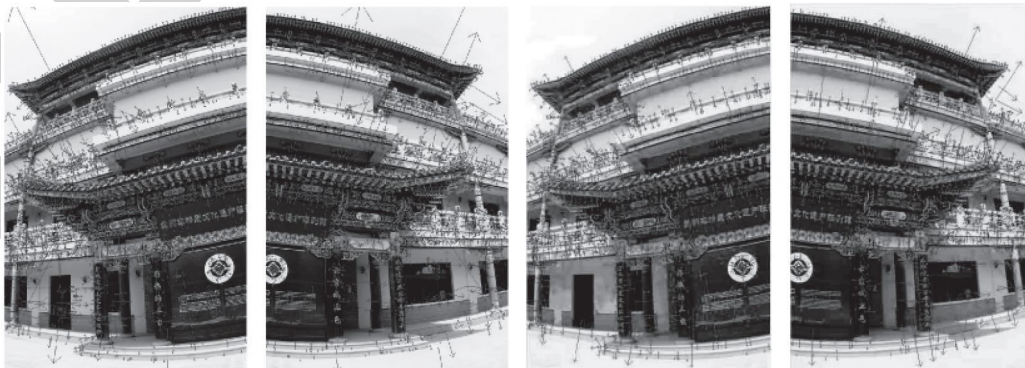
FIGURE 8: The ORB algorithm feature detection results.

Tversky and WBCE loss function. When $\beta = 0.8$, the WBCE Tversky loss function achieves the best segmentation performance. Science Tversky can adjust the accuracy and recall rate by adjusting the beta value. It can inhibit false negative and improve false positive, so as to improve the recall rate.

Therefore, with β, the accuracy decreases and the recall increases, which is consistent with the experimental results in Figures 12(c) and 12(d).

As mentioned earlier, the use of tolerance loss on the secondary network of our network can obtain moderate

FIGURE 9: The panoramic video image after stitching.
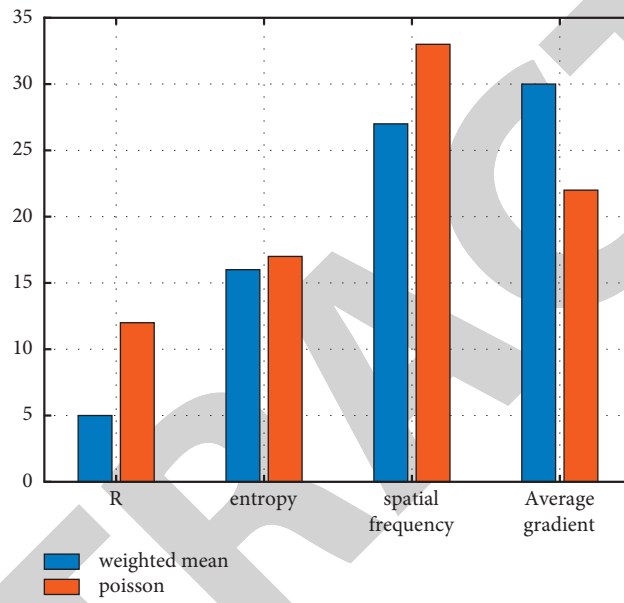
FIGURE 10: Convergence distribution of different methods.
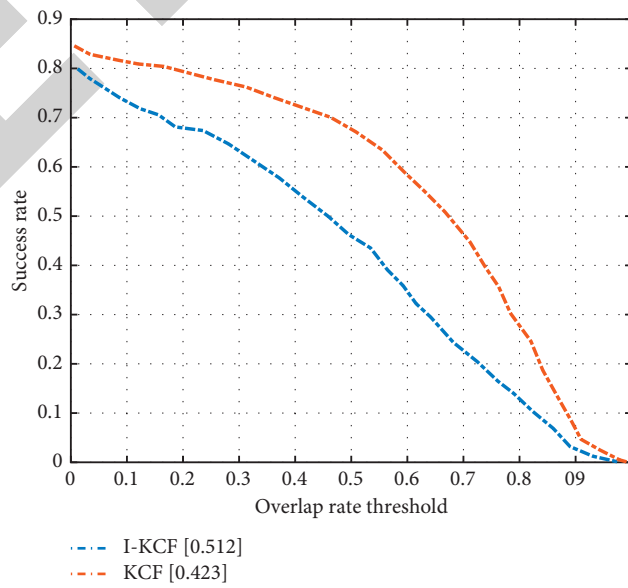
FIGURE 11: Success rate comparison chart.
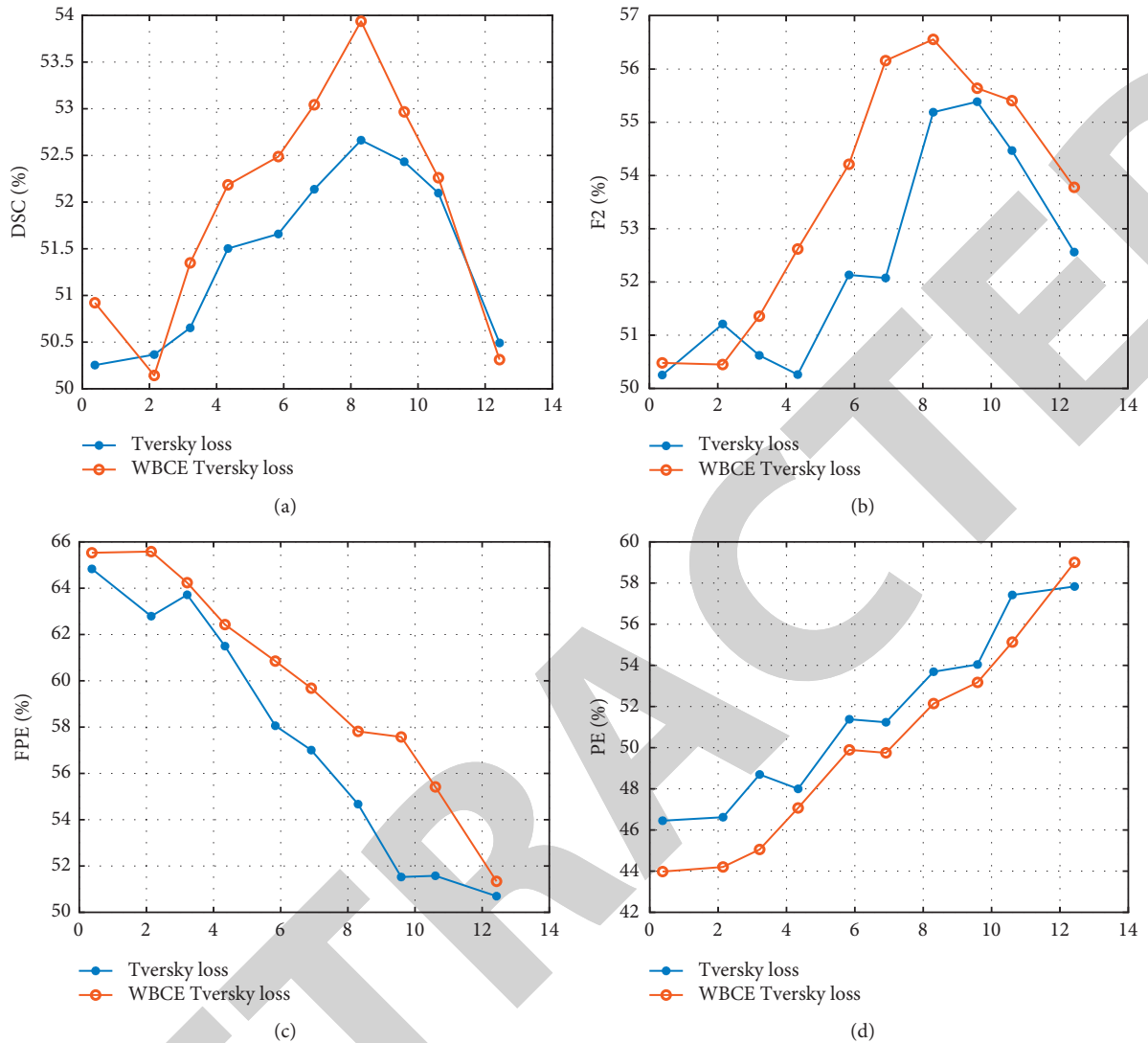
FIGURE 12: The performance comparison of WBCE Tversky and Tversky loss functions under different parameter configurations.

model constraints and a larger range of loss. Therefore, the parameters need to be adjusted, β, λ, and δ. Take appropriate values.

The training is still carried out on gau-a-unet, and the experimental results are drawn as a broken line diagram, as shown in Figure 13. In Figure 13(a), when the δ value remains unchanged, the higher the λ value, the higher the false-positive rate because the larger λ values provide greater weight for series items. And when the λ value remains the same, the smaller the δ value, the higher the false-positive rate because the specificity is δ value, which is considered as the training target. Hence, the smaller the δ value, the

smaller the specificity value. $F = 1 - s$, so the false-positive rate will increase with the decrease of specificity.

The false-positive rates corresponding to different values of λ and δ were sorted in the ascending order, as shown in Figure 13(b), and it can be seen that the false-positive rate gradually increased with the change of parameter configuration. When $\lambda = 5$ and $\delta = 0.6$, the false-positive rate achieves the maximum value, which is as high as 15.06%. From Figure 13(a), it can be seen that, for a pair of λ and δ values, the resulting false-positive rate is sometimes closer in value to that produced by using adjacent larger (or smaller) λ value and adjacent smaller (or larger) δ value, which means that
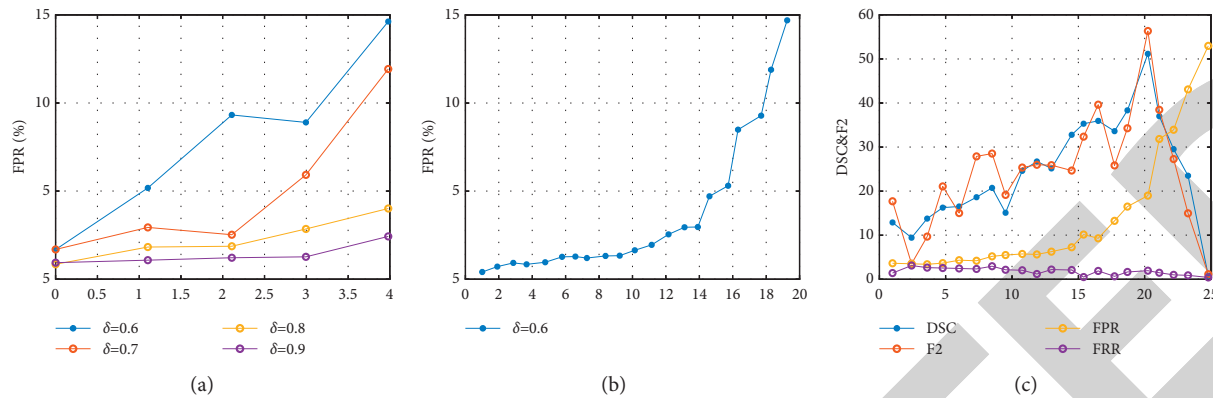
Figure 13: Different δ and λ values configure the false-positive rate curve under the tolerance loss function.

the magnitude of the false-positive rate is the result of the combined effect of λ and δ.

7. Conclusions

This paper proposes a new augmented reality system—the panoramic video augmented reality system. The two key technologies of panoramic video generation and augmented reality 3D registration to realize this system are studied, and the panoramic video stitching technology based on the improved ORB feature detection algorithm and the augmented reality 3D registration method based on I-KCF tracking are proposed. By adopting and improving the fast ORB feature detection algorithm and fast KCF target tracking algorithm, the panoramic video augmented reality system is more real time and noise-resistant, and the augmented reality technology is applied to the panoramic video imaging in this paper so that the scenes with added virtual information in the panoramic video can be viewed anytime and anywhere, which broadens the application fields of panoramic video and augmented reality and has certain practicality and innovation. The next step is to perform GPU acceleration on the graphics image processing algorithm.

Data Availability

The datasets used in this paper are available upon request to the author.

Conflicts of Interest

The author declares that there are no conflicts of interest regarding this work.

References

[1] L. Kunming and D. Lina, "Research on the application of environmental art design based on digital media technology," *Journal of Physics: Conference Series*, vol. 1915, no. 2, Article ID 022072, 2021.

[2] C. Xue, "Research on the expression of environmental art design based on virtual technology," *Revista de la Facultad de Ingenieria*, vol. 32, no. 9, pp. 560–565, 2017.

[3] J. Zhao, "Research on the application of BIM in environmental art design," *IOP Conference Series: Earth and Environmental Science*, vol. 676, no. 1, Article ID 012018, 2021.

[4] Z. Bo, X. Haitao, L. Lulu, J. Jun, Z. Jianyue, and T. Yiping, "Development of coordinate-machine of the scene of road traffic accident based on binocular stereo omni-directional vision sensors," *Procedia Engineering*, vol. 24, pp. 262–266, 2011.

[5] J. Multisilta, "Editorial on mobile and panoramic video in education," *Education and Information Technologies*, vol. 19, no. 3, pp. 565–567, 2014.

[6] "A Study on emotional touchpoint analysis by Eye-Tracking technology in dynamic environment - focused on automotive interior space," *Journal of Digital Design*, vol. 14, no. 2, pp. 285–294, 2014.

[7] X. Yu and W. Wang, "Research ON the application OF waste residue IN environmental artisitc design[j]," *China Ceramics*, vol. 44, no. 11, pp. 85–87, 2008.

[8] "Research on environmental art design system based on virtual reality technology," *Journal of Physics: Conference Series*, vol. 1992, no. 2, Article ID 022129, 2021.

[9] R. Rosariastuti, S. Sudadi, and F. S. Prasasti, "A bioremediation process based on the application of Rhizobium sp. I3 and Ramie (Boehmeria nivea L.) in lead contaminated soils," *Journal fur Kulturpflanzen*, vol. 72, no. 2/3, pp. 40–48, 2020.

[10] C. Zhang, T. Xie, K. Yang et al., "Positioning optimisation based on particle quality prediction in wireless sensor networks," *IET Networks*, vol. 8, no. 2, pp. 107–113, 2019.

[11] S. Hao, "Research on the application of GIS technology in the preparation of village planning--Take Feng Sheng Village as an example," *IOP Conference Series: Earth and Environmental Science*, vol. 791, no. 1, Article ID 012150, 2021.

[12] C. Krzysztof and J. Paweł, "The threat of environmental pollution with harmful substances, on the example of uncontrolled fires in landfills and actions aimed at reducing it," *Safety & Fire Technology*, vol. 57, no. 1, pp. 80–98, 2021.

[13] J. W.-M. Yuen, I. Y.-P. Wong, P. K.-F. Chiu et al., "A comprehensive community-based prevalence study on nocturia in Hong Kong male adults," *International Journal of Environmental Research and Public Health*, vol. 18, no. 17, Article ID 9112, 2021.

[14] C. Wang and R. Pawlowicz, "Internal wave generation from tidal flow exiting a constricted opening,," *Journal of Geophysical Research: Oceans*, vol. 122, no. 1, pp. 110–125, 2017.

Research Article

Energy Control Strategy for Parallel Hybrid Electric Vehicle Based on Terminal Neural Network

Haitao Yan [1,2] and Yongzhi Xu [2]

[1]*School of Mechanical and Precision Instrument Engineering, Xi'an University of Technology, Xi'an, Shaanxi 710048, China*
[2]*Sanmenxia Polytechnic, Sanmenxia, Henan 472000, China*

Correspondence should be addressed to Haitao Yan; yht0211@stu.xaut.edu.cn

Energy control strategy is a key technology of hybrid electric vehicle, and its control effect directly affects the overall performance of the vehicle. The current control strategy has some shortcomings such as poor adaptability and poor real-time performance. Therefore, a transient energy control strategy based on terminal neural network is proposed. Firstly, based on the definition of instantaneous control strategy, the equivalent fuel consumption of power battery was calculated, and the objective function of the minimum instantaneous equivalent fuel consumption control strategy was established. Then, for solving the time-varying nonlinear equations used to control the torque output, a terminal recursive neural network calculation method using BARRIER functions is designed. The convergence characteristic is analyzed according to the activation function graph, and then the stability of the model is analyzed and the time efficiency of the error converging to zero is deduced. Using ADVISOR software, the hybrid power system model is simulated under two typical operating conditions. Simulation results show that the hybrid electric vehicle using the proposed instantaneous energy control strategy can not only ensure fuel economy but also shorten the control reaction time and effectively improve the real-time performance.

1. Introduction

Energy crisis and environmental pollution are two major problems that need to be solved in today's social development. With the development of society, environmental pollution has become increasingly prominent, especially the recent haze weather, seriously affecting people's daily life and even threatening life and health, and automobile exhaust emission is one of the main sources of this severe weather. Traditional fuel vehicles have been difficult to meet increasingly demanding energy saving standards and environmental protection indicators, while new energy vehicles such as fuel cell vehicles have encountered bottlenecks due to constraints of technology, cost, or infrastructure [1–5]. In addition, the power battery technology of pure electric vehicle is not mature enough to meet the requirements of long-distance continuous driving. In contrast, hybrid electric vehicle technology is advancing steadily.

From the current situation, it is necessary and feasible to take hybrid electric vehicle as a transitional model to relieve energy and environmental pressure.

As hybrid electric vehicles have two energy sources, engine and battery, it is particularly important to make reasonable use of the two energy sources to provide the vehicle with good fuel economy and emission performance [6–9]. Therefore, the energy control of the vehicle becomes a crucial part of the hybrid power system. Hybrid electric vehicles can be divided into series, parallel, and hybrid according to the type, quantity, and connection relationship of parts [10–13]. Parallel hybrid power system can reduce fuel consumption and pollutant emission of the vehicle without greatly increasing the cost and the complexity of the vehicle structure. Therefore, the research focus of this paper is on parallel hybrid electric vehicle (HEV).

Energy control strategy refers to the torque distribution and coordination control strategy between the engine and

motor. At present, there are four basic types of energy control strategies [14,15]: (1) logic threshold control strategy; (2) instantaneous control strategy; (3) global optimal control strategy; (4) fuzzy control strategy. The full name of instantaneous control strategy is instantaneous equivalent minimum fuel consumption control strategy. In each control cycle of the hybrid power system, the optimal power output of the engine and motor is calculated with the goal of obtaining the lowest fuel consumption. Instantaneous control strategies can also combine fuel economy and emission performance. Instantaneous control strategy can achieve the minimum equivalent fuel consumption within each control cycle, but it requires a lot of calculation and is difficult to achieve. In addition, according to the optimization theory, the sum of minimum values is not equal to the minimum value of sum, so the instantaneous control strategy cannot achieve global optimization. The global optimal control strategy [16] can theoretically obtain the lowest fuel consumption in the real sense, but the outstanding disadvantage is that the driving condition of the vehicle must be known during the whole operation process, which is obviously inconsistent with the reality. At present, this control strategy can only be used to evaluate other control effects, and it is difficult to be applied in real-time vehicle control.

2. Literature Review

In view of the problems of the above control strategies, many scholars have improved and optimized the energy control strategy of HEV by using different research methods.

In the existing research results, Yin et al. [17] introduced the condition recognition technology into the fuzzy control strategy to solve the shortcoming that it could not adapt to different working conditions and expand the application scope of fuzzy control. However, the paper did not give a clear conclusion about which working conditions it was applicable to. Han et al. [18] used the stochastic dynamic programming theory to solve the problem that the global optimal control strategy requires to know future driving conditions in advance. The simulation results show that the control method can achieve good results in a variety of driving conditions, but the calculation is large and the simulation time is long. Calafiore et al. [19] proposed a stochastic model predictive control method to try to solve the problem that future driving conditions of hybrid electric vehicles are difficult to predict. Ming et al. [20] established an intelligent control strategy of hybrid electric vehicle considering the influence of driving conditions by using fuzzy neural network. The actual vehicle verification results show that the established control strategy reduces the influence of driving condition fluctuation on the performance of electric vehicles. However, this method takes too long to obtain recognition results. These research results have improved the existing energy control strategies to a certain extent and improved the performance of hybrid power system, but so far there is no optimal control scheme that can be used for practical control. The optimization and improvement of HEV energy control strategy need further research.

The terminal neural network is a kind of neural network model with bounded convergence error in finite time. The effectiveness of the network solving model usually includes two aspects, that is, higher solving accuracy of the model (for example, in the field of chaotic system control) and faster solving speed (for example, in the field of terminal sliding mode control). The finite-time convergent dynamic characteristics of the terminal state greatly improve the convergence efficiency of the dynamic system and, at the same time, obtain higher system convergence accuracy. Chai et al. [21] proposed a class of terminal system with dynamic convergence, which has the dynamic property of finite-time convergence and was applied to the solution of time-varying equations and the repetitive motion planning of redundant manipulator, both of which achieved high convergence accuracy. Different from the finite-time convergent dynamic network, the finite-state convergent neural network proposed by Yuan et al. [22] can adopt any form of nonlinear activation function, which is applicable to a wider range and more in line with actual needs.

Therefore, a terminal neural network model for controlling torque output is established in this paper. Firstly, the objective function of the instantaneous control strategy is defined, and the nonlinear equations used to control the torque output are solved by using terminal neural network models. Then, an example is given based on the neural network model, and the simulation results are obtained. The simulation results show that the terminal network controller can achieve good fuel economy and improve the real-time control performance.

3. Principle of Instantaneous Control Strategy

3.1. Transient Control Policy Definition. The full name of instantaneous control strategy is instantaneous equivalent fuel consumption minimum control strategy [23,24], so it can be seen that this control rule mainly includes two aspects:

(1) Equivalent fuel consumption: for a general hybrid electric vehicle, its power battery electric energy needs to be maintained within a certain range, that is, the power battery electric energy consumed in a certain stage (except the recovered braking energy) needs to be supplemented by a certain amount of fuel consumed by the engine in the future operation of the hybrid electric system. Therefore, an equivalence relationship between the energy consumed by the power battery and the fuel required to replenish it is needed.

(2) Instantaneous optimization: in each control period, according to the power requirements of the hybrid power system, the optimal power output of the motor and engine is calculated in real time with the minimum equivalent fuel consumption as the standard, and the two are controlled based on this rule.

The minimum control strategy of instantaneous equivalent fuel consumption of HEV can be simply expressed as

$$M = \sum \mathrm{Min}\{m_{fc}[T_{fc}(t), \omega_{fc}(t)]\Delta t + m_{mc\_eq}[T_{mc}(t), \omega_{mc}(t)]\Delta t\}. \tag{1}$$

At the same time, the following constraints should also be met:

$$0 < T_{fc}(t) < T_{fc-\max},$$
$$0 < \omega_{mc}(t) < \omega_{mc-\max}, \tag{2}$$
$$SOC_{\mathrm{final}} - SOC_{\mathrm{initial}} = 0,$$

where ω_{fc} is the engine speed, ω_{mc} is the motor speed, $\omega_{mc\_\max}$ is the maximum motor speed, $T_{fc\_\max}$ is the maximum engine torque, m_{fc} is the engine fuel consumption at every instant, m_{mc_eq} is the equivalent fuel consumption of power battery, T_{fc} is the engine torque, T_{mc} the is motor torque, SOC_{initial} is the initial state of charge of power battery, and SOC_{final} is the state of charge at the end of power battery.

3.2. Calculation Method of Equivalent Fuel Consumption of Power Battery.

This paper establishes the relationship between power battery and fuel energy consumed by engine. Specific methods are as follows:

(1) Establish the relationship between engine fuel consumption rate and motor output torque:

$$b = f(T_{mc}). \tag{3}$$

When the total demand torque is known, the relation curve between the motor output torque and engine fuel consumption rate can be established according to engine universal characteristic curve. As shown in Figure 1, points A, B, C, and D correspond to the working conditions of engine driving alone and charging for power battery, engine driving alone and not charging for power battery, engine and motor codriving vehicle, and motor driving alone.

(2) Establish the corresponding relationship between power battery variation and motor output torque:

The qualitative relation curve between the variation of power battery and the output torque of the motor is shown in Figure 2, where three points A, B, and C, respectively, represent the working conditions in which the engine drives the motor to charge the power battery, the engine drives the system without the power battery participating in the operation, and the engine and the power battery jointly drive the hybrid power system.

According to the relation curve between the state of charge and voltage of the power battery and the relation curve between the current of the motor and the output torque, the power battery variation corresponding to the output torque of each motor can be calculated, and the relation between the variation of the power battery and the output torque of the motor can be established.
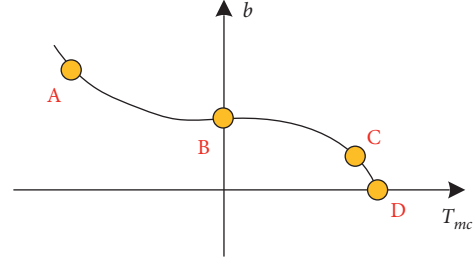
FIGURE 1: Qualitative relation curve between engine fuel consumption rate and motor output torque.
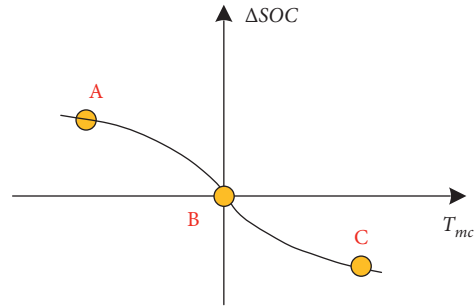
FIGURE 2: Qualitative relation curve between power battery variation and motor output torque.

$$\Delta SOC = f(T_{mc}). \tag{4}$$

(3) Establish the corresponding relationship between engine fuel consumption rate and actual variation of power battery:

$$b = f(\Delta SOC_{\mathrm{act}}), \tag{5}$$

where $\Delta SOC_{\mathrm{act}}$ is the actual change of the power battery.

(4) Calculate the fuel energy corresponding to the actual change of power battery:

According to the relationship between the actual change of power battery obtained in Step 3 and the engine fuel consumption rate, the corresponding relationship between the actual change of power battery and the fuel energy consumed by the engine can be established, as shown in Figure 3.

The above method makes use of the equal relationship between the electric energy consumed by the power battery and the electric energy replenished by the power battery to establish the relationship between the actual change of the power battery and the fuel energy consumed by the engine. The disadvantage of this approach is that it does not take into account that the operating state of the hybrid system is changing from moment to moment. The above method is based on the assumption that the demand speed of vehicles at the future moment is the same as that at present, which is contrary to the actual situation. Based on the analysis of the energy flow of parallel hybrid electric vehicle, the instantaneous control strategy is calculated.
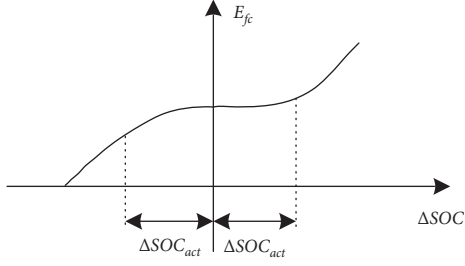
FIGURE 3: Qualitative relation curve between actual variation of power battery and engine energy consumption.

3.3. Calculation of Instantaneous Equivalent Fuel Consumption of Power Battery.

The working mechanism of the instantaneous control strategy can be briefly described as follows: in order to achieve the lowest fuel consumption of the normal running of the vehicle, the distribution ratio of the power required by the vehicle between the engine and the motor is optimized in real time in each control cycle. Therefore, a transient optimal objective function which can accurately reflect the actual operation characteristics of HEV should be established. The energy flow diagram of the hybrid power system is shown in Figure 4.

The engine power, the power allocated by the engine to the motor for charging the power battery, and the total required power of the system satisfy the following relational expression:

$$P_{fcN} = P_{fcY} - P_{mc\_chg} = P_{req}, \quad (6)$$

where P_{req} is the driving power required by the vehicle at a certain time, P_{mc_chg} is the charging power of the motor, and P_{fcY} is the engine power.

Therefore, the fuel consumption rate of the engine charging the power battery at a certain time in the future can be expressed by the fuel consumption rate when the engine only participates in driving the vehicle. The equivalent fuel consumption is

$$m_{mc-qq} = \frac{b_{chg} P_{mc-chg}}{\eta_{mc} \eta_{chg} \eta_{dischg}}, \quad (7)$$

where b_{chg} is the fuel consumption rate of the engine for charging the power battery, η_{mc} is the motor efficiency, η_{chg} is the charging efficiency of the power battery, and η_{dischg} is the discharging efficiency of the power battery.

3.4. Objective Function of Instantaneous Control Strategy.

According to the objective function of the minimum control strategy for instantaneous fuel consumption given previously, equation (1) can be rewritten as

$$M = \sum \text{Min}\{m_{fc}[P_{fc}(t)]\Delta t \\ + m_{mc\_q}[P_{mc}(t)]\Delta t\}, \quad (8)$$

$$m_{mc\_eq}[P_{mc}(t)] = f_{eq\_dischg} P_{mc}. \quad (9)$$

In order to explain the different functions of the motor at different times in the process of deriving instantaneous fuel

consumption, charge and discharge can be unified when the objective function is expressed. Therefore, equation (8) can be rewritten as

$$M = \sum \text{Min}\{m_{fc}[P_{fc}(t)]\Delta t + [\lambda f_{eq\_dischg} + (1-\lambda)f_{eq\_chg}]P_{mc}\Delta t\}, \quad (10)$$

where $\lambda = [1 + \text{sign}(T_{mc})]/2$.

According to equation (10), the control effect of the instantaneous control strategy depends on the size of the two variables. These two variables are respectively called the equivalent fuel consumption coefficient of the power battery under the discharge condition and the charging condition, and their specific values are determined by the hybrid electric vehicle configuration, efficiency characteristics, and driving conditions.

4. Instantaneous Control Strategy Based on Terminal Neural Network

The control rule of instantaneous control strategy can be regarded as the nonlinear mapping of single output, and the power battery output refers to the motor output torque. Consider the following dynamic characteristic equation to construct the recursive neural network model [25]:

$$\dot{E} = -\rho S(E), \quad (11)$$

where E is the $n \times n$ dimensional error matrix, $\rho > 0$ is the adjustable gain coefficient, and $S(E)$ is the activation function, which is strictly monotonically increasing odd function.

It is known that the recursive neural network is asymptotically stable by the properties of activation function. The asymptotic convergence performance is analyzed below. According to equation (11),

$$\frac{dE}{dt} = -\rho E \Longrightarrow \frac{1}{E}\,dE = -\rho\,dt\,, \quad (12)$$

$$\int_{E_0}^{E} \frac{1}{E}\,dE = \int_0^t -\rho dt \Longrightarrow \ln E - \ln E_0 = -\rho t, \quad (13)$$

where E_0 is a constant. According to equation (13), when $E = 0$, $t \longrightarrow \infty$, so the error converges to zero in infinite time; the recursive neural network model characterized by the dynamic characteristics of asymptotic stability is called asymptotic convergence network.

In order to realize the finite-time convergence of energy control of parallel hybrid electric vehicles, the control rules of instantaneous control strategy are regarded as the solution of time-varying nonlinear equations of single output (motor output torque), and the optimization objective function is equation (12), which can be expressed as follows:

$$\begin{cases} f_1(x_1, x_2, \ldots, x_n) = 0, \\ f_2(x_1, x_2, \ldots, x_n) = 0, \\ \qquad \vdots \,, \\ f_n(x_1, x_2, \ldots, x_n) = 0, \end{cases} \quad (14)$$
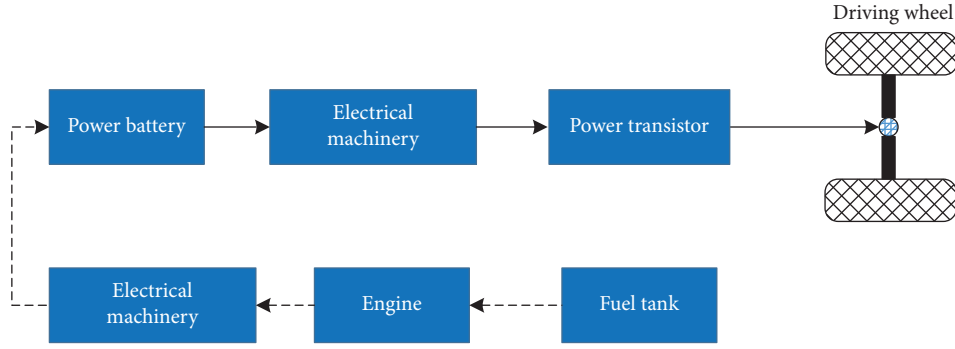
FIGURE 4: Energy flow diagram of hybrid power system.

where x_n is the n-th unknown variable and f_n is a smooth differentiable nonlinear function.

To solve the nonlinear equations, the error equation is defined as follows:

$$e(t) = f(x) - 0. \qquad (15)$$

The time needed to obtain the solution of nonlinear equations by using the terminal neural network is finite, but the asymptotic network needs enough time. In order to simplify the analysis, the time-varying nonlinear equations are set as follows:

$$\begin{cases} \dfrac{\ln x_1(t) - 1}{(t-1) = 0}, \\[2mm] {}_1(t)x_2(t) - \exp\left(\dfrac{1}{(t+1)}\right)\sin(t) = 0, \\[2mm] x_1^2(t) - \sin(t)x_2(t) + x_3(t) - 2 = 0, \\[2mm] x_1^2(t) + x_2^2(t)x_3(t) + x_4(t) - t = 0. \end{cases} \qquad (16)$$

Then, the test calculation result is

$$X^*(t) = \begin{bmatrix} x_1^* \\ x_2^* \\ x_3^* \\ x_4^* \end{bmatrix} = \begin{bmatrix} \exp\left(\dfrac{1}{(t+1)}\right) \\[2mm] \sin(t) \\[2mm] 2 - \exp\left(\dfrac{2}{(t+1)}\right) + \sin(t) \\[2mm] t - 2 - 2\sin^2(t) \end{bmatrix}. \qquad (17)$$

The initial value can be any value, and $X(0) = \begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix}^T$ in this paper.

A recursive neural network model is constructed to solve the time-varying nonlinear equations. The convergence characteristics are analyzed according to the activation function graph, and then the stability of the model is analyzed respectively and the time of error convergence to zero is deduced.

5. Terminal Neural Network Calculation

The terminal neural network based on BARRIER function is used to calculate the solution of nonlinear equations. The neural network equation is as follows:

$$\dot{e} = -\varepsilon \mathrm{Bar}(e, \delta). \qquad (18)$$

BARRIER excitation function [26] is

$$S(e) = \mathrm{Bar}(e, \delta) = \sqrt{\dfrac{|e|}{|e| + 1/\delta}}\,\mathrm{sgn}(e). \qquad (19)$$

The parameters in the terminal neural network are $\varepsilon = 5$ and $\delta = 0.5$. The calculation results are shown in Figure 5, and the calculate value $X(t)$ converges to the ideal value $\mathbf{X}^*(t)$.

Figure 6 shows the comparison between the two norms of $f(x)$ in the terminal network and the asymptotic network. It can be seen from the figure that $\|f(x)\|_2$ obtained by using the terminal network has a faster convergence speed and converges to zero in finite time.

6. Simulation Experiment and Result Analysis

6.1. Simulation Software and Parameters. In this paper, the hybrid electric vehicle simulation software ADVISOR is used as a test platform to test whether the neural network controller can improve the response speed of instantaneous control. ADVISOR, which stands for advanced vehicle simulator, was developed by the U.S. National Renewable Energy Laboratory. ADVISOR adopts modular programming idea and MATLAB/Simulink programming language, and the code is completely free to the public so that it has good secondary development and cosimulation with other software functions. The simulation process of ADVISOR is shown in Figure 7. First, the actual speed and fuel consumption rate of the vehicle are calculated according to the model data and component data, and then simulation experiments are carried out. Finally, simulation graphics and data are output.

In order to increase the authenticity of the simulation experiment, the parameters of the vehicle in this simulation experiment were formulated according to the data of a listed hybrid model. The main parameters are shown in Table 1.
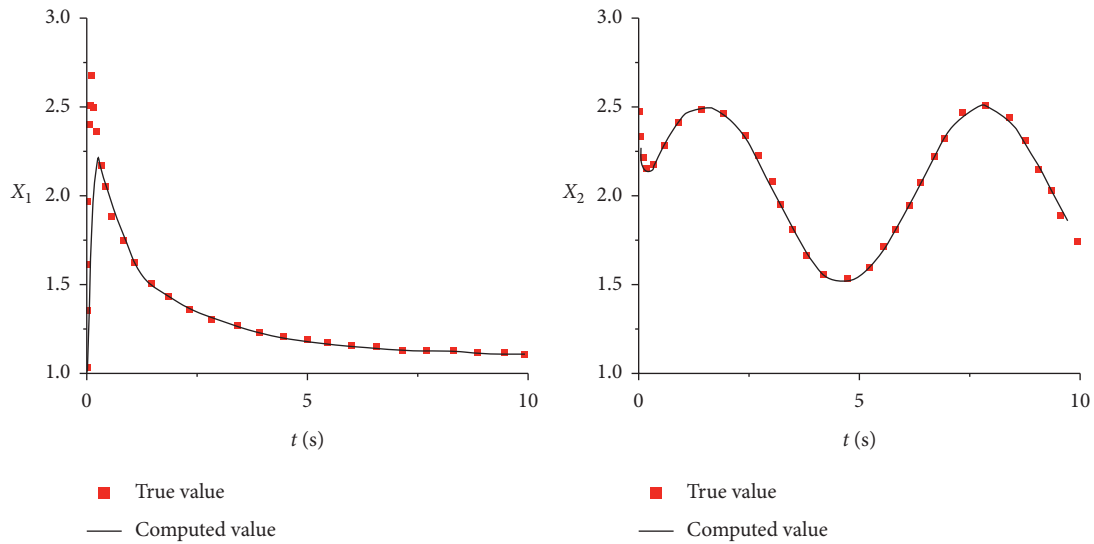
FIGURE 5: Comparison between the true result and resultant values of nonlinear equations.
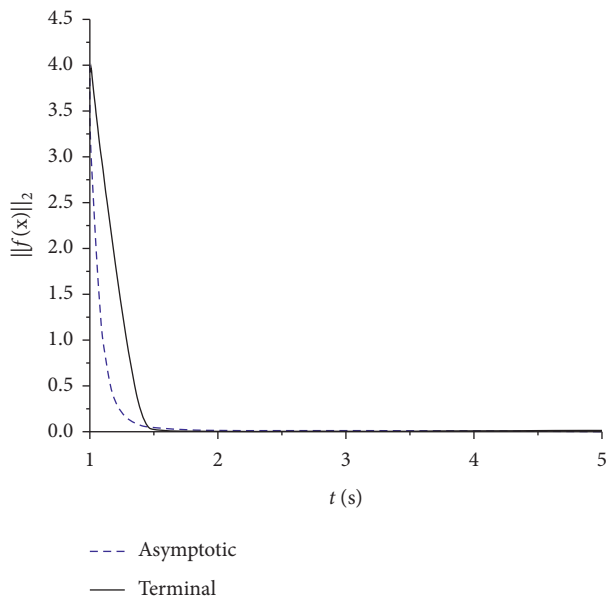
FIGURE 6: Comparison of $\|f(x)\|_2$ under terminal network and asymptotic network.
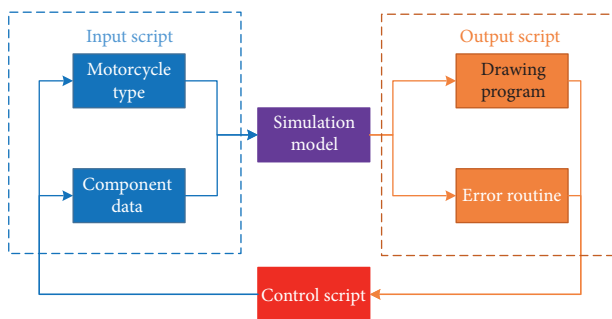
FIGURE 7: ADVISOR simulation process.

The choice of driving conditions will have a great impact on the performance of the test vehicle. The new European Urban Driving Cycle (NEDC) and Highway Fuel Economy Test (HWFET) were selected to carry out simulation tests. Table 2 shows the relevant parameters of the two commonly used driving conditions.

6.2. Comparison and Analysis of Simulation Results. After all the simulation work is completed, the simulation results of traditional instantaneous control, fuzzy neural network control, and terminal neural network control are compared. The three control strategies are compared in *SOC*, fuel consumption, and simulation time.

In Figure 8, NEDC_mdf, NEDC_fnn, and NEDC_tnn, respectively, represent the simulation results corresponding to the three schemes of traditional instantaneous control, fuzzy neural network control, and terminal neural network control of the hybrid electric vehicle model under NEDC conditions. HWFET_mdf, HWFET_fnn, and HWFET_tnn, respectively, represent the simulation results corresponding to the traditional instantaneous control, fuzzy neural network control, and terminal neural network control schemes of the hybrid electric vehicle model under HWFET conditions. As can be seen from Figure 8, the terminal neural network controller can approximately simulate the traditional instantaneous control rules, and the SOC value is maintained in the high efficiency range.

As can be seen from Table 3, the terminal neural network controller realizes the advantages of low fuel consumption of the fuzzy neural network control strategy and can bring good economy to the vehicle. As can be seen from Table 4, the simulation time of the terminal neural controller is significantly shortened compared with the other two control strategies, thus improving the real-time performance of the control.

TABLE 1: Parameters of HEV for simulation.

Automobile key module	Parameter	Value
Full vehicle	Mass	1605 kg
	http://dict.youdao.com/w/drag%20coefficient/	0.32
http://dict.youdao.com/w/engine/	http://dict.youdao.com/w/displacement/	2.5 L
	http://dict.youdao.com/w/maximum%20power/	118 kW
	Type	Hydrogen ballast
http://dict.youdao.com/w/battery%20pack/	Number of modules	34
	Capacity	6.5 Ah
	http://dict.youdao.com/w/nominal%20voltage/	244.8/650 V
Motor	Type	Permanent magnet synchronous
	http://dict.youdao.com/w/maximum%20power/	105 kW

TABLE 2: Two typical driving conditions.

Parameter	NEDC	HWFET
Cycle time (s)	1184	765
Maximum speed (km·h^{-1})	10.93	16.51
http://dict.youdao.com/w/operating%20range/ (km)	120	96.4
Average velocity (km·h^{-1})	33.21	77.58
http://dict.youdao.com/w/maximum%20acceleration/ (m·s^{-2})	1.06	1.43
Maximum deceleration (m·s^{-2})	−1.39	−1.48
Idle time (s)	298	6
Number of stopping	13	1
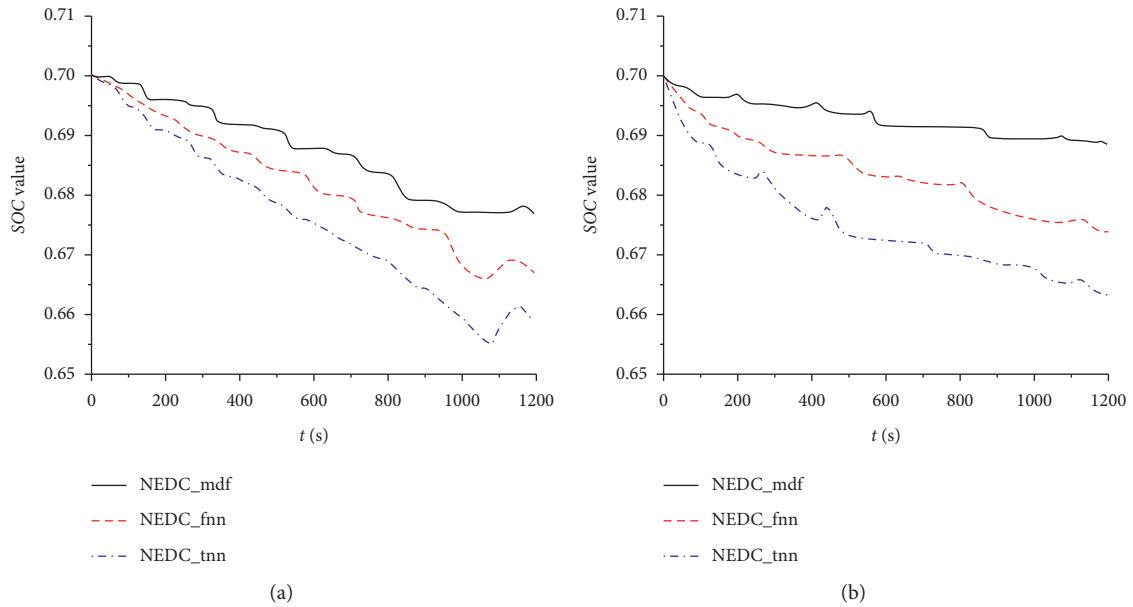
FIGURE 8: Comparison of power battery SOC. (a) NEDC. (b) HWFET.

TABLE 3: Fuel consumption per 100 kilometers.

	Conventional instantaneous control	Fuzzy neural network control	Terminal neural network control
NEDC (L·100 km^{-1})	10.1	9.8	9.8
HWFET (L·100 km^{-1})	9.2	7.8	8.0

TABLE 4: Simulation time.

	Conventional instantaneous control	Fuzzy neural network control	Terminal neural network control
NEDC (s)	45.6	102.3	15.6
HWFET (s)	35.2	89.4	11.3

7. Conclusions

In view of the shortcomings of traditional instantaneous control strategy, such as slow response and weak real-time performance, this paper uses neural network controller as torque control device to seek for more reasonable control of engine and motor output power of parallel strong hybrid electric vehicle. Under two typical operating conditions, ADVISOR software was used to simulate the hybrid electric vehicle model embedded with traditional instantaneous control, fuzzy neural network control, and terminal neural network control, respectively. The results show that the reaction time of the instantaneous control can be greatly reduced by the control strategy of the terminal neural network under the premise of ensuring the fuel economy and emission performance of the vehicle [27–30].

Data Availability

The experimental data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] S. Zhang, R. Xiong, and F. Sun, "Model predictive control for power management in a plug-in hybrid electric vehicle with a hybrid energy storage system," *Applied Energy*, vol. 185, pp. 1654–1662, 2017.

[2] J. Cao and R. Xiong, "Reinforcement learning-based real-time energy management for plug-in hybrid electric vehicle with hybrid energy storage system," *Energy Procedia*, vol. 142, pp. 1896–1901, 2017.

[3] S. Suganya, S. Charles Raja, and P. Venkatesh, "Smart management of distinct plug-in hybrid electric vehicle charging stations considering mobility pattern and site characteristics," *International Journal of Energy Research*, vol. 41, no. 14, pp. 2268–2281, 2017.

[4] M. G. Arend and T. Franke, "The role of interaction patterns with hybrid electric vehicle eco-features for drivers' eco-driving performance," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 59, no. 2, pp. 314–327, 2017.

[5] A. Cerofolini, "Optimal supervisory control strategy for a transmission-mounted electric drive hybrid electric vehicle," *International Journal of Automotive Technology*, vol. 20, no. 4, pp. 663–677, 2019.

[6] C. Xiang, F. Ding, W. Wang, and W. He, "Energy management of a dual-mode power-split hybrid electric vehicle based on velocity prediction and nonlinear model predictive control," *Applied Energy*, vol. 189, no. 5, pp. 640–653, 2017.

[7] C.-M. Lai, Y.-H. Cheng, M.-H. Hsieh, and Y.-C. Lin, "Development of a bidirectional DC/DC converter with dual-battery energy storage for hybrid electric vehicle system," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 2, pp. 1036–1052, 2018.

[8] Z. D. Asher, D. A. Baker, and T. H. Bradley, "Prediction error applied to hybrid electric vehicle optimal fuel economy," *IEEE Transactions on Control Systems Technology*, vol. 26, no. 6, pp. 1–14, 2017.

[9] X. Zhou, D. Qin, and J. Hu, "Multi-objective optimization design and performance evaluation for plug-in hybrid electric vehicle powertrains," *Applied Energy*, vol. 208, no. 12, pp. 1608–1625, 2017.

[10] Y. Chao, X. Jiao, L. Liang, Z. Yuanbo, and C. Zheng, "A robust H∞ control-based hierarchical mode transition control system for plug-in hybrid electric vehicle," *Mechanical Systems and Signal Processing*, vol. 99, no. 6, pp. 326–344, 2018.

[11] R. Koubaa and L. Krichen, "Double layer metaheuristic based energy management strategy for a Fuel Cell/Ultra-Capacitor hybrid electric vehicle," *Energy*, vol. 133, no. 8, pp. 1079–1093, 2017.

[12] S. Ahmadi, S. M. T. Bathaee, and A. H. Hosseinpour, "Improving fuel economy and performance of a fuel-cell hybrid electric vehicle (fuel-cell, battery, and ultra-capacitor) using optimized energy management strategy," *Energy Conversion and Management*, vol. 160, no. 5, pp. 74–84, 2018.

[13] B. Duan, Q. Wang, X. Zeng, Y. Gong, D. Song, and J. Wang, "Calibration methodology for energy management system of a plug-in hybrid electric vehicle," *Energy Conversion and Management*, vol. 136, pp. 240–248, 2017.

[14] Z. Zhao, D. Lei, J. Chen, and H. Li, "Optimal control of mode transition for four-wheel-drive hybrid electric vehicle with dry dual-clutch transmission," *Mechanical Systems and Signal Processing*, vol. 105, pp. 68–89, 2018.

[15] M. Kohli, L. M. Prevedello, R. W. Filice, and J. R. Geis, "Implementing machine learning in radiology practice and research," *American Journal of Roentgenology*, vol. 208, no. 4, pp. 1–7, 2017.

[16] Z. Lei, D. Sun, Y. Liu et al., "Analysis and coordinated control of mode transition and shifting for a full hybrid electric vehicle based on dual clutch transmissions," *Mechanism and Machine Theory*, vol. 114, pp. 125–140, 2017.

[17] X.-x. Yin, Y.-g. Lin, W. Li, Y. J. Gu, H. W. Liu, and P. F. Lei, "A novel fuzzy integral sliding mode current control strategy for maximizing wind power extraction and eliminating voltage harmonics," *Energy*, vol. 85, pp. 677–686, 2015.

[18] L. Bai, J. Guo, X. Tianyou, and Y. Minghai, "Emotional Monitoring of Learners Based on EEG Signal Recognition," *Procedia Computer Science*, vol. 174, pp. 364–368, 2020.

[19] G. Han, X. Pu, Z. He, and C. Liu, "Integrated planning and allocation: a stochastic dynamic programming approach in container transportation," *Chaos, Solitons & Fractals*, vol. 114, pp. 264–274, 2018.

[20] G. C. Calafiore and L. Fagiano, "Stochastic model predictive control of LPV systems via scenario optimization," *Automatica*, vol. 49, no. 6, pp. 1861–1866, 2013.

[21] L. Ming, Y. Ying, L. Liang, L. Yao, and W. Zhou, "Energy management strategy of a plug-in parallel hybrid electric vehicle using fuzzy control," *Energy Procedia*, vol. 105, pp. 2660–2665, 2017.

[22] J. Chai, J.-E. Guo, L. Meng, and S.-Y. Wang, "Exploring the core factors and its dynamic effects on oil price: an application on path analysis and BVAR-TVP model," *Energy Policy*, vol. 39, no. 12, pp. 8022–8036, 2011.

[23] Z. Yuan and L. Yuan, "Existence and global convergence of periodic solution of delayed neural networks," *Mathematical and Computer Modelling*, vol. 48, no. 1, pp. 101–113, 2017.

[24] J. C. Juinne-Ching Liao and S. N. Sheng-Nian Yeh, "A novel instantaneous power control strategy and analytic model for integrated rectifier/inverter systems," *IEEE Transactions on Power Electronics*, vol. 15, no. 6, pp. 996–1006, 2000.

[25] K. Furukawa, H. M. Tarafdar, T. Ise, and S. Kumagai, "A control strategy for active filters using quasi-instantaneous positive sequence extraction filters," *Electrical Engineering in Japan*, vol. 148, no. 1, pp. 54–65, 2010.

[26] L. Wu, Y. Wang, X. Li, and J. Gao, "Deep attention-based spatially recursive networks for fine-grained visual recognition," *IEEE Transactions on Cybernetics*, vol. 49, pp. 1791–1802, 2018.

[27] W. Bai, B. Zhang, Q. Zhou, and R. Lu, "Multigradient recursive reinforcement learning NN control for affine nonlinear systems with unmodeled dynamics," *International Journal of Robust and Nonlinear Control*, vol. 30, no. 8, pp. 17–29, 2019.

[28] H. Fathabadi, "Fuel cell hybrid electric vehicle (FCHEV): novel fuel cell/SC hybrid power generation system," *Energy Conversion and Management*, vol. 156, no. 6, pp. 192–201, 2018.

[29] T. K. Nizami and A. Chakravarty, "Neural network integrated adaptive backstepping control of DC-DC boost converter," *IFAC-PapersOnLine*, vol. 53, no. 1, pp. 549–554, 2020.

[30] K. Khnissi, C. B. Jabeur, and H. Seddik, "A smart mobile robot commands predictor using recursive neural network," *Robotics and Autonomous Systems*, vol. 131, pp. 84–90, 2020.

Research Article

Design and Implementation of Data Sharing Traceability System Based on Blockchain Smart Contract

Yang Kang and Qiang Li [ID]

College of Artificial Intelligence, Chongqing University of Arts and Sciences, Yongchuan 402160, Chongqing, China

Correspondence should be addressed to Qiang Li; 20160006@cqwu.edu.cn

"Traceability" is an important method often used in modern supply management. The traceability system is a system based on the Internet of Things technology. In this process, users will share resources through cloud service providers, so how to ensure data security is also one of the issues we consider. Blockchain technology is an emerging technology in the field of information technology. Its decentralized nature, distributed storage, and difficult data modification provide us with fair exchange and sharing of data. Feasible solutions: in this article, we have studied the key issues of fair exchange and safe sharing of data based on blockchain and designed a multigroup data sharing scheme based on alliance chain. To solve the various existing traceability systems problem, this paper designs a new traceability system based on blockchain technology and implements a system prototype to verify the feasibility of the system.

1. Introduction

Imagine we went to a Chinese restaurant for dinner. The waiter led us to the self-service ordering area, there are all kinds of vegetables and meat in front of us, and there is a miniature two-dimensional code in front of these dishes. We scan the two-dimensional code through our mobile phones to obtain the production place, processing process and parameters, logistics information, and so on. This experience undoubtedly increases our dining experience. With the awareness of food safety deeply rooted in people's hearts, many agricultural products-related enterprises began to establish their own agricultural products tracking systems to help consumers track the production and processing data of agricultural products [1], enhance consumers' purchasing confidence, and enhance their own market competitiveness.

Because of its characteristics, blockchain can provide a security framework for transactions on the network [2], and it is also a trusted cryptosystem. Blockchain network system is an infrastructure, which can protect content and track network operations, so it has attracted the attention of many companies or managers. The Personal Health Record (PHR)

system is problematic in some cases because it contains a large amount of private data of patients, so it is troublesome for patients to approve medical personnel to enter the system in an emergency. Therefore, the main purpose of a health care management framework based on blockchain technology is tamper protection. In view of the lack of product data sharing, fake and shoddy goods will flood, and how to prevent them has become an important issue [3]. How to use blockchain technology to share product data end-to-end in cross-border environment is to use a structure to track the process of manufacturing goods and prevent counterfeit goods. Although data sharing is practical. Researchers put forward a new CP-ABE scheme with multiauthority and large universe, which is the sharing of multigroup fine-grained data in the cloud Internet of Things system based on blockchain, which can ensure the integrity of data, introduce public audit, and realize tracking [4]. Experiments show that the scheme can be used in large-scale Internet of Things systems in terms of security analysis. Ground rescue needs data transmission support to be efficient and safe. Un-manned Aerial Vehicle (UAV) can perform urgent rescue tasks and share data with the ground, which also involves a

network, unreliable behavior tracking, and so on, which makes the data sharing between vehicles and UAVs pose a security threat [5]. To solve this problem, a lightweight vehicle blockchain-supported secure (LVBS) data-sharing framework is proposed, which not only improves the security but also optimizes the quality of data sharing strategies. In the process of placing orders and arriving goods in online shopping, goods are often exchanged for no reason. Therefore, we propose a decentralized database that can be maintained collectively by using the high reliability and high confidentiality of blockchain, which provides a way for logistics confidentiality and produces a modern logistics environment [6]. An intelligent antiswitching package for logistics system traceability based on blockchain is proposed. Its function is to enable customers to track their location, temperature, and humidity at any time, thus realizing intelligent logistics. Alliance blockchain technology is a multigroup data sharing scheme based on blockchain [7], which enables users to verify each other whether the shared data is valid without a third party. Some existing schemes cannot realize data sharing among different users, and then this scheme solves the problem of data sharing among different groups and has security and reliability. Combined with this epidemic, the potential application and opportunities of epidemic data deposition and sharing are combined to fight against novel coronavirus pneumonia epidemic, and the potential application of blockchain in this area is found. The advantage of blockchain is that it can provide high performance and security for the stored data and can be accessed globally so that the stored related medical data can be shared [8]. Blockchain data sharing is one of the most influential and practical methods available for health care data management. The medical data sharing and privacy protection eHealth system based on blockchain, namely, SpChain [9], has higher security and higher data retrieval rate in EMR sharing and management scheme than the traditional eHealth system based on blockchain. In addition, key blocks and micro blocks are established to store EMRS of patients.

However, these agricultural product traceability systems also have shortcomings, mainly including the following problems:

(1) Low data security:

Traditional agricultural products traceability system mostly adopts B/S design mode; resources are concentrated in a certain institution or organization; system centralization is serious; and data is easy to be tampered with, resulting in low data security.

(2) Low trustworthiness:

Because each node of the agricultural products supply chain is independent, it is easy to produce information gap and form information island, and the trust between them is greatly reduced.

(3) It is difficult to obtain source data:

The core of the agricultural product traceability system is data, which comes from every link of the agricultural product circulation process. Under the

traditional supply chain mode, there are no clear requirements and regulations for the participants in the supply chain, leading to great differences among the participants, unbalanced distribution of resources and technologies, unified management standards among each link, and lack of some key technologies, which make it difficult to collect data and information of agricultural products source.

2. Characteristics of Blockchain Technology and Introduction of Related Technologies

Blockchain is a way to organize data skillfully, while blockchain technology [10] is a technology that integrates many outstanding achievements of human intelligence, including database technology and network technology related to distributed books, network technology related to consensus algorithm, and asymmetric encryption technology and software technology related to intelligent contracts. Blockchain technology, which integrates these cutting-edge technologies, has challenged and confronted the mechanism based on central node credit since it was put forward. With the development of time, this characteristic of blockchain technology is becoming more and more obvious, which makes it no longer limited to digital cryptocurrency [11], but widely used in many fields of social and economic life, such as supply chain finance, data authentication, asset management, election voting, and fair security traceability introduced in this paper. Nowadays, blockchain technology has become an important technology that cannot be ignored among many high-end cutting-edge technologies.

2.1. Distributed Book. Unlike the traditional system network structure, the network architecture and data storage of the blockchain system choose distributed books with a decentralized architecture. The transmission of information in the network does not pass through the central node, and each member node retains complete data, which is embodied in the distribution of nodes in entity and the distribution of nonrelational databases on each node in logic. Account book is a storage form of data and a management unit of data. Nonrelational database is the logical carrier of data and account book, and the physical carrier of data and account book is the member node. Blockchain is the underlying form of an account book, that is, the data form of a chain of blocks strung together. After the consensus process, members will reach a consensus and write the same data into the blockchain system, so each member keeps a copy of the data. Because the data in each ledger cannot be deleted, the total amount of data will only increase, which is a great test for the ability of data storage. Still, with the rapid development of technology, this will not be a problem. The breeding nodes, slaughtering nodes, processing nodes, transportation nodes, and sales nodes in the livestock products supply chain are distributed in different places in the physical sense, in line with the distributed ledger technology in the blockchain. In the ideal traceability model, the status of nodes where each link is located is expected to be equal, the information flow

does not pass through the central node, and its network architecture naturally fits the characteristics of distributed books in blockchain technology, so it is reasonable and efficient for livestock products traceability system to select the network architecture based on blockchain technology.

2.2. Cryptographic Algorithm.
Cryptography is the core of the blockchain data layer, which escorts the whole data transmission and access of blockchain. Cryptographic algorithms will be used in three places in the blockchain platform of cold chain drug traceability, including hash algorithm and asymmetric encryption algorithm. This section will discuss the principle and selection of these two algorithms.

2.2.1. Hash Algorithm.
The secure hash algorithm is an alias of the hash algorithm that maps data of arbitrary length into hash strings of fixed length through certain rules. The more common ones are SHA-256, SHA-384, SHA-512, and so on, which are widely used in various technologies requiring encryption, including blockchain technology [12].

2.2.2. Asymmetric Encryption Algorithm.
Asymmetric encryption algorithm comes from the secret way of the key in the cryptographic algorithm. The key consists of a public key and a private key. The keys are generated at the same time, in which the public key is public as its name implies, while the private key is private and needs to be saved by itself. When encryption is needed, one of the same pairs of keys is used to complete the encryption operation, and only the other key can complete the decryption of the message. Asymmetric encryption algorithm plays an important role in blockchain and is the cornerstone of the security of the whole blockchain system. The asymmetric encryption algorithm is mainly used in encrypted communication and digital signature in blockchain [13].

2.3. SM2 Digital Signature Algorithm.
SM2 algorithm is an improved national cryptographic algorithm based on ECC published in China in 2010, which includes digital signature algorithm, key exchange protocol, and public key encryption algorithm. It has become the current public key algorithm standard GM/T 0003.2-2012 in China and entered the international standard ISO/IEC14888-3.

2.3.1. Principle of SM2 Digital Signature Algorithm.
The elliptic curve equation E, (A, B) selected by the SM2 algorithm is as follows:|

$$y^2 = x^3 + ax + b. \tag{1}$$

The Abelian group $(G, +)$ is defined on Ep (a, b), where the unit element O of the group is defined as an infinity point. Let the elliptic curve be $y^2 = x^3 - x$. Take two points P and Q on the elliptic curve $(G, +)$ arbitrarily, make a straight line intersecting the other point R' of the elliptic curve, make a parallel line of y-axis intersecting R through R′, and define

$P + Q = R$. In this way, the sum of the additions made by two points on the Abelian group will also be on the elliptic curve and also satisfy the properties of the Abelian group. When the same points P are added, that is, when the two points P and Q coincide, the tangent of the point P is made. Repeat the above operation, and when k points P are added, it is recorded as k_P.

Ordinary elliptic curves are continuous and not suitable for encryption. In the national standard, the elliptic curves with 256 bits in prime number field are used to transform the elliptic curves into discrete points. Let the prime number be P, the prime number field formed by the elliptic curves Fp, and the elliptic equation Erp (a, b):

$$y^2 = x^3 + ax + b \,(\mathrm{mod}\, p), \tag{2}$$

where a and b are elements in the prime field and satisfy $4a^3 + 27\, b^2 \neq 0 \,(\mathrm{mod}\, P)$. SM2 can be encrypted by assuming that, in the elliptic curve $P = d_G$, where P and G are two points on Ep (a, b), n is the order of G ($nG = O \,\infty$), and d is an integer and its value is not higher than order n. Given D and G, it is easy to compute P in the front direction but difficult to compute D in the reverse, according to the definition of Abelian group addition rule. Let the set $\{O, G, 2G, 3G, (n-1) G\}$ be a cyclic subgroup generated by G on E: (a, b). If d is selected as the private key and $P = dG$ as the public key, the problem of finding d from P is the discrete logarithm problem (ECDLP) on elliptic curve group, which is the mathematical basis of elliptic curve encryption algorithm [14].

2.3.2. SM2 Digital Signature Scheme.
The digital signature scheme designed by SM2 encryption principle is as follows: key generation: input SM2 elliptic curve parameter *parms* (elliptic curve equation Ep, large prime number p, base point G, order n of base point); the private key d is randomly generated and kept secret, and the public key is generated by using the relationship between public and private keys:

$$P = [d]G. \tag{3}$$

In the formula, P is the obtained SM2 public key, which is an important basis for SM2 to be used in encryption and signature. SM2 signature process: input SM2 elliptic curve parameter *parms*, private key d, and message M to be signed, and calculate hash value Z_A at the same time:

$$Z_A = H_{256}\left(I\ DL_A \,\|\, ID_A \| a \| b \| x_G \| y_G \| x_A \| y_A\right). \tag{4}$$

ID_A is the user's distinguishable identification, IDL_A is the length of ID_A, a and b are the coefficients of elliptic curve, x_G and y_G are the horizontal and vertical coordinates of base point G, respectively, and $x_A \| y_A$ is the horizontal and vertical coordinates of public key, respectively.

The hash value Z_A is obtained, and then the hash digest e with the message M to be signed is calculated:

$$e = H_{256}\left(Z_A \| M\right). \tag{5}$$

Randomly generate $k \in [1, n-1]$ and calculate the elliptic curve point $X1$ from it:

$$X_1 = (x_1, y_1) = [k]G. \tag{6}$$

Calculate the signature parameters r and s to output the signature (r, s):

$$
\begin{aligned}
r &= (e + x_1)\bmod n, \\
s &= \left((1 + d)^{-1}(k - rd)\right)\bmod n.
\end{aligned}
\tag{7}
$$

SM2 verification process: input $parms$, the public key P owned by the verifier, the message M to be verified, and the signature (R'', S') sent by the signer to obtain the hash digest E of the message M to be verified according to formula (5), and calculate the T value:

$$t = (r' + s')\bmod n. \tag{8}$$

Verify whether t is equal to zero. If it is equal to zero, the verification fails; otherwise, calculate the elliptic curve point $X'1$:

$$X_1' = (x_1, y_1) = [s']G + [r']P_A. \tag{9}$$

Verify whether $r' = (e' + x1')$ mode is true. If it is true, the verification is successful.

2.4. Consensus Algorithm.

At present, the consensus algorithms commonly used in blockchain are Pow, Pos, DPos, and PBFT. Pow uses the workload proof mechanism to determine the node accounting right and ensures the consistency of data through a large number of calculations. Pos adds the concept of currency age to the workload proof mechanism and uses currency age instead of computing power to reduce the difficulty of block solution. DPOS uses digital cash holders to generate node sets with bookkeeping rights through elections and adopts rotation bookkeeping to generate blocks. PBFT is a distributed consensus algorithm based on state machine replication. Consensus is completed through three-stage communication between nodes. For the four consensus mechanisms in throughput, the performance of delay and scalability is compared. The application of the consensus mechanism in blockchain ranges from PoW to PoS and then to DPOS and PBFT. In this process, computing power competition is gradually replaced by equity competition. The cost of obtaining node bookkeeping rights is gradually reduced. With the development of the blockchain application, the transaction speed gradually increases. At the same time, the consensus mechanism has gradually evolved from the initial decentralization to the current weak centralization. Each consensus mechanism has different advantages and disadvantages. PoW can realize decentralization to the greatest extent, with safe, reliable, and low consumption of network resources. Still, it consumes too much computing resources. The computational attack is easy to occur, and the consensus time is longer. The emergence of PoS solves the problem of excessive waste of PoW computing power, but it makes the pressure of network traffic increase. The implementation is more complicated. DPOS has made appropriate compromises on centralization. The consumption of network resources is reduced, the consensus time is greatly shortened, and the system throughput is improved. However, due to its weak centralization, the security is reduced, which is prone to security loopholes. The PBFT algorithm solves the problem of a distributed system with Byzantine error nodes and improves the fault tolerance rate of the distributed system. However, due to the existence of a large number of point-to-point communications, it occupies a large number of communication resources [15].

2.5. Smart Contracts.

More than ten years before blockchain technology was proposed, the idea of smart contract existed. However, because there was no corresponding underlying technical support at that time, it had been unknown until the blockchain developed into the 2.0 era. Combining intelligent contract with virtual digital cash technology becomes an important part of programmable blockchain technology [16, 17]. Intelligent words in intelligent contracts reflect their ability to judge and execute independently; the word contract shows that it needs certain rules, just like contracts, as the standard of independent judgment. When intelligent contracts judge that a certain behavior conforms to this rule, they will perform corresponding operations accordingly. Self-judgment is the task that computer programming language is good at accomplishing, and self-execution is also well realized by programming language. Because the intelligent contract takes the programming language as the carrier, that is to say, it is a written program in practical application, whether the data triggers its conditions and executes corresponding behaviors is completely determined by the machine, and human intervention is impossible. This is very consistent with the design purpose of blockchain technology to ensure that data cannot be tampered with artificially. After the intelligent contract is integrated into the blockchain technology, all kinds of business can be written into it to judge and execute spontaneously, which greatly reduces the influence of human factors, improves the credibility of transaction data, and makes the blockchain technology not limited to the currency circulation field but gradually integrated into many other leading cities in social and economic life.

3. The Basic Principle of Blockchain Uses Double-Chain Traceability System Design

3.1. Design of Traceability System

3.1.1. System Requirements Analysis.

This paper describes the traceability of information in the traditional blockchain production process and eliminates the risk of information tampering in the blockchain production process. Using the decentralized characteristics of blockchain, a decentralized traceability system is constructed.

Simply using an existing high-reliability public chain or a license chain maintained by multiple system participants as a carrier for storing traceability information has drawbacks. It will multiply the cost, and the processing speed will be slow. Storing traceability information in the license chain can

better solve the cost and speed problems. However, storage in the license chain faces the risk of small nodes and easy attack. This requires a much cost to solve the security problem. The public blockchain has the advantage of reliability, but it needs to invest many computing resources. Therefore, this paper will design a traceability scheme combining double chain, public chain, and license chain. The traceability system needs to be able to locate the responsible person; that is, it needs to be able to find the input person of each traceability information. In the traceability system described in this paper, the traceability information is entered by publishing the traceability information of transaction records and storing the traceability information in the blockchain using one's own account. Because publishing transactions in the blockchain requires an account signature, in the license chain, you only need to find the account that published the traceability record to know who has entered the traceability information into the traceability system; that is, the traceability system described in this paper has the function of locating the responsible person.

3.1.2. Overall Design.

In the traceability system designed in this paper, the blockchain operated by nodes related to the traceability system is used to store the product traceability records recorded by each traceability node. In order to enable the traceability system to process a large amount of traceability information, a license chain is used to store traceability information in the traceability system described herein. In order to improve the security of the above license chain, license chain block authentication information must be stored in the public chain to check whether the data in the license chain has changed. Universal chains use highly reliable chains, such as Ethereum and Bitcoin. The overall design of the traceability system is shown in Figure 1.

As shown in Figure 1, each node of the source nodes sequentially adds source information to the authorization chain. Hashes of blocks of several license chains are stored in a database in the form of a Merkel tree, the roots of which are stored in the public chain. How the license chain stores traceability information will be described in detail later.

Figure 1 shows that the manufacturer can generate labels for each product. The user can find the location of the product corresponding to the last trace information through the tag, and the trace information can find the faster source information in the license chain. For each source information stored in the license chain, the user can obtain a block title corresponding to the license chain and can verify whether the source information is actually stored in the license chain by Merkel authentication. For a block title in the license chain, the user can compare the block title with the information stored in the shared chain and use Merkel to verify whether the block has been tampered with.

3.1.3. License Chain Design.

In the source tracking system described herein, source information is stored in the license chain through transactions. In the source tracking system described herein, source information is stored in the form shown in Table 1. Trace information includes information

recorded at a trace node and the transaction number for which the previous trace information is recorded in the license chain. In addition to the above information, the traceability information will also include the signature of the person in charge and the information of the person in charge holding the public key for locating the relevant person in charge.

In this tracking system, the person entering the source information has an account of the license chain. The inputs of the source information start the transaction, which is all through the traceability information recorded in the account. So, the account signature must be used at the beginning of the transaction. At the beginning of the transaction, the mining node of the license chain will verify and protect the legitimacy of the transaction. In the license chain, as long as the account of the transaction with additional record tracking information is found, it can be determined that the recorder of the transaction information is responsible for positioning. The MPT tree stores the correspondence between the product number and the last transaction tracked by the product. This tree is similar to the Ethereum state tree. Each worksheet node records the status of each product, including transaction numbers and book labels. Actions that record the latest product tracking information can be found in the license chain through transaction numbers. Operations that record information from other sources can be found through transactions. Bookmarks are used to prevent labels from being copied. Bookmarks are initialized to false. If the product has been tracked and questioned, its value should be changed. If the value is true, the information stored in the corresponding form node will not change. In this way, consumers who have restored the origin information of the product know that the product has been restored.

3.1.4. Public Chain Design.

This paper discusses two methods of storing hash validation of license chain blocks in the public chain.

(1) In the open chain, a hash transaction storing the title of the license chain block is published.

(2) The smart contract expanded in the public chain stores the hash of the license chain block according to this contract.

In this paper, Scenario 2 hopes that, by expanding the intelligent contract on the open blockchain, anyone can observe the logic of the contract and the hash verification of each block of the license chain stored in the contract. If a transaction for a burst column storing a license key block is started directly in the shared chain, other systems need to provide a method of finding a transaction storing a hash of the corresponding block in the shared chain. However, for a blockchain such as Ethereum, the cost of Scheme 2 is too high. The license chain generates a transaction that calls the smart contract every 10 seconds, starts the transaction at the public chain store every 10 seconds, and calls the saved source information of the smart contract. After reducing the occurrence speed of license blocks, the response speed of stored information of the source system will slow down. That
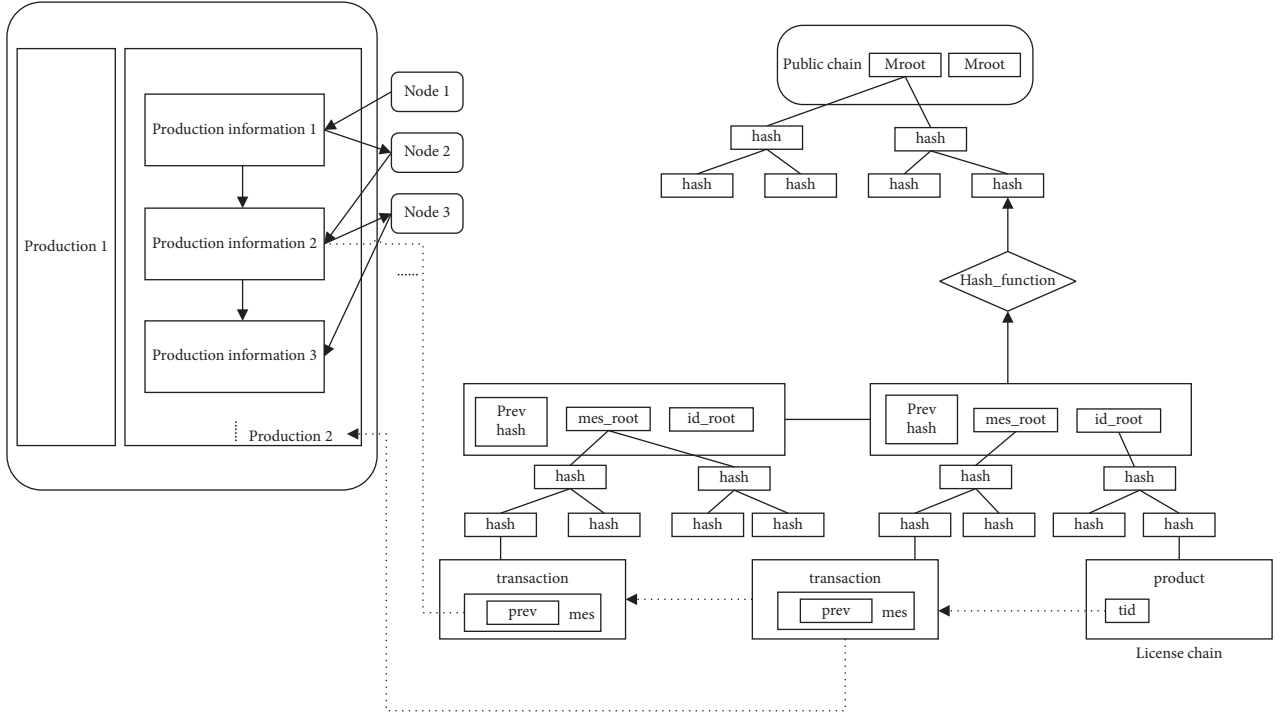
Figure 1: Overall design drawing of the traceability system.

Table 1: Storage structure of traceability records.

Traceability information
Production information
Signature of the relevant responsible person on production information
Record the transaction number of the previous traceability information

$$H_c = H(H_B + H_A), \qquad (10)$$

$$H_B = H(H_D + H_C), \qquad (11)$$

$$H_G = H(H_F + H_E), \qquad (12)$$

is, the source node can store the source information in the tracking system for a long time after starting the tracking information request.

In order to ensure the response speed of the information stored in the tracking system and reduce the cost, the source tracking system in this paper forms a lava tree from several blocks of continuous authorization chains and stores the roots of this lava in the public chain. When verifying whether the block has been tampered with, the block's validity is verified by using the Merkel tree [18] to which the Merkel verification block belongs. In the license chain, the block head acts as the leaf nodes of two Merkel trees, which is used to verify whether a block in the license chain has been tampered with. The block needs to extract all nodes and their siblings on the root node path from the leaf nodes of the corresponding Merkel tree and its leaf nodes. Then, verify that the hash values stored by these nodes in the parent node are all child node information and hash values. If the above conditions are satisfied and Merkel's root coincides with Merkel's root stored in the shared chain, the block corresponding to the permission chain is considered not to have been tampered with. As shown in Figure 2, to verify node a, it is necessary to verify the accuracy of the following formulas:

3.1.5. Entry and Query System Requirements. In the source tracing system described in this article, desktop software is used to enter tracing information. The operation method is that the operator fills in the trace information or selects trace information from the drop-down menu. Then, the QR code recording the source information of the commodity is scanned, and the source information of the commodity is recorded in the tracking system. The specific implementation is described in the next chapter. For the search system, users can query the specific tracking information of products through the Internet.

3.2. Realization of Traceability Prototype System

3.2.1. Implementation of License Chain. Because the amount of complete and used blockchain is too large, in this paper, we build a blockchain that takes Taifang's private chain as a prototype system to store tracking information. This paper will implement and verify the license chain P2P network [19] and consistency algorithm designed [20]. The license chain of the prototype system is shown in Figure 3.

In the prototype system, to save the space of the private chain of Ethernet Square, the tracking information is stored in the coupon chain by using summary information, which
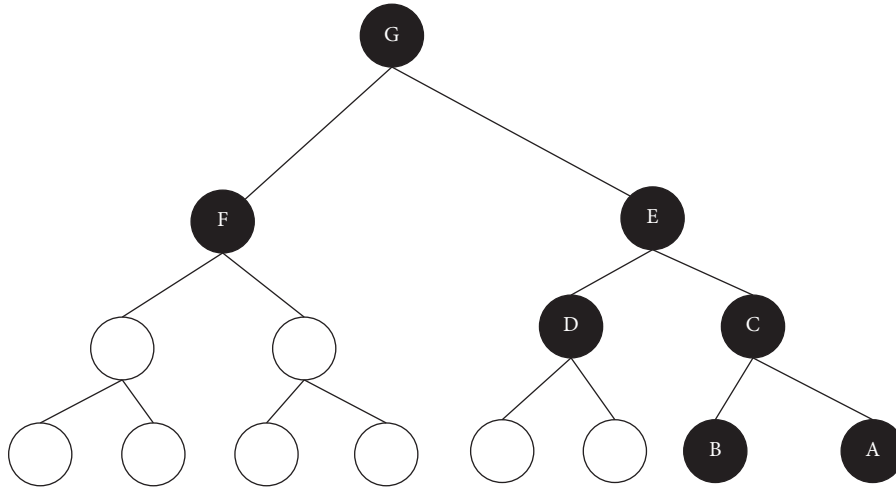
Figure 2: Merkel verification.
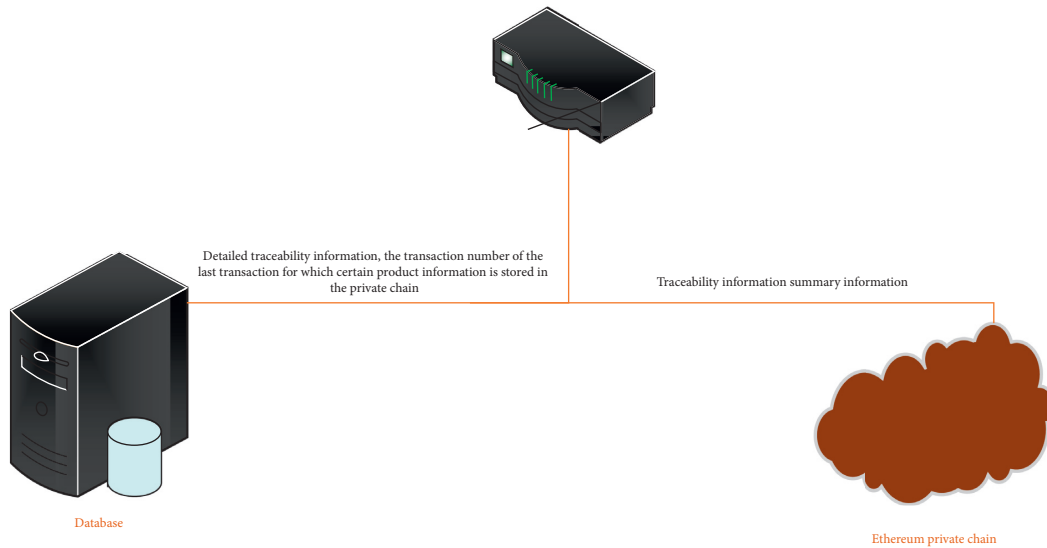
Figure 3: License chain in the prototype system.

includes hash containing source information and link of source information to the database. Detailed source information obtains the above link in the database and stores the detailed source information in a hash value in an Ethernet private chain for verifying the authenticity of the source information stored in the database. The key value of the link value is stored in the database. Here, link represents the detailed source information link, and value represents the detailed source information, both using MySQL VARCHAR type.

3.2.2. Realization of Public Chain.

The source tracing system [21] in this paper is to organize the blocks of the permission chain into the form of tag tree and store each private chain block in the Merkel tree with $2n-1$ nodes (including 2 $(n-1)$ leaf nodes). In the Merkel tree described above, each leaf node stores a hash value of one block of the permission chain, and the root node of the Merkel tree is stored in the public chain. The license chain block header is continuously

located in the database, which is in the form of key value, and the key value is the node number and the hash value of the value node. For each Merkel tree, assuming that the starting number of Merkel tree is start, the leaf node number of recording the first hash is start $+2n-1$, and then the cut-off hash leaf nodes allowed to be stored are added to the Merkel tree according to the increasing order of leaf node number. In the lava tree described in this chapter, the parent node number of n nodes is $(start + n)/2$, the sibling node number is $n + 1 - 2 * (n\% 2)$, and the parent node stores the hash value of all the child nodes. If each Merkel tree has $2N$ leaf nodes, the block of block number N calculates the start (start position) corresponding to the Merkel tree using the following, where int0 represents a downward modification:

$$start = int\left(\frac{n}{2^n}\right) * 2^{n+1} + 1. \tag{13}$$

The label key corresponding to the leaf node can also be obtained from

$$key = start + 2^n - 1. \tag{14}$$

In the prototype system, Ethernet Square [22] is adopted as a shared chain, and the intelligent contract deployed in Ethernet Square is used to access data. Smart contracts use solity to store the root node of each Merkel. The above contract only allows a specific account to perform storage operations and prohibits changes to data stored in mapping. If each Merkel tree has $2n$ leaf nodes, the block with block number n corresponds to the Merkel tree number: int ($n/2n$).

3.2.3. Double-Stranded Interaction.

In this paper, when investigating trace information. The background will confirm whether the block holding trace information in the verification license chain has been tampered with.

Merkel's verification process is as follows:

Step 1. In the database, find the leaf node of the Merkel tree of the storage block according to the method in Section 3.2.1.

Step 2. Calculate the key values of the sibling node and the parent node from the key values of the leaf node in the first step in the way of Section 3.2.1, and obtain the value values stored by the sibling node and the parent node in the database.

Step 3. Set the node to store the hash value $H1$ and the sibling node to store the hash value $H2$ and verify whether the value H stored by the parent node satisfies $H = H (H1, H2)$, if not, Merkel authentication fails.

Step 4. After obtaining the key value of the parent node, repeat the second step until the root node of the Merkel tree is reached.

Step 5. Compare the hash values stored in the Merkel tree root node to the hash values stored in the common chain.

In Step 3, the hash value is calculated using sha256, and in the prototype system, the sha256 function of the OpenSSL library is used. The calculation method of formula (15) is used to calculate $H (H1, H2)$. In equation (15), STR1 is a hash value stored in a node with a smaller number and STR2 is a hash value stored in a larger node. When validating, if the stored STR1 node number is odd, you need to exchange the hash value STR2 stored in STR1 and sibling nodes:

$$H (STR 1, STR 2) = sha 256 (STR1 + STR 2). \tag{15}$$

3.2.4. Entry of Traceability Information.

The logistics information module in the agricultural product traceability system mainly includes agricultural product ID, industrial and commercial information of sales enterprises, and other data. Upload is when the agricultural product ID is entered through the page, and the industrial and commercial information is obtained through the third-party interface to automatically fill in the interface. Click the upload button and upload the sales information to the blockchain network through the method PutState as shown in Figure 4.

Figure 4 is to realize the data uploading function of different traceability information, which has been mentioned in the article, including data uploading such as agricultural product ID and industrial and commercial information of sales enterprises, among which information such as agricultural product ID can be uploaded to the traceability system through Invoke interface.

3.2.5. Query of Traceability Information.

The traceability information query module in the agricultural product traceability system mainly queries the information of each link of the agricultural product supply chain on the blockchain network through the agricultural product ID as shown in Figure 5.

3.2.6. Detection System of Heavy Metal Ions in Soil.

For traceability of agricultural products, heavy metal concentration information of soil with crops is usually stored as traceability information in the traceability system. According to the blockchain technology described in this paper, the heavy metal inspection node is used as a tracking node of the source system, and the soil heavy metal ion information is recorded in the source system as an item of tracking information.

As shown in Figure 6, the soil heavy metal detection system described in this section mainly consists of three parts:

(a) Heavy metal ion concentration sensors, including copper and chromium ion concentration sensors.

(b) An analog-to-digital converter.

(c) A heavy metal ion concentration analysis system running on a single chip microcomputer.

4. System Test

The previous chapter discusses mainly building the environment and functional development of detailed design and analysis. This chapter discusses the traceability system of agricultural products function test and performance test, to ensure that the system can run normally and stably. It mainly includes unit, function, and performance tests.

4.1. System Unit Test.

System unit test is mainly to test the current chain code file method. The chain code of agricultural products traceability system studied in this paper uses go programming language, so the go test tool used in unit test and unit test file is shown in Table 2.

Let us take product.go as an example. In the chaincode directory, create a new product _ test.go file. The test function is as shown in Algorithm 1.

Execute the go test command, and the test result "pass" indicates that the test succeeded; otherwise, it failed.

4.2. System Function Test.

System function test is mainly to verify whether the blockchain network in the agricultural product traceability system is normal, and whether the data
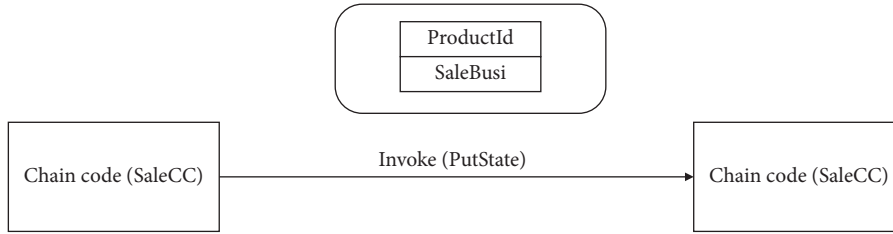
FIGURE 4: Upload model of agricultural product sales data.
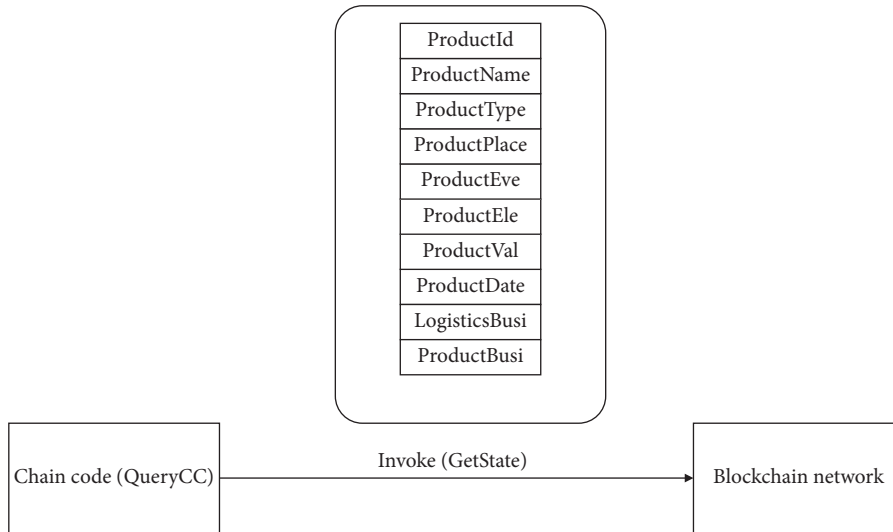
FIGURE 5: Traceability model of agricultural products.
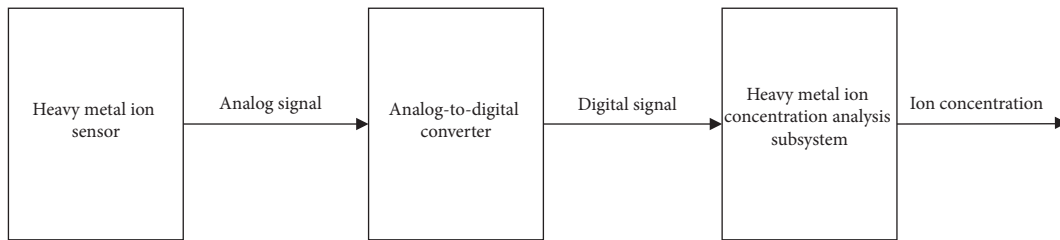
FIGURE 6: Soil heavy metal inspection system.

TABLE 2: List of unit test files.

Test content	Test results
product.go	Pass
process.go	Pass
logistics.go	Pass
sale.go	Pass
source.go	Pass

uploading and query function modules are normal. Specific verification data are shown in Table 3.

4.3. System Performance Test. The system performance test, also called the stress test or load test, mainly tests whether the system can normally work under a certain load. The performance test of the agricultural product traceability system studied in this paper mainly starts with two indicators: user concurrency and response time, through LoaderRunner software to simulate online users and through LoaderRunner software controller module for multiuser simulation.

In the controller interface, you can choose to set the number of users manually and then increase the test index in a gradual way to test the limit of the system. The traceability system of agricultural products studied in this paper adopts the working state of 100 people, 400 people, and 1000 people, and its test results are shown in Tables 4–6.

We also compared the performance of agricultural products traceability system in Heilongjiang Province. Under the same experimental environment, the results are as follows.

To show the superiority of performance more clearly, the blockchain system is compared with the traditional multiple traceability systems, and its average response time is plotted with the number of users. The results are as shown in Figure 7.

```
 package main
import(
   "fmt"
   "eccoding/json"
   "github.com/hyper ledger/fabric/core/chaincode/shim"
   pb "github.com/hyperledger/fabric/protos/peer"
   )
func TestUploadProductInfo(t *testing.T){//The test function name must begin with test and must accept a * testing.T type parameter
   got: = uploadProductinfo("orange", "orange01", "fruit", "sichuan", "xxx", "xxx")//Program output result
   want: = shim.Success(nil)//Expected result
   if !reflect.DeepEqual(want, got){
      t.Error("excepted:%v, got:%v", want, got)//Test failure output error prompt
      }
}
```

ALGORITHM 1: The product _ test Function.

TABLE 3: System function test table.

Test	Test content	Test results
Functional test project	Does the container in the blockchain network operate normally	Normal
	Is the interface of each link in the agricultural product supply chain displayed normally	Yes
Functional interface test	Inquire whether the traceability information interface is displayed normally	Yes
	Is page jump normal	Yes
	Whether the required entry is validated	Yes
Form validation test	Prompt for input errors	Yes
	Prompt for success or error of submission	Yes
Browser compatibility test	Is the display of different mobile phones normal	Normal

TABLE 4: Performance test of agricultural product traceability system based on blockchain.

Number of users	Average response time (ms)	Peak traffic response time	Test conclusion
100	0.925	1.145	Pass
400	1.318	1.923	Pass
1000	2.466	2.786	Pass

TABLE 5: Performance test of Heilongjiang agricultural products traceability system.

Number of users	Average response time (ms)	Peak traffic response time	Test conclusion
100	2.012	2.203	Pass
400	3.318	3.567	Pass
1000	4.466	4.498	Pass

TABLE 6: Performance test of the traditionally developed traceability system.

Number of users	Average response time (ms)	Peak traffic response time	Test conclusion
100	3.523	3.125	Pass
400	6.451	6.874	Pass
1000	7.355	7.845	Pass

As can be seen from Figure 7, the designed data sharing traceability system based on blockchain has a better response time, which means that our performance can meet the general requirements of the traceability system and is better.

4.4. *Comparison of Specific Functional Tests of Block Systems.* ART: average response time per request; TPS: the number of concurrent requests that can be responded to per second.
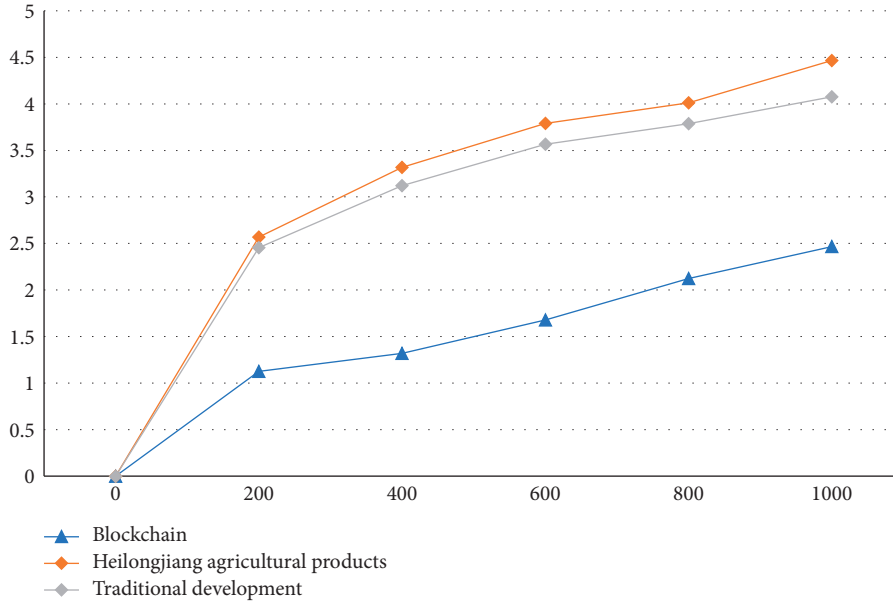
FIGURE 7: Performance diagram of various traceability systems.

TABLE 7: System query test results.

System	Number of query requests	1000	1500	2000	2500	3000	3500	4000
Blockchain	Mean response time (MS)	22	35	52	97	125	215	278
	Success rate (%)	100	100	100	100	100	100	100
Heilongjiang agricultural products	Mean response time (MS)	25	40	70	125	189	264	387
	Success rate (%)	100	99.7	99.6	99.5	99.4	98.4	99.3
Traditional development	Mean response time (MS)	32	54	87	165	221	312	452
	Success rate (%)	100	99.4	99.3	99.2	99.2	99.1	99
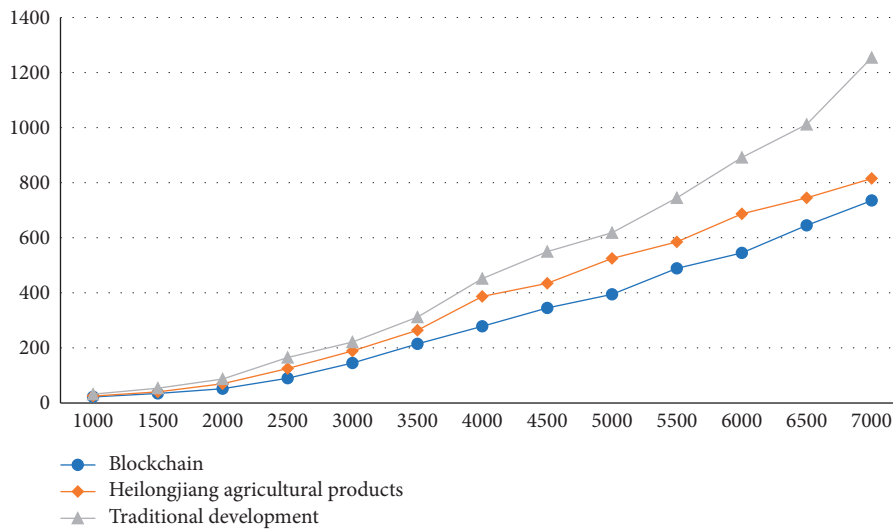
FIGURE 8: ART-concurrent information query request curve.
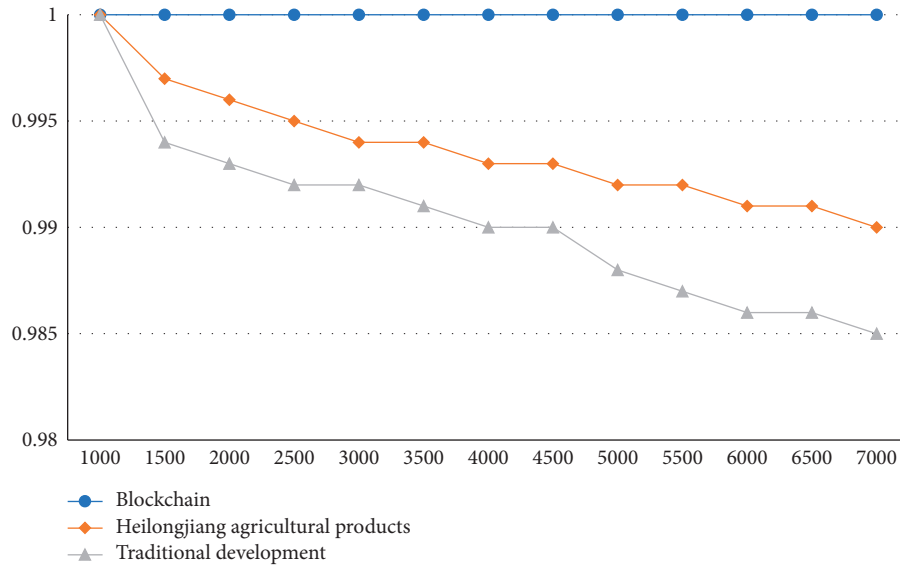
Figure 9: ART-concurrent information query request success rate.

Table 8: Business transaction processing function test.

System	Transaction volume	30	60	90	120	150	180	210
Blockchain	Mean response time (MS)	74	156	227	345	489	654	801
	Success rate (%)	100	100	100	100	100	100	100
Heilongjiang agricultural products	Mean response time (MS)	85	174	310	415	612	1021	1250
	Success rate (%)	100	99.5	99.4	99.4	99.3	99.2	99.1
Traditional development	Mean response time (MS)	112	225	398	521	787	1254	1545
	Success rate (%)	99.6	99.4	99.3	99.2	99.1	89.9	89.7
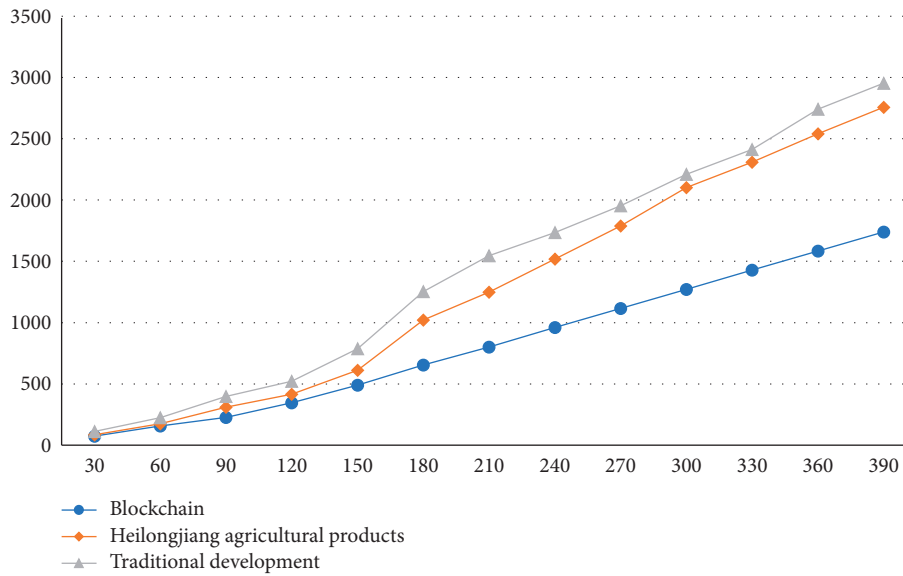
Figure 10: ART-concurrent completion transaction volume curve.
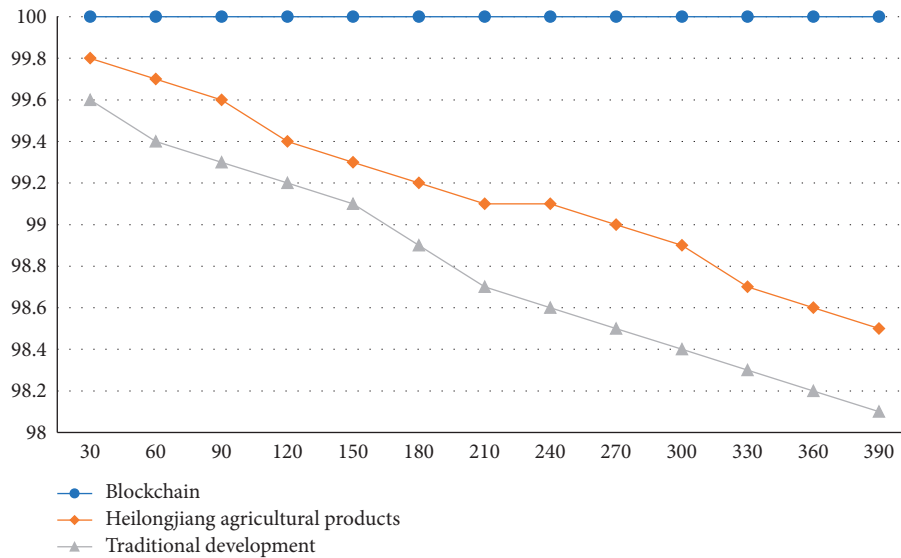
FIGURE 11: ART-success rate of concurrent business transaction.

Through the influence of ART and TPS indexes on the load performance of blockchain system because it is a simulation system, considering many factors, such as cost, time, and experimental conditions, according to the experimental test process, the system is reduced in equal proportion, which is convenient for the experiment.

4.4.1. Query Function Test. Grouped according to the number of query requests, there are seven groups for 1000, 1500, 2000, 2500, 3000, 3500, and 4000. The data are as shown in Table 7.

From the query time and success rate of concurrent information in Figures 8 and 9, the system based on the blockchain can have a very low average response time and the highest query success rate. In the other two methods, the average response time increases obviously with the increase in concurrency, but the success rate is getting lower and lower.

4.4.2. Business Transaction Processing Functional Testing. There are seven groups, 30, 60, 90, 120, 150, 180, and 200, which are grouped according to the transaction volume of concurrent requirements. The data are as shown in Table 8.

From the response time and success rate of completing business transactions in Figures 10 and 11, the system based on the blockchain can have a very low average response time and the highest query success rate. In the other two methods, the average response time increases obviously with the increase of concurrency, but the success rate is getting lower and lower.

5. Conclusion

In this paper, aiming at the shortcomings of the traditional agricultural product traceability system, such as low security, untrustworthiness, and difficulty in information collection, a data-sharing traceability system based on blockchain technology is proposed. The traceability system uses a license chain with high throughput to store traceability information and

realizes the function of locating the responsible person. To reduce the security overhead of the license chain, the verification information of the license chain is stored in the public chain with great computing power to ensure security, and its legitimacy is verified by Merkle verification. The security of the public chain is fully guaranteed. Future work considers different intelligent optimization algorithms [23, 24] in blockchain to further optimize the system and improve its performance.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] H. Lei, "Empirical analysis of the influence of tax on the agricultural products processing industry development under the constraint of financial index," *Boletin Tecnico/Technical Bulletin*, vol. 55, no. 15, pp. 570–578, 2017.

[2] A. R. Rajput, Q. Li, and M. T. Ahvanooey, "A blockchain-based secret-data sharing framework for personal health records in emergency condition," *Healthcare*, vol. 9, no. 2, pp. 206–215, 2021.

[3] V. Jaiman and V. Urovi, "A consent model for blockchain-based distributed data sharing platforms," *IEEE Access*, vol. 8, pp. 143734–143745, 2020.

[4] T. Li, J. Zhang, Y. Lin, S. Zhang, and J. Ma, "Blockchain-based fine-grained data sharing for multiple groups in Internet of Things," *Security and Communication Networks*, vol. 2021, Article ID 6689448, 13 pages, 2021.

[5] Z. Su, Y. Wang, Q. Xu, and N. Zhang, "LVBS: lightweight vehicular blockchain for secure data sharing in disaster rescue," *IEEE Transactions on Dependable and Secure Computing*, no. 99, p. 1, 2020.

[6] C.-L. Chen, Y.-Y. Deng, W. Weng, M. Zhou, and H. Sun, "A blockchain-based intelligent anti-switch package in tracing logistics system," *The Journal of Supercomputing*, vol. 77, no. 7, pp. 7791–7832, 2021.

[7] H. Huang, X. Chen, and J. Wang, "Blockchain-based multiple groups data sharing with anonymity and traceability[J]," *Science China Information Sciences*, vol. 63, no. 3, pp. 1–13, 2020.

[8] B. Zhang, C. W. Chao, Y. Tsybovsky et al., "A platform incorporating trimeric antigens into self-assembling nanoparticles reveals SARS-CoV-2-spike nanoparticles to elicit substantially higher neutralizing responses than spike alone," *Scientific Reports*, vol. 10, no. 1, Article ID 18149, 2020.

[9] R. Zou, X. Lv, and J. Zhao, "SPChain: blockchain-based medical data sharing and privacy-preserving eHealth system," *Information Processing & Management*, vol. 58, no. 4, Article ID 102604, 2021.

[10] Z. Zheng, S. Xie, H. Dai, X. Chen, and H. Wang, "An overview of blockchain technology: architecture, consensus, and future trends," in *Proceedings of the 2017 IEEE International Congress on Big Data (BigData Congress)*, pp. 557–564, Honolulu, HI, USA, June 2017.

[11] R. Saxena, D. Arora, V. Nagar, and S. Mahapatra, "Bitcoin: a digital cryptocurrency," in *Blockchain Technology: Applications and Challenges*, S. K. Panda, A. K. Jena, S. K. Swain, and S. C. Satapathy, Eds., vol. 203, , pp. 13–28, Intelligent Systems Reference Library, 2021.

[12] C. S. Shih and K. W. Yang, "Design and implementation of distributed traceability system for smart factories based on blockchain technology," in *Proceedings of the Conference on Research in Adaptive and Convergent Systems*, pp. 181–188, Chongqing, China, September 2019.

[13] O. I. Konashevych, "Data insertion in blockchain for legal purposes. How to sign contracts using blockchain," *Elektronnoe Modelirovanie*, vol. 41, no. 5, pp. 103–120, 2019.

[14] R. Shanthakumari and S. Malliga, "Dual layer security of data using LSB inversion image steganography with elliptic curve cryptography encryption algorithm," *Multimedia Tools and Applications*, vol. 79, no. 4, pp. 3975–3991, 2020.

[15] Y. Zhang, W. Yang, Q. Xue, J. Huang, and W. Che, "Broadband dual-polarized differential-fed filtering antenna array for 5G millimeter-wave applications," *IEEE Transactions on Antennas and Propagation*, no. 99, p. 1, 2021.

[16] T. G. Volkova, "Reasons and features of implementing blockchain technology in the pension system of the Russian federation," *Bulletin of Udmurt University. Series Economics and Law*, vol. 30, no. 3, pp. 333–339, 2020.

[17] P. Zhu, J. Hu, X. Li, and Q. Zhu, "Using blockchain technology to enhance the traceability of original achievements," *IEEE Transactions on Engineering Management*, no. 99, pp. 1–15, 2021.

[18] L. Tian and Y. Sun, "Research summary of blockchain fragmentation propagation mechanism based on Merkel tree," *Journal of Physics: Conference Series*, vol. 1914, no. 1, Article ID 012010, 2021.

[19] P. Podduturi, P. Ahmadi, K. Islam, and T. Maco, "RFID implementation in supply chain management using P2P network overlays," in *Proceedings of the 2019 Wireless Telecommunications Symposium (WTS)*, New York, NY, April 2019.

[20] P. Wang, P. Liu, and F. Chiclana, "Multi-stage consistency optimization algorithm for decision making with incomplete probabilistic linguistic preference relation," *Information Sciences*, vol. 556, pp. 361–388, 2020.

[21] A. L. Nitka, W. M. DeVita, and P. M. McGinley, "Evaluating a chemical source-tracing suite for septic system nitrate in household wells," *Water Research*, vol. 148, pp. 438–445, 2019.

[22] X. Liu, Z. Cai, H. Fan, and M. Yu, "Experimental studies on the rtEthernet-based centralized fault management system for smart grids," *Electric Power Systems Research*, vol. 181, pp. 106163.1–11, 2020.

[23] G. Chen and S. Li, "Research on location fusion of spatial geological disaster based on fuzzy SVM," *Computer Communications*, vol. 153, pp. 538–544, 2020.

[24] G. Chen, L. Wang, M. Alam, and M. Elhoseny, "Intelligent group prediction algorithm of GPS trajectory based on vehicle communication," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 7, pp. 3987–3996, 2020.

Retraction

Retracted: Research on the Generation and Design Method of Residential Community Scheme Based on Improving Design

Scientific Programming

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope

(2) Discrepancies in the description of the research reported

(3) Discrepancies between the availability of data and the research described

(4) Inappropriate citations

(5) Incoherent, meaningless and/or irrelevant content included in the article

(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] C. Li and L. Wang, "Research on the Generation and Design Method of Residential Community Scheme Based on Improving Design," *Scientific Programming*, vol. 2021, Article ID 5053494, 9 pages, 2021.

Research Article

Research on the Generation and Design Method of Residential Community Scheme Based on Improving Design

Chunyu Li[1] and Lei Wang [iD][2]

[1]*School of Civil Engineering and Mechanics, Yanshan University, Qinhuangdao 066004, Hebei, China*
[2]*Qinhuangdao Vocational and Technical College, Qinhuangdao 066100, Hebei, China*

Correspondence should be addressed to Lei Wang; ichbinqq@qvc.edu.cn

Along with the urban renewal and development, the urban living environment has given rise to various problems that need to be solved. With an eye on the future development model of residential communities, an experimental preliminary design for the construction of architectural space, public space, and landscape space based on people's actual needs is carried out in an attempt to alleviate the more urgent symbiotic relationship between people and urban environment. To this end, this paper proposes a planning and design generation framework for the constructed external spatial environment of building groups based on a recursive double-adversarial network model. Firstly, we extract the features of the constructed external spatial environment of the building group in depth and generate the expression feature map, which is used as a supervisory signal to generate an expression seed image of the constructed external spatial environment of the building group; then we use the generated seed image together with the constructed external spatial environment of the original target building group as the input to generate a feature-holding image as the output of the current frame, and the feature-holding image is also used as the input for the next. Finally, the seed image generation network and the feature-holding image generation network are recursively used to generate the next frame, and the video sequence of the expressions of the constructed external spatial environment of the building group with the same feature-holding expressions as the original input is recursively obtained several times. The experimental results on the building group database show that the proposed method can generate clear and natural video frames of the constructed external spatial environment of the building group, which can be gradually derived from the design of building units to the construction of the building group and penetrate into the planning and design of the external spatial environment in order to comprehensively improve the living environment of urban population and provide a design method and theoretical support for the design of future urban residential communities.

1. Introduction

At present, along with the rapid development of the global economy and the leapfrogging take-off of China's economy, the overall social environment is also facing various problems such as economic, environmental, and urbanization. On the economic front, prices are rising; housing prices are still high; employment pressure is high; and various resources are becoming increasingly scarce due to various uncertainties. On the environmental side, the negative effects of climate and environmental pollution continue [1]; on the urbanization side, the rapid development of cities and population expansion have become the norm, which also leads to the lack of management resources, social conflicts, urban environmental degradation, and other serious problems [2]. There are four major problems in urban development as follows.

The tendency for higher demand for living areas: people have continued to pursue new housing in all periods of social development. Although the living area and per capita area of urban households have increased, the average housing area of typical urban households continues to be lower than the national urban average, and the gap has widened [3]. As the urban household population continues to increase, urban housing tends to develop more and more highly.

Continuous expansion of urban form: the first level of urban areas is the distribution of town clusters in the macroregion; the second level is the external spatial form of the city, that is, the plan form and facade form of the city; and the third level is the zoning form within the city [4]. As the urban form continues to expand, the scope of the "urban fringe" area continues to expand.

Traffic congestion and environmental pollution in cities: with the increasing number of cars and population in cities, the problems arising from traffic in cities are becoming more and more prominent. In many cities today, the excessive number of cars often leads to traffic jam, frequent traffic accidents, and atmospheric pollution [5]. Traffic problems have brought a serious impact on the socioeconomic development of cities.

Continued high urban energy consumption: China's urban energy consumption is exploding, and China's electric energy mainly comes from coal-based thermal power, thus causing a surge in CO_2 emissions. In the face of the above-mentioned situations, the expansion of urban areas cannot solve the fundamental problems, and a good urban development cycle requires targeted spatial construction solutions and attempts [6]. Along with the urban renewal and development, the urban living environment has given rise to various problems that need to be solved. Focusing on the future development mode of residential communities, we try to alleviate the urgent symbiotic relationship between people and the urban environment by conducting experimental preliminary design for the construction of architectural space, public space, and landscape space based on the actual needs of people. To this end, this paper proposes a planning and design generation framework for the constructed external spatial environment of building groups based on a recursive double-adversarial network model. Firstly, we extract the features of the constructed external spatial environment of the building group in depth and generate the expression feature map, which is used as a supervisory signal to generate an expression seed image of the constructed external spatial environment of the building group; then we use the generated seed image together with the constructed external spatial environment of the original target building group as the input to generate a feature-holding image as the output of the current frame, and the feature-holding image is also used as the input for the next. Finally, the seed image generation network and the feature-holding image generation network are recursively used to generate the next frame, and the video sequence of the expressions of the constructed external spatial environment of the building group with the same feature-holding expressions as the original input is recursively obtained several times [7]. The experimental results on the building group database show that the proposed method can generate clear and natural video frames of the constructed external spatial environment of the building group, which can be gradually derived from the design of building units to the construction of the building group and penetrate into the planning and design of the external spatial environment in order to comprehensively improve the living

environment of urban population and provide a design method and theoretical support for the design of future urban residential communities [8].

2. Possible Ways of Constructing the Architectural Space Form of Future Residential Communities

2.1. Functional Requirements for the Construction of Architectural Forms. At this stage, the structure of urban buildings is basically frame type; however, the existing environment, area restrictions, traffic requirements, energy-saving, functional changes, style outlook and people's lifestyles, the future building energy system, integrated cleaning system, green ecological system, media interface system, air system, lifestyle, and other systems need to have an overall comprehensive consideration [9].

Dutch architect Rem Koolhaas has made it clear that the global economic crisis will lead to a decrease in showy buildings and an increase in buildings that are more "socially responsible" and better serve the public. In the current economic downturn, architectural design is placing more emphasis on space utilization. Therefore, the future structure and form of buildings will pay more attention to ecology and environmental protection, focusing on the construction of a small ecological environment. Innovative new materials are used to build units, and different forms of construction styles are tried [10]. The space formed between different shaped units will be used for greening, which can greatly save the land on the one hand and form a good integrated microcirculation system on the other hand and finally make the whole building form a unique ecological group [11].

2.2. The Generation of Architectural Space: From Unitary to Group Structure. It is a source of ideas in today's architectural space design to form new forms of architectural space by continuously repeating and combining unitary buildings. The process of constructing a group of units is similar to building a box, and the form of the building constructed in this way is similar to children's building blocks. The space formed between the different shaped units can be used for landscaping, and the whole building can thus form an independent ecological cluster [12]. The design structure of this building unit can meet the functional needs of the occupants on the one hand and increase the green space on the other.

Nowadays, if you need 3 acres of land to build a community, you may have to set aside 1 acre for green space and public space, while in the future, this modular building will form more public space between its units, which will not only save land resources but also increase the area of greenery and make it dispersed in the building system [13].

2.3. Generation and Design Evolution of Architectural Units. In this experimental construction, the initial single building consists of two overlapping square boxes, with the upper box picking out a part and its common part running through the two upper and lower spaces, and then the single building is continuously stacked and combined to form a group

structure and finally form a group of buildings, that is, the whole building, as shown in Figure 1 [14].

2.4. Generation of Building Plan Layout.

The whole building plan was linear in the predesign period and evolved into art for the sake of light and visual needs. The curved plan layout makes the building more vivid. Each building has 96° of curvature; each floor has 12 unit buildings; the whole building has 96 degrees of curvature; each group of unit buildings has 8° of curvature; and 96 divided by 12 is exactly equal to 8, thus forming the final initial plane of the single building, as shown in Figure 2. The three circles in the middle are the public space, and the two buildings are interconnected through this group of public spaces [15].

2.5. Design Analysis of Building Unit A.

Building unit A (see Figure 3) has single floors 1-8 in a curved block consisting of 12 individual units combined and each with a driveway of 7 meters in width. Two elevator shafts are provided for easy access to the floors. The second floor of each house is a living room, bathroom, kitchen, and so on, and the second floor is mainly a study and bedroom [16]. The composition of the double-layer structure of 1–8 floors in the designed building is the same as the composition of the single-layer structure of 1–8 floors, but with the opposite house type; each house also has a driveway with a width of 7 meters and an elevator shaft, but the difference from the single story is that the double story has an extended balcony. The living room, bathroom, and kitchen are on the second floor of each house, and the study and bedroom are on the second floor, as shown in Figure 4. Building A's 9-25 single floors are the same as 1-8 single floors in terms of layout and composition, but the biggest difference is that there is no driveway on 9-25 single floors, but there are still two separate automatic elevator passages and corridors, as shown in Figure 5.

2.6. Design Analysis of Building Unit B.

Building B is also curved, as opposed to building A, which is more youthful and dynamic, as shown in Figure 5. The curved form of the building is more convenient to dye and so on. The traffic problem has brought a serious impact on the socioeconomic development of the city [17].

2.7. Continued Climb in Urban Energy Consumption.

Energy consumption in China's cities is currently exploding, and the fact that China's electricity is mainly generated by coal-based thermal power has led to a surge in CO_2 emissions. In the face of the above-mentioned situation, the expansion of urban areas cannot solve the fundamental problems, and a good urban development cycle requires targeted spatial construction solutions and attempts [18].

3. Dual Network Model

The key idea of the method in this paper is to construct a double-adversarial network structure model for recursively generating the constructed external spatial environment of building groups by using a constructed external spatial
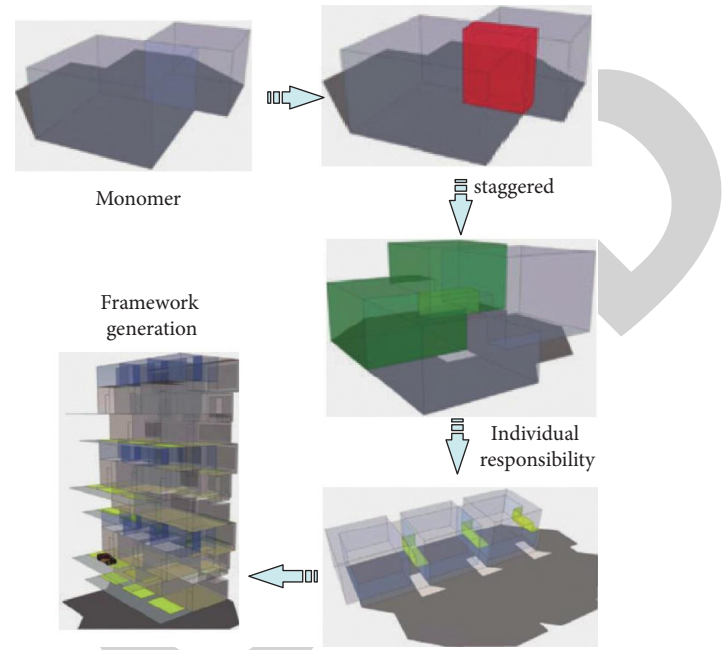
Figure 1: Generation and design evolution of building units.
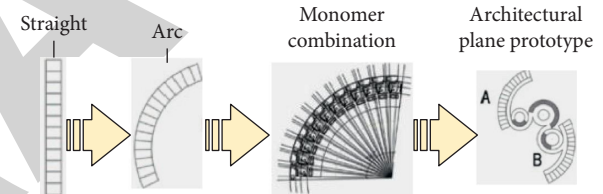
Figure 2: Generation of building plan layout.

environment generator for building groups focusing on expression features and a constructed external spatial environment generator for building groups focusing on features such as texture features and image light and dark details and to connect the two networks in series to form a recursive model [19]. The structure is recursive, and the U-net-like jump connection is added to the model, which can effectively maintain the texture information of the underlying image in the image generation, and is used to synthesize fine expressions of the constructed external spatial environment of the building group. In expression video generation, the conventional algorithm uses a separate image for all expression synthesis, which is not conducive to the continuous maintenance of the generated video frames. It is to be noted that the changes in the expressions of the built external spatial environment of the building group in the video frames are often reversible and small, using the image generated in the previous frame as the input for the next frame can generate the same quality of images using a smaller model and effectively reduce the number of model parameters [20].

3.1. Dual Network Model.

Our proposed dual network model consists of two deep learning-based generative adversarial networks, including a network φ_1 (FaceGAN) for

Convolution layer 2 + leakvrelu + pooling layer

Deconvolution * 2 + leakyrelu
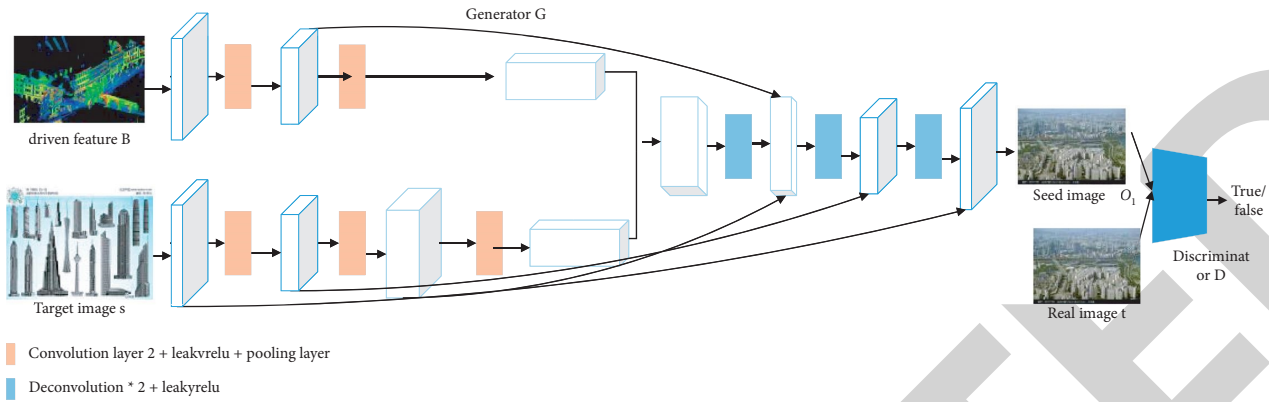
Figure 3: Structure of network φ_1 (GAN) generator for initial generation.

Figure 4: Building A rendering.

Non-change

Low vegetation

N.v.g. surface

Tree

Water

Building

Playground
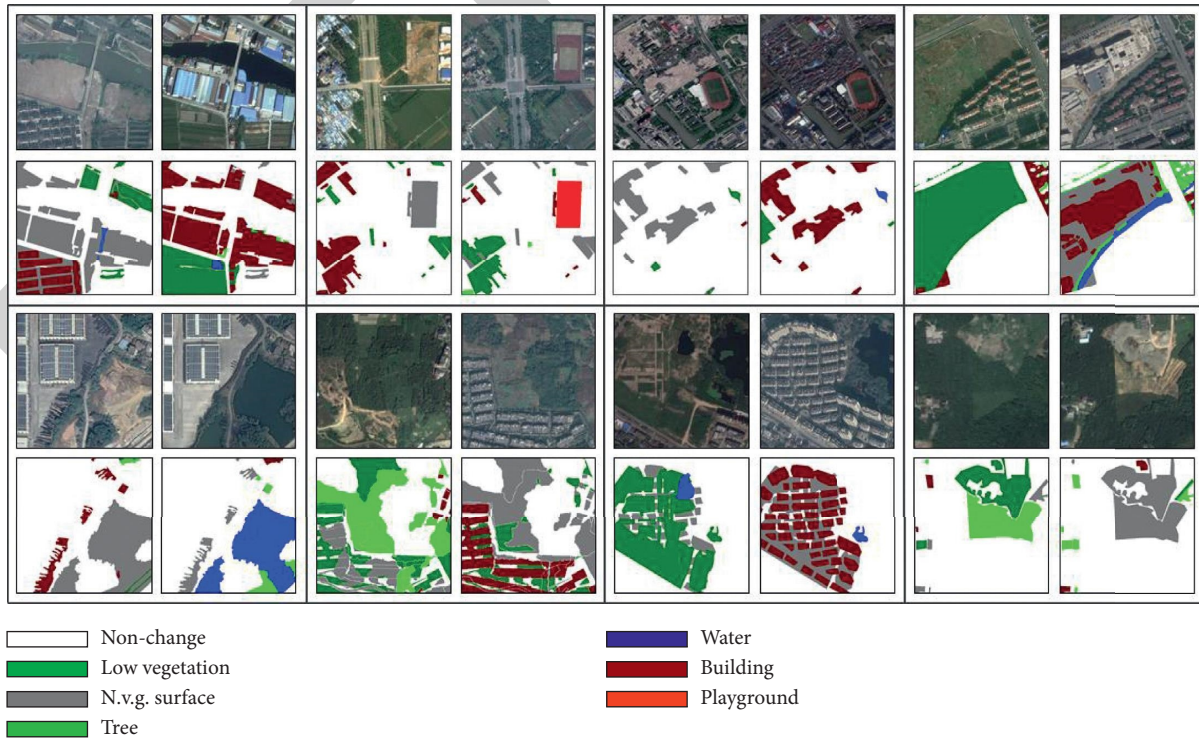
Figure 5: Building B flow analysis diagram.

seed image generation and a network φ_2 (FineGAN) for generating feature-preserving images, and the two networks are connected in series to form a dual network model.

3.2. Model Architecture.

For the constructed external spatial environment image S of the target building group and the expression feature image B, nonreciprocal codecs are used for the two images since they provide different information, where the constructed external spatial environment feature image B of the building group is used as supervisory information in the generation of seed image O_1^1 to generate the seed image corresponding to the expression feature, while the original target image S provides texture and other image detail features in the generation of the seed image and the feature-holding image. Each component of the model is composed of a convolutional layer and a pooling layer. In the φ_1 network, unlike the traditional generative network, the U-net [21] network structures are borrowed, and jump connections are used between different feature layers for maintaining the texture information of the underlying target image in the generated expression image [22].

φ_2 network structure is shown in Figure 3; for the expression feature image B of the constructed external spatial environment of the building group, multiple convolutions and pooling layers are used to extract the depth features, and the expression feature image extractor is used to extract the expression information of the constructed external spatial environment of the building group, which is used as the supervisory information in the image reconstruction; relatively speaking, it contains less information. In order to reduce the number of network parameters and improve the system speed, the depth of the feature extractor should be reduced appropriately [23].

The image S of the external environment of the target building group can be constructed by the feature extractor, since the constructed external spatial environment target image of the building group provides texture and individual features in generating the constructed external spatial environment expression animation of the building group; the constructed external spatial environment target image feature extraction and feature decoding of the building group need to preserve more details of the constructed external spatial environment of the building group [24]. Extracting different features in the image S of the external environment of the building complex can increase the depth of the encoder network; using jump connection, the low-level image texture features are preserved, and the constructed external spatial environment feature image of the building group and the constructed external spatial environment of the target building group are connected separately and simultaneously across layers of features, and finally more information is fused in the decoder stage to improve the image fineness [25].

As shown in Figure 6, two inputs are used in φ_2, one of which is the seed image O_1 output by φ_1, and the other input is the original constructed external spatial environment image S of the target building group, so two feature extractors with the same structure are used in φ_2 but without sharing parameters. In the decoding stage, more jump connections are added.

In addition, the enhanced discriminator of the VGG network with stronger classification power is used in the φ_2 discriminator, which enables φ_2 to distinguish of more subtle differences between the generated and real images. Since the stronger discriminator makes the generative network more difficult to train, some φ_1 network parameters are shared in φ_2 when training φ_2.

3.3. Recursive Video Frame Generation.

The consistency between different frames in the generated video is not considered in the literature related to the expression of the constructed external spatial environment of the building group based on deep learning [26], and most of them are generated only for single graphics, although the static image generation has achieved a very fine and realistic effect, but once the related generation of the video is performed, it exposes the generation of different image frames with the problem of large differences in brightness and darkness, that is, some frames are brighter and some frames are darker, and has the problem of uneven distribution, which leads to flicker and video fidelity degradation during video playback [27]. The recursive generation method proposed in this paper uses the seed image generation network to obtain the expression seed image, then uses the feature retention network to generate the feature retention image, and then uses the generated feature retention image as the input of a new frame for a recursive generation. Finally, the color, brightness, and texture of all the video frames are more uniform, and the generated video is more realistic and natural [28].

Two networks are used to generate the constructed external spatial environment of the building group separately: first, the seed image of the constructed external spatial environment of the building group is generated by φ_1, and then the generated image is further optimized for recursive generation using φ_2 to obtain the output image of the constructed external spatial environment of the building group with features maintained.

In the recursive generation, let the input sequence of driving image frames be B, the corresponding i-th frame driving image be b^i, and the constructed external spatial environment image of the target building group be S. The flow of recursive generation is shown in Figure 4. The seed O_1^1 image generation of the constructed external spatial environment expression of the building group is performed by φ_1. O_1^1 will be used as the input of φ_2. At the same time, the constructed external space environment of the target building group will also be used as the input of φ_2 to obtain the feature-holding image O_1^2 of the constructed external space environment of the first frame building group and output it. Then, input the construction external space environment boundary sequence B^2 of the building group in the second frame, take the feature holding Image O_1^2 finally output in the first frame and the construction external space environment feature image B^2 of the building group corresponding to the second frame
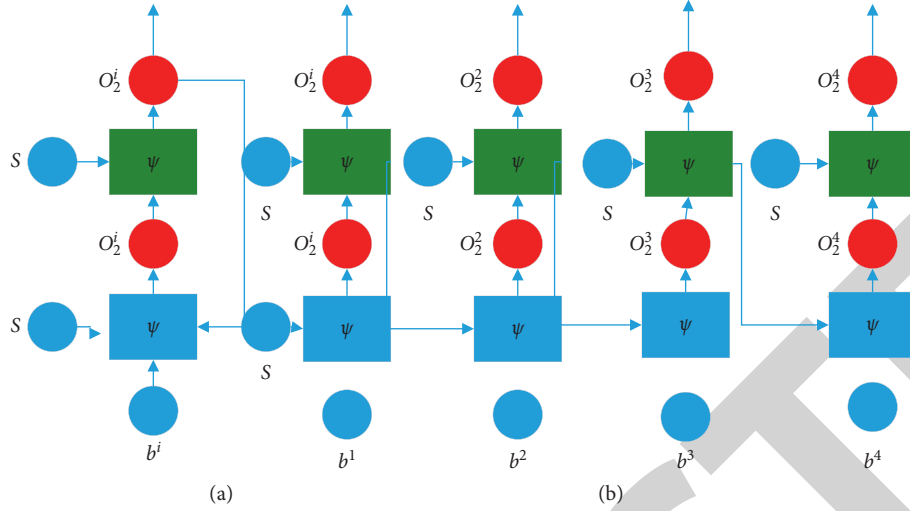
Figure 6: The framework of expression generation of the external spatial environment for the construction of architectural groups: (a) recursive generation flowchart and (b) recursively generate flow chart data flow.

as φ_1 input, and generate the second frame seed image O_1^2. Then, the second frame feature-holding image O_2^2 is generated by O_1^2. With this loop, the output of all expression frames is generated recursively, as shown in the following equations:

$$\psi_1\left(S, b^i\right) = O_1^i, \tag{1}$$

$$\psi_2\left(S, O_1^i\right) = O_2^i, \tag{2}$$

where O_2^i is the output of the i-th frame.

3.4. Residential Community Architectural Space Form Generation.

In this experimental construction, the initial single building consists of two overlapping square boxes, with the upper box picking out a part and its common part running through the two upper and lower spaces, and then the single building is continuously stacked and combined to form group structure and finally form a group of buildings, that is, the whole building, as shown in Figure 4.

Building B also has an arc-shaped unit, as opposed to building A, which is more energetic like young people. The arc-shaped building form is more convenient for lighting, and the sunlight from the second floor can shine through the second-floor windows to the first floor, which is good for lighting and sufficient light. The elevator shaft on both sides leads from the first-floor garage to the top floor. The 9-25 floors of building B are residential space, but the single-story units are the opposite of the double-story units [29–31].

In addition, unlike building A, building B is lit mainly through the windows on the back of the building; in order to make up for the lack of light on the first floor, the light can reach the first floor through the second floor for better lighting effect, and the curved building form is more convenient for lighting [32].

3.5. Performance Evaluation.

In this paper, two experiments are designed to investigate the dichotomous (normal and anomaly) and five-class performance of our model ($K = 100$). To compare with other machine learning methods, comparison experiments were also designed to compare the classification performance of our model model with machine learning methods such as C4.5 decision trees, random forests, random trees, support vector machines, and recurrent neural networks [32].

(1) Binary Classification. Table 1 shows the classification results of the our model model on the test set in the binary classification experiment. The experiments show that after 100 iterations, the our model has a high detection accuracy of 97.44% in the training set and 91.88% in the test set.

The results compared with previous researchers' proposed C4.5 decision trees, random forests, random trees, support vector machines, and recurrent neural networks on the same benchmark data set are shown in Figure 7. It is clear that the our model outperforms all the other classification algorithms in the binary classification experiments [33].

(2) Five Classifications. In the five classification experiments, as shown in Figure 8 that our detection model achieves an accuracy of 97.47% on the training set and 82.40% on the test set. The results of our model on the test set are shown in Table 2. Table 3 shows the detection accuracy and recall rates for different attack types.

As shown in Figure 9, the detection accuracy of the our model model is higher than that obtained by classification algorithms such as C4.5 decision trees, random forests, random trees, support vector machines, and recurrent neural networks. Due to the uneven distribution of samples in the dataset, the detection accuracy of the model with five classifications decreases compared to the two classifications [34].

Table 1: Classification results of the our model model in the binary classification experiment.

Correct category	Prediction results	
	Anomaly	Normal
Anomaly	11,507	1,326
Normal	505	9,206
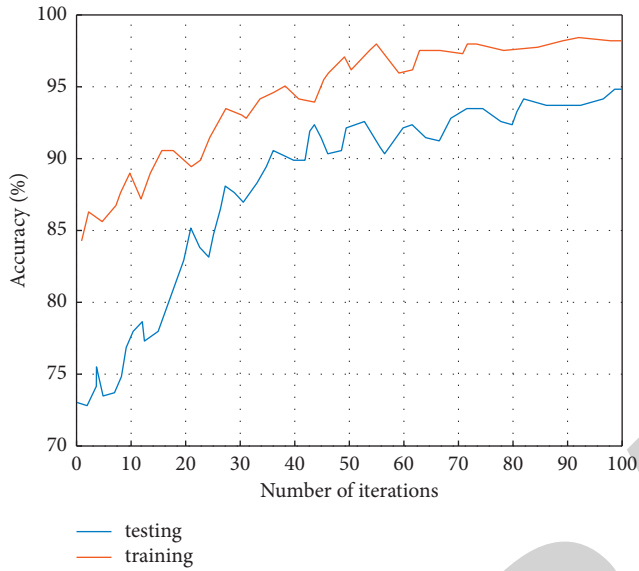
Figure 7: Detection accuracies of the binary classification model on the training and test sets.
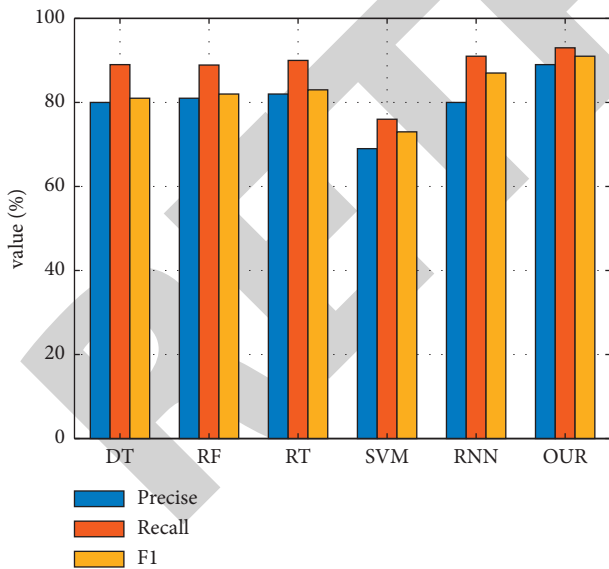
Figure 8: Comparison of the performance of the models for the second classification.

Table 2: Classification results of our model in the five classification experiments.

Correct category	Prediction results				
	Normal	Dos	Probe	R2L	U2R
Normal	9,278	152	267	4	10
Dos	938	6,398	93	8	21
Probe	216	107	2,095	3	0
R2L	1,969	10	0	772	3
U2R	152	0	9	5	34

Table 3: Detection accuracy and recall for different attack types.

Type	Accuracy (%)	Recall rate (%)
Dos	95.97	85.79
Probe	85.02	86.53
R2L	97.47	38.95
U2R	50.00	17.00
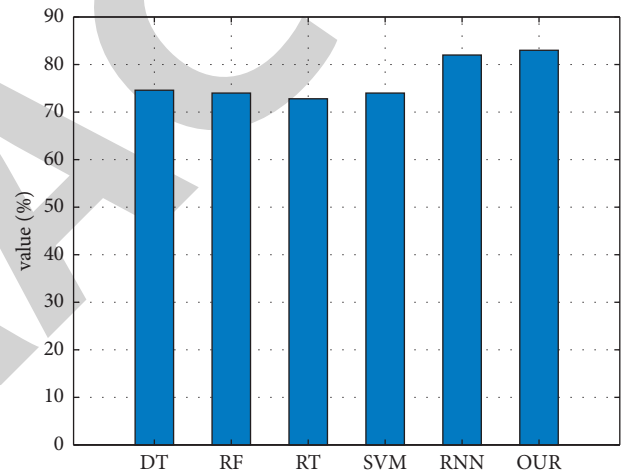
Figure 9: Accuracy rate of each model in five categories.

4. Conclusions

The design of urban residential communities focuses on people's living requirements, emphasizes community participation, and puts itself in the shoes of users, especially the user groups' usage requirements, living customs, and emotional psychology. In this paper, we propose a planning and design generation framework for constructing external spatial environments of building groups based on a recursive dual-adversarial network model. The experimental results on the building group database show that the proposed method can generate clear and natural video frames of the external spatial environment of the building group, which can be gradually derived from the design of the building

units to the building group and penetrate into the planning and design of the external spatial environment.

Data Availability

The data used in this paper are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest regarding this work.

References

[1] Y. Suzuki, "Proximity-based non-contact perception and omnidirectional point-cloud generation based on hierarchical information on fingertip proximity sensors," *Advanced Robotics*, no. 2, pp. 1–17, 2021.

[2] I. Dincer and M. A. Rosen, "A worldwide perspective on energy, environment and sustainable development," *International Journal of Energy Research*, vol. 22, no. 15, pp. 1305–1321, 2015.

[3] J. M. M. Engels, H. Dempewolf, and V. Henson-Apollonio, "Ethical considerations in agro-biodiversity research, collecting, and use," *Journal of Agricultural and Environmental Ethics*, vol. 24, no. 2, pp. 107–126, 2011.

[4] C. Brian and L. Cook, "The ethics of exercise in eating disorders: can an ethical principles approach guide the next generation of research and clinical practice?" *Journel of Sports Health Science*, vol. 6, no. 3, pp. 295–298, 2017.

[5] I. D. Shorts, "The prediction of in-programme failure among delinquent youths in a non-residential community-based programme," *British Journal of Clinical Psychology*, vol. 24, no. 4, pp. 301-302, 2011.

[6] F. A. C. Wright, G. Law, S. K. Chu, J. S Cullen, and D. G Le Couteur, "Residential age care and domiciliary oral health services: reach-OHT-The development of a metropolitan oral health programme in Sydney, Australia," *Gerodontology*, vol. 34, no. 1 Suppl, pp. 420–426, 2017.

[7] A. Zarling and R. Scheffert, "Implementation of ACT in correctional and forensic settings," *Journal of Contextual Behavioral Science*, vol. 22, pp. 44–51, 2021.

[8] J. F. R. B. Graddipappcomp, G. Fshp, R. H. R. Mapplsci et al., "Programme to improve the use of drugs in older people and involve general practitioners in community education," *Journal of Clinical Pharmacy and Therapeutics*, vol. 24, no. 4, pp. 289–297, 2010.

[9] J. Robertson, E. Emerson, L. Pinkney et al., "Community-based residential supports for people with intellectual disabilities and challenging behaviour: the views of neighbours," *Journal of Applied Research in Intellectual Disabilities*, vol. 18, no. 1, pp. 85–92, 2005.

[10] C. J. Uneke, C. D. Ndukwe, A. A. Ezeoha, H. C. Urochukwu, and C. T. Ezeonu, "Improving maternal and child healthcare programme using community-participatory interventions in ebonyi state Nigeria," *International Journal of Health Policy and Management*, vol. 3, no. 5, pp. 283–287, 2014.

[11] M. Suzuki, H. Aizawa, and M. Umeda, "Bringing UP community successors through events IN residential area," *Journal of Architecture and Planning (Transactions of AIJ)*, vol. 67, no. 560, pp. 185–192, 2002.

[12] B. Highton, "Residential mobility, community mobility, and electoral participation," *Political Behavior*, vol. 22, no. 2, pp. 109–120, 2000.

[13] K. Frensch, G. Cameron, and M. Preyde, "Community adaptation of youth accessing residential programs or a home-based alternative: school attendance and academic functioning," *Child and Youth Care Forum*, vol. 38, no. 6, pp. 287–303, 2009.

[14] R. Austin, B. Hunter, and L. Hollywood, "Supporting community cohesion through ICT: the epartners programme in Northern Ireland," *Computers in Human Behavior*, vol. 52, pp. 508–514, 2015.

[15] H. Chalmers, P. Tyrer, and P. Aggleton, "Sex and relationships education in schools - evaluation of a pilot programme for the certification of community nurses," *Health Education Journal*, vol. 65, no. 1, pp. 28–40, 2006.

[16] T. Xie, C. Zhang, Z. Zhang, and K. Yang, "Utilizing active sensor nodes in smart environments for optimal communication coverage," *IEEE Access*, vol. 7, pp. 11338–11348, 2018.

[17] M. Fried, "Residential attachment: sources of residential and community satisfaction," *Journal of Social Issues*, vol. 38, no. 3, pp. 107–119, 2010.

[18] M. Chiesa, P. Fonagy, J. Holmes, and C. Drahorad, "Residential versus community treatment of personality disorders: a comparative study of three treatment programs," *American Journal of Psychiatry*, vol. 161, no. 8, pp. 1463–1470, 2004.

[19] P. Glassman, C. Miller, T. Wozniak, and C. Jones, "A preventive dentistry training program for caretakers of persons with disabilities residing in community residential facilities," *Special Care in Dentistry*, vol. 14, no. 4, pp. 137–143, 2010.

[20] Li Wang et al., "A communication strategy of proactive nodes based on loop theorem in wireless sensor networks," in *Proceedings of the 2018 Ninth International Conference on Intelligent Control and Information Processing (ICICIP). IEEE*, Wanzhou, China, November 2018.

[21] N. K. Meena, J. Yang, and E. Zacharis, "Optimisation framework for the design and operation of open-market urban and remote community microgrids," *Applied Energy*, vol. 252, p. 113399, 2019.

[22] M. N. Kabir, Y. Mishra, G. Ledwich, Z. Y. Dong, and K. P. Wong, "Coordinated control of grid-connected photovoltaic reactive power and battery energy storage systems to improve the voltage profile of a residential distribution feeder," *IEEE Transactions on industrial Informatics*, vol. 10, no. 2, pp. 967–977, 2014.

[23] Z. Ni and A. Das, "A new incentive-based optimization scheme for residential community with financial trade-offs," *IEEE Access*, vol. 6, pp. 57802–57813, 2018.

[24] G. Di Lorenzo et al., "Modeling and design of a residential energy community with PV sharing," in *Proceedings of the 2020 IEEE International Conference on Environment and Electrical Engineering and 2020 IEEE Industrial and Commercial Power Systems Europe (EEEIC/I&CPS Europe). IEEE*, Madrid, Spain, June 2020.

[25] Z. Huang, H. Yu, Z. Peng, and M. Zhao, "Methods and tools for community energy planning: a review," *Renewable and Sustainable Energy Reviews*, vol. 42, pp. 1335–1348, 2015.

[26] S. Liu, J. Liu, Q. Yang et al., "Coupled simulation of natural ventilation and daylighting for a residential community design," *Energy and Buildings*, vol. 68, pp. 686–695, 2014.

[27] S. M. A. Eben, "Learning from tradition: the planning of residential neighborhoods in a changing world," *Habitat International*, vol. 28, no. 4, pp. 625–639, 2004.

Research Article

Intelligent Recognition Method of Athlete Wrong Movement Based on Image Vision

Wang Lu[1] and JiangYuan Hou [ID][2]

[1]*The School of Physical Education of Yantai University, Yantai, Shandong 264005, China*
[2]*Sports Training Department, Hebei Sport University, Shijiazhuang 050000, Hebei, China*

Correspondence should be addressed to JiangYuan Hou; 1996004@hepec.edu.cn

Current methods of human body movement recognition neglect the depth denoising and edge restoration of movement image, which leads to great error in athletes' wrong movement recognition and poor application intelligence. Therefore, an intelligent recognition method based on image vision for sports athletes' wrong actions is proposed. The basic principle, structure, and 3D application of computer image vision technology are defined. Capturing the human body image and point cloud data, the three-dimensional dynamic model of sports athletes action is constructed. The color camera including CCD sensor and CMOS sensor is selected to collect the wrong movement image of athlete and provide image data for the recognition of wrong movement. Wavelet transform coefficient and quantization matrix threshold are introduced to denoise the wrong motion images of athletes. Based on this, the feature of sports athlete's motion contour image is extracted in spatial frequency domain, and the edge of the image is further recovered by Canny operator. Experimental results show that the proposed method can accurately identify the wrong movements of athletes, and there is no redundancy in the recognition results. Image denoising effect is good and less time-consuming and can provide a reliable basis for related fields.

1. Introduction

With the increasing importance attached to sports events, athletes need to train according to various standard movements in the training process. Referees cannot recognize the wrong movements because of the fast speed or the large number of people. With the development of computer vision technology, it is widely used in the analysis of human body structure. In order to realize the intelligent recognition of athlete's wrong action, computer vision technology is applied. It can not only improve the athlete's movement level but also quickly and accurately judge whether the athlete has the wrong movement, thus enhancing the competition fairness. Computer vision technology is one of the key researches in the field of graphics and computer vision because there are many sports items and there are some differences in recognizing athletes' wrong actions. There are a variety of organs and tissues in the human body. Different combinations of these tissues can make the human body complete

specific behavior, integrate the information of human motion, and provide a strong basis for the analysis of human behavior. Therefore, the action recognition based on image human body has very critical research value [1, 2]. At present, the commonly used methods of action recognition are usually disturbed by illumination, occlusion, or shaking, which leads to the difficulty of action feature extraction and the inability to get useful features accurately, so that the subsequent segmentation and recognition effect is not ideal.

Liu [3] proposed a method of human posture recognition based on multifeature fusion. Aiming at the problem that the existing posture recognition algorithms cannot reflect the dynamic characteristics of athletes' posture, this paper proposes a posture recognition algorithm based on multifeature fusion. Firstly, the image is captured by optical image collector, and then the image is transformed to gray level to improve the image quality. Then, the body contour and the motion region are obtained based on shadow elimination and interframe difference. Finally, posture

region and body contour are extracted based on Radon transform and discrete wavelet transform. Chen [4] proposed a moving image contour feature extraction method based on multithreshold optimization. Aiming at the problems of long extraction time and low extraction accuracy in traditional moving image contour feature extraction methods, a moving image contour feature extraction method based on multi-threshold optimization is proposed. Through moving image contour feature analysis, the membership function is obtained by using the maximum interclass variance fuzzy constraint method, and multiple thresholds of target contour in moving image are calculated by using fuzzy membership. The geometric center values of the two contour points adjacent to the center point in the image contour range are calculated by using multiple constrained thresholds. The curvature symbol is obtained by calculating the curvature angle, and the contour features of the moving image are extracted according to the curvature symbol. Shen et al. [5] proposed an action recognition method of sports athletes based on deep learning. This method processes the image data of sports athletes through dense optical flow method and extracts the characteristics of wrong actions of sports athletes by combining short-term memory neural network and convolution neural network. This method cannot obtain key frames in the process of data processing, resulting in low feature extraction rate of the method.

Aiming at the above problems, this paper proposes a new intelligent recognition method of sports athletes' wrong actions based on image vision.

2. Intelligent Recognition Method of Sports Athletes' Wrong Actions Based on Image Vision

2.1. Computer Image Vision Technology. With the development of computer technology and computer vision technology, people begin to study 3D vision gradually. Computer vision technology is mainly derived from photogrammetry and is mainly used in 2D image recognition and analysis. Today, computer vision technology is powerful enough to be used in a variety of fields [6, 7]. The main principle of computer vision system is to obtain the target image first, then extract the feature, and finally analyze, process, and calculate the feature, in order to make a reasonable decision. Figure 1 is the basic structure of computer vision system, in which the computer is the core part, needs to control the normal operation of each module, and also needs to calculate and output the results.

2.2. Establishment of a Three-Dimensional Dynamic Model of Sports Action. High realistic 3D dynamic model requires not only the deformation of skeletal joints, but also the movement of the associated skin driven by the joints, so as to produce reasonable movement. Therefore, based on the above captured and processed motion images and point cloud data, a 3D dynamic model of sports motion is built, in which the former builds the skeleton model of the sportsman and the latter adds the display appearance to the skeleton model to make the virtual human model more realistic.
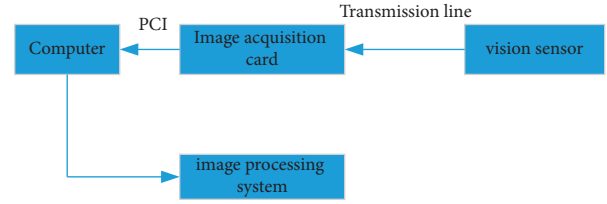
FIGURE 1: Structure of computer vision system.

2.2.1. Establishment of Surface Models. After the establishment of the skeleton model, in order to make the 3D human model more three-dimensional and realistic, we need to build the skin model outside the bone. Based on the point cloud data, the surface model is built by triangle mesh method. The process is as follows: first, select any point as the initial point, connect it with the nearest two points to form a triangle, and then extend along the three vertices to make the mesh grow continuously; finally all points are connected to form a triangle network. In meshing, the meshing of the parts that often produce movement should be more detailed, while the meshing of the parts that do not produce large movement should be larger.

After building the skeleton model and surface model, it is necessary to determine which segments of the skeleton affect a point on the skin mesh and then bind the two together to form a complete 3D human model. The specific expression is

$$F(n) = R_1 + R_2 + R_3. \tag{1}$$

In the formula, R_1, R_2, and R_3 are expressed as the fuzzy recognition parameters of the three-dimensional dynamic model of athletes' movements under the hybrid architecture.

2.2.2. Establishment of Skeletal Models. Skeleton is the most basic support for the human body to complete all kinds of movement. Only skeleton movement can make the whole 3D model of human body move correspondingly. The skeleton model was established according to the optical motion images. Since human motion is mainly embodied in 16 main parts, the established skeleton model is a simplified skeleton model. This is shown in Figure 2.

When describing these 16 parts, the corresponding joints connecting the bones are also included, and the rotation operation of the bones is also the operation of the corresponding joints. The bone names corresponding to the numbers are shown in Table 1.

Only when the skeleton of the 3D model is driven to move can the corresponding movement of the surface model be generated; that is, the construction of the 3D dynamic model of sports movement is realized.

2.3. Error Action Image Acquisition and Processing. This research selects the color camera which includes the CCD sensor and the CMOS sensor to collect the sports athlete wrong movement image. The camera can capture color images, depth images, and bone images of the wrong motion simultaneously [8].
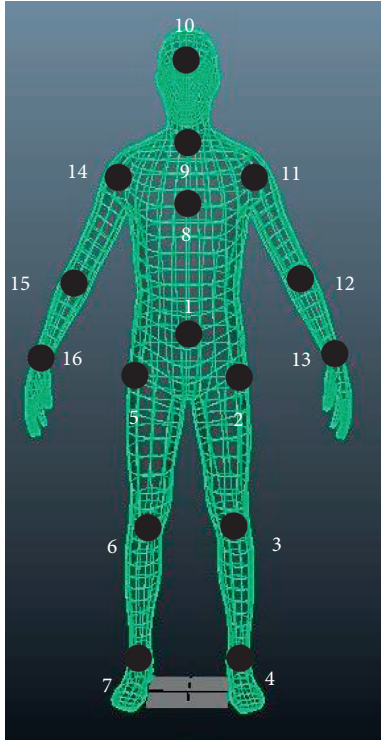
Figure 2: Bone model.

Table 1: Correspondence between bone name and serial number.

Number	Name
1	Abdomen
2	Right thigh (right hip)
3	Right lower leg (right knee)
4	Right foot (right ankle)
5	Left thigh (left hip)
6	Left lower leg (left knee)
7	Left foot (left ankle)
8	Chest
9	Neck
10	Head
11	Right upper arm (right shoulder)
12	Right forearm (right elbow)
13	Right hand (right wrist)
14	Left upper arm (left shoulder)
15	Left forearm (left elbow)
16	Left hand (left wrist)

The color image and the depth image are transmitted in the form of data stream. The color image resolution is $640 * 480$, the frame number is 30 Fps, the format is Bayer format, and the color data can be encoded as RGB-32 bits. The depth image acquisition process is consistent with color image. The effective position information is 13 bits higher and the user ID information is 3 bits lower. Skeletal images are captured from depth image data and contain 3D coordinates of 20 nodes, visually displaying skeletal maps of athletes [9].

In order to facilitate the application of error motion images, the spatial coordinates of color images, depth images, and bone images are analyzed. The color space, depth space, and skeleton space coordinates are shown in Figure 3.
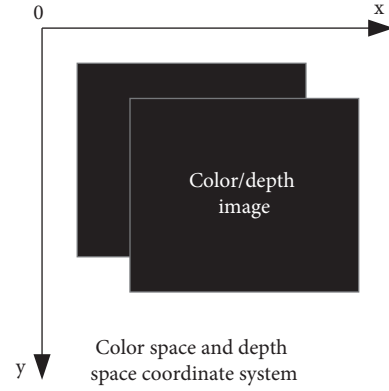
Figure 3: Schematic diagram of color space, depth space, and bone space coordinate system.

Set color space pixel coordinates to $(x(t), y(t), z(t))$, depth space pixel coordinates to $(x_s(t), y_s(t), z_s(t))$, and bone space pixel coordinates to $(x_d(t), y_d(t), z_d(t))$.

The skeletal space and depth space coordinate system conversion formula is

$$\begin{cases} x_d(t) = x_s(t) \dfrac{320}{z_s(t)\tan(a/2)}, \\[2mm] y_d(t) = y_s(t) \dfrac{240}{z_s(t)\tan(b/2)}, \\[2mm] z_d(t) = z_s(t). \end{cases} \quad (2)$$

In formula (2), a represents the horizontal angle of view of the camera with a value of $57°$ and b represents the vertical angle of view of the camera with a value of $43°$.

The conversion formula between depth space and color space coordinate system is

$$\begin{cases} x_s(t) = \dfrac{\Delta l}{z(t)\tan a} + x(t), \\[2mm] y_s(t) = y(t), \\[2mm] z_s(t) = z(t). \end{cases} \quad (3)$$

In formula (3), Δl represents the displacement of the camera.

The color image, depth image, and skeleton image are transformed into the same coordinate system. In order to facilitate image processing and ignore z direction information, the error motion image is $I(x, y)$, which provides image data for the following error motion feature extraction.

2.3.1. Wavelet Transform of Noisy Motion Image.
An action image wavelet transform with noise can be described by

$$f(x, y) = s(x, y) + e(x, y). \quad (4)$$

In the expression, the image information can be described as $s(x, y)$, the image noise signal can be described as

$f(x, y)$, and the variance of Gaussian white noise as ϖ^2 is described as $e(x, y)$, subject to $N(0, \varpi^2)$. If there is multiplicative noise signal in the image, it needs to be processed by logarithmic conversion, and the multiplicative noise can be converted to Gaussian white noise in logarithmic dimension.

Image wavelet transform with noise has the following characteristics:

(1) Wavelet transform coefficients have certain spatial orientation characteristics. LH represents edge data in horizontal direction, HL represents edge data in vertical direction, and HH represents edge data in diagonal direction. These edge data will provide a strong basis for the denoising of subsequent images [10, 11].

(2) The series of white noise can be transformed by wavelet base coefficients so that it can be represented by zero mean white noise.

(3) In the wavelet transform domain of the noisy image, the signal energy is mostly near the coefficient with higher absolute value, while the noise is completely the opposite [12]. Therefore, a threshold value is set, the coefficient not exceeding the threshold value is set to 0, and the wavelet coefficients beyond the threshold value are stored. Depending on the wavelet coefficients after processing, it can be understood that this part of the coefficient is a normal signal in the image, while the residual coefficient has noise to obtain the specific location of the noise.

2.3.2. Image Denoising under Quantization Matrix Threshold.

The multiresolution of image is decomposed by decomposition algorithm, and the wavelet coefficient matrix is constructed, then the threshold of matrix is quantified, and the new wavelet coefficient matrix is obtained. Finally, under the threshold of quantization matrix, the image in the matrix is reconstructed by reconstruction algorithm to obtain the denoised image [13, 14]. The specific procedures are as follows:

(1) According to the decomposition method, at the quantization matrix threshold, the orthogonal wavelet basis is generated in the wavelet function, and the initial image is decomposed into three layers of wavelet to obtain HH, LH, HL detail component and LL smooth component of the image signal

(2) Wavelet thresholds are quantified

How to select threshold and quantize threshold is the key step in denoising. In a certain sense, at the quantization matrix threshold, this step is related to the quality of the image signal reconstruction. Therefore, this paper uses soft threshold algorithm, quantization threshold γ defined as

$$\gamma = \begin{cases} \text{sign}(\gamma), \\ 0. \end{cases} \tag{5}$$

2.4. Spatial and Frequency Domain Feature Extraction of Sports Athletes' Action Contour Image.

Based on the above notable image, the number of pixels in the recognized sports athletes action image is calculated to get the target pixel number and the number of frames in the period. Suppose a sportsman has a frame n image in his movement, the formula of the energy generated by the sportsman's movement is expressed as follows:

$$G(x, y) = \frac{\sum_n^{i=1} B_i(x, y)}{n}. \tag{6}$$

In formula (6), $G(x, y)$ represents the gray image and $B_i(x, y)$ represents the gesture parameters of the motion image.

After the above basic processing, the high-frequency part and low-frequency part of the athlete's movement are distinguished by the discrete cosine transformation method to effectively extract the frequency domain characteristics of the athlete's movement posture [15]. Five sets of motion images, in which $\{f(x, y), x, y = 0, 1, 2, \ldots, N - 1\}$ is represented as $N \times N$, are represented by the DCT formula as follows:

$$F = f(x, y) \cdot \frac{(2x + 1)u\pi}{2N} \cdot \frac{(2y + 1)v\pi}{2N}. \tag{7}$$

In formula (7), $f(x, y)$ represents the gray value of the coordinates of pixels $f(x, y)$ in the sports athletes' action image, u and v, respectively, represent the horizontal and vertical conversion rates of pixels in the image, and N represents the DC part of the image features.

After the above calculation, the same transformation coefficient matrix as that of the original athlete can be obtained; that is, the frequency domain characteristics of the athlete's movement can be reasonably reflected. Based on the above transformation, a human posture model can be established. Its structure diagram is shown in Figure 4.

The human contour model is represented as follows:

$$v = \{N, R, F(D)\}. \tag{8}$$

In formula (8),

$$\begin{cases} N = \dfrac{\overline{W}}{H}, \\[2mm] R = \dfrac{A_r}{W \cdot H}, \\[2mm] D = [d_1, d_2, \cdots, d_n], \\[2mm] F(D) = R \cdot D. \end{cases} \tag{9}$$

In formula (9), $\overline{W}$ represents the average width parameter of the contour of the human body, H and W, respectively, represent the height and width values of the athlete's motion image, A_r represents the area of the athlete's motion target, and R represents the characteristic value in the athlete's motion model [16].
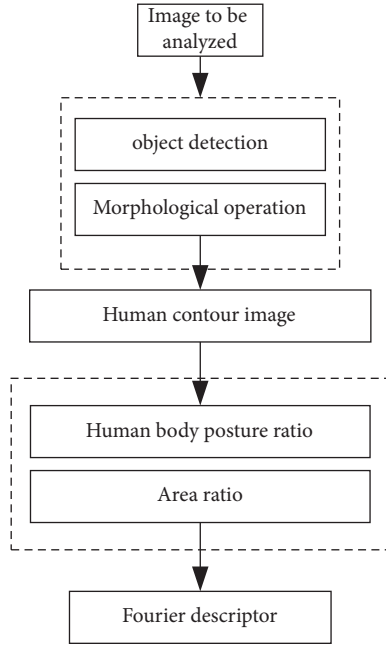
FIGURE 4: Human posture model structure.

2.5. Edge Restoration Based on Canny Operator.

The Canny method uses a suitable Gaussian function to smooth the image by columns and rows and also to convolve the image signals [17]. The Gaussian functions used are

$$G(x, y) = \frac{e^{-\left((x^2 + y^2)/2\sigma^2\right)}}{2\pi\sigma_2}. \tag{10}$$

In formula (10), σ represents the Gaussian curve, which controls the smoothing intensity.

Canny operator is constructed on the basis of 2D convolution $\nabla G(x, y) \cdot f(x, y)$, obtains edge direction and intensity, and identifies edge features through threshold [18, 19].

The two-dimensional convolution of $\nabla G(x, y)$ is divided into two one-dimensional convolvers and the result is

$$\frac{\partial G(x, y)}{\partial x} = \frac{y^2}{20^2},$$
$$\frac{\partial G(x, y)}{\partial y} = \frac{x^2}{20^2}. \tag{11}$$

Convolution is then performed with $f(x, y)$ for each convolver to obtain

$$E_x = \frac{\partial G(x, y)}{\partial x} * f,$$
$$E_y = \frac{\partial G(x, y)}{\partial y} * f. \tag{12}$$

The following conditions are met:

$$A(i, j) = E_x^2 + E_y^2, a(i, j) = \arctan \frac{E_y(i, j)}{E_x(i, j)}, \tag{13}$$

where the edge strength is described as $A(i, j)$ and the direction perpendicular to the edge is described as $a(i, j)$.

In the process of extracting edge features by Canny operator [20, 21], the selection of threshold is very important. If the threshold is too large, the edge feature recognition will be intermittent, and the low threshold will lead to false contour of the image. In this paper, we use double threshold method to solve the problem of threshold selection. Firstly, we propose two thresholds $r1$ and $r2$, and $2^{r1} = r2$, so that we can obtain two threshold edge images $N_1[i, j]$ and $N_2[i, j]$ because $N_2[i, j]$ is obtained by using high threshold. Therefore, the double threshold algorithm needs to link the edge into the contour within $N_2[i, j]$, and when the contour is linked with each other, the method can connect the inner edge of the contour by searching for the coordinates in the 8 fields of $N_1[i, j]$, so that the algorithm can continuously collect the edges within $N_1[i, j]$ until $N_2[i, j]$ is connected. The flow of the algorithm is as follows:

(1) Calculate the derivative Gx, Gy of the image gray scale according to the reciprocal operator and simultaneously calculate the gradient direction and size of the derivative.

(2) If the gray value of an image pixel in a certain direction is low, it is necessary to set the pixel coordinate to 0, that is, nonedge pixels.

(3) The threshold is calculated based on the histogram of the image. If the gray level value exceeds the threshold, the gray level area is the edge of the image, and vice versa. Then the continuity between the point above the threshold and the previous point is found. If the gray level value is not continuous, then the neighborhood coordinates of the point are found in the stack of the low threshold, and these coordinates are connected, so that the search is iterated until the overall contour is complete [22].

2.6. Intelligent Recognition of Athletes' Wrong Actions.

Bayesian classifier is the key to intelligent recognition of sports athletes' wrong movements. Therefore, a classifier is designed according to Bayesian algorithm [23–25].

Bayesian classifier is a kind of classification method designed according to Bayesian algorithm on the premise of conditional independence assumption. For the training sample set, the joint probability distribution function of the input and output of the training set is calculated firstly, based on which, the maximum posterior probability output of the input data is calculated by Bayes algorithm [26, 27].

Suppose the training dataset is $R = \{(x_1, y_1), (x_2, y_2), \cdots, (x_N, y_N)\}$, and the j feature of the i training sample is $x_i^{(j)}$, which is composed of a plurality of values and recorded as a_{jl}.

For input value x, the priori probabilities and conditional probabilities are calculated as follows:

$$
\begin{cases}
P(Y = c_k) = \sum_{i=1}^{N} I(y_i = c_k), \quad k = 1, 2, \ldots, \\
P(X^{(j)} = a_{jl} \mid Y = c_k) = \sum_{i=1}^{N} I(x_i^{(j)} = a_{jl}, y_i = c_k).
\end{cases}
\tag{14}
$$

In formula (14), $Y = c_k$ represents the output space corresponding to the input space; N represents the number of training datasets; and $P(X^{(j)} = a_{jl} \mid Y = c_k)$ represents the joint probability distribution function [28, 29].

For a given input x, the corresponding output space is expressed as

$$
P(Y = c_k) \prod_{j=1}^{n} P(X^{(j)} = x^{(j)} \mid Y = c_k).
\tag{15}
$$

Determine the category of input x according to formula (15), and the determined formula is expressed as

$$
y = \arg\max P(Y = c_k) \prod_{j=1}^{n} P(X^{(j)} = x^{(j)} \mid Y = c_k).
\tag{16}
$$

Using Bayes to estimate the conditional probability, the result is

$$
P(X^{(j)} = a_{jl} \mid Y = c_k) = \frac{\sum_{i=1}^{N} I(x_i^{(j)} = a_{jl}, y_i = c_k)}{\sum_{i=1}^{N} I(y_i = c_k) \cdot S_j}.
\tag{17}
$$

In formula (17), S_j represents the total number of characteristic values [30, 31].

Based on formula (17), a classifier that maximizes a posteriori probability is obtained, and its expression is

$$
f(x) = \arg\max P\left(\frac{Y = c_k}{X = x}\right).
\tag{18}
$$

The above process completes the design of Bayesian classifier, which provides a solid support for intelligent recognition of sports athletes' wrong movements.

After the above basic processing, the feature vectors and their tags are given to the classifier for action recognition. In general, not all of the data contained in the feature data is useful but also contains data that is not relevant. To avoid this, the feature dimension of the data needs to be subtracted [32] by classifying it as shown in Figure 5.

Based on the above process, SVM is used to classify the input features in advance. When there is an error classification, there is a support vector near the hyperplane of the classification [33, 34]. In classification, the test set is represented by T, T_{su} is the support vector set, and k is the number of classifiers. The process is as follows:

Step 1: support vector machine algorithm is used to calculate the corresponding support vector and to solve the coefficient and constant b.

Step 2: if T is not an empty set, take $x \in T$, and if T is an empty set, stop [35].
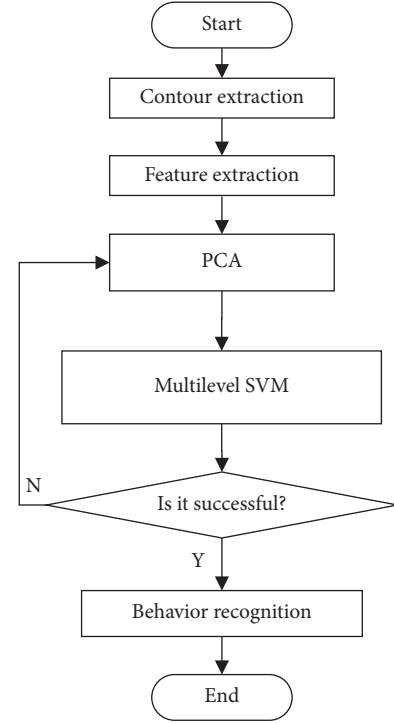
Figure 5: Flowchart of hybrid classification model.

Step 3: calculate $g(x) = \sum y_i \alpha_i K(x_i, x) - b$.

Step 4: if $g(x) > \varepsilon$, then $f(x) = \mathrm{sgn}(g(x))$ is directly used as the output of the classifier, and if $g(x) > \varepsilon$, it is subsumed into the classifier for classification [36].

Based on the above process, we classify the support vectors and complete the recognition of the athlete's movement.

3. Simulation Experiment Design and Result Analysis

3.1. Image Acquisition of Experimental Samples. The main equipment used in the experiment is the image acquisition device, which can make the wrong movement of sports transient and fast. The experiment adopts a short time acquisition and storage system to realize the acquisition and storage of the experimental image, which is composed of the camera, the acquisition card, the cable, the computer, and the acquisition software. The parameters of the camera are shown in Table 2.

The experimental sample selects two sports videos as the experimental objects. The two videos are football match and gymnastics match, respectively. In order to verify the reliability and effectiveness of the model, the following experiments are designed. In view of the experimental object sports video frequency to carry on the athlete wrong movement recognition, the comparison method selects reference [3] proposed based on the multicharacteristic fusion athlete posture recognition method and reference [4] proposed based on the multithreshold optimization movement image outline characteristic extraction method. The recognition results of different methods are given in the

TABLE 2: Camera parameters.

Parameter name	Numerical value	Company
Image type	Mono/color	—
Pixel depth	8/10	Bit
Pixel size	5.5×5.5	μm
Data output type	Camera link	-
Maximum frame rate	340	fps
Maximum resolution	2048×1088	Pixels
Overall dimension	$63.5 \times 63.5 \times 44.1$	mm

(a)   (b)

FIGURE 6: Detection results of the algorithm studied in this paper. (a) Test sample 1. (b) Test sample 2.

(a)   (b)

FIGURE 7: Recognition results of athlete's human posture recognition method based on multifeature fusion. (a) Test sample 1. (b) Test sample 2.

experimental results. The specific recognition results are shown in Figures 6–8.

Through the analysis of the two images, it can be seen that the algorithm studied in this paper can clearly identify the key wrong actions of athletes. There are some redundant results in the recognition results of athletes' human posture recognition method based on multifeature fusion and moving image contour feature extraction method based on multithreshold optimization, and the recognition results are not unique, and the error is relatively large compared with the actual action.

3.2. Image Denoising Experiment. To verify the performance of the proposed method, within 250 frames, two frames are extracted: as shown in Figures 9 and 10, there is uneven noise distribution in the two frames, the contrast is too high, the image is denoised by the proposed method, and the result is shown in Figures 9 and 10.

It can be seen from Figures 9 and 10 that after the image denoising is completed, the noise elimination effect in the initial image is good, which provides a strong basis for subsequent action image recognition.

Figure 8: Recognition results of moving image contour feature extraction method based on multithreshold optimization. (a) Test sample 1. (b) Test sample 2.

Figure 9: Takeoff image with noise and back high jump image. (a) Test sample 1. (b) Test sample 2.

Figure 10: Two-frame image denoising effect. (a) Test sample 1. (b) Test sample 2.

3.3. Accuracy Test of Athletes' Wrong Movement Recognition. In addition, in order to further verify the better accuracy of the proposed algorithm, we compare it with the traditional algorithm, establish a 3D visual inspection model, and then test two different methods for many times, so as to get the results shown in Figure 11. As can be seen from Figure 5, the accuracy is above 90%, while the progress of the traditional algorithm is between 70% and 77%. So the algorithm studied in this paper has better accuracy and can control the error in a reasonable range.
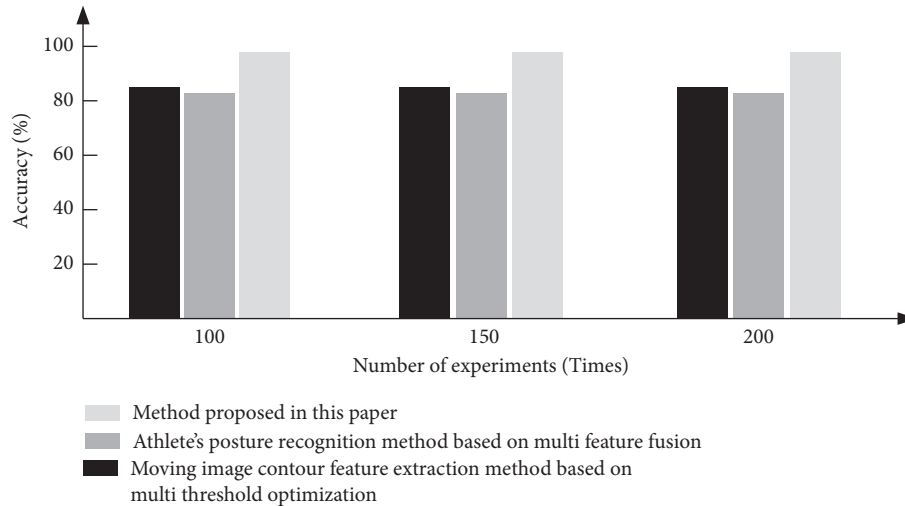
Figure 11: Comparison of recognition accuracy.
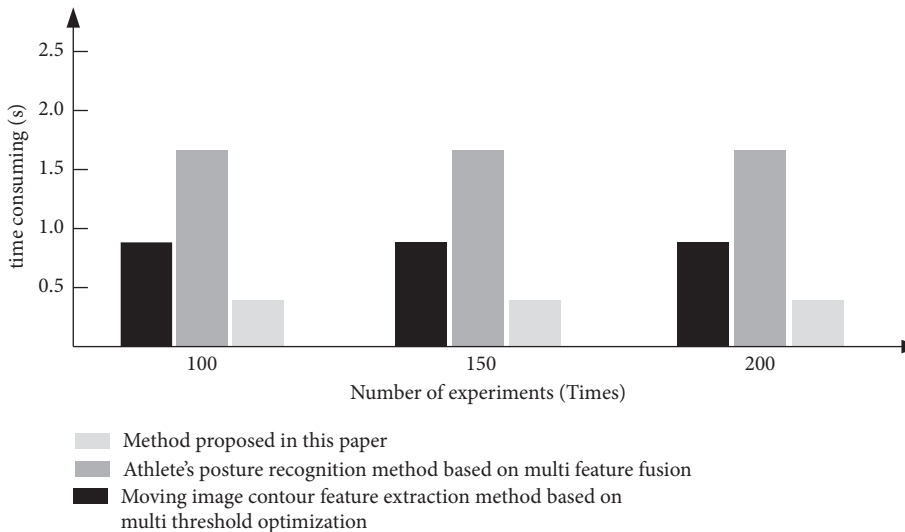
Figure 12: Time-consuming intelligent recognition of wrong actions.

3.4. Time-Consuming Comparative Test of Different Methods. Comparing the three methods for the intelligent recognition of erroneous actions of 200 successful samples of erroneous actions takes time, and the comparison results are shown in Figure 12.

As shown in Figure 12, the intelligent recognition time of reference [3] method for each wrong action is 0.25 s on average, the intelligent recognition time of reference [4] method for each wrong action is 0.39 s on average, and the intelligent recognition time of each wrong action is 0.08 s on average. Therefore, the proposed method has high speed and accuracy of intelligent recognition of athlete's wrong movement, which fully shows that the model has good intelligent recognition performance.

To sum up, the intelligent recognition method of sports athletes' wrong actions based on image vision has good effect and high recognition accuracy. It can complete the recognition of sports athletes' wrong actions in a shorter time, and the recognition effect is ideal.

4. Conclusion

(1) This paper puts forward an intelligent recognition method of athletes' wrong action based on image vision, which improves the application disadvantages of traditional athletes' wrong action recognition methods

(2) The intelligent recognition method of sports athletes' wrong actions based on image vision has ideal recognition accuracy

(3) The image noise is low, which can complete the recognition of sports athletes' wrong actions in a shorter time

Data Availability

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Conflicts of Interest

The authors declare that they have no conflicts of interest regarding this work.

References

[1] E. Chen and J. fan, "Method of human action feature extraction and recognition based on MEM-LBP [J]," *Application Research of Computers*, vol. 35, no. 4, pp. 1277–1280, 2018.

[2] X. Luo, H. Li, X. Yang, Y. Yu, and D. Cao, "Capturing and understanding workers' activities in far-field surveillance videos with deep action recognition and bayesian nonparametric learning," *Computer-Aided Civil and Infrastructure Engineering*, vol. 34, no. 4, pp. 333–351, 2019.

[3] S. Liu, "Player's posture recognition algorithm based on multi-feature fusion," *Information & Technology*, vol. 23, no. 8, pp. 17–19, 2019.

[4] C. Chen, "Extraction method of contour features by multi-threshold optimization for motion images," *Journal of Shenyang University of Technology*, vol. 41, no. 3, pp. 315–319, 2019.

[5] Y. P. Li, T. T. Liu, and L. Zhang, "Human action recognition based on deep learning," *Application Research of Computers*, vol. 41, no. 4, pp. 261–265, 2020.

[6] S. Cagnoni, H. Al-Sahaf, Y. Sun, B. Xue, and M. Zhang, "Special issue on evolutionary computer vision, image processing and pattern recognition," *Applied Soft Computing*, vol. 97, no. 12, Article ID 106675, 2020.

[7] L. Antoniomeira, L. E. Techiopereira, M. E. Rozalinosantos et al., "USPLeaf: automatic leaf area determination using a computer vision system1," *Revista de Ciencias Agronomicas*, vol. 51, no. 4, pp. 1–10, 2020.

[8] Z. Liao, H. Hu, and Y. Liu, "Action recognition with multiple relative descriptors of trajectories," *Neural Processing Letters*, vol. 51, no. 1, pp. 287–302, 2020.

[9] C. Cao, Y. Zhang, C. Zhang, and H. Lu, "Body joint guided 3-D deep convolutional descriptors for action recognition," *IEEE Transactions on Cybernetics*, vol. 48, no. 3, pp. 1095–1108, 2018.

[10] S. Chakraborty, S. H. Shaikh, A. Chakrabarti, and R. Ghosh, "An image denoising technique using quantum wavelet transform," *International Journal of Theoretical Physics*, vol. 59, no. 11, pp. 3348–3371, 2020.

[11] A. Muthukrishnan, J. Charles Rajesh kumar, D. Vinod Kumar, and M. Kanagaraj, "Internet of image things-discrete wavelet transform and Gabor wavelet transform based image enhancement resolution technique for IoT satellite applications," *Cognitive Systems Research*, vol. 57, no. 10, pp. 46–53, 2019.

[12] C.-J. Zhang, X.-Y. Huang, and M.-C. Fang, "MRI denoising by NeighShrink based on chi-square unbiased risk estimation," *Artificial Intelligence in Medicine*, vol. 97, no. 6, pp. 131–142, 2019.

[13] C. Rodrigues, Z. M. Assis Peixoto, and F. Magalhaes Freitas Ferreira, "Ultrasound image denoising using wavelet thresholding methods in association with the bilateral filter," *IEEE Latin America Transactions*, vol. 17, no. 11, pp. 1800–1807, 2019.

[14] C. Vimala and P. A. Priya, "Artificial neural network based wavelet transform technique for image quality enhancement," *Computers & Electrical Engineering*, vol. 76, no. 6, pp. 258–267, 2019.

[15] A. Ullah, J. Ahmad, K. Muhammad, M. Sajjad, and S. W. Baik, "Action recognition in video sequences using deep Bi-directional LSTM with CNN features," *IEEE Access*, vol. 6, no. 99, pp. 1155–1166, 2018.

[16] H. Wu and Z. Cheng, "Action recognition algorithm based on complexity measure and multi-scale motion coding," *Optical Technique*, vol. 44, no. 04, pp. 427–434, 2018.

[17] I. Al-Nahhal, O. A. Dobre, E. Basar, C. Moloney, and S. Ikki, "A fast, accurate, and separable method for fitting a Gaussian function [tips & tricks]," *IEEE Signal Processing Magazine*, vol. 36, no. 6, pp. 157–163, 2019.

[18] X. Lu and Y. Zhang, "Human body flexibility fitness test based on image edge detection and feature point extraction," *Soft Computing*, vol. 24, no. 12, pp. 8673–8683, 2020.

[19] F. Yuan, G. Li, X. Xia, B. Lei, and J. Shi, "Fusing texture, edge and line features for smoke recognition," *IET Image Processing*, vol. 13, no. 14, pp. 2805–2812, 2019.

[20] Z. Wu and Z. Zheng, "Action recognition algorithm based on deep learning and motion information," *Computer Engineering and Design*, vol. 39, no. 08, pp. 2668–2674, 2018.

[21] Y. Shi, M. Sun, Z. Li, J. Luo, and M. Yang, "Action recognition based on motion history image and convolution neural network," *Natural Science Journal of Xiangtan University*, vol. 041, no. 02, pp. 109–117, 2019.

[22] N. Slavkovic and M. Bjelica, "Risk prediction algorithm based on image texture extraction using mobile vehicle road scanning system as support for autonomous driving," *Journal of Electronic Imaging*, vol. 28, no. 3, pp. 1–033034.13, Article ID 033034, 2019.

[23] D. Ludl, T. Gulde, and C. Curio, "Enhancing data-driven algorithms for human pose estimation and action recognition through simulation," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 9, pp. 3990–3999, 2020.

[24] M. Hajizadeh and H. Ebrahimnezhad, "Eigenspace compression: dynamic 3D mesh compression by restoring fine geometry to deformed coarse models," *Multimedia Tools and Applications*, vol. 77, no. 15, pp. 19347–19375, 2018.

[25] P. X. Fuchs, A. Fusco, J. W. Bell, S. P. von Duvillard, C. Cortis, and H. Wagner, "Movement characteristics of volleyball spike jump performance in females," *Journal of Science and Medicine in Sport*, vol. 22, no. 7, pp. 833–837, 2019.

[26] K. Hirokatsu, S. Yutaka, A. Yoshimitsu, O. Shoko, and M. Yasuhiro, "Temporal and fine-grained pedestrian action recognition on driving recorder database," *Sensors*, vol. 18, no. 2, pp. 627–641, 2018.

[27] F. Meng, B. Xu, T. Zhang, B. A. Muthu, and C. B. Sivaparthipan, "Application of AI in image recognition technology for power line inspection," *Energy Systems*, vol. 248, no. 1, pp. 1–23, 2021.

[28] U. B. Angadi, A. Rai, and G. Uma, "MBFerns: classification and extraction of actionable knowledge using Multi-Branch Ferns-based Naive Bayesian classifier[J]," *Soft Computing*, vol. 25, no. 6, pp. 1–13, 2021.

[29] T. T. T. Nguyen, T. T. Nguyen, R. Sharma, and A. W.-C. Liew, "A lossless online bayesian classifier," *Information Sciences*, vol. 489, no. 7, pp. 1–17, 2019.

[30] H. Zhang, B. Wen, J. Liu, and Y. Zeng, "The prediction and error correction of physiological sign during exercise using bayesian combined predictor and naive bayesian classifier," *IEEE Systems Journal*, vol. 13, no. 12, pp. 4410–4420, 2019.

[31] S. Wang, S. Zhang, T. Wu, Y. Duan, L. Zhou, and H. Lei, "FMDBN: a first-order Markov dynamic Bayesian network classifier with continuous attributes," *Knowledge-Based Systems*, vol. 195, no. 3, Article ID 105638, 2020.

[32] A. Maa, B. Ac, C. Sv, and K. Abdessamad, "A single Bayesian network classifier for monitoring with unknown classes," *Engineering Applications of Artificial Intelligence*, vol. 85, no. 10, pp. 681–690, 2019.

[33] C. Mao, L. Lu, and B. Hu, "Local probabilistic model for Bayesian classification: a generalized local classification model [J]," *Applied Soft Computing*, vol. 93, Article ID 106379, 2020.

[34] Z. Shen, Z. Man, Z. Cao, and J. Zheng, "A new intelligent pattern classifier based on structured sparse representation," *Computers & Electrical Engineering*, vol. 84, no. 5, Article ID 106641, 2020.

[35] J. Liu, H. Rahmani, N. Akhtar, and M. Ajmal, "Learning human pose models from synthesized data for robust RGB-D action recognition," *International Journal of Computer Vision*, vol. 127, no. 8, pp. 1545–1564, 2019.

[36] L. Dong, D. Moc, E. Kmc et al., "A computer vision-based method for spatial-temporal action recognition of tail-biting behaviour in group-housed pigs," *Biosystems Engineering*, vol. 195, no. 7, pp. 27–41, 2020.

Research Article

Research on Music Emotion Intelligent Recognition and Classification Algorithm in Music Performance System

Chun Huang and Diao Shen [ID]

General Education and International College, Chongqing College of Electronic Engineering, ChongQing 400031, China

Correspondence should be addressed to Diao Shen; 200402007@cqcet.edu.cn

The music performance system works by identifying the emotional elements of music to control the lighting changes. However, if there is a recognition error, a good stage effect will not be able to create. Therefore, this paper proposes an intelligent music emotion recognition and classification algorithm in the music performance system. The first part of the algorithm is to analyze the emotional features of music, including acoustic features, melody features, and audio features. Then, the three kinds of features are combined together to form a feature vector set. In the latter part of the algorithm, it divides the feature vector set into training samples and test samples. The training samples are trained by using recognition and classification model based on the neural network. And then, the testing samples are input into the trained model, which is aiming to realize the intelligent recognition and classification of music emotion. The result shows that the kappa coefficient k values calculated by the proposed algorithm are greater than 0.75, which indicates that the recognition and classification results are consistent with the actual results, and the accuracy of recognition and classification is high. So, the research purpose is achieved.

1. Introduction

Watching entertainment programs has become one of the main leisure activities in our daily lives. When watching the program, we can often see that the lights will change with the performance of the performers, so as to render the performance scene and drive the scene atmosphere, which helps the performers to complete their performance better. The conversion of the performance lighting is mainly completed under the control of the music performance system. Namely, the control principle of the music performance system is to control the lighting by classifying and identifying the emotions contained in the performers' music. For example, if the emotions expressed in the music are cheerful, joyful, and enthusiastic, the corresponding rhythm of lighting will change fast, and the light color will be red or other bright colors. On the contrary, if the emotional factors expressed in music are depressive and serious, the corresponding rhythm of lighting will change slowly. And, the light color will be dark blue or other deep cool colors [1]. Once the emotion

recognition is wrong, it is very likely to ruin a performance. Based on the above background, music emotion recognition and classification has become the most critical part of the music performance system and the focus of researches.

There are many researches on emotion recognition and classification contained in music. For example, in literature [2], Wang Jinhua, Ying Na, Zhu Chendu, and others extracted emotion strong correlation feature spectrogram through the hybrid convolution neural network model and recognized emotion in music on this basis. In literature [3], Wang Jie and Zhu Beibei took Chinese lyrics as the main object, extracted the emotional feature keywords contained in lyrics, and then calculated the similarity between words and Chinese emotional dictionary, to achieve music emotional classification. In literature [4], Li Qiang and Liu Xiaofeng constructed a probabilistic neural network model (PNN) to classify music emotion, extracted feature parameters in the process of music playing, and then input the feature parameters into the PNN model to complete emotion classification.

Combined with the experience in previous researches, in order to improve the accuracy of recognition and classification, this study extracts the emotional features contained in music from multiple aspects and then constructs a multifeature space vector. Finally, the classification of music emotion is achieved by using the constructed recognition and classification model, which helps the completion of the control of lighting in the music performance system. The purpose of this study is to help performers complete stage music performance and improve the performance appeal.

2. Research on Music Emotion Intelligent Recognition and Classification Based on Multifeatures

A perfect music performance is not only achieved by the help of necessary musics but also the complementary scene atmosphere. In the music performance, the setting of the scene atmosphere is mainly realized by lighting, which often changes with the emotional factors expressed in the music and assists the music to create a good stage effect [5]. Based on this, in order to control the lighting in the performance, realizing the music emotion recognition is very important. Therefore, to achieve a more suitable music emotion, it is necessary to construct the music recognition and classification model and complete the research on intelligent recognition and classification of music emotion in the music performance system.

2.1. Analysis of the Features of Music Emotion. The realization of music emotion recognition and classification is based on music emotion features, so the extraction of music emotion features is analyzed in the former part of this study. In the previous classification of music emotions, most studies were just based on one kind of music feature. Although this kind of study can also complete the classification task, its accuracy cannot be guaranteed [6]. In order to solve the above problem, this paper analyzes all the factor features of music emotion and then combines them together into a feature vector, which will do a great help to carry out the purpose.

Considering that the premise of extracting emotional features in music is to understand the composition of music, among which the music-related factors can show emotional features more obviously. It includes acoustic features, melody features, and audio features [7]. In view of these three aspects, the following content will analyze it specifically.

Among them, acoustic characteristics, melodic characteristics, and audio characteristics are the general music-related factors that can show emotional characteristics more obviously [7]. According to the above analysis, the Thayer emotion model constructed mainly includes two dimensions: energy and pressure, as shown in Figure 1. The above two dimensions representing the abscissa and ordinate, respectively, correspond to Figure 1 that can reflect the measurement of standard strength. From left to right, the abscissa corresponds to emotion from happiness to sadness,
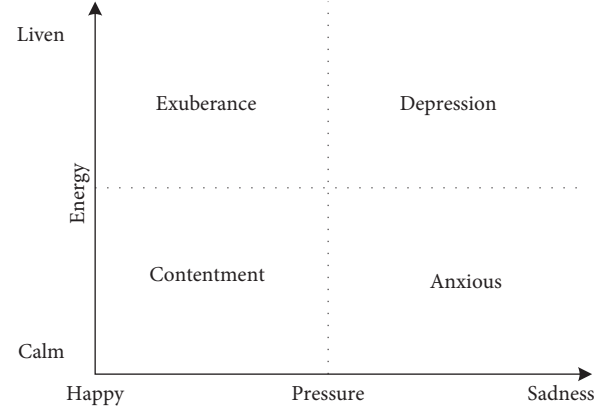
FIGURE 1: Thayer emotion model.

and from bottom to top, the ordinate corresponds to emotion from calm to vitality.

Based on Figure 1, the following content specifically analyzes acoustic features, melodic features, and audio features.

2.1.1. Acoustic Features. Acoustic factor is the most basic component of music. Music with different emotions has different acoustic features. The basic corresponding relationship is shown in Table 1 [8].

2.1.2. Melody Features. Melody is the overall beat and tune of music, which can be described by the following five characteristic parameters [9].

(1) Balance Parameter Y1. Balance refers to the proportion of volume in the left and right channels. The calculation formula is as follows:

$$Y1 = P(k) - \frac{64}{127}, \quad k = 1, 2, \ldots, 16. \tag{1}$$

In the formula, $P(k)$ represents the equilibrium value, $P(k) \in (0 \sim 127)$.

(2) Volume Parameter Y2. Volume refers to the loudness of sound that can be heard by human [10]. The calculation formula is as follows:

$$Y2 = \frac{V}{127}. \tag{2}$$

In the formula, V represents the total volume loudness, and its range is $V \in (0 \sim 127)$.

(3) Pitch Parameter Y3. Pitch refers to the vibration frequency of the note fundamental frequency. Music with fast rhythm has fast vibration frequency; on the contrary, it has slow vibration frequency [11]. The calculation formula is as follows:

$$Y3 = \frac{1}{127n} \sum_{i=1}^{n} P_{ti}. \tag{3}$$

In the formula, P_{ti} represents the pitch value, n represents the number of note fundamental frequency.

TABLE 1: The corresponding relationship between acoustic features and music emotion.

Acoustic features	Pronunciation	Tone quality	Speech speed	Intensity	Mean pitch	Pitch range
Happy	Normal	Breathing sound, chirping sound	Fast or slow	High	Very high	Very wide
Hate	Normal	Mumble, chest voice	Extremely fast	Low	Extremely low	Slightly wider
Anger	Tight	Breathing sound	Slightly faster	High	Extremely high	Very wide
Sadness	Vague	Resonance sound	Slightly slower	Low	Slightly lower	Slightly narrow
Fear	Clear	Shrill voice	Very fast	Normal	Extremely high	Very wide

(4) Average Strength Parameter Y4. Strength refers to the strength of the power generated by music. The soothing music has weak strength, while the more shocking music has strong strength [12]. The calculation formula is as follows:

$$Y4 = \sum_{i=1}^{N} \frac{U_{el(m,i)}}{N}, \quad m = 1, 2, \ldots, 16. \quad (4)$$

In the formula, $U_{el(m,i)}$ represents the sound's intensity factor; m represents the sort number of musical notes; and N indicates the total number of musical notes.

(5) Energy Parameter of Musical Notes. Energy of musical notes refers to the sum of the product of the pitch and the length of a note. The calculation formula is as follows.

$$Y5 = \sum_{i=1}^{n} \left(p_{ji} \times d_{ji} \right), \quad j = 1, 2, \ldots, 16. \quad (5)$$

In the formula, p_{ji} represents the pitch and length of i notes in track j.

2.1.3. Features of Audio Frequency. Audio frequency is one of the important factors in music, which affects the rhythm of music. The faster the rhythm is, the more obvious the audio is, and the happier the music emotion will be. On the contrary, the music emotion is more dull or depressed [13]. Audio features are described from two aspects, the following part is a detailed analysis.

Time domain characteristics are as follows.

(1) Zero Crossing Rate Z_n. Zero crossing rate refers to the frequency of the audio signal waveform passing through the zero level. Generally speaking, in a piece of music, the zero crossing rate of high frequency band will be higher; otherwise, the zero crossing rate will be lower. Through this parameter, we can well distinguish between unvoiced and voiced music. Generally, unvoiced music is mostly used in cheerful music, while voiced music is often used in low and deep music [14]. The zero crossing rate is calculated as follows:

$$Z_n = \frac{\sum_{m=1}^{N} \text{sgn}\left[s_n x(m)\right] - \text{sgn}\left[s_n x(m-1)\right]}{2N}. \quad (6)$$

In the formula, $s_n x(m)$ represents the sign function of audio signal of $x(m)$ and N represents the effective width of the window.

(2) Range M_n. Range refers to the width of the waveform vibration of the audio signal. The more passionate the music is, the greater the audio amplitude is. The more soothing the

music is, the smoother the audio amplitude is. The audio amplitude is described as follows:

$$M_n = \sum_{m=n-(N-1)}^{n} |x(m)w(n-m)| = \sum_{m=n-(N-1)}^{n} |x(m)|w(n-m). \quad (7)$$

In the formula, $w(n-m)$ represents the moving window function.

(3) Frequency Domain Characteristics. The frequency domain characteristics of audio include two parts: spectrum centroid R_t and spectrum flux F_t. The calculation formula is as follows:

$$\text{spectrum centroid } R_t = \frac{\sum_{n=1}^{N} q_t[n] \times n}{\sum_{n=1}^{N} q_t[n]},$$

$$\text{spectrum flux } H_t = \sum_{n=1}^{N} \left(S_t[n] - S_{t-1}[n] \right)^2. \quad (8)$$

In the formula, $q_t[n]$ represents the amplitude of the short-time spectrum of the t frame at frequency point n and $S_t[n]$ and $S_t[n-1]$ represents the amplitude value of the spectrum at frame t at frequency n and $n-1$.

Based on the above three types of 14 music emotional features, the feature vector is constructed to describe the emotional factors of a part of music or a piece of music. It is described as follows:

$$\begin{aligned} U &= \{U_1, U_2, U_3\}, \\ U_1 &= \{S_1, S_2, S_3, S_4, S_5\}, \\ U_2 &= \{Y_1, Y_2, Y_3, Y_4, Y_5\}, \\ U_3 &= \{Z_n, M_n, C_t, F_t\}. \end{aligned} \quad (9)$$

In the formula, U_1 represents acoustic features; U_2 represents the characteristics of melody; U_3 represents audio features; S_1 represents the speed of speech; S_2 represents the pitch; S_3 represents strength; S_4 represents sound quality; and S_5 stands for pronunciation.

2.2. Construction of Music Emotion Recognition and Classification Model. Based on the emotional features contained in the above music, this chapter uses the BP neural network to establish a recognition and classification model to realize the recognition and classification of music emotion.

BP neural network is an intelligent algorithm that simulates the neural work of brain, which mainly includes three layers, and the classification processing is realized through the operation of each layer [15]. The recognition and classification model constructed by this algorithm is shown in Figure 2.
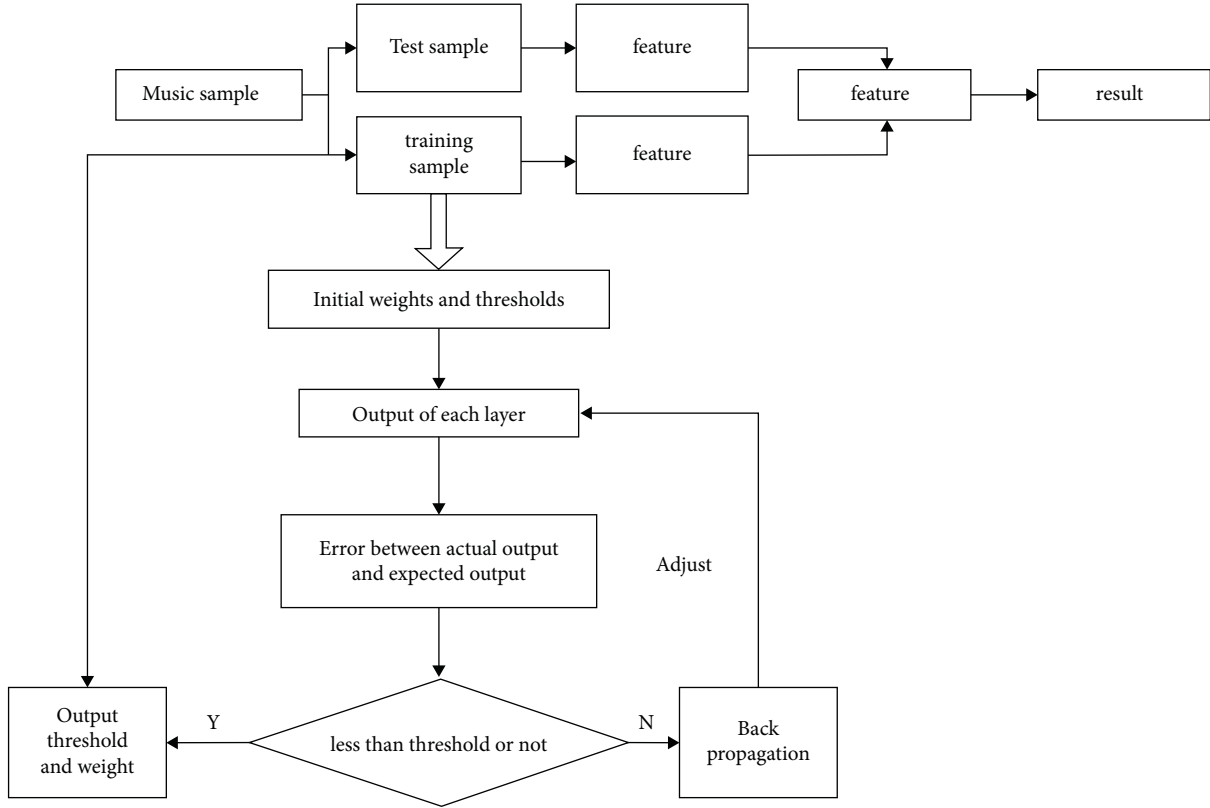
FIGURE 2: Schematic diagram of construction and working principle of recognition and classification model.

In Figure 2, the model needs to be trained before practical application. The specific process is as follows: through feedforward operation of each layer, it can obtain the output result and subtract the output result from the expected result. When the difference between the two is less than the set threshold, the difference is propagated backward after training, and the process is repeated from output to input until the weight and threshold reach the optimal [16]. The purpose of the BP neural network training is to adjust and optimize the weights and thresholds of every two levels in the model. Therefore, the formula is given as follows:

(1) Connection weight w_{ij} and threshold θ_j between the first input layer and the second hidden layer, adjustment formula is as follows:

$$\begin{cases} w_{ij}(N+1) = w_{ij}(N) + \beta \cdot g_j^k \cdot c_i \\ \theta_j(N+1) = \theta_j(N) + \beta \cdot g_j^k \\ i = 1, 2, \ldots, n, \\ j = 1, 2, \ldots, m, \\ 0 < \beta < 1. \end{cases} \tag{10}$$

In the formula, g_j^k represents error value of the second layer; c_i represents input feature vector; N represents iterations; and k represents the number of training samples.

(2) Connection weight v_{jt} and threshold γ_t between the second hidden layer and the third output layer, adjustment formula is as follows:

$$\begin{cases} v_{jt}(N+1) = v_{jt}(N) + \alpha \cdot t_t^k \cdot y_j \\ \gamma_t(N+1) = \gamma_t(N) + \alpha \cdot t_t^k \\ j = 1, 2, \ldots, p, \\ t = 1, 2, \ldots, q, \\ 0 < \alpha < 1. \end{cases} \tag{11}$$

In the formula, t_t^k represents error value and y_j represents the output value of the second layer node.

After training of the model based on the BP neural network, the music emotion classification can be realized by inputting test music samples.

3. Example Analysis

In order to test the application effect of the multifeature recognition and classification algorithm in music emotion recognition of the music performance system, the following research takes MATLAB software as the algorithm operation platform, selects specific calculation examples, and carries out simulation test and analysis.

Based on the five emotions of happiness, disgust, anger, sadness, and fear in Table 1, this study constructs a simulation model of music emotion recognition process in the music performance system, as shown in Figure 3.
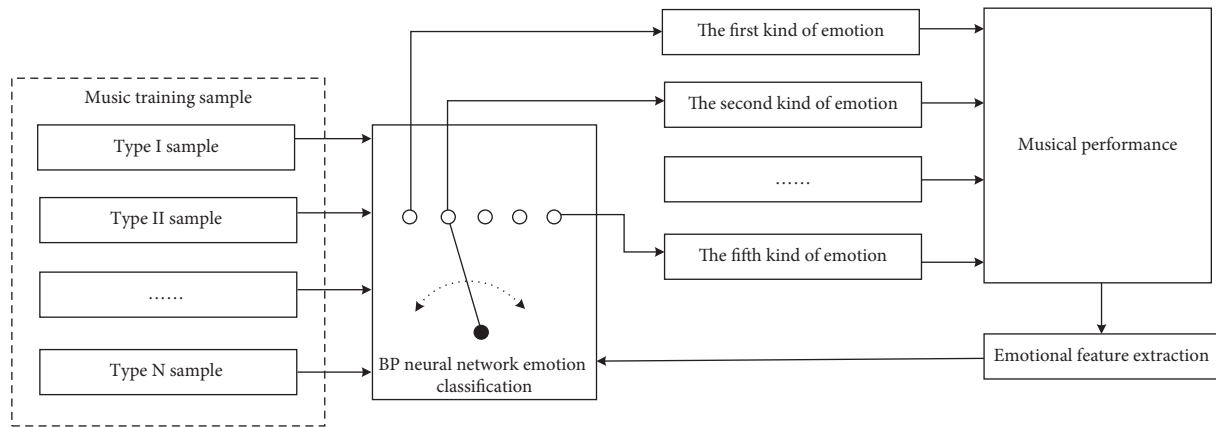
Figure 3: Simulation model.

As can be seen from Figure 3, the example analysis steps based on the simulation model are as follows: firstly, input different music training samples in MATLAB software, secondly classify music emotion through the BP neural network emotion classifier and then determine the music performance lighting pertinently, so as to achieve the purpose of analyzing music emotion.

3.1. Sample Selection.

The samples selected in the test come from three emotional corpora, namely, EMO-DB, Belfast, and *e* NTERFACE. According to the different emotions of the selected samples, the samples can be divided into five categories, and the specific distribution of the samples is shown in Table 2.

3.2. Neural Network Classification Model Training.

Input the training samples in Table 2 above into the neural network classification and recognition model for model training. The parameters of model training are set as follows: the number of nodes in the three-layer structure is 3-32-5; the maximum training times are set to 500 times; the target of convergence accuracy is less than 0.0002; tansig function is the transfer function of the hidden layer, and purelin function is the transfer function of output layer. The training page based on Simulink is shown in Figure 4.

It can be seen from Figure 4 that the convergence accuracy of the classification model based on the neural network finally stabilizes at 0.000125, meeting the convergence accuracy target (the set convergence accuracy is less than 0.0002), indicating that the performance of the constructed model meets the needs of subsequent classification and can be used in the actual music emotional intelligence classification test.

3.3. Recognition Results of Melody Feature.

Melody is the soul of music, and interval difference is the most basic element of melody movement. Interval difference refers to the formation of different melody combinations according to the differences between the high and low tones, so that people can perceive different musical images, thoughts, and emotions. Figure 5 shows five examples of interval difference statistics for music with five different emotions: happiness, sadness, tenderness, anger, and fear.

Figure 5 shows statistical results of interval difference. The smaller the interval difference, the higher the percentage of data, the more interesting and relaxed the main emotion of the music. The smaller the interval difference, the more intense and depressed the main emotion of the music. The frequent occurrence of large interval difference cannot increase the fluency of music, but will cause a sudden feeling. Therefore, in the main melody of music, there are also some differences in the ratio range of the same interval difference for music with different emotions.

3.4. Extraction Result of Audio Features.

With the acceleration of music rhythm in the music performance system, the more obvious the audio is, and the more cheerful the musical emotion is expressed. On the contrary, the musical emotion is dull or depressed, and the extraction results of audio features are shown in Figure 6.

According to the analysis results in Figure 6, the extraction area of a music audio feature is regular under a single emotion change, but the extraction area of audio feature is irregular under different emotion changes, which is consistent with the music emotion reflected by the actual rhythm.

3.5. Recognition Result of Spectral Flux.

Spectral flux refers to the difference mean value of the spectrum of all two adjacent audio frames, which reflects the dynamic characteristics of the music signal. The music signal includes three parts: unvoiced, voiced, and mute, which determines that the spectrum flux of music has a large range of variation. The change of performers' performance will also cause the change of spectrum flux. Therefore, the following takes music tracks containing multiple emotions as an example and carries out the spectrum flux identification test. The test results are shown in Figure 7.

According to the analysis result in Figure 7, with the different changes of music emotions in the music performance system, the output spectral density shows an irregular change trend. The main reason is that, there are many emotional

TABLE 2: Distribution of test samples.

Emotional style	Name of music segment	Feature dimension	Sample attribute
Joy	Segment of *The Red Head Rope*	455	Test sample
	Segment of *It Rains on Our Love*	355	Test sample
	Segment of *Carmen*	784	Test sample
	Segment of *The Trout*	232	Test sample
Grief	Segment of *Theme From Schindler's List*	534	Test sample
	Segment of *Butterfly Lovers*	454	Test sample
	Segment of *"Pathetique"*	234	Test sample
	Segment of *Farewell*	545	Test sample
Gentleness	Segment of *Lullaby*	215	Test sample
	Segment of *Blue Danube waltz*	313	Test sample
	Segment of *Little Star*	534	Test sample
	Segment of *For Elise*	341	Test sample
Indignation	Segment of *Polonaise*	132	Test sample
	Segment 1 of *Destinies*	431	Test sample
	Segment of *The Internationale*	453	Test sample
	Empty Crazy	315	Test sample
Fear	Segment of *Gloomy Sunday*	341	Test sample
	Segment of *One Missed Call*	345	Test sample
	Segment of *Thirteen Pairs of Eyes*	422	Test sample
	Segment of *Step by Step*	244	Test sample
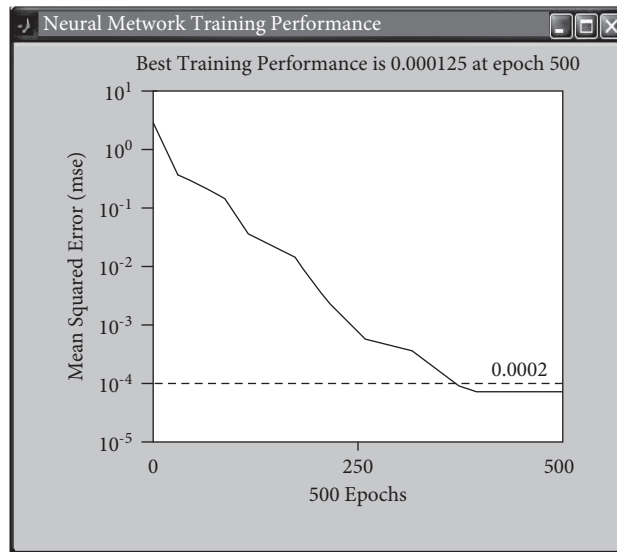
FIGURE 4: Training page of neural network recognition and classification model.

expressions in a music, and the music content (such as lyrics and background ect.) will affect the expression of emotion, which further leads to the irregular change of output spectral density.

3.6. Algorithm Evaluation Index. The kappa coefficient is the index to evaluate the performance of the algorithms. This coefficient is usually used to evaluate the consistency between the application results and the actual results of the algorithm, and its calculation formula is as follows:

$$k = \frac{p_o - p_e}{1 - p_e}. \tag{12}$$

In the formula, p_o represents observation consistency rate; p_e represents expected consistency rate; $k \in [-1, 1]$, the larger the value of K, the better the consistency is, and the closer the algorithm application result is to the actual result. Generally speaking, when $k \geq 0.75$, the recognition and classification is more accurate. If <0.4, it indicates that there is a lack of consistency, and the accuracy of recognition and classification is poor.

The test samples of Table 2 are tested by inputting them into the training results in Chapter 2.2, and the classification results of output recognition are shown in Figure 8.

Figure 5: Statistical results of interval difference.

Figure 6: Extraction result of audio features.

Figure 7: Recognition result of spectral flux.

Figure 8: Classification diagram of output identification.

4. Conclusion

In order to achieve the effect of stage performance, generally in the process of performance, the light will change with the artistic conception created by the performance under the control of the music performance system. Therefore, it can be seen that music emotion recognition is crucial to the lighting control. Based on this, this paper proposes an intelligent music emotion recognition and classification algorithm in the music performance system. The following conclusions are drawn from the study:

(1) The proposed algorithm mainly extracts emotional features from music, and then inputs them into the constructed recognition and classification model so as to clarify the emotion to be expressed in music and control the lighting.

(2) This study verifies the algorithm's performance by an example that proves the proposed algorithm can accurately recognize the music interval difference. The extraction area of audio feature is regular, and the output spectral density shows an irregular change trend, which is consistent with the actual trend of music expression. The kappa coefficient values are greater than 0.75, indicating that the recognition and classification results are in good agreement with the actual results, and the research goal is achieved.

(3) However, this study did not apply the algorithm to the actual music performance system and lacks some practical application, which needs to be further verified and analyzed in the future.

Data Availability

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Conflicts of Interest

The authors declare that they have no conflicts of interest regarding this work.

Based on Figure 8, the test results are counted according to formula (12) to calculate the kappa coefficient. The results are as follows:

$$k_{\text{joy}} = 0.855,$$
$$k_{\text{grief}} = 0.870,$$
$$k_{\text{gentleness}} = 0.912, \quad (13)$$
$$k_{\text{indignation}} = 0.825,$$
$$k_{\text{fear}} = 0.811.$$

In the formula, k_1, k_2, k_3, k_4, and k_5 represent the kappa coefficients of happiness, sadness, tenderness, anger, and fear, respectively. The kappa coefficient K values calculated by this method are all greater than 0.75, which indicates that the recognition and classification results are in agreement with the actual results, that is, the application results of algorithm are close to the actual results, and the recognition and classification accuracy is high. Then, the research purpose is achieved.

References

[1] Z. Duan and J. Yan, "Research on stage lighting control method based on music emotion recognition," *Computer Measurement and Control*, vol. 28, no. 11, pp. 95–100, 2020.

[2] J. Wang, N. Ying, C. Zhu, Z. Liu, and Z. Cai, "Speech emotion recognition algorithm on extraction of deep space attention characteristics based on spectrogram," *Telecommunication Science*, vol. 35, no. 07, pp. 100–108, 2019.

[3] J. Wang and B. Zhu, "Classification of musical emotions oriented to Chinese lyrics," *Computer Systems & Applications*, no. 8, pp. 24–29, 2019.

[4] L. Qiang and X. Liu, "Sentiment classification of music features based on PNN," *Computer Engineering and Design*, vol. 40, no. 2, pp. 528–532, 2019.

[5] X. Deng, G. F. Xing, M. I. Jianxun, D. Li, J. Wang, and Y. Tang, "Classifying emotional EEG using sparse representation method," *Application Research of Computers*, vol. 36, no. 3, pp. 801–806, 2019.

[6] G. Li, T. Yun, and L. Qi, "Multi-feature speech emotion recognition based on random forest classification optimization," *Microellectronics & Computer*, vol. 36, no. 1, pp. 70–73, 2019.

[7] C. E. Li and L. Zhi, "Research on classification of electronic music signals based on particle swarm optimization and support vector machine," *Modern electronic technology*, vol. 43, no. 21, pp. 51–54, 2020.

[8] H. Li, H. Li, and M. Lin, "Research on long term music emotion based on dynamic brain network," *Journal of Fudan University (Natural Science)*, vol. 59, no. 3, pp. 330–337, 2020.

[9] X. Li and C. Yun, "UGC automatic emotion recognition method based on joint optimization of feature selection and dispositional analysis," *Journal of Industrial Engineering and Engineering Management*, vol. 33, no. 2, pp. 61–71, 2019.

[10] Z. Zhu, J. Tao, H. Ge, and J. Jia, "Passive sonar target classification and recognition technique based on BPSO-KNN algorithm," *Acoustic technology*, vol. 38, no. 2, pp. 219–223, 2019.

[11] N. Lu, "Design of music emotion classification method based on audio and lyrics dual mode," *Techniques of Automation and Applications*, vol. 39, no. 5, pp. 166–169, 2020.

[12] X. Jing, "Design of intelligent detection algorithm for electronic music signal in complex noise scene," *Modern electronic technology*, vol. 43, no. 19, pp. 49–52, 2020.

[13] L. Yi, "Electronic music classification model based on multi feature fusion and machine learning algorithm," *Microcomputer Applications*, vol. 36, no. 9, pp. 117–119, 2020.

[14] T. Li, S. Ye, G. Ye, and Y. Zhu, "Research on speech emotion recognition algorithm based on ensemble learning," *Computer Technology and Development*, vol. 30, no. 6, pp. 82–86, 2020.

[15] T. Feng and Z. Yang, "Speech emotion recognition algorithm based on multi task learning and recurrent neural network," *Signal Processing*, vol. 35, no. 7, pp. 1133–1140, 2019.

[16] C. Chuang, R. chellali, and X. Yin, "Speech emotion recognition based on BP neural network optimized by improved genetic algorithm," *Application Research of Computers*, vol. 36, no. 2, pp. 344–346, 2019.

Research Article

Research on Action Recognition Method of Dance Video Image Based on Human-Computer Interaction

FenTian Peng[1] and Hongkai Zhang [ID][2]

[1]*Shangqiu Normal University College of Music, Hhnan 476000, China*
[2]*Physical Education Department of Guilin Teachers College, Guilin 541199, China*

Correspondence should be addressed to Hongkai Zhang; zhg@mail.glnc.edu.cn

Human-computer interaction technology simplifies the complicated procedures, which aims at solving the problems of inadequate description and low recognition rate of dance action, studying the action recognition method of dance video image based on human-computer interaction. This method constructs the recognition process based on human-computer interaction technology, constructs the human skeleton model according to the spatial position of skeleton, motion characteristics of skeleton, and change angles of skeleton, describes the dance posture features by generating skeleton node graph, and extracts the key frames of dance video image by using the clustering algorithm to recognize the dance action. The experimental results show that the recognition rate of this method under different entropy values is not less than 88%. Under the test conditions of complex, dark, bright, and multiuser interference, this method can make the model to describe the dance posture accurately. Furthermore, the average recognition rates are 93.43%, 91.27%, 97.15%, and 89.99%, respectively. It is suitable for action recognition of most dance video images.

1. Introduction

Action recognition is one of the focus in current research studies. Many scholars have optimized and innovated the action recognition technology according to the characteristics of action changes and human body structure [1, 2]. Among them, the documents [3] optimized the action recognition method by using the improved deep convolution neural network and built a new recognition network by combining the Google Net network model with the idea of batch normalization transformation. The 4 documents [4] introduced the MEMS sensor network to collect the acceleration and angular velocity of gymnastics and recognized gymnastic movements by the classification model based on standard deviation, mean square error, and other classification feature of parameter setting, which has a high recognition rate.

However, when the background of the video image is complex, the light is weak, or there is multiuser interference,

the recognition rate of the previous research methods will be greatly reduced. In the broadest sense, human-computer interaction includes the interaction between human and machine, human and computer, and human and robot. In the past few years, human-computer interaction technology has been widely used in various industries and has good development in education, medical treatment, manufacturing, and other aspects, which can improve the display effect of action recognition effect. So, it is necessary to develop a new action recognition method of dance video image based on human-computer interaction technology.

2. Action Recognition Method of Dance Video Image Based on Human-Computer Interaction

2.1. Construction of Human Skeleton Model Based on Human-Computer Interaction. Our human skeleton consists of more than 200 skeleton nodes, and each node has its own degree of

freedom. In order to recognize dance action, the human skeleton model has a certain complexity. This study adds Kinect device to the action recognition method and replaces all skeleton information with joint points and then generates the joint point image which is consistent with the contour of the human body by human-computer interaction technology, which all aims to feedback the correlation between skeletons and joint nodes.

Generated by Kinect device, the nodes of hands, wrists, feet, and ankles are closer. In order to simplify the calculation, the nodes such as wrist and ankle are removed to optimize the human skeleton model. In view of the spatial position of skeletons, this paper studies the changes of skeleton position caused by dance movements at a certain time [5, 6]. Taking the two actions of raising hands and kicking as examples, the lower body is still when raising hands, while the upper body is relatively still when kicking. Based on this, this study regards the neck as a central node, takes the neck node as the center, constructs the spatial position structure of the model, establishes the space coordinate system by taking other skeleton nodes as the center, and then uses the coordinates of other points to do subtraction. It can get the following results:

$$H_{ab} = \left\{ r_i - r_j |, \quad i, j = 1, 2, \ldots, N, i \neq j \right\}. \tag{1}$$

In the formula, H_{ab} represents an eigenvector whose coordinate is (a, b); r_i and r_j represent the horizontal coordinate of two adjacent skeleton nodes i and j on the same frame, respectively; and N represents the total number of nodes. Because there are differences in dancers' figure, the position from head to spine center is different. In order to minimize the influence, this design method calculates the absolute length from head to spinal center of identified dancer according to the calculation result of formula (1) and reduces the effects of figure difference. The formula is as follows:

$$H_{ab}' = \frac{H_{ab}}{L}. \tag{2}$$

In the formula, a represents the actual eigenvector under adaptive change and L represents the absolute length from head to spinal center. It is known that dancing is a dynamic movement process, so the constructed model needs to have the ability of dynamic change, which can generate matching recognition results according to the change of each frame in the video sequence [7–10]. Therefore, this method can calculate every skeleton node by subtraction based on the principle of difference operation between frames.

$$H_{bb}' = \left\{ r_i^\alpha - r_j^\beta |, \quad i, j = 1, 2, \ldots, N \right\}. \tag{3}$$

In the formula, α and β represent the frame sequence of two continuous changing actions, respectively. The generation of dance action depends on the skeleton angle, so the model also needs to have the basic characteristics of angle changes of skeleton. If the azimuth angle and elevation angle of the skeleton are set as θ and γ, it also can obtain the skeleton recognition angle of the model through the following equations:

$$\begin{cases} \theta = \arctan\left(\dfrac{r_i^z - r_j^z}{r_i^x - r_j^x} \right), \\[3mm] \gamma = \arctan\left(\dfrac{r_i^y - r_j^y}{\sqrt{\left(r_i^x - r_j^x \right)^2 + \left(r_i^z - r_j^z \right)^2}} \right). \end{cases} \tag{4}$$

Substituting the calculation results of formulae (2)–(4) into the human skeleton model, the recognition of human skeleton motion features in the dance video images under the human-computer interaction technology [11, 12] can be realized.

The model minimizes the influence of dancers' body difference through formula (2) and obtains more accurate action recognition results. With the support of formula (3), the model has the ability of dynamic change and generates matching recognition results according to the change of each frame of video sequence. The model has the basic characteristics of bone angle change under formula (4), so it can recognize the characteristics of human bone movement in dance video images based on human-computer interaction technology [11, 12].

2.2. Description of Dance Posture Characteristics.
Dance movement can be seen as a combination of multiple postures on the timeline, and its complexity and duration determine the length of the posture sequence. The skeleton feature recognition nodes of different dance postures can be obtained based on the above model. Under the control of human-computer interaction technology, the human skeleton model generates the posture according to the distance characteristics between the specific joint point and the central joint, among which 10 nodes were the most stable feature joint points.

According to the characteristic joint node $g_i^t = (x_i^t, y_i^t, z_i^t)$ and the "SpineBase" node $g_0^t = (x_0^t, y_0^t, z_0^t)$, it can calculate the relative distance between them and describe the dance posture at t time according to the result, in which $i = 1, 2, \ldots, 10$. The formula is as follows:

$$\left| \overrightarrow{G_0^t G_i^t} \right| = \sqrt{\left(x_i^t, x_0^t \right)^2 + \left(y_i^t, y_0^t \right)^2 + \left(z_i^t, z_0^t \right)^2}. \tag{5}$$

In the formula, G_i^t represents the spatial coordinates of the i-th characteristic joint point at the time t and G_0^t represents the initial node coordinate. In order to eliminate the movement differences caused by height, body shape, and other aspects, this model deals with the skeleton distance characteristics of different people by the equal proportion method and distinguishes the directional characteristics of dance movements of left and right, up and down, and so on [13, 14]. It is known that the ratio of length of joint skeleton to height is constant. If the distance feature is multiplied by the height scale factor λ, then $\lambda = 1/\Delta y$ can be obtained, in which Δy represents the absolute value of the y coordinate difference between "Head" joint and "FootLeft" joint [15–18]. Assuming that frame number of dance video image

is M, then the distance characteristic matrix of $n * M$ dimension is as follows:

$$
f = \begin{bmatrix} G_1 \\ G_1 \\ \vdots \\ G_t \\ \vdots \\ G_M \end{bmatrix}
$$

$$
= \lambda \begin{bmatrix} \left|\overrightarrow{G_0^1 G_1^1}\right| & \left|\overrightarrow{G_0^1 G_2^1}\right| & \cdots & \left|\overrightarrow{G_0^1 G_n^1}\right| \\ \left|\overrightarrow{G_0^2 G_1^2}\right| & \left|\overrightarrow{G_0^2 G_2^2}\right| & \cdots & \left|\overrightarrow{G_0^2 G_n^2}\right| \\ \cdots & \cdots & \cdots & \cdots \\ \left|\overrightarrow{G_0^t G_1^t}\right| & \left|\overrightarrow{G_0^t G_2^t}\right| & \cdots & \left|\overrightarrow{G_0^t G_n^t}\right| \\ \cdots & \cdots & \cdots & \cdots \\ \left|\overrightarrow{G_0^M G_1^M}\right| & \left|\overrightarrow{G_0^M G_2^M}\right| & \cdots & \left|\overrightarrow{G_0^M G_n^M}\right| \end{bmatrix}. \tag{6}
$$

According to the above process, this model realizes the recognition of dance video movements and describes the characteristics of dance posture. The description of recognition node of dancers' skeleton feature on the timeline is shown in Figure 1.

According to Figure 1, it can predict human activity intention and performance with high accuracy by determining the posture and motion of the specific dancer's skeleton feature recognition node on a given timeline. Although different people have different sizes and shapes of skeleton, their motion performance is similar when they perform similar movements. Therefore, for a given movement, the speed and acceleration of joint points are more stable than the position information of dancers' skeleton feature recognition node.

2.3. Image Action Recognition by Extracting Key Frame.

The sample video sequence of dance movements is established based on the description results of dance posture features, which is represented by $\{p^{(1)}, p^{(2)}, \ldots, p^{(M)}\}$. And there is $p^{(i)} \in W^{(M)}$. Here, $p^{(i)}$ represents the 3D coordinate position vector of the joint point in the ith frame of the sequence; $W^{(M)}$ represents the set of vectors; and $\{c^{(1)}, c^{(2)}, \ldots, c^{(k)}\}$ represents the extracted sample centers whose number is k. k cluster centroids are selected randomly and represented by $\{o_1, o_2, \ldots, o_K\}$, respectively,
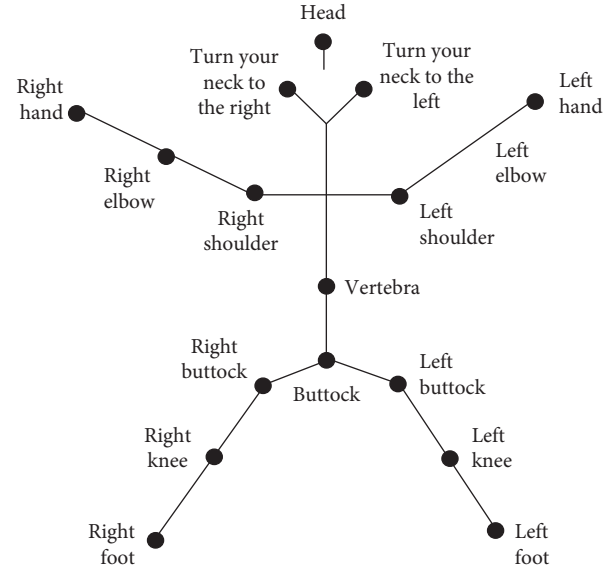
FIGURE 1: Identify nodes of dancer skeleton feature.

$\{o_1, o_2, \ldots, o_K\} \in W^{(M)}$ [19, 20]. It can calculate that the minimum distance between the sample and k cluster centroid in the random sample $p^{(i)}$, and the formula is as follows:

$$
D = \arg\min \sum_{i=1}^{M} \sum_{j=1}^{K} \left\| p^{(i)} - o_j \right\|^2. \tag{7}
$$

In the formula, D represents the minimum distance between the sample and the centroid. It can classify the samples into a class set that should belong to a certain feature according to the above calculation results, and then the centroid is recalculated for each class.

$$
o_j = \frac{\sum_{i=1}^{M} l_{ij} p^{(i)}}{\sum_{i=1}^{M} l_{ij}}. \tag{8}
$$

In the formula, l_{ij} represents the feedback constant when the vector $p^{(i)}$ is classified; $l_{ij} = 1$ exists when the class is j; otherwise, $l_{ij} = 0$ [21, 22]. k sample centers are returned according to the calculation results, and the above steps are repeated until the convergence is completed.

The frame clustering is completed according to the above process, and the Euclidean distance is calculated between the sample and the mth frame. The formula is as follows:

$$
I = \left[\text{dis}\left| (x_{i1}, y_{i2}, z_{i3}), (x_{m1}, y_{m2}, z_{m3}) \right|, \ldots, \text{dis}\left| (x_{ij}, y_{ij}, z_{ij}), (x_{mj}, y_{mj}, z_{mj}) \right| \right]. \tag{9}
$$

After calculating the Euclidean distance for all the frames of the dance video image, the minimum value of $\text{dis}|(x_{ij}, y_{ij}, z_{ij}), (x_{mj}, y_{mj}, z_{mj})|$ is recorded as 1, otherwise as 0. The extraction of key frames is completed so far. According to the deep fusion feature of the result, namely, the fusion of the location feature and the angle feature of the joint points, the movement recognition of the dance video image is realized based on the human-computer interaction [23–25].

3. Case Testing

3.1. Setup Human-Computer Interaction Test Environment. The design of the interactive interface can be realized through Windows Presentation Foundation. In order to ensure the stability of the test, this example selects a four-wheel robot for man-machine interaction, and the upper limb of the robot can simulate human arm movement. This example sets 10 kinds of dance movements according to the human-computer interaction technology. The action picture is shown in Figure 2.

The robot can recognize actions and make corresponding postures according to the ten groups of dance action images in Figure 2. The recognition device of the robot is shown in Figure 3.

The robot can recognize the movements of these dance video images and make corresponding gestures. The recognition device is shown in Figure 3.

In Figure 3, the speed control is a very important link in the robot driving process. The four-wheel self identification robot can adjust driving force according to the position signal collected by the recognition device. If the driving force is greater than the resistance, the robot will continue to accelerate at a certain acceleration. And it is easy to rush out as turn the corner when the speed is too high. In the same way, if the driving force is less than the resistance, the robot will continue to slow down. The robot tested in this example must keep the corresponding speed according to the actual situation, which is to ensure the unique variable in the process of human-computer interaction and simulation.

The human-computer interaction interface is shown in Figure 4.

In the human-computer interaction display interface, the first step is to record the dance video. During this process, the recorder should stand within the effective recognition range of the Kinect sensor. At this time, the skeleton image and dance video image can be shown in the interface. And the name of the recorded action in ComboBox control is selected. The second step is to click "Motion capture," and the human-computer interface display comes into 3S countdown. The recorder can start to show the dance movements after timing. The interface uploads the action feature data to the buffer in real time and starts to enter the action recognition program, while recorder can check action recognition effect. If the dance movements are not standard or the collected data are not comprehensive, it needs to record again. The last step is to record 10 kinds of dance movements one by one according to the above process, then click "Save to Template. txt" to store the data, establish the action template library, update the system interface, open "Load Ation Template," and finally enter the action recognition test stage after generating all the data.

In order to setup human-computer interaction test environment, it is necessary to turn on the switch of the robot, connect the Bluetooth of the robot's main controller with the Bluetooth adapter of the computer, and select the communication serial port and baud rate. Then, it is also important to click the "Open/Close COM" button to establish effective communication between the robot and the computer. The recorder can control the robot through its own actions. According to the recognition results of the video image of the recorder's dance movements, this test evaluates the action recognition method based on human-computer interaction and analyzes the overall performance and robustness of the method.

3.2. Overall Performance Evaluation. It is known that different joint points have different abilities to describe actions. So, its contribution can be described by information entropy, among which the joint points are lower than the threshold value that needs to be eliminated. This study determines the threshold value according to the recognition rate and gets information entropy threshold. The formula is as follows:

$$s = \frac{n_1}{n_1 + n_2}. \tag{10}$$

In the formula, s represents the average recognition rate of 10 dance movements and n_1 and n_2 represent the times of recognition successes and recognition failures, respectively. Assuming that the threshold is K, the interval is set to 0.05 according to the equal interval value method when $K \in [0.05, 0.95]$. The recognition rate under different entropy values is shown in Figure 5.

According to the test results in Figure 5, the recognition rate is the highest when $K = 0.25$ and $s = 0.9762$. When the value of K continues to increase, the value of s decreases. Because the entropy value of some joint points with large contribution will be less than the threshold value when the value of K is too large, it always filters out these joint points. So, it will influence the action display effect of the robot in the process of action recognition. According to the test results shown in Figure 5, with the increase in K value, the overall result is above 0.88 although the value of average recognition rate s decreased somewhat, which indicates that this method is practical and can be applied to the action test.

3.3. Robustness Testing. In order to achieve a good test effect of action recognition in different dance video images, the test verifies whether the recognition method can successfully recognize the dance movement under the condition of different background environment, different ambient light, and the presence of multiuser interference.

Figure 5 shows the action recognition effect based on the above test conditions.

According to Figure 6, it can be seen that under different background environments, the recognition methods based on human-computer interaction can use the constructed human skeleton model to describe the characteristics of dance posture accurately.

This test sets 10 groups of actions and calculates the average recognition effect of 10 groups of actions in four test environments. Taking the example of lifting the left arm, swinging the right hand backward, and lifting the left leg, the simulation diagram is shown in Figure 7.

For the simulation recognition results obtained under the human-computer interaction technology and the quantization processing, it is processed quantitatively. The quantization results are shown in Table 1.

FIGURE 2: 10 groups of dance actions in the experiment.
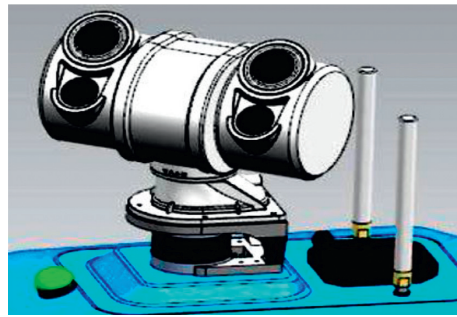
FIGURE 3: Recognition device of four-wheel robot.
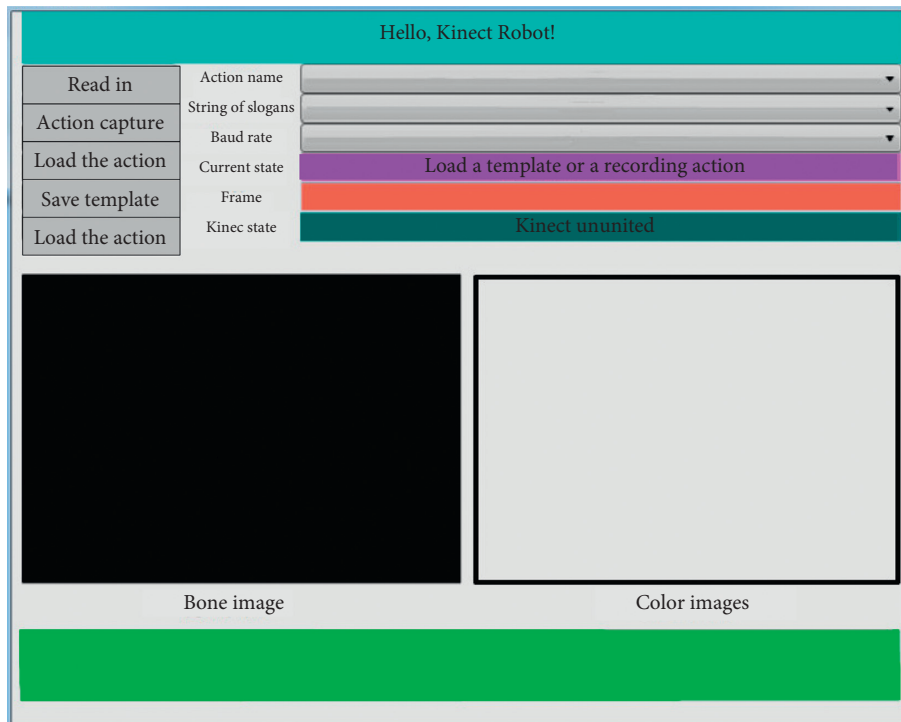
FIGURE 4: Display interface of human-computer interaction.

FIGURE 5: Recognition rate under different entropy values.

(a)

(b)

FIGURE 6: Continued.

(c)

(d)
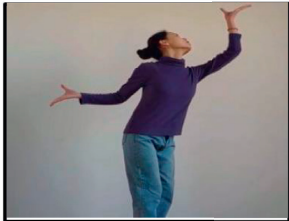
FIGURE 6: Recognition effect: (a) complex environment; (b) dark environment; (c) bright environment; (d) multiuser interference environment.

(a)

(b)

(c)
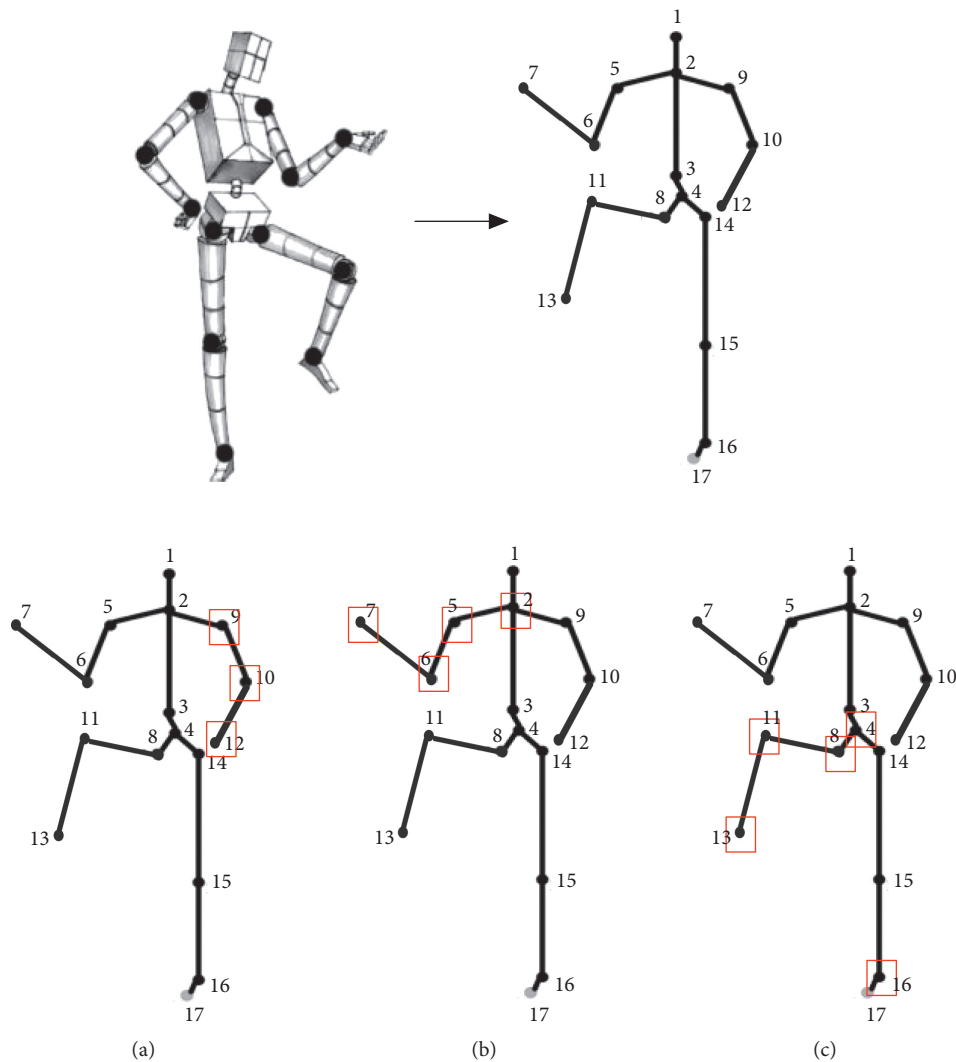
FIGURE 7: Simulation recognition results based on human-computer interaction technology: (a) the right hand swings backward; (b) the left arm up; (c) the left leg up.

TABLE 1: Test result of average recognition rate (%).

Dance movements	Complex environment	Dark environment	Bright environment	Multiuser interference environment
Left arm raised	95.2	90.3	97.5	90.4
Right arm raised	92.6	92.4	97.4	89.8
Arms raised	93.8	90.5	96.8	91.4
Arms flatted	94.1	90.3	97.2	90.5
Arms downward	92.5	90.4	97.6	89.2
Arms crossed	93.3	90.5	96.8	88.7
The left hand swings forward	93.0	92.7	96.9	89.5
The right hand swings backward	92.5	91.4	96.9	90.1
Clap	94.7	93.6	97.5	92.3
Arms akimbo	93.2	91.5	97.2	88.4
Arms on the shoulder	92.8	90.4	96.9	89.6

According to Table 1, it can be seen that the average values of four groups of test results are 93.43%, 91.27%, 97.15%, and 89.99%, respectively. Based on the above test results, it shows that the action recognition method of dance video image based on human-computer interaction can obtain more accurate recognition results under different test backgrounds.

4. Conclusion

This research takes human-computer interaction technology as the innovation point, studies a new action recognition method, and completes the recognition of specific dance video image action through human-computer interaction. The experimental results show that

(1) The joint points of the human are mainly distributed in the limbs. The human skeleton model can be built combined with the joint points of the upper limbs, lower limbs, head, neck, and shoulders. Its principle is to accurately record the movement of each joint point and recognize the movements of the dancer in the process of various movements so as to output the correct human motion skeleton. With the human motion skeleton model, the recognition accuracy and efficiency of the computer vision system can be significantly improved. The human skeleton model based on human-computer interaction technology can accurately describe the characteristics of dance posture and make the recognition rate of the method in the range of 0.8865–0.9762 under different entropy values.

(2) Using the clustering algorithm to extract key frames of dance video image can reduce the influence of interference environment on recognition effect. In the experimental results, the average recognition rates in complex environment, dark environment, bright environment, and multiuser interference environment are 93.43%, 91.27%, 97.15%, and 89.99%, respectively. The above data prove that the proposed method is robust and suitable for motion recognition of most dance video images.

This research takes human-computer interaction technology as the innovation point, studies a new action recognition method, and recognizes the action in dance video image by human-computer interaction. However, as a result of the test time is insufficient and experimental test preparation is hasty, there are some certain limitations in terms of test location and setting different test environment. In the future research, we can set more experimental test conditions, but not limited to the different gender, body shape of the recording object, and dance actions of different difficulties, so as to provide more detailed and accurate research results for the development of new recognition technology.

Data Availability

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Conflicts of Interest

The authors declare that they have no conflicts of interest regarding this work.

References

[1] K. Zhang, T. Liu, Z. Liu, Y. Zhuang, and Y. Chai, "Multimodal human-computer interaction technology for emotion regulation," *China Journal of Image and Graphics*, vol. 25, no. 11, pp. 2451–2464, 2020.

[2] X. Fan, J. Fan, F. Tian, and G. Dai, "Human-computer interaction and artificial intelligence: from competition to integration," *Scientia Sinica Informationis*, vol. 49, no. 3, pp. 361–368, 2019.

[3] S. Chen, W. Wei, B. He, S. Chen, and J. Liu, "Action recognition based on improved deep convolutional neural network," *Application Research of Computers*, vol. 36, no. 3, pp. 945–949+953, 2019.

[4] J. Sun, X. Meng, H. Liang, H. Duan, and Z. Zhan, "Gymnastics action recognition based on MEMS sensor," *Journal of Electronic Measurement and Instrument*, vol. 34, no. 3, pp. 94–99, 2020.

[5] D. Zhao, L. Yu, G. Liu, and Z. Dai, "Human-machine interaction intention recognition technology based on D-S evidence theory," *Modular machine tool and automatic processing technology*, no. 3, pp. 60–63, 2019.

[6] L. Jing, L. Zhu, and X. Gou, "Kinematics analysis of lower limb rehabilitation exoskeleton mechanism based on human-

machine closed chain," *Journal of Engineering Design*, vol. 26, no. 1, pp. 65–72+109, 2019.

[7] G. Zuo, Z. Xu, J. Lu, and D. Gong, "A structure-optimized DDAG-SVM action recognition method for upper limb rehabilitation training," *Acta Automatica Sinica*, vol. 46, no. 3, pp. 549–561, 2020.

[8] B. Jia and P. Li, "Number gesture recognition method in human-computer interaction," *Journal of Huaqiao university (Natural Science Edition)*, vol. 41, no. 2, pp. 260–267, 2020.

[9] H. Zhang, D. Fu, and K. Zhou, "Video action recognition method based on timing enhancement," *Pattern Recognition and Artificial Intelligence*, vol. 33, no. 10, pp. 951–958, 2020.

[10] G. Liu, N. Wang, Y. Zhou, W. Wang, and Y. Yue, "Dumbbell action recognition based on improved ReliefF algorithm," *Science technology and engineering*, vol. 19, no. 32, pp. 219–224, 2019.

[11] J. Yang, M. Shi, F. Chao, and C. Zhou, "The dancing-robot for action imitation based on deep learning," *Journal of Xiamen University*, vol. 58, no. 5, pp. 759–766, 2019.

[12] L. Chen, "Research on recognition technology of human-body action in dance video image," *Modern Electronic Technology*, vol. 40, no. 3, pp. 51–53, 2017.

[13] H. Li, "Research on motion recognition method in dance video images," *Video Engineering*, vol. 42, no. 7, pp. 34–37, 2018.

[14] Y. Xu, "Research on dynamic recognition of gait silhouettes of dance movement based on AVI video," *Modern Electronic Technology*, vol. 43, no. 16, pp. 119–121, 2020.

[15] D. Li, "Design of retrieval system of specific action segments in high dynamic dance video," *Modern Electronic Technology*, vol. 41, no. 5, pp. 97–101, 2018.

[16] S. Lu, Q. Cao, and M. Sun, "Hardware design of mechatronic integration robot joint and its driving control system," *Journal of Chongqing University of posts and Telecommunications (Natural science edition)*, vol. 47, no. 1, pp. 45–49, 2021.

[17] N. Ma, X. Shi, D. Qin, and C. Liu, "A method of key frame extraction for music and dance video," *Journal of System Simulation*, vol. 30, no. 7, pp. 2801–2807, 2018.

[18] X. Chen, Y. Wang, Z. Wu, and J. Li, "Research and realization of a stage augmented reality system," *Computer Engineering and Applications*, vol. 54, no. 3, pp. 11–17, 2018.

[19] Q. Cai, E. Li, J. Jiang et al., "Study on the tea identification of near-infrared hyperspectral image combining spectra-spatial information," *Spectroscopy and Spectral Analysis*, vol. 39, no. 8, pp. 2522–2527, 2019.

[20] E. Yang, A. Zhang, R. Yang, and C. Wang, "Automatic recognition algorithm of three-dimensional image of surface cracks on ballastless track of high speed railway," *Railway Transaction*, vol. 41, no. 11, pp. 95–99, 2019.

[21] S. Dong, X. Sun, S. Xie, and M. Wang, "Automatic defect identification technology of digital image of pipeline weld," *Natural Gas Industry*, vol. 39, no. 1, pp. 113–117, 2019.

[22] W. Liu, H. Tian, J. Xie, E. Zhao, and J. Zhang, "Identification methods for forest pest areas of UVA aerial photography based on fully convolutional networks," *Journal of Agricultural Machinery*, vol. 50, no. 3, pp. 179–185, 2019.

[23] X. Zhou, X. He, and C. Zheng, "Recognition of radio signal based on image deep learning," *Journal of Communications*, vol. 40, no. 7, pp. 114–125, 2019.

[24] 魏. 彤. Wei Tong and 周. Zhou Yin-he, "Blind sidewalk image location based on machine learning recognition and marked watershed segmentation," *Optics and Precision Engineering*, vol. 27, no. 1, pp. 201–210, 2019.

[25] Z. Zhong, "A method of image graying for roadway line identification," *Journal of Tongji University*, vol. 47, no. 1, pp. 178–182, 2019.

Research Article

Selection of Outline Descriptors Based on LightGBM with Application to Infrared Image Target Recognition

Xiaohong Hu [iD][1] **and Ziyang Yao**[2]

[1]*College of Artificial Intelligence, Wuxi Vocational College of Science and Technology, Wuxi 214028, China*
[2]*School of Internet of Things Technology, Wuxi Vocational College of Science and Technology, Wuxi 214028, China*

Correspondence should be addressed to Xiaohong Hu; 1201701@wxsc.edu.cn

Infrared sensing technology can be well used for night observation, which is becoming an important measure for battlefield reconnaissance. It is a powerful way to implement precision strikes and situational awareness by improving the ability of target recognition based on infrared images. For the problem of infrared image recognition, the Light Gradient Boosting Machine (LightGBM) is employed to select the outline descriptors extracted based on the elliptic Fourier series (EFS), which is combined with sparse representation-based classification (SRC) to achieve target recognition. First, based on the target outlines in the infrared image, the multi-order outline descriptors are extracted to characterize the essential characteristics of the target to be recognized. Then, the LightGBM feature selection algorithm is used to screen the multi-order outline descriptors to reduce redundancy and improve the pertinence of features. Finally, the selected outline descriptors are classified based on SRC. The method effectively improves the effectiveness of the final features through the feature selection of LightGBM and reduces the computational complexity of classification at the same time, which is beneficial to improve the overall recognition performance. The mid-wave infrared (MWIR) dataset of various targets is employed to carry out verification experiments for the proposed method under three different conditions of original samples, noisy samples, and partially occluded samples. By comparing the proposed method with several types of existing infrared target recognition methods, the results show that the proposed method can achieve better performance.

1. Introduction

The development and maturity of infrared imaging technologies provides an important tool for night observation [1–4]. In the civil field, traffic monitoring can be performed based on infrared images or videos, which effectively avoids accidents. In the military field, infrared imaging can be used for battlefield monitoring and precision strikes. The target recognition method based on infrared image aims to confirm the target category in the measured samples, so as to obtain valuable information for technical support. As a supervised classification problem, infrared image target recognition usually adopts a two-stage process of feature extraction and classification. Feature extraction acquires features such as intensity distribution, local texture, key points, and regional distributions of the target through the analysis of infrared images. In [5], the boundary features were extracted for target recognition. In [6], the image moments were employed to analyze the regional characteristics for discriminating different categories of targets. Local textual descriptors, i.e., histogram of oriented gradient (HOG) features, were applied to infrared image target recognition in [7]. The multi-scale HOG features were developed in [8] to describe the intensity distribution properties of infrared images for target recognition. The classification process employs mature classifiers to train the extracted features and obtain the corresponding categories of the unknown samples. Typical classifiers used in infrared image target recognition include support vector machine (SVM) [8], sparse representation-based classification (SRC) [9], neural networks [10], etc. Since the emergence of deep learning theory and algorithms, the classification models

represented by convolutional neural network (CNN) have been widely used in image processing and have become a powerful tool for infrared target recognition [10–13].

The results of feature extraction are used as the input of subsequent decision making. In this sense, the effectiveness of the extracted features directly determines the validity of the final decision. According to the existing literatures, most of the features used in infrared target recognition were directly applied based on mature algorithms in the field of pattern recognition, lacking targeted analysis and screening. In fact, some of the extracted features may have little contribution to the final decision or even cause negative influences on the classification process. For example, the redundant features would definitely increase the computational load and possibly cause some wrong decisions. In [14, 15], the genetic algorithm and particle swarm optimization algorithm were used to select the extracted Zernike moment features for synthetic aperture radar (SAR) target recognition. A basic comparison between the results of using all the features and the selected features verified the necessity and validity of feature selection. Inspired by this work, this paper employs the Light Gradient Boosting Machine (LightGBM) [16–20] for the selection of outline descriptors extracted by elliptic Fourier series (EFS) [21–23] and applies it in infrared image target recognition. As a new boosting algorithm, LightGBM is operated on the gradient boosting decision tree (GBDT) developed in the early stage and optimizes learning through additive models and forward step-by-step algorithms. According to relevant research studies, LightGBM can effectively improve the learning speed and model accuracy. Based on LightGBM, the extracted multi-order outline descriptors can be analyzed and screened. Only the effective components are maintained, and redundant ones are eliminated, so as to ensure the effectiveness of the final features. In the classification stage, SRC is used as the basic classifier to characterize and recognize the final selected outline descriptors. As a classifier based on the compressive sensing theory, SRC could adapt to different kinds of features and achieve robust performance [24–31]. In the experiments, the public mid-wave infrared (MWIR) target image dataset is employed to test and verify the proposed method, and the results reflect its effectiveness and robustness.

2. Outline Descriptors

The ellipse Fourier descriptor uses ellipse superposition to approach the boundary of a specific object. It can approach the boundary infinitely in the form of multiple harmonics and has the advantages of scale transformation, rotation transformation, and invariance of starting point transformation [21–23]. This paper first uses the method in [32–35] to obtain the outline point set of the target. For a closed boundary C, it can be expressed as a vector form $v(t) = [x(t) \ y(t)]^T$, where $t \in [0, 2\pi]$. Therefore, $v(t)$ is the periodic function of the parameter t, which can be expressed in the form of Fourier series as shown in the following equation:

$$\begin{bmatrix} x(t) \\ y(t) \end{bmatrix} = \sum_{k=0}^{\infty} \mathbf{F_k} \begin{bmatrix} \cos(kt) \\ \sin(kt) \end{bmatrix}, \tag{1}$$

where $\mathbf{F_k} \in R^{2\times 2}, k = 0, 1, 2, \ldots$, represents the EFS outlier descriptors, which are calculated as follows:

$$\begin{aligned} \mathbf{F_k} &= \begin{bmatrix} a_k \ b_k \\ c_k \ d_k \end{bmatrix} \\ &= \frac{1}{2\pi} \int_0^{2\pi} \begin{bmatrix} x(t) \\ y(t) \end{bmatrix}^T \begin{bmatrix} \cos(kt) \\ \sin(kt) \end{bmatrix} dt. \end{aligned} \tag{2}$$

In equation (2), $\mathbf{F_0}$ is the 0-order elliptic Fourier descriptor, which represents the DC component of the harmonic. It denotes the center of the closed boundary, which contains two parameters $[a_0 c_0]^T$. For $k \geq 1$ subharmonics, there are four coefficients $[a_k b_k c_k d_k]^T$. This paper adopts the first 6 orders of the EFS outline descriptors for the following classification.

3. Proposed Target Recognition Method

3.1. Feature Selection Based on LightGBM. The multi-layer features from the same images are usually not completely uncorrelated, and the contribution of different features to the target classification is also distinguished. As the number of extracted features increases, the amount of repetitive redundancy will increase, and the noise and error will also increase. The increase will not only increase the calculation amount of the algorithm but also reduce the recognition rate of the target. Therefore, it is necessary to select features that are beneficial to target classification. In this paper, LightGBM is used for feature selection [16–20], the training samples are put into LightGBM for training, the importance of each feature is calculated, and the key features are selected according to the importance to reduce the number of features required for classification. The classification performance reduces the recognition time at the same time.

The LightGBM algorithm measures the importance of feature attributes according to the total number of times the feature is used to split in all decision trees, sorts the feature elements in descending order, starts from the full set of sample features, and judges whether to exclude according to the accuracy of the result. The current feature with the lowest degree of importance is looped in this way to achieve feature selection. The specific process is described as follows (Algorithm 1).

It can be seen that the proposed LightGBM feature selection algorithm sorts according to the importance of features and preliminary evaluation of the model results, which greatly reduces the volatility of features. It is beneficial to the improvement of the accuracy of target recognition in the later stage. At the same time, because the features of relatively low importance are deleted, the redundancy of the features is reduced, so the time consumption required for target recognition can also be reduced.

Input: dataset D, feature set $F = \{T_i \mid i = 1, 2, \ldots, d\}$;
Output: optimal feature subset F_{best};
Step 1: LightGBM is used to calculate the importance I_i of sample feature elements T_i, respectively.
Step 2: the feature elements T_i are arranged in descending order according to the results I_i obtained in the first step.
Step 3: LightGBM is used to evaluate and calculate the accuracy a_i of the current feature subset F.
Step 4: the sorted feature subset is selected in step 2 by deleting the feature element T_i with the least importance in F and calculating the
   recognition accuracy a_{best}.
Step 5: if $a_i < a_{\text{best}}$, then go back to step 3 for execution; otherwise, recycle the feature elements T_i deleted in the previous step and then
   go back to step 3 for execution.
Step 6: when all the feature elements T_i have been traversed, the optimal feature subset F_{best} is output.

ALGORITHM 1: Feature selection using LightGBM.

3.2. SRC for Classification. SRC is a non-parametric classifier developed from the compressive sensing theory, which was first applied to face recognition with excellent performance. In that work, the effectiveness and robustness of SRC were validated. Afterwards, SRC was widely used in pattern recognition like SAR target recognition [25–27]. For a sample from the ith class, SRC assumes that it can be linearly represented by the samples from the same class as follows:

$$y_{i,\text{test}} = A_i * x_i + \varepsilon_i, \tag{3}$$

where $y_{i,\text{test}}$ is a test sample assumed from the ith class; A_i comprises the training samples of the ith class; x_i denotes the coefficient vector with only a few non-zero elements; and ε_i is the reconstruction error.

When the test sample y is from an unknown class, SRC is performed on the global dictionary $A = [A_1, A_2, \ldots, A_C]$ as follows:

$$\begin{aligned} \widehat{x} &= \arg \min \|x\|_0, \\ \text{s.t. } &\|y - Ax\|_2^2 \leq \varepsilon, \end{aligned} \tag{4}$$

where $\widehat{x}$ denotes the coefficient vector over the global dictionary.

The optimization task in (4) can be robustly solved by the ℓ_1-norm relaxation or greedy algorithm like orthogonal matching pursuit (OMP). With an optimal solution of $\widehat{x}$, the reconstruction errors related to different classes can be computed as equation (5) for decision making.

$$\begin{aligned} r(i) &= \|y - A * \delta_i(\widehat{\alpha})\|_2, \\ \text{Class}(y) &= \arg \min_i r(i), \end{aligned} \tag{5}$$

where $\delta_i(\widehat{\alpha})$ corresponds to the coefficients of ith training class.

3.3. Implementation Details. Figure 1 shows the basic process of the proposed method, which can be defined as two stages of feature extraction and classification decision according to traditional idea. In the feature extraction stage, the first step is to use EFS outline descriptors to analyze the target contour in the infrared image to obtain multi-order features. On this basis, LightGBM is used for training and learning, the best feature subset is selected, and the redundant components are eliminated. For the training samples, the same operation is followed to construct the corresponding global dictionary, which contains the outline descriptors corresponding to all the samples. Finally, SRC is used to classify the outline descriptors of the test sample and obtain its corresponding target category.

4. Experiments and Discussion

4.1. Dataset. The experiments are carried out on the MWIR dataset, which contains images of 10 typical military targets acquired by the mid-wave infrared sensors. Figure 2 shows the exemplar images of the 10 targets. According to the relevant experimental settings in [21], 120 images are randomly selected as the training set and 100 images are used as the test sample. During the experiment, four types of methods are selected for comparison, which are denoted as SVM, SRC, EFS, and CNN methods. Among them, the EFS method does not use the feature selection by LightGBM in the proposed method and directly classifies the extracted multi-order outline descriptors based on SRC.

4.2. Original Test Samples. First, experiments are conducted on the basis of the original training and test samples. At this time, the test sample and the training sample come from similar conditions and their difference is small. So, the recognition problem is relatively simple. Table 1 shows the detailed recognition results of the ten types of targets in this paper. The recognition rates of all types of targets are above 97%, and the final average recognition rate is 97.9%, reflecting the effectiveness of the proposed method. Table 2 further compares the average recognition rates of different methods on the original samples. The method in this paper has the highest performance and shows its advantages. The performance of the CNN method under the current conditions is second only to the method in this paper, indicating the effectiveness of the deep learning model. Compared with the method that directly uses outline descriptors, this paper introduces LightGBM to analyze and select them, and the retained features further improve the recognition performance, which shows the effectiveness of the proposed method.

FIGURE 1: Procedure of infrared target recognition based on selected outline descriptors by LightGBM.

FIGURE 2: Images of targets in MWIR dataset. (a) Pickup. (b) SUV. (c) BTR70. (d) BRDM2. (e) BMP2. (f) T72. (g) ZSU23-4. (h) 2S3. (i) MTLB. (j) D20.

TABLE 1: Recognition results of the proposed method on the original test samples.

Class	Pickup	SUV	BTR70	BRDM2	BMP2	T72	ZSU23/4	2S3	MTLB	D20	Recognition rate (%)
Pickup	97	0	2	0	1	0	0	0	0	0	97
SUV	0	98	0	1	0	0	0	0	0	1	98
BTR70	0	0	100	0	0	0	0	0	0	0	100
BRDM2	1	1	0	98	0	0	0	0	0	0	98
BMP2	0	0	1	0	97	0	0	1	0	1	97
T72	0	0	0	2	1	96	0	1	0	0	96
ZSU23/4	2	0	1	0	0	0	97	0	0	0	97
2S3	0	0	0	0	0	0	0	100	0	0	100
MTLB	0	0	0	0	0	0	0	0	99	1	99
D20	0	1	0	1	0	1	0	0	0	97	97
Average (%)					97.9						

TABLE 2: Performance of different methods on original samples.

Method	Average recognition rate (%)
Proposed	97.9
SVM	94.5
SRC	95.1
EFS	96.9
CNN	97.2

4.3. Noisy Test Samples. Natural noise, perceived interference, and sensor thermal noise may all cause the signal-to-noise ratio (SNR) of the final infrared image to be reduced. Therefore, it is necessary to improve the performance of the identification method under noise interference conditions. This paper firstly generates test sets

with different signal-to-noise ratios based on the original test samples by adding white noise and then tests various methods to obtain the results shown in Figure 3. It can be seen that noise corruption has caused the performance of various methods to decrease, but the proposed method maintains the strongest robustness. Compared with SVM, the SRC method has performance advantages and reflects the adaptability of the sparse representation mechanism to the influence of noise. Despite the lack of secondary feature selection, the EFS method can still maintain good noise robustness, reflecting the stability of outline descriptors under noise interference. In detail, EFS outline descriptors are mainly related to the target contour, and the overall regional characteristics of the target can be kept robust under noise corruption conditions, so this

FIGURE 3: Performance of different methods under noise corruption.

FIGURE 4: Performance of different methods under partial occlusions.

type of method can achieve stable performance under noise conditions. The proposed method combines the advantages of outline descriptors, LightGBM-based feature selection, and sparse representation, so the final recognition result has advantages.

4.4. Partially Occluded Test Samples. There may be two reasons to cause partially occluded infrared images. One is that the target itself is partially absence, and the other is that it is partially occluded during the imaging process. In both cases, there are some missing target characteristics

in the final image, which makes the recognition problem more complicated. In this paper, based on the original test samples, the partial occlusion algorithm is used to simulate the construction of test sets under different missing levels (measured by the ratio of the occluded area to the target area). Figure 4 shows the average recognition rate curve of various methods under partial missing conditions. It can be seen that the performance of all methods has been significantly reduced under the partial missing condition, and the method in this paper can maintain the highest recognition rate at all levels, indicating its robustness. The SRC method has a performance advantage over the SVM method, which shows the effectiveness of the sparse representation mechanism for partial deletions. The EFS method is at a disadvantage compared to the proposed method due to the lack of secondary feature screening, but it has certain advantages over other methods. This shows that the outline descriptors have certain adaptability to the damage of the target area caused by partial missing. In summary, this method combines outline descriptors, LightGBM-based feature screening, and SRC-based decision making to jointly improve the recognition performance under the current test conditions.

5. Conclusion

This paper proposes an infrared image target recognition method based on feature selection of EFS outline descriptors. LightGBM is used to screen the multi-order outline descriptors of infrared targets to eliminate redundant components. On this basis, SRC is used to classify the selected outline descriptors to determine the target category of the test sample. The proposed method improves the final recognition performance through effective extraction, secondary feature selection, and SRC-based decision making. Experiments are carried out on the MWIR dataset to classify 10 typical military targets. According to the experimental results, the following conclusions can be drawn. (1) The proposed method achieves better recognition performance on the original test samples, noisy samples, and partially occluded samples than the four types of comparison methods, showing its performance advantages. (2) The introduction of LightGBM for the selection of the original multi-order outline descriptors effectively improved the overall recognition performance in the proposed method. (3) The LightGBM-based feature selection can be regarded as a general framework for other types of features for different kinds of pattern recognition problems.

Data Availability

The dataset can be accessed upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

References

[1] X. Dai, Y. Duan, J. Hu et al., "Near infrared nighttime road pedestrians recognition based on convolutional neural network," *Infrared Physics & Technology*, vol. 97, pp. 25–32, 2019.

[2] C. Gao, Y. Du, J. Liu et al., "InfAR dataset: infrared action recognition at different times," *Neurocomputing*, vol. 212, pp. 36–47, 2016.

[3] H. Deng, X. Sun, M. Liu, C. Ye, and X. Zhou, "Small infrared target detection based on weighted local difference measure," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 7, pp. 4204–4214, 2016.

[4] S. Kim, W.-J. Song, and S.-H. Kim, "Infrared variation optimized deep convolutional neural network for robust automatic ground target recognition," in *Proceedings of the IEEE Conference Computer Vision Pattern Recognit Workshops (CVPRW)*, pp. 195–202, Honolulu, HI, USA, August 2017.

[5] H. Wang, X. Yang, and H. Ding, "Method of features extraction for infrared image recognition based on image moment," in *Proceedings of the International Conference on Computer Application and System Modeling*, pp. 443–446, Taiyuan, China, October 2010.

[6] A. d'Acremont, R. Fablet, A. Baussard, and G. Quin, "CNN-based target recognition and identification for infrared imaging in defense systems," *Sensors*, vol. 19, p. 2040, 2019.

[7] Q. Kang, H. Zhao, D. Yang, H. S. Ahmed, and J. Ma, "Lightweight convolutional neural network for vehicle recognition in thermal infrared images," *Infrared Physics & Technology*, vol. 104, p. 1030, 2020.

[8] X. Li, V. Monga, and A. Mahalanobis, "Multiview automatic target recognition for infrared imagery using collaborative sparse priors," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 10, pp. 6776–6790, 2020.

[9] Y.-R. Cho, S. Shin, S.-H. Yim, K. Kong, H.-W. Cho, and W.-J. Song, "Multistage fusion with dissimilarity regularization for SAR/IR target recognition," *IEEE Access*, vol. 7, pp. 728–740, 2019.

[10] N. M. Nasrabadi, "Deeptarget: an automatic target recognition using deep convolutional neural networks," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 55, no. 6, pp. 2687–2697, 2019.

[11] A. Akula, A. K. Shah, and R. Ghosh, "Deep learning approach for human action recognition in infrared images," *Cognitive Systems Research*, vol. 50, pp. 146–154, 2018.

[12] F. He, X. Hu, B. Liu, H. Wang, and D. Zhang, "Infrared image recognition technology based on visual processing and deep learning," in *Proceedings of the Chinese Automation Congress (CAC)*, pp. 641–645, Shanghai, China, November 2020.

[13] A. Akula and H. K. Sardana, "Deep CNN-based feature extractor for target recognition in thermal images," in *Proceedings of the 2019 IEEE Region 10 Conference (TENCON)*, pp. 2370–2375, Kochi, India, October 2019.

[14] M. Amoon and G. A Rezai-Rad, "Automatic target recognition of synthetic aperture radar (SAR) images based on optimal selection of Zernike moments features," *IET Computer Vision*, vol. 8, no. 2, pp. 77–85, 2014.

[15] M. Amoon, G.-A. Rezai-Rad, and M. R. Daliri, "PSO-based optimal selection of Zernike moments for target discrimination in high-resolution SAR imagery," *Journal of the Indian Society of Remote Sensing*, vol. 42, no. 3, pp. 483–493, 2014.

[16] G. Ke, Q. Meng, T. Finley et al., "LightGBM: a highly efficient gradient boosting decision tree," in *Proceedings of the Advanced Neural Information Process System*, pp. 3146–3154, Long Beach, CA, USA, December 2017.

[17] F. Alzamzami, M. Hoda, and A. E. Saddik, "Light gradient boosting machine for general sentiment classification on short texts: a comparative evaluation," *IEEE Access*, vol. 8, pp. 101840–101858, 2020.

[18] T. Chen, J. Xu, H. Ying et al., "Prediction of extubation failure for intensive care unit patients using light gradient boosting machine," *IEEE Access*, vol. 7, pp. 150960–150968, 2019.

[19] X. Chen, K. Zhou, H. Pan, and X. Chen, "Anode effect prediction based on light gradient boosting machine," in *Proceedings of the 2019 IEEE 8th Data Driven Control and Learning Systems Conference (DDCLS)*, pp. 499–503, Dali, China, May 2019.

[20] M. H. Abdurrahman, B. Irawan, and C. Setianingsih, "A review of light gradient boosting machine method for hate speech classification on twitter," in *Proceedings of the 2020 2nd International Conference on Electrical, Control and Instrumentation Engineering (ICECIE)*, pp. 1–6, Kuala Lumpur, Malaysia, November 2020.

[21] G. C. Anagnostopoulos, "SVM-based target recognition from synthetic aperture radar images using target region outline descriptors," *Nonlinear Analysis*, vol. 71, no. 2, pp. e2934–e2939, 2009.

[22] Y. Z. Hu, Y. Hu, X. B. Ai et al., "Stability evaluation of fault diagnosis model based on elliptic fourier descriptor," *Journal of Control Science and Engineering*, vol. 2018, Article ID 1238231, 10 pages, 2018.

[23] R. Reig-Bolaño, P. Marti-Puig, E. Gallego et al., "Feature selection for analyzing and retrieving fish otoliths using elliptic fourier descriptors of shapes," in *Proceedings of the 2011 7th International Conference on Next Generation Web Services Practices*, pp. 290–295, Salamanca, Spain, October 2011.

[24] J. Wright, A. Y. Yang, A. Ganesh, and S. S. Sastry, "Robust face recognition via sparse representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 2, pp. 210–227, 2009.

[25] J. J. Thiagaraianm, K. N. Ramamurthy, P. Knee, S. Spanias, and V. Berisha, "Sparse representations for automatic target classification in SAR images," in *Proceedings of the 4th International Symposium Communication, Control Signal Process*, pp. 1–4, Limassol, Cyprus, March 2010.

[26] H. Song, K. Ji, Y. Zhang, X. Xing, and H. Zou, "Sparse representation-based SAR image target classification on the 10-class MSTAR data set," *Applied Sciences*, vol. 6, no. 1, p. 26, 2016.

[27] S.-G. Sun, "Automatic target recognition using boundary partitioning and invariant features in forward-looking infrared images," *Optical Engineering*, vol. 42, no. 2, pp. 524–534, 2003.

[28] G. Lin, G. Fan, L. Yu, X. Kang, and E. Zhang, "Heterogeneous structure fusion for target recognition in infrared imagery," in *Proceedings of the IEEE Conference Computer Vision Pattern Recognition Workshops (CVPRW)*, pp. 118–125, Boston, MA, USA, June 2015.

[29] A. Apatean, A. Rogozan, and A. Bensrhair, "SVM-based obstacle classification in visible and infrared images," in *Proceedings of the IEEE European Conference Signal Process*, pp. 293–297, Glasgow, UK, August 2009.

[30] M. N. A. Khan, G. Fan, D. R. Heisterkamp, and L. Yu, "Automatic target recognition in infrared imagery using dense hog features and relevance grouping of vocabulary," in *Proceedings of the IEEE Conference Computer Vision Pattern Recognition Workshops*, pp. 293–298, Columbus, OH, USA, June 2014.

[31] Z. Ding, N. M. Nasrabadi, and Y. Fu, "Deep transfer learning for automatic target classification: MWIR to LWIR," *Proceedings SPIE International Society Optics Photon*, vol. 9844, Article ID 984408, 2016.

[32] C. Mu, J. Wang, Z. Yuan, X. Zhang, and C. Han, "The research of the ATR system based on infrared images and L-M BP neural network," in *Proceedings of the IEEE International Conference Image Graph*, pp. 801–805, Qingdao, China, July 2013.

[33] S. Zhang, J. Gong, D. Chen, L. Xu, and L. Yan, "Sparsity-motivated multi-scale histograms of oriented gradients feature for SRC," in *Proceedings of the IEEE International Conference Unmanned System (ICUS)*, pp. 389–393, Beijing, China, October 2017.

[34] J. Tan, X. Fan, S. Wang et al., "Target recognition of SAR images by partially matching of target outlines," *Journal of Electromagnetic Waves and Applications*, vol. 33, no. 7, pp. 865–881, 2019.

[35] X. Zhu, Z. Huang, and Z. Zhang, "Automatic target recognition of synthetic aperture radar images via Gaussian mixture modeling of target outlines," *Optik*, vol. 194, Article ID 162922, 2019.

Retraction

Retracted: An Intelligent Evaluation Method of Information Course Teaching Effect Based on Image Analysis

Scientific Programming

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope

(2) Discrepancies in the description of the research reported

(3) Discrepancies between the availability of data and the research described

(4) Inappropriate citations

(5) Incoherent, meaningless and/or irrelevant content included in the article

(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] H. Chen and J. Zhang, "An Intelligent Evaluation Method of Information Course Teaching Effect Based on Image Analysis," *Scientific Programming*, vol. 2021, Article ID 3200865, 9 pages, 2021.

Research Article

An Intelligent Evaluation Method of Information Course Teaching Effect Based on Image Analysis

HaiDong Chen [1] **and JuFang Zhang**[2]

[1]*Dean's Office, Shang Hai East Sea College, Shang Hai 201206, China*
[2]*Dean's Office, Shang Hai Jiao Tong University, Shang Hai 200241, China*

Correspondence should be addressed to HaiDong Chen; 1168@esu.edu.cn

Due to its own limitations, the traditional teaching quality evaluation method has been unable to adapt to the development of information-based curriculum teaching. Therefore, the establishment of a scientific and intelligent teaching effect evaluation method will help to improve the teaching quality of college teachers. To solve the above problems, a student fatigue state evaluation method based on the quantum particle swarm optimization artificial neural network is proposed. Firstly, face detection is realized by adding three Haar-like feature blocks and improving the AdaBoost algorithm of a weak classifier connection. Secondly, in order to effectively improve the image imbalance, the MSR algorithm is used to enhance the face data image, which is effectively suitable for network training. Then, by readjusting the connection mode, the DenseNet is improved to fully reflect the local detail feature information of the low level. Finally, quantum particle swarm optimization (QPSO) is used to optimize the DenseNet structure, which makes the optimization of network structure more automatic and solves the uncertainty of manual selection. The experimental results show that the proposed method has a good detection effect and prove the effectiveness and correctness of the proposed method.

1. Introduction

With the rapid development of Educational Informatics, it has been widely used in many aspects of the higher education field and achieved good results, but research and practice have lagged behind in measuring teaching quality, such as the teaching quality evaluation system and evaluation model; it is well known that the evaluation of teaching quality analysis is a very complex nonlinear process [1–4]; there are multiple influencing factors and dynamic variables involved, so that the traditional model of the teaching quality has become less fully competent in work addressing this ambiguity.

At present, cameras are used in most classrooms of information courses to monitor students' status in real time, resulting in a large amount of video data. The traditional teaching quality evaluation method needs to monitor each student's sitting posture, facial expression, and other information in the video manually to judge whether each student has fatigue. This method has the problems of low efficiency and high labor cost [5–8], so it cannot be popularized and applied on a large scale. How to find the fatigue state of students in time and effectively by means of automation in practical application, so as to effectively prevent the occurrence of students' inattention in class, has a strong practical significance and helps to ensure the teaching quality of information-based courses.

Artificial intelligence uses computers to simulate some human thinking processes and intelligent behaviors, so as to endow machines with human intelligence [9–11]. Machine learning is a branch of artificial intelligence, and its most successful application is in the field of computer vision [12, 13]. As a technology to realize machine learning, deep learning has attracted great attention. Deep learning adopts a multilevel network structure to model the human brain autonomous learning. The advantages of deep learning are

mainly in the following two aspects. First, deep learning mostly adopts a deep neural network. With the deepening of network level, it can often get better learning effect, which is far from being achieved by shallow learning [14, 15]. Second, in-depth learning attaches great importance to the learning of the characteristics of the training object and has a very strong feature learning ability [16–18]. Many researchers have applied deep learning to fatigue state detection tasks and achieved good recognition accuracy, among which the most representative is the artificial neural network. However, the essence of the artificial neural network is a gradient descent algorithm, which will fall into a local optimal solution. Theoretically, it still has room for optimization.

In order to solve the above problems, this study proposed a student fatigue state evaluation method based on the quantum particle swarm optimization artificial neural network. The proposed method improves Dense-Net [19] by reducing the number of redundant connections, so as to fully reflect low-level local detail feature information. Finally, QPSO [20] is used to optimize the DenseNet structure and increase the number of hyperparameters, which makes the optimization of network structure more automatic and solves the uncertainty problem of artificial selection. Experimental results show that, on CAS_PEAL and self-built datasets, the accuracy of the QPSO-DenseNet algorithm is higher than that of the optimal DenseNet structure selected manually.

2. Literature Review

For students in the information class, physical fatigue will lead to distraction, leading to students being unable to concentrate in the normal teaching process. The research technology on fatigue state detection at home and abroad began in the 1930s and has not made great progress. It is not until the last two decades that the research results have made great progress, which is mainly due to the development of information processing technology, sensor technology, and deep learning technology in recent years. Research on fatigue state detection can be divided into two aspects: the method based on physiological parameters and the method based on human behavior characteristics.

The fatigue state detection method based on physiological parameters receives signals from electrode patches on the forehead, heart, and muscle through medical special instruments and equipment, and the signals are transmitted to the system for real-time analysis. The threshold value of physiological parameter signals defined clinically is used as the basis for fatigue discrimination. Luo et al. [21] evaluated driving fatigue based on frontal EEG signals. Wang et al. [22] proposed the method of deep learning and convolution neural network combined with EEG, which improved the detection accuracy compared with previous methods. However, physiological parameters are mostly collected by special human body signal acquisition instruments, and the volume of medical equipment is relatively large, which cannot be used in the classroom. In addition, because one end of the data acquisition equipment must be connected to the skin of the human body, it is not conducive to normal teaching activities, so the adaptability is poor.

The driver fatigue state detection method based on driver behavior characteristics is a research method combining pattern recognition and image processing technology. With the facial state image collected with the camera as the input data, after image processing and computer vision-related technology processing, the dynamic data of the eyes and mouth on the face are analyzed to judge whether the fatigue limit is reached and then the fatigue state is judged. At present, there are many research results using this method. Qian et al. [23] used Haar-like features and AdaBoost classifier to conduct real-time detection of the face and analyzed the optimal sampling rate and minimum possible target size to improve detection accuracy and calculation efficiency. Eye localization and iris localization are performed on the detected face. Finally, PERCLOS was used as the criterion to judge fatigue. You et al. [24] used elliptic curve fitting and image enhancement methods, respectively, to determine fatigue according to eye opening. Guo et al. [25] used the convolutional neural network and Bayesian network, respectively, to extract face features and the relationship between features.

In order to improve the accuracy of student fatigue state evaluation, this paper proposes a student fatigue state evaluation method based on the quantum particle swarm optimization artificial neural network. Firstly, face detection is realized by the improved AdaBoost algorithm. Secondly, in order to effectively improve the image imbalance, Multiscale Retinex (MSR) [26] algorithm is used to enhance the face data image. Then, DenseNet is improved by reducing the number of redundant connections. Finally, the DenseNet structure was optimized by QPSO.

3. Evaluation Method of Students' Fatigue State

3.1. Face Detection Based on the Improved AdaBoost Classifier. The purpose of this stage of face detection is to determine the position of the human face region in the whole image from the input image and output coordinates and other information, such as head tilt angle. Face detection is a key step in the study of student fatigue state detection. Although the relative positions of facial features are roughly the same, it is still very difficult to detect faces under the influence of different expressions and external light.

In the face detection stage, the cascaded AdaBoost classifier [27] can quickly exclude a large number of nonface images at the initial stage of training, which will greatly help improve the efficiency of face detection in the end. Haar-like features are inspired by Haar wavelet transform, which can describe feature edges and feature changes in each direction of the image in detail. Haar-like feature is an important tool for AdaBoost to realize face detection. In order to improve the accuracy of face detection, this paper adds three rectangular features on the basis of the original rectangular features (Figure 1), as shown in Figure 2.

In Figure 2, the left rectangle is used to represent the left cheek, the middle rectangle is used to represent the right cheek, and the right rectangle is used to represent the local
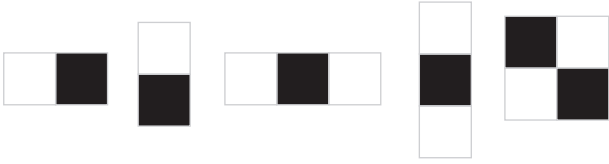
Figure 1: Original rectangular feature.
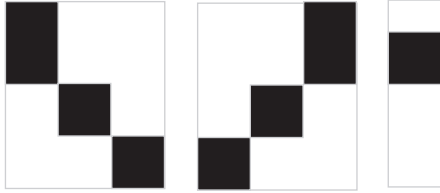
Figure 2: Haar-like feature block proposed.

features of the face. The white part of local features represents the forehead, and the rest represents the eyes.

The improved AdaBoost method obtains the strong classifier $G(x)$ by combining several weak classifiers.

$$
\begin{aligned}
f(x) &= \sum_{n=1}^{N} e^{a_n G_n(x)}, \\
G(x) &= \text{sign}(f(x)) = \text{sign}\left(\sum_{n=1}^{N} e^{a_n G_n(x)}\right),
\end{aligned}
\tag{1}
$$

where a_n represents the coefficient of $G_n(x_i)$, which is how important it is in $G_n(x_i)$.

$$
a_n = \frac{1}{2} \log \frac{1 - e_n}{e_n},
\tag{2}
$$

where e_n represents the error rate on the training set.

$$
e_n = \sum_{i}^{M} w_{ni} I(G_n(x_i) \neq y_i),
\tag{3}
$$

where w_{ni} represents the weight.

Then, the corresponding weights in the training set are constantly updated so that the incorrectly classified samples have increasing weight and the correctly classified samples have decreasing weight. In this way, the training can make the samples that are difficult to classify easily get attention.

3.2. Data Enhancement Algorithm. The Retinex algorithm is based on the consistency of colors and is determined by an object's ability to reflect light. It can strike a balance between dynamic compression, edge image enhancement, and color constancy. At the same time, the lighting and color of objects can be kept consistent, that is, the influence of information transmission between the two is independent. A schematic diagram of Retinex is shown in Figure 3.

The original image is enhanced after illumination estimation and correction. The R and L branches in the above process can be expressed by the following formula:

$$
I(x, y) = R(x, y) \cdot L(x, y),
\tag{4}
$$

where $I(x, y)$ represents the received image information, $R(x, y)$ represents the reflection parameter of the target object in the image information distribution, and $L(x, y)$ represents the irradiation parameter.

In the field of image processing, the principle of image enhancement is often used in order to obtain better adaptive images. Retinex differs from traditional image enhancement algorithms, such as linear and nonlinear transformations, random cropping, and flipping, in that they can only locally enhance certain features of the image.

The Multiscale Retinex (MSR) algorithm is used to enhance the face data image, in order to solve the problem of the complexity of the Retinex method in operation and the decrease of operation speed and to better adapt to the artificial neural network. Gaussian filtering is completed after the face image is processed in batch, and then, the filtering results at different scales are averaged to obtain the best face image. The calculation formula is as follows:

$$
r_i(x, y) = \sum_{K=1}^{N} w_k \left(\log_2 (I_i(x, y)) - \log_2 (I_i(x, y) * F_k(x, y)) \right),
\tag{5}
$$

where N represents the number of scale parameters, w_k represents the average weight of the k-th scale, and $F_k(x, y)$ represents the Gaussian filter function on the k-th scale.

$$
\sum_{K=1}^{N} w_k = 1,
$$

$$
F_k(x, y) = \frac{1}{\sqrt{2\pi c_k}} \exp\left(-\frac{x^2 + y^2}{2c_k^2}\right),
\tag{6}
$$

where c_k represents the calculation coefficient on the k-th scale.

3.3. The Improved DenseNet Network Model. DenseNet is an artificial network with a new connection mode. In the DenseNet network, each layer is connected with all other layers; each layer receives its input from all previous layers and propagates the feature mapping of this layer to all subsequent layers to ensure the maximum transmission of information between the layers of the network. Compared with other neural networks, this network has the advantages of encouraging feature reuse, strengthening feature transfer, reducing the amount of calculation, and alleviating the disappearance of a gradient. It is more suitable for fatigue state detection. Therefore, this paper improves on the basis of DenseNet to improve the accuracy of fatigue state evaluation. The network structure of DenseNet is shown in Figure 4.

Because the correlation between the training features is increased between the layer and layer connections of the network, in the network with dense connections, the similar features generated by the later layer connections are better than those generated by the previous layer connections. The last layer of dense connection often accepts the output of all previous layers as input for aggregation, but there is
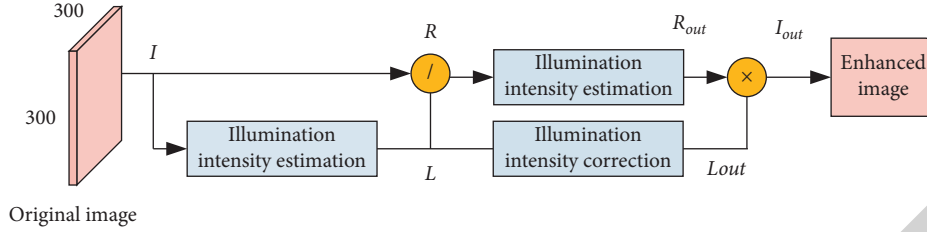
Figure 3: Retinex principle process.
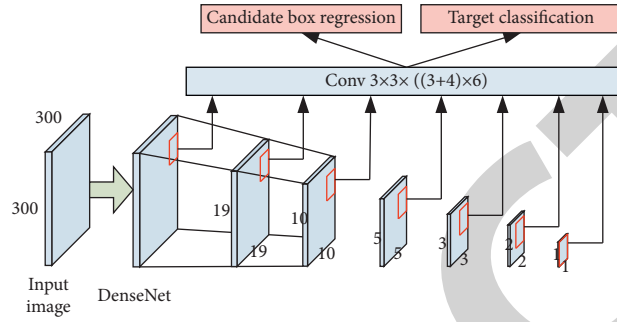
Figure 4: Network structure of DenseNet.

redundant information between these features in the last layer, resulting in the easy loss of feature information in the lower layer.

DenseNet has the defect that feature information is easy to be lost, so it should be improved. The main idea of this paper is to abandon the traditional idea that each layer in DenseNet accepts the output of all previous layers as the input of this layer, in order to reduce the number of interlayer connections. The connection of the first part of the network maintains a dense connection state; the later part is connected to the next layer with only one connection. The structure diagram of the improved new module is shown in Figure 5:

As shown in Figure 5, firstly, 1×1 convolution is used to change their dimensions to 256. For 38×38 sizes, the feature map needs to use a step size of 2 and filter size of 2×2 and downsample it to 19×19. The bilinear interpolation method should be used to upsample the feature graph with a scale smaller than 19×19 to enlarge it to 19×19. They are then concatenated together to form a pixel layer; Batch Norm processing is used to make the activation values of different hierarchical features with the same order of magnitude. Finally, this layer is used as the base layer to generate subsequent connections through the simple stacking of 3×3 convolution blocks.

3.4. Optimization of DenseNet Structure Based on QPSO.
Because the velocity of PSO particle has upper limit, its search space is limited. From the perspective of quantum mechanics, QPSO enables the particle to have an uncertain search trajectory, searching in the whole feasible region, but with global convergence. The iterative expression of QPSO algorithm is as follows:

$$
\begin{aligned}
p_{id}^k &= \varphi_{id}^k \text{pbest}_{id}^k + \left(1 - \varphi_{id}^k\right)\text{gbest}_d^k, \\
\varphi_{id}^k &\sim U(0, 1), \\
x_{id}^{k+1} &= p_{id}^k \pm \frac{L_{id}^k}{2} \ln \frac{1}{u_{id}^k}, \\
u_{id}^k &\sim U(0, 1),
\end{aligned}
\tag{7}
$$

where x_{id}^{k+1} is the d-dimensional component of the i-th particle position in the k+1-th iteration, φ_{id}^k and u_{id}^k are the random numbers evenly distributed on $(0, 1)$, pbest_{id}^k is the d-dimensional component of the individual optimal position (local optimal value) of particle i in the k-th iteration, and gbest_d^k is the d-dimensional component of the optimal particle position (global optimal value) of all particles in the population.

The update mode of pbest_{id}^k and gbest_d^k in the QPSO algorithm is exactly the same as that of the PSO algorithm.

$$
L_{id}^k = 2\alpha \left| x_{id}^k - C_d^k \right|,
\tag{8}
$$

where α is the compression expansion factor. When $\alpha < 1.781$, the QPSO algorithm converges globally. C_d^k is the average value of the d-dimensional component of its own optimal position gbest_d^k in the search process of all particles in the population.

$$
C_d^k = \frac{1}{N} \sum_{i=1}^N \text{pbest}_{id}^k, \quad (1 \le d \le M),
\tag{9}
$$

where N is the population size and M is the particle dimension.
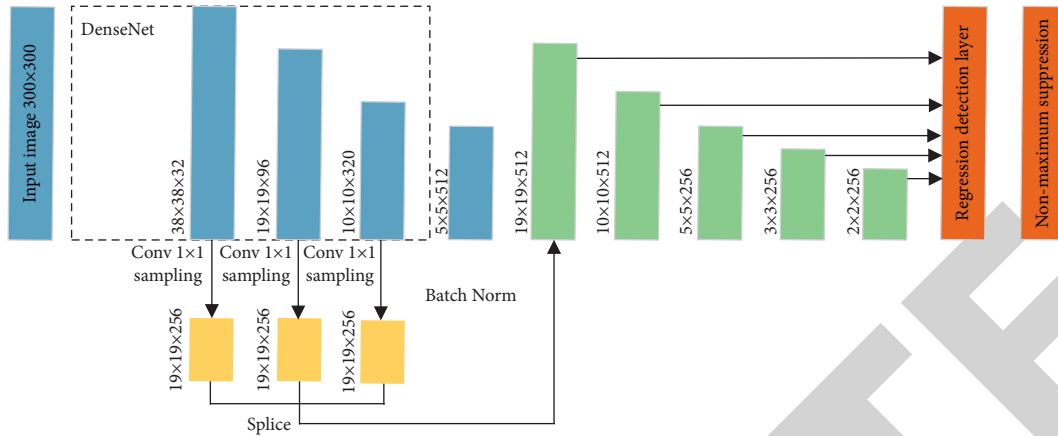
FIGURE 5: Improved DenseNet network structure.

The flowchart of the QPSO-DenseNet model is shown in Figure 6.

4. Experiment and Analysis

4.1. Experimental Environment and Parameter Setting. The experimental simulation environment includes Inter core i7-8750h, 8 GB memory, and graphics card GTX 1060. The operating system is Windows 10. Deep learning framework is TensorFlow, and we use Microsoft Visual Studio 2010 as the development environment. The maximum number of iterations of QPSO is set to 30, the population size is 10, and the dimension of particles is 11.

4.2. Fatigue Discrimination Parameters.

(1) The parameter PERCLOS is the ratio of the time of eye closure to the time of monitoring. Within a certain period of time, when the number of eye closure frames exceeds 70% or 80%, it can be judged as fatigue. Since the number of frames of the image collected using the camera in a fixed time is unchanged, the PERCLOS value is expressed by the ratio of the number of frames with eyes closed in the video to the total number of frames in the video during this period, which can be expressed by the following formula:

$$PERCLOS = \frac{eyes\ closed\ frames}{total\ frames\ in\ detection\ period} \times 100\%. \tag{10}$$

(2) The parameter BF is the number of blinks per unit time. There are two types of blinking: protective blinking in response to external stimuli and human involuntary behavior. Normal nonfatigue people will blink quickly; especially when you are energetic, the BF value is 2–4 times per second and the duration of each blink is short about 0.25–0.3 seconds; when people are tired, their blinking speed will slow down, BF will decrease, and the duration will be prolonged after each eye closure. In order to study the BF

characteristics under fatigue state, 10 sample data under different fatigue states are randomly selected from the dataset for analysis and the mean and standard deviation under nonfatigue state and fatigue state with time windows of 20 s and 40 s are calculated, respectively, as shown in Table 1.

It can be seen that BF is significantly different in the 20 s and 40 s time windows, so BF can be used as the detection parameter of fatigue driving. Based on the above experimental analysis, the CAS_PEAL dataset [28] was selected, and each face in the dataset had 27 different posture images. A partial example of the CAS_PEAL dataset is shown in Figure 7.

Two parameters, PERCLOS and BF, were used to discriminate fatigue. w was used to represent the proportion of PERCLOS parameter in the discrimination (the proportion of BF parameter was $1 - w$). Figure 8 shows the accuracy of the PERCLOS parameter discriminating fatigue with different proportions.

It can be seen from Figure 8 that when $w = 1$, that is, only the PERCLOS parameter is used to distinguish fatigue, the accuracy rate is 92.4%. When $w = 0$, that is, only the BF parameter is used to distinguish fatigue, the accuracy rate is 82%. When $w = 0.6$, the accuracy is 91%. It can be concluded that the comprehensive detection result of PERCLOS and BF is better than that of a single parameter.

4.3. Comparative Analysis of Experimental Results under Different Loss Functions. In this experiment, the artificial neural network model was used to conduct experiments on the CAS_PEAL and self-built datasets and the accuracy of fatigue state discrimination was compared horizontally. During the test, the test set randomly included images with stronger blocking effects such as environment, light, and angle. In order to verify the validity and correctness of the proposed improved loss function, the settings of the loss function are constantly changed under the condition that other parameters in the network model remain unchanged, so as to test the performance of different loss functions in the network, as shown in Table 2.
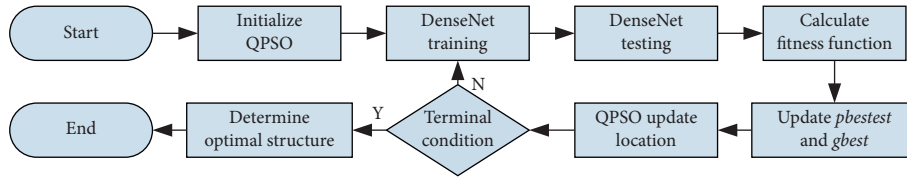
FIGURE 6: Flowchart of the QPSO-DenseNet model.

TABLE 1: Analysis table of blink frequency in nonfatigue state and fatigue state.

Time window (s)	Not fatigued		Fatigued	
	Mean value	Standard deviation	Mean value	Standard deviation
20	15.10	5.97	15.46	5.65
40	15.10	3.65	15.46	3.64
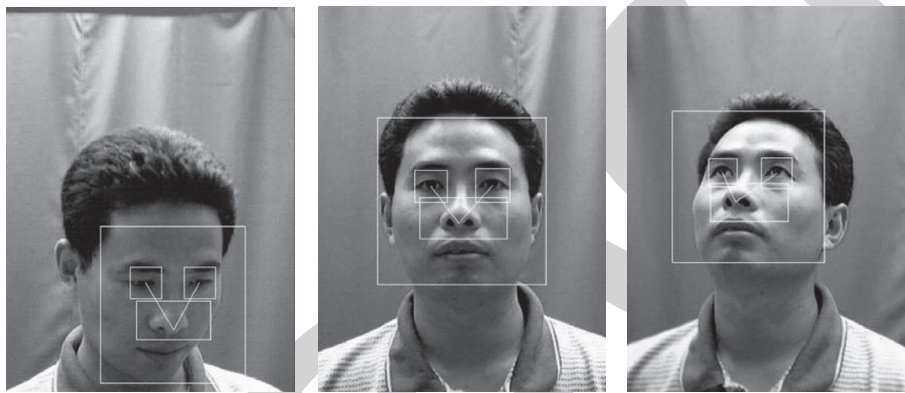
FIGURE 7: Partial example of the CAS_PEAL dataset.
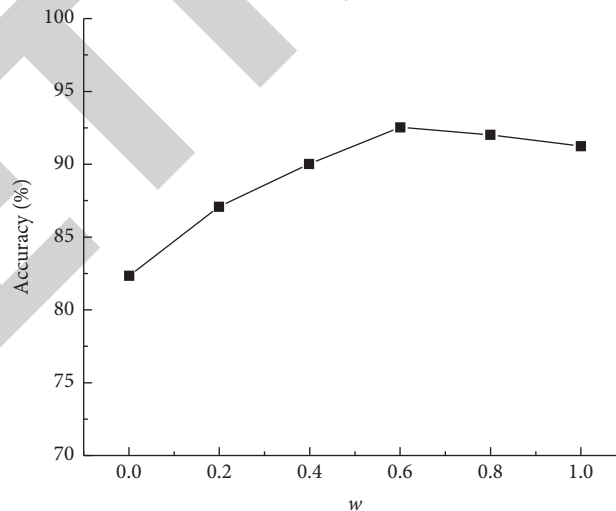
FIGURE 8: The accuracy rate varies with the parameter ratio.

As can be seen from Table 2, Softmax loss function has a good performance in both the original dataset and the dataset containing affected images. Therefore, Softmax loss function is uniformly used in subsequent experiments.

4.4. Comparative Analysis of Experimental Results under Different Connection Modes. In order to verify the

correctness and effectiveness of the improved connection mode in this paper, the network with unchanged connection mode and the network with improved connection mode in this experiment are tested and compared on different datasets, and the results are shown in Table 3.

As can be seen from Table 3, in terms of detection accuracy, the network with improved connection mode in this

TABLE 2: Correct recognition rate under different loss functions.

Loss function	Original dataset	Datasets with affected images
Hinge loss	0.866	0.821
Softmax loss	**0.891**	**0.844**
Square loss	0.882	0.779
Exponential loss	0.831	0.798
Smooth L1 loss	0.843	0.801

TABLE 3: The verification accuracy of different networks.

Dataset	Accuracy of detection under different connection modes		Average time consumption under different connection modes (s)	
	Original link mode	Improved link mode	Original link mode	Improved link mode
CAS_PEAL	0.891	**0.897**	0.52	**0.47**
Self-built dataset	0.837	**0.852**	0.73	**0.65**
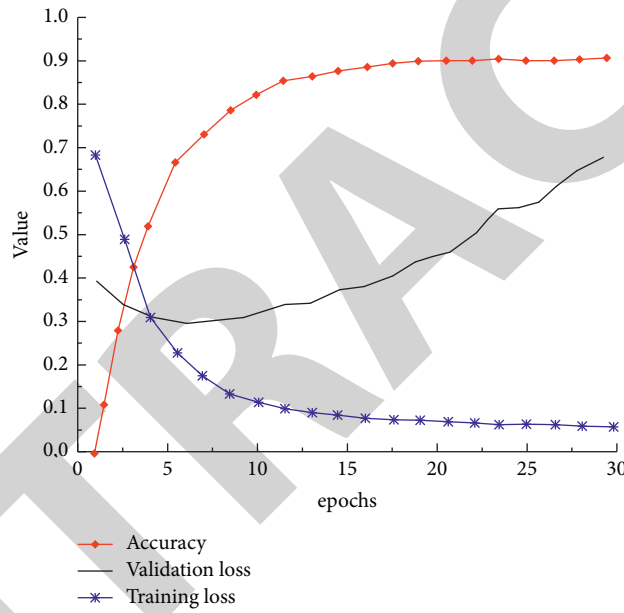
FIGURE 9: Network training process diagram.

TABLE 4: Accuracy of detection under different datasets.

	DenseNet	QPSO-DenseNet
CAS_PEAL	0.891	**0.921**
Self-built dataset	0.837	**0.872**
CAS_PEAL + self-built datasets	0.781	**0.813**

The bold values given in all tables indicate better performance.

TABLE 5: Comparison of accuracy of different models.

	CNN	PSO-CNN	QPSO-DenseNet
CAS_PEAL	0.886	0.903	**0.921**
Self-built dataset	0.824	0.865	**0.872**
CAS_PEAL + self-built datasets	0.767	0.809	**0.813**

paper has a better performance in CAS_PEAL and self-built datasets. In terms of average time consumption, the network with improved connection mode in this paper has low time consumption in both datasets. The results show that the improved network not only improves the detection accuracy but also reduces time consumption.

4.5. Comparative Analysis of Comprehensive Experimental Results. In order to verify the overall efficiency of the QPSO-DenseNet model, the detection accuracy curve and loss function curve with the change of training times were plotted. It can be seen from the curve that the accuracy rate changes with the training times, when it is close to 30 epochs, the accuracy rate is basically stable, and the loss rate also tends to be stable. The network training process is shown in Figure 9.

The accuracy results of the QPSO-DenseNet model and the original DenseNet model on CAS_PEAL and self-built datasets are shown in Table 4.

It can be seen that the accuracy of the QPSO-DenseNet model is higher than that of the original DenseNet model in a comprehensive view, no matter in a certain dataset or two datasets. In order to fully verify the advancement of the proposed model, comparison with the CNN and PSO-CNN models was carried out under the same experimental conditions. The accuracy comparison results are shown in Table 5.

It can be seen from Table 5 that the QPSO-DenseNet model in this paper is superior to the CNN and PSO-CNN models in terms of final detection accuracy. When QPSO-DenseNet trains face images, feature reuse produces a network model with easy training and a high parameter efficiency. By connecting the eye feature maps learned in different layers, the input changes of subsequent layers are increased and the efficiency is improved. This is the main reason why the proposed model is superior to other models. From the above results, it is feasible to use the QPSO-DenseNet model to judge the eye state to describe the fatigue state of students in class.

5. Conclusions

In this paper, based on the traditional DenseNet, an improved DenseNet network framework is constructed by reducing the number of redundant connections and the DenseNet structure is optimized by QPSO. Firstly, experiments are carried out for different loss functions and structures with reduced connection numbers, and the detection accuracy and time consumption are analyzed and compared on different datasets. Finally, the QPSO-DenseNet model was compared with CNN and PSO-CNN models, and the experimental results showed that the QPSO-DenseNet showed a good detection effect.

Data Availability

The experimental data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest regarding the present study.

References

[1] Y. Jiang and Y. Wang, "Evaluation of teaching quality of public physical education in colleges based on the fuzzy evaluation theory," *Journal of Computational and Theoretical Nanoscience*, vol. 13, no. 12, pp. 9848–9851, 2016.

[2] J. Cheng and Y. Xiong, "The quality evaluation of classroom teaching based on FOA-GRNN," *Procedia Computer Science*, vol. 107, pp. 355–360, 2017.

[3] Q. Jian, "Multimedia teaching quality evaluation system in colleges based on genetic algorithm and social computing approach," *IEEE Access*, vol. 7, pp. 183790–183799, 2019.

[4] K. Dong, D. Chu, K. Wang, L. Guo, and S. Gu, "A classroom teaching quality evaluation system for the master of software engineering in China based on engineering education accreditation," *Computer Education*, vol. 288, no. 12, pp. 125–130, 2018.

[5] A. Abdelhadi and M. Nurunnabi, "Engineering student evaluation of teaching quality in Saudi arabia," *International Journal of Engineering Education*, vol. 35, no. 1A, pp. 262–272, 2019.

[6] M. L. Rocca, M. L. Parrella, I. Primerano, I. Sulis, and M. P. Vitale, "An integrated strategy for the analysis of student evaluation of teaching: from descriptive measures to explanatory models," *Quality and Quantity*, vol. 51, no. 2, pp. 1–17, 2016.

[7] S. Liqiong, Y. Jing, J. Xiaoyan, H. Luoya, S. Shoimai, and Z. Yan, "Based on Delphi method and Analytic Hierarchy Process to construct the Evaluation Index system of nursing simulation teaching quality," *Nurse Education Today*, vol. 79, pp. 67–73, 2019.

[8] K. Dreiling, D. Montano, H. Poinstingl et al., "Evaluation in Undergraduate Medical Education: Conceptualizing and Validating a Novel Questionnaire for Assessing the Quality of Bedside teaching," *Medical Teacher*, vol. 39, pp. 1–8, 2017.

[9] A. Buczak and E. Guven, "A survey of data mining and machine learning methods for cyber security intrusion detection," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 2, pp. 1153–1176, 2017.

[10] M. Raissi and G. E. Karniadakis, "Hidden physics models: machine learning of nonlinear partial differential equations," *Journal of Computational Physics*, vol. 357, pp. 125–141, 2018.

[11] C. Voyant, G. Notton, and S. Kalogirou, "Machine learning methods for solar radiation forecasting: a review," *Renewable Energy*, vol. 105, no. 5, pp. 569–582, 2017.

[12] M. P. Pound, J. A. Atkinson, A. J. Townsend et al., "Deep Machine Learning provides state-of-the-art performance in image-based plant phenotyping," *GigaScience*, vol. 6, no. 10, pp. 1–10, 2018.

[13] I. Kavakiotis, O. Tsave, A. Salifoglou, N. Magalveras, L. Vaglavas, and L. Chouvarda, "Machine learning and data mining methods in diabetes research," *Computational and Structural Biotechnology Journal*, vol. 15, no. C, pp. 104–116, 2017.

[14] S. Liu, X. Wang, M. Liu, and J. Zhu, "Towards better analysis of machine learning models: a visual analytics perspective," *Visual Informatics*, vol. 1, no. 1, pp. 48–56, 2017.

[15] M. Kohli, L. M. Prevedello, R. W. Filice, and J. R. Geis, "Implementing machine learning in radiology practice and research," *American Journal of Roentgenology*, vol. 208, no. 4, pp. 1–7, 2017.

[16] X. Lin, X. Wang, and L. Li, "Intelligent detection of edge inconsistency for mechanical workpiece by machine vision

Research Article

Simulation Analysis of the Evolution of Sustainable Operation of Transport Infrastructure Projects under Government Regulation Based on Prospect Theory and BP Neural Network

Chongsen Ma [ID]**, Yun Chen** [ID]**, and Yinghui Zhang** [ID]

College of Transportation Engineering, Changsha University of Science and Technology, Hu Nan 410000, China

Correspondence should be addressed to Chongsen Ma; machongsen@stu.csust.edu.cn

The decisions and actions of operators in the operation of transport infrastructure play a crucial role in the sustainability of the project but are influenced by factors such as the strength of government regulation and the frequency of use by users. The influence of social recognition and acceptance by the authorities on the decisions of the various parties involved in a project is becoming increasingly significant. To address this issue, this paper attempts to apply prospect theory to characterize the impact of changes in recognition on the decisions of project managers and the government from the perspective of recognition and to construct a tripartite evolutionary game model for the government, the operator, and the users, taking into account the combined effect of multiple factors, to explore the evolutionary law of the operator's strategy choice. Evolutionary game theory, in which each person is considered irrational and behavior is changeable, is more realistic. The addition of prospect theory allows the model to more realistically reflect the decisions of each participant in the game process when faced with risk. The results of the study show that there is an optimal level of external regulation to maximize the benefits for all three parties in the game, strong government regulation does not necessarily improve service quality, operators tend to provide low-quality services in the game process and incentives should be increased, and that attempts should be made to provide users with a variety of transport infrastructure options to ensure that users' interests are maximized. The paper further analyzes the indicators established by sensitive factors using BP neural networks on the basis of the analysis of transportation infrastructure operation and impact sensitive factors using evolutionary games and finds that the conclusions obtained by deep learning algorithms are more consistent with those obtained using evolutionary games, achieving cross-validation of the results. The reliability of the result is improved, and it is demonstrated that deep learning algorithms can be introduced as a supplement in the process of future analysis of transportation infrastructure operations. Finally, management suggestions are made in light of the actual situation.

1. Introduction

As an important part of urban infrastructure, transport infrastructure has a large impact on the development and economic growth of cities. How to achieve efficient transport infrastructure operations and guarantee the sustainability of the project during its life cycle is a problem that the government must face [1]. Besley and Ghatak studied how to allocate the public and private sector controllers for purely public goods [2]. At the same time, due to the existence of uncertainty in the project operation process and the incompleteness in the agreement, the project may generate risks, and the government must prevent the operator from damaging the public interest through certain means to prevent the operator from damaging the public interest [3]. Some scholars believe that risks should be avoided through supplementary mechanisms [4]. Through the above-mentioned scholars' studies, it can be found that, in the operation process, the interests of the various parties involved in the project are not consistent, and they may damage the overall interests for their own benefit. Therefore, it is of practical significance to study the behavior of the project parties in the operation process and the reasons for their occurrence.

At the level of operational regulation of projects, Xia and Gao argue that as social regulation gradually becomes mainstream, a credit will become the core of social club

regulation of infrastructure operations [5]. Jiang argues that the main problems of China's infrastructure currently focus on government departments' negligence in supervision, and operators usually choose unreasonable pricing and negligent management strategies in order to expand their interests [6]. He argues that the lack of government policies, regulatory systems, and reasonable risk allocation systems at the regulatory level leads to project risks [7]. El-Gohary et al. argue that when the private sector only cares about its own interests, the probability of adopting a breach of trust strategy to operate rises [8]. The above-mentioned studies show that there are many problems in the process of infrastructure operation and it is difficult to achieve sustainable operation of projects. The lack of government regulation and the private sector's focus on its own interests usually result in reduced benefits for users and the government. The parties involved in the project are unable to maximize the overall benefits.

To address the difficulty of achieving sustainability in the operation of transport infrastructure, Wu argues that the PPP model should be introduced for the operation of stock assets to expand the sources of financing to revitalize the project [9]; Sun proposes promoting "hybridization" to enhance the benefits of the project through the analysis of two typical projects [10]; Xia et al. used an evolutionary game approach to analyze the impact of different levels of social regulation on the behavior of government departments and the private sector.

However, in the course of many scholars' research, the following two main problems exist.

(1) All parties involved in the project are imperfectly rational in the process of participating in the decision-making game, and they do not precisely perceive the risks of the project and the risks of their behavior, and their perception of risks is influenced by the environment they are in. However, there are relatively few papers in the existing literature that address this point.

(2) Although the existing studies take a variety of approaches, they usually only consider the game between the government and the operator in the analysis process and rarely include users in the study [11–15]. Although transport infrastructure is a quasi-public benefit facility, in the actual operation process, if only government subsidies are considered in the benefit calculation process, without considering the impact of users and social recognition on the project decision in the operation process, it cannot fully meet the actual situation; at the same time, in the current research, less psychological factors such as recognition are included in consideration of the dynamic game, and it is more practical to include recognition in the model [16–20].

(3) In the traditional analysis, the operation of transportation infrastructure is mostly analyzed in the form of game, regression, or expert scoring [21–25]. In the research process, machine learning and deep learning [26, 27], which are more studied in recent years, are less involved, and there is less research on whether such methods can be used to predict the operation effect of transportation infrastructure.

The innovation points of this paper are mainly reflected in the following two points:

(1) Based on prospect theory and imperfectly rational decision-making in the study of sustainable operation of transport infrastructure projects under government regulation, it more realistically reflects the choices of the parties involved in the project when facing different degrees of risk.

(2) The modeling process takes into account the impact of behavior on social acceptance and the acceptance of higher regulatory authorities and quantifies it. More realistic and appropriate recommendations are made in response to the simulation results.

(3) The paper uses BP neural networks to predict the operation of transportation infrastructure and finds that the conclusions obtained are closer to those of the traditional evolutionary game approach, which can be better implemented with the traditional approach and facilitate the operation of transportation infrastructure.

2. Simulation Model Construction

2.1. Simulation Analysis Model Construction for Sustainable Operation Evolution of Transportation Infrastructure Projects. In recent years, transport infrastructure projects have become more difficult to operate sustainably due to their large scale of construction, high capital investment, and long payback period. However, regardless of the mode of operation, the main interest groups involved can be summarized as follows: the government, the construction operator, and the user. From an economic point of view, the three parties have a mutual influence on each other. The government commissions the developer to build or operate the project, supervises the project with preagreed contracts, and monitors the performance of the project in terms of actual usage; the developer provides services to the users through construction and operation, and the users' satisfaction and frequency of use influence the government's decisions on similar projects. Due to the asymmetry of information in the whole process and the imperfect rationality of the stakeholders in the decision-making process, the construction and operation parties may be driven by their own technical and information advantages to act opportunistically, causing the government to pay more costs and the users to reap higher profits without receiving the services they deserve. To ensure that transport infrastructure projects meet their construction objectives and achieve sustainable operation, the government and user groups should monitor the behavior of the construction operators. The relationship between the main stakeholders in the sustainable operation of transport infrastructure is shown in Figure 1.
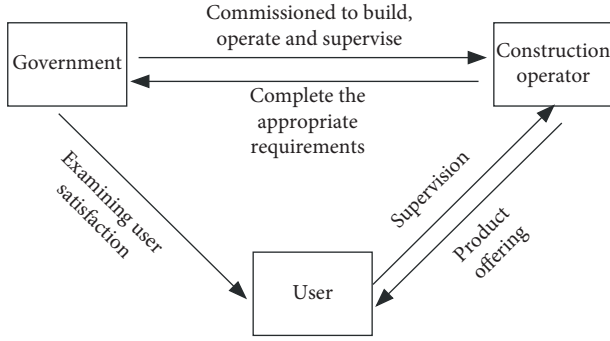
FIGURE 1: Tripartite relationship map of the game.

2.2. Model Assumptions

Assumption 1. In transport infrastructure, the government entrusts the operation of the project to the construction operator. The government regularly reviews the construction and operational effectiveness of the project and adjusts the operation subsidy according to the operational effectiveness to ensure that the social benefits of the project are maximized. The construction operator is responsible for the operation of the project and receives a return through user fees and operating subsidies.

Assumption 2. In the operation of transport infrastructure, the government, the construction operator, and the user are all effectively rational "economic agents;" i.e., all parties involved in the project are not fully rational. In the game process, the three parties play a finite number of repeated games.

Assumption 3. Evolutionary games are based on expected utility theory and do not take into account the influence of psychological perception factors on the decision outcome of the game subjects during the decision-making process. In this paper, we use the prospect theory proposed by Kahnerman et al. to correct for the inconsistent behavior of risk preferences of decision subjects. Prospect theory states that one cannot perceive losses and gains in absolute terms but rather the relative value of perceived losses. This value is expressed using $\Delta\omega_1$, which is the difference between the actual loss or gain ω_1 and a reference point ω_0. This reference point is subject to the subjective influence of the decision-maker. The choice of reference point varies in different fields of study. In the field of finance and investment, the average return is mostly chosen as the reference point, whereas in the literature on regulatory research using prospect theory, 0 is mostly chosen as the reference point. In this paper, 0 is chosen as the reference point. In prospect theory, the expected total utility of a decision is measured using a value function $v(\Delta\omega_1)$ and a weight function π. The prospect value is

$$V = \sum_t \pi(p\iota)v(\Delta\omega\iota). \tag{1}$$

Each participating subject in the game makes a judgment on its next move based on its perceived value of the lost gain, where a value function is

$$V(\Delta\omega_i) = \begin{cases} (\Delta\omega_i)^\theta, & \Delta\omega_i \geq 0, \\ -\lambda(-\Delta\omega_i)^\theta, & \Delta\omega_i < 0, \end{cases} \tag{2}$$

where θ is the risk attitude coefficient, indicating the marginal degree of diminishing perceived value of profit and loss by the game subject; λ is the loss avoidance coefficient, indicating the sensitivity of the game subject to loss; the larger the value, the greater the sensitivity of the game subject to loss. At the same time, the game subject judges the weights according to the actual situation of the event, and the judgment formula is

$$\pi(p_i) = \frac{p^\gamma}{(p^\gamma + (1-p)^\gamma)^{1/\gamma}}, \tag{3}$$

with the exception of very small probability events, $\pi(p_i) < p_i$, $\pi(p_i) + \pi(1 - p_i) \leq 1$ and $\pi(1) = 1$, $\pi(0) = 0$. In prospect theory, the probability of a low-probability event occurring is usually overestimated, and the probability of a high-probability event occurring is usually underestimated.

Assumption 4. The government has two strategic choices based on its own and society's overall benefits: a strong control model and a weak control model. The building operator has two strategies: to provide high-quality services and to provide low-quality services. Users have two strategic choices: high frequency of use and low frequency of use. The probability of the government, the construction operator, and the user choosing the strong control model and providing high-quality services and high frequency of use are x, y, z, $(0 \leq x \leq 1, 0 \leq y \leq 1, 0 \leq z \leq 1)$, respectively; then, the probability of choosing the weak control model and providing low-quality services and low frequency of use is $1 - x$, $1 - y$, $1 - z$.

Assumption 5. The government's choice of the "strong control model" means that the government will take appropriate measures to correct any problems that arise during the construction and operation of the project and will make real-time adjustments to the project's subsidies if the operational performance is not up to standard. The "weak control model" means that the government will not adjust the incentives and subsidies according to the operation of the enterprise and will provide a uniform standard subsidy. The construction operator chooses to "provide high-quality services," which means that the operator, in the course of operation, strictly abides by the contractual agreement and gives full play to the initiative of the enterprise, providing services that meet or exceed the requirements of the government and the needs of the applicable parties. To "provide low-quality services" means that the operator has chosen to partially or fully violate the contractual agreement and in some cases has even colluded with government officials.

2.3. Parameter Setting and Model Construction. If the government chooses the "strong control model," it incurs additional regulatory costs C_3 and receives incentives from higher authorities and social benefits SR_1; if the government chooses the "weak control model," it incurs no additional regulatory costs and receives no additional benefits. If the government chooses the "weak control model," there are no additional regulatory costs and no additional benefits. However, when the operator chooses to "provide a low-quality service," it will be challenged by society, resulting in a combined loss of SR2 (SR2 < SR1, SR1 > C3).

Assuming a fixed revenue of V to be realized by the building operator, an additional cost of C_1 and an incentive subsidy of B are incurred by the operator if it chooses to "provide a high-quality service." The cost of choosing to "provide a low-quality service" is C_2 (including loss of goodwill, loss of potential customers, etc.), $C_3 < C_1 < C_2$, but if the operator chooses to provide a low-quality service, it will incur a penalty of F in the event of government regulation.

There is no additional gain for the user when the operator "provides a high-quality service," regardless of which strategy is chosen. In the case of low-quality services, there is an additional loss of Y_1 if "high usage" is chosen, and a loss of Y_2 if "low usage" is chosen. If the government adopts a "strong control model" when the operator is providing a "low-quality service," it will gain an additional SR_3 due to the increased trust in the government as a result of the increased government control over the project ($SR_1 > SR_3$, $Y_2 > Y_1$).

It is assumed that the probability of the operator being detected by the government for "providing low-quality services" is α under the "weak control model" and 100% when the government adopts a strong control model.

In this paper, the cost of government regulation and the operating cost of the builder are assumed to be deterministic, while the rest of the parameters are related to subjective perceptions and are calculated using prospective values. The benefit matrix of the evolutionary game for the operation of transport infrastructure projects is shown in Table 1.

2.4. Model Solution. According to Table 1, the prospective and average expectations of the government's "strong control model" and "weak control model" strategies are

$$
\begin{aligned}
E_{11} &= yz*(-C_3 + V(SR_1)) + y(1-z)*(-C_3 + V(SR_1)) \\
&\quad + (1-y)z*(-C_3 + V(SR_1) + V(SR_3)) \\
&\quad + (1-y)*(1-z)*(-C_3 + V(SR_1) + V(SR_3)), \\
E_{12} &= yz*0 + y(1-z)*0 + (1-y)z \\
&\quad *(-V(SR_2)) + (1-y)*(1-z)*(-V(SR_2)), \\
\overline{E}_{13} &= xE_{11} - (1-x)E_{12}.
\end{aligned}
\tag{4}
$$

The prospective expectations and mean expectations for the "high-quality service" and "low-quality service" strategies of the operating service providers are

$$
\begin{aligned}
E_{21} &= xz*(V - C_1 + B) + x(1-z)*(V - C_1 + B) \\
&\quad + (1-x)z*(V - C_1 + B) + (1-x)*(1-z)*(V - C_1 + B), \\
E_{22} &= xz*(V - C_2 - V(F)) + x(1-z)*(V - C_2 - V(F)) \\
&\quad + (1-x)z*(V - C_2 - V(F)) + (1-x)*(1-z) \\
&\quad *(V - C_2 - V(F)), \\
\overline{E}_{23} &= yE_{21} - (1-y)E_{22}.
\end{aligned}
\tag{5}
$$

The prospective expectations and mean expectations of users adopting the "high use" and "low use" strategies are

$$
\begin{aligned}
E_{31} &= xy*0 + x(1-y)* - (V(Y_1)) + (1-x)y*0 \\
&\quad + (1-x)*(1-y)* - (V(Y_1)), \\
E_{32} &= xy*0 + x(1-y)* - (V(Y_2)) + (1-x)y*0 \\
&\quad + (1-x)*(1-y)*(-V(Y_2)), \\
\overline{E}_{33} &= zE_{31} - (1-z)E_{32}.
\end{aligned}
\tag{6}
$$

The replicated dynamic differential equation for the government side, the building operator, and the user side choosing an active strategy can be expressed:

$$
\begin{aligned}
F(x) = \frac{dx}{dt} &= x*(E_{11} - \overline{E}_{13}) = x*(1-x)*(yz*V(SR_1) - C_3) + y*(1-z)*(V(SR_1 - C_3)) + (1-y)*z \\
&\quad *(V(SR_1) + V(SR_2) + V(SR_3) - C_3) + (1-y)*(1-z)*(V(SR_1) + V(SR_2) + V(SR_3)) - C_3, \\
F(y) = \frac{dy}{dt} &= y*(E_{21} - \overline{E}_{23}) = y*(1-y)*(xz*(B - C_1 + C_2 + V)(F)) + x*(1-z)*(B - C_1 + C_2 + V)(F) \\
&\quad + (1-x)*z*(B - C_1 + C_2 + \alpha)V(F) + (1-x)*(1-z)*(B - C_1 + C_2 + \alpha V(F)), \\
F(z) = \frac{dz}{dt} &= z*(1-z)*(x*(1-y))*(V(Y_2) - V(Y_1)) + (1-x)*(1-y)*(V(Y_2) - V(Y_1)).
\end{aligned}
\tag{7}
$$

TABLE 1: Revenue matrix.

Strategy		Government			
		Strong control model		Weak control model	
		High-quality service	Low-quality services	High-quality service	Low-quality services
User	High usage	$-C_3 + V(\mathrm{SR}_1)$ $V - C_1 + B$ 0	$-C_3 + V(\mathrm{SR}_1) + V(\mathrm{SR}_3)$ $V - C_2 - V(F)$ $-V(Y_1)$	0 $V - C_1 + B$ 0	$-V(\mathrm{SR}_2)$ $V - C_2 - \alpha V(F)$ $-V(Y_1)$
	Low usage	$-C_3 + V(\mathrm{SR}_1)$ $V - C_1 + B$ 0	$-C_3 + V(\mathrm{SR}_1) + V(\mathrm{SR}_3)$ $V - C_2 - V(F)$ $-V(Y_2)$	0 $V - C_1 + B$ 0	$-V(\mathrm{SR}_2)$ $V - C_2 - \alpha V(F)$ $-V(Y_2)$

When the probability of the government side choosing the strong control model is 1, the government gains an additional SR_1; when the government side chooses the weak control model, the government gains an additional 0. The prospective value of SR1 is

$$
\begin{aligned}
V(\mathrm{SR}_1) &= \pi(1) * V(\mathrm{SR}_1) \\
&= (\mathrm{SR}_1)^\theta.
\end{aligned}
\tag{8}
$$

When the probability of the construction operator choosing a low-quality service is 1, the penalty imposed by the higher level of government is F; when a high-quality service is chosen, the penalty received is 0. The prospective value of F is

$$
\begin{aligned}
V(-F) &= \pi(1) * (-F) \\
&= -\lambda F^\theta.
\end{aligned}
\tag{9}
$$

When the probability of a user taking a high frequency of use is 1, the operator gains a loss of Y_1 if it provides a low-quality service; when a low frequency of use is chosen, the loss is reduced to Y_2; the prospect value is

$$
\begin{aligned}
V(Y_1) &= \pi(1) * V(-Y_1) \\
&= -\lambda Y_1^\theta, \\
V(Y_2) &= \pi(1) * V(-Y_2) \\
&= -\lambda Y_2^\theta.
\end{aligned}
\tag{10}
$$

Similarly, the prospect values of SR_2 and SR_3 can be obtained as

$$
\begin{aligned}
V(\mathrm{SR}_2) &= \pi(1) * V(\mathrm{SR}_2) \\
&= -\lambda (\mathrm{SR}_2)^\theta, \\
V(\mathrm{SR}_3) &= \pi(1) * V(\mathrm{SR}_3) \\
&= (\mathrm{SR}_3)^\theta.
\end{aligned}
\tag{11}
$$

Substituting the prospect values yields a dynamic equation for the replication of the active strategy of the three-way game.

$$
\begin{aligned}
F(x) &= x * (1-x) * (yz^*) * \left((\mathrm{SR}_1)^\theta - C_3 \right) \\
&\quad + y * (1-z) * \left((\mathrm{SR}_1)^\theta - C_3 \right) + (1-y) * z * \\
&\quad + \left((\mathrm{SR}_1)^\theta - \lambda (\mathrm{SR}_2)^\theta + (\mathrm{SR}_3)^\theta - C_3 \right) + (1-y) \\
&\quad * (1-z) * + \left((\mathrm{SR}_1)^\theta - \lambda (\mathrm{SR}_2)^\theta + (\mathrm{SR}_3)^\theta - C_3 \right),
\end{aligned}
\tag{12}
$$

$$
\begin{aligned}
F(y) &= y * (1-y) * \left(xz * \left(B - C_1 + C_2 - \lambda F^\theta \right) \right) + x * (1-z) \\
&\quad * \left(B - C_1 + C_2 - \lambda F^\theta \right) + (1-x) * z * \left(B - C_1 + C_2 - \alpha \lambda F^\theta \right) \\
&\quad + (1-x) * (1-z) * \left(B - C_1 + C_2 - \alpha \lambda F^\theta \right),
\end{aligned}
\tag{13}
$$

$$
\begin{aligned}
F(z) &= z * (1-z) * (x * (1-y)) * \left(\lambda Y_1^\theta - \lambda Y_2^\theta \right) \\
&\quad + (1-x) * (1-y) * \left(\lambda Y_1^\theta - \lambda Y_2^\theta \right).
\end{aligned}
\tag{14}
$$

3. Equilibrium Analysis of the Tripartite Game for the Operation of Transport Infrastructure Projects

3.1. Unilateral Stabilization Strategies for Game Subjects

3.1.1. Government-Side Progressive Stability Analysis.
Let equation (12) equal 0 and solve for $x = 0$, $x = 1$, $y = \lambda (\mathrm{SR}_2)^\theta - (\mathrm{SR}_3)^\theta - (\mathrm{SR}_1)^\theta + C_3/\lambda (\mathrm{SR}_2)^\theta - (\mathrm{SR}_3)^\theta - 2ZC_3 = Y^*$. It follows from the stability theorem for replicating dynamic differential equations that $F(x) = 0$, and $\partial F(x)/\partial x < 0$, x is an evolutionary stabilization strategy.

When $y = Y^*$, $F(x) = 0$ constantly established. The stability point is $x = 0$, $x = 1$. Any value of x is a steady state; i.e., the strategy of the construction unit does not change over time.

When $y < Y^*$, $F(x) = 0$ constantly established. And $\begin{cases} \partial F(x)/\partial x > 0, & x = 0 \\ \partial F(x)/\partial x < 0, & x = 1 \end{cases}$ established. The stability point is $x = 1$. Any value of x is a steady state. That is, this suggests that the perceived benefits of strong control measures are greater than the costs of regulation. In this regard, the benefits include multiple gains, and it is clear from prospect theory that game players are usually reluctant to take losses when faced with

gains; i.e., government departments usually tend to choose the strong control model to avoid incurring losses.

When $y > Y^*$, $F(x) = 0$ constantly established. And $\begin{cases} \partial F(x)/\partial x > 0, \ x = 1 \\ \partial F(x)/\partial x < 0, \ x = 0 \end{cases}$ established. The stability point is $x = 0$. Any value of x is a steady state. That is, it shows that government departments are more sensitive to the perceived cost of inputs than the returns they receive when taking strong control measures. They prefer to bear penalties with uncertainty rather than invest more in regulatory costs. That is, the government tends to choose a weak control model. In this case, the higher level of government should take measures to assign responsibility and increase control, making the government strengthen its control over the project.

3.2. Construction Operator Progressive Stability Analysis.
Let equation (13) equal 0 and solve for $y = 0$, $y = 1$, $z = B - C_1 + C_2 - \alpha\lambda F^\theta - x\lambda F^\theta + \alpha x\lambda F^\theta / 2x(B - C_1 + C_2 - x\lambda F^\theta)$ $= Z^*$. It follows from the stability theorem for replicating dynamic differential equations that $F(y) = 0$, $\partial F(y)/\partial y < 0$, and x is an evolutionary stabilization strategy.

When $z = Z^*$, $F(y) = 0$ constantly established. The stability point is $y = 0$, $y = 1$. Any value of y is a steady state; i.e., the strategy of the construction unit does not change over time.

When $z < Z^*$, $F(y) = 0$ constantly established. And $\begin{cases} \partial F(y)/\partial y > 0, \ y = 0 \\ \partial F(y)/\partial y < 0, \ y = 1 \end{cases}$ established. The stability point is $y = 1$; Any value of x is a steady state. That is, it shows that the sum of the penalties and losses received by the building operator for providing a low-quality service is greater than the benefits under this strategy. If the building operator adopts a speculative strategy, it will face inevitable losses. In this case, the building operator tends to choose to provide a high-quality service.

When $z > Z^*$, $F(y) = 0$ constantly established. And $\begin{cases} \partial F(y)/\partial y > 0, \ x = 1 \\ \partial F(y)/\partial y < 0, \ x = 0 \end{cases}$ established. The stability point is $y = 0$. It indicates that the penalties and losses received by building operator in providing low-quality services are less than the gains under this strategy. The building operator

tends to choose a speculative strategy to obtain higher returns in this situation.

3.3. User Progressive Stability Analysis.
Let equation (14) equal 0 and solve for $z = 0$, $z = 1$, $x = \lambda(SR_2)^\theta - (SR_3)^\theta - (SR_1)^\theta + C_3/\lambda(SR_2)^\theta - (SR_3)^\theta - 2ZC_3 = X^*$. It follows from the stability theorem for replicating dynamic differential equations that $F(z) = 0$, $\partial F(z)/\partial z < 0$, and x is an evolutionary stabilization strategy.

When $x = X^*$, $F(z) = 0$ constantly established. The stability point is $z = 0$, $z = 1$. Any value of z is a steady state; i.e., the strategy of the construction unit does not change over time.

When $x < X^*$, $F(z) = 0$ constantly established. And $\begin{cases} \partial F(z)/\partial z > 0, \ z = 0 \\ \partial F(z)/\partial z < 0, \ z = 1 \end{cases}$ established. The stability point is $z = 1$. Any value of z is a steady state. That is, this situation indicates that the user receives more losses than gains when adopting a low frequency of use, in which case the user tends to choose a high frequency of use.

When $x > X^*$, $F(z) = 0$ constantly established. And $\begin{cases} \partial F(z)/\partial z > 0, \ z = 1 \\ \partial F(z)/\partial z < 0, \ z = 0 \end{cases}$ established. The stability point is $z = 0$. This situation indicates that the loss suffered by the user by adopting a low frequency of use is less than the gain, in which case the user tends to choose a low frequency of use.

3.4. Mixed Stabilization Strategy for Gaming Subjects.
Let $F(x) = F(y) = F(z) = 0$. The equilibrium points can be obtained as follows: $E_1 = (0,0,0)$, $E_2 = (1,0,0)$, $E_3 = (0,1,0)$, $E_4 = (0,0,1)$, $E_5 = (1,1,0)$, $E_6 = (1,0,1)$, $E_7 = (0,1,1)$, $E_8 = (1,1,1)$, and $E_9 = (x^*, y^*, z^*)$, where E_9 is the mixed strategy equilibrium point. If the equilibrium point in the three-way evolutionary game is ESS, it must be satisfied that the equilibrium point is a pure strategy equilibrium, and therefore, only the asymptotic stability of E_1–E_8 needs to be discussed. The asymptotic stability of the system can be obtained from the analysis of the Jacobian matrix, as proposed by Friedman:

$$J = \begin{bmatrix} \dfrac{\partial F(x)}{\partial x} & \dfrac{\partial F(x)}{\partial y} & \dfrac{\partial F(x)}{\partial z} \\[2mm] \dfrac{\partial F(y)}{\partial x} & \dfrac{\partial F(y)}{\partial y} & \dfrac{\partial F(y)}{\partial z} \\[2mm] \dfrac{\partial F(y)}{\partial x} & \dfrac{\partial F(y)}{\partial y} & \dfrac{\partial F(y)}{\partial z} \end{bmatrix}$$

$$= \begin{bmatrix} \left[(SR_2)^\theta\lambda(y-1)+(1-y)(SR_3)^\theta+(SR_1)^\theta-C_3\right](1-2x) & x(1-x)\left[\lambda(SR_2)^\theta-(SR_3)^\theta\right] & 0 \\[2mm] y(1-y)F^\theta\lambda(\alpha-1) & \left[F^\theta\alpha\lambda(x-1)-F^\theta\lambda x+B-C_1+C_2\right](1-2y) & 0 \\[2mm] 0 & z(1-z)\lambda\left(Y_2{}^\theta-Y_1{}^\theta\right) & (1-y)\lambda\left(Y_1{}^\theta-Y_2{}^\theta\right)(1-2z) \end{bmatrix}.$$

$$(15)$$

Each of the eight equilibrium points is substituted into the Jacobian matrix, and if the eigenvalues of the corresponding matrix are all negative, then the equilibrium point is the ESS of the system, and the eigenvalues of each point are shown in Table 2.

This paper makes the following assumptions, $(SR_2)^\theta \lambda < (SR_1)^\theta < (SR_3)^\theta$. Under prospect theory, the loss to social challenge is less than the reward to government regulation; i.e., the government tends to adopt an active strategy to control the project. The stabilization strategy of the game is analyzed in the following cases:

Situation 1: $(SR_2)^\theta \lambda + C_3 < (SR_3)^\theta + (SR_1)^\theta$, $B < F^\theta \alpha \lambda + C_1 - C_2$. In this scenario, the operator chooses to provide a low-quality service, resulting in a combined loss that is less than the cost of government regulation, while the benefit to the operator of reducing the quality of the service is greater than the government's incentive for high-quality service. From the table, we can find that this is the game model evolution equilibrium point, and evolution equilibrium strategy is strong control mode, low-quality service, and low frequency of use.

Situation 2: $(SR_2)^\theta \lambda + C_3 > (SR_3)^\theta + (SR_1)^\theta$, $B < F^\theta \alpha \lambda + C_1 - C_2$. In this case, the combined loss caused by the operator choosing to provide a low-quality service is greater than the cost of government regulation, while the gain to the operator from reducing the quality of the service is greater than the government's incentive for a high-quality service. Table 3 shows that this is the equilibrium point for the evolution of the game model, and the evolutionary equilibrium strategy is a weak control model, low-quality service, and low frequency of use.

Situation 3: $(SR_2)^\theta \lambda + C_3 < (SR_3)^\theta + (SR_1)^\theta$, $B > F^\theta \alpha \lambda + C_1 - C_2$. In this case, the combined losses incurred by the operator when it chooses to provide a low-quality service are less than the cost of government regulation, while the revenue gained by the operator for reducing the quality of the service is less than the government's incentive for a high-quality service.

Situation 4: $(SR_2)^\theta \lambda + C_3 > (SR_3)^\theta + (SR_1)^\theta$, $B > F^\theta \alpha \lambda + C_1 - C_2$. In this case, the combined losses incurred when an operator chooses to provide a low-quality service are greater than the cost of government regulation, while the revenue gained by the operator for reducing the quality of the service is less than the government's incentive for a high-quality service.

There is no stable equilibrium strategy under situation 3 and situation 4.

The above analysis shows that, in the actual operation of the project, the operator usually chooses to provide low-quality services out of self-interest, in order to obtain high revenue by the number of operations. Therefore, in the process of gaming the project, the government should set up reasonable penalties to enhance the willingness of the operators to provide high-quality services.

4. Analysis of Simulations

In order to verify the validity of the evolutionary stability analysis, the model was assigned according to realistic conditions, and numerical simulations were carried out using Python. The parameters set in this paper are shown in Table 4.

Firstly, in order to analyze the impact of the change in the regulatory cost C_3 on the process and outcome of the evolutionary game, C_3 is assigned to 10, 14, and 18, respectively, and the simulation results of replicating the system of dynamic equations evolving 50 times over time are shown in Figure 2. In order to analyze the impact of loss avoidance coefficients on the process and outcome of the evolutionary game, the values of C_3 are assigned as 1, 2, and 3, respectively, and the simulation results are shown in Figure 3.

The graph of the impact of regulatory costs shows that as the cost of regulation rises, it accelerates the rate at which the operating companies and the government side evolve to a point of stability. As the cost of regulation rises, the probability of the government choosing a strong control model decreases. Therefore, when choosing a strong control model, the government needs to be careful to control the cost of control so that it is within a reasonable range. For projects in an unstable state (e.g., in the early stages of operation or within a special period), the regulatory requirements should be appropriately relaxed to facilitate the project to get on track and achieve sustainable operation as soon as possible.

The risk aversion coefficient influence diagram shows that as the risk aversion coefficient rises in the evolutionary process, the probability of the government side choosing a strong control model and the operator choosing a high-quality service rises. Therefore, during the operation of the project, priority should be given to risk-sensitive operators, and the penalties for accidents should be increased to raise the risk awareness of operators and increase the probability of operators choosing high-quality services.

The government incentive subsidy B was then analyzed by assigning B values of 6, 16, and 26 and replicating the simulation results for the system of dynamic equations evolving 50 times over time as shown in Figure 4.

The graph of the impact of incentive subsidies shows that as the incentive subsidy rises, it reduces the rate at which the operator and the government side evolve to a stable point. The evolutionary process shows that if the incentive subsidies are increased, to a certain extent, it will reduce the incentive of the operator to improve the service quality. Therefore, the government should formulate a reasonable subsidy incentive policy to avoid the negative effects of high subsidies.

Substituting the parameters so that they evolve 50 times from the initial strategy combination over time, the results are shown in Figures 5–7.

Figures 5–7 show that there is a stable point of evolution of the system (strong control model, low-quality service, and low usage frequency) under the condition that scenario 1 is satisfied. In the case of scenario 2, there is a stable

TABLE 2: Eigenvalues.

Equalization points	Eigenvalue λ_1	Eigenvalue λ_2	Eigenvalue λ_3
$E_1\,(0,0,0)$	$-(SR_2)^\theta\lambda + (SR_3)^\theta + (SR_1)^\theta - C_3$	$-F^\theta\alpha\lambda + B - C_1 + C_2$	$\lambda(Y_1^\theta - Y_2^\theta)$
$E_2\,(1,0,0)$	$(SR_2)^\theta\lambda - (SR_3)^\theta - (SR_1)^\theta + C_3$	$-F^\theta\alpha\lambda + B - C_1 + C_2$	$\lambda(Y_1^\theta - Y_2^\theta)$
$E_3\,(0,1,0)$	0	$(SR_1)^\theta - C_3$	$F^\theta\alpha\lambda - B + C_1 - C_2$
$E_4\,(0,0,1)$	$-(SR_2)^\theta\lambda + (SR_3)^\theta + (SR_1)^\theta - C_3$	$\lambda(Y_2^\theta - Y_1^\theta)$	$-F^\theta\alpha\lambda + B - C_1 + C_2$
$E_5\,(1,1,0)$	0	$-(SR_1)^\theta + C_3$	$F^\theta\lambda - B + C_1 - C_2$
$E_6\,(1,0,1)$	$(SR_2)^\theta\lambda - (SR_3)^\theta - (SR_1)^\theta + C_3$	$-F^\theta\lambda + B - C_1 + C_2$	$\lambda(Y_2^\theta - Y_1^\theta)$
$E_7\,(0,1,1)$	0	$(SR_1)^\theta - C_3$	$F^\theta\alpha\lambda - B + C_1 - C_2$
$E_8\,(1,1,1)$	0	$-(SR_1)^\theta + C_3$	$F^\theta\lambda - B + C_1 - C_2$

TABLE 3: The equilibrium stability of the nonlinear dynamic system for eight situations.

Equalisation points	Situation 1				Situation 2				Situation 3				Situation 4			
	λ_1	λ_2	λ_3	Stability	λ_1	λ_2	λ_3	Stability	λ_1	λ_2	λ_3	Stability	λ_1	λ_2	λ_3	Stability
$E_1\,(0,0,0)$	+	−	−	−	−	−	−	ESS	+	+	−	−	−	+	−	−
$E_2\,(1,0,0)$	−	−	−	ESS	+	−	−	−	−	+	−	−	+	+	−	−
$E_3\,(0,1,0)$	0	+	+	−	0	+	+	−	0	+	−	−	0	+		−
$E_4\,(0,0,1)$	+	+	−	−	−	+	−	−	+	+	+	−	−	+	+	−
$E_5\,(1,1,0)$	0	−	+	−	0	−	+	−	0	−	−	−	0	−	−	−
$E_6\,(1,0,1)$	−	−	+	−	+	−	+	−	−	+	+	−	+	+	+	−
$E_7\,(0,1,1)$	0	+	+	−	0	+	+	−	0	+	−	−	0	+	−	−
$E_8\,(1,1,1)$	0	−	+	−	0	−	+	−	0	−	−	−	0	−	−	−

TABLE 4: Assignment table.

Parameters	Assignment
SR_1	20
SR_2	10
SR_3	10
θ	1
B	12
λ	2
C_1	30
C_2	50
C_3	18
F	10
Y_1	20
Y_2	25
α	0.5

evolution point (weak control model, low-quality of service, and low frequency of use). In the case of scenario 3, there is a stable evolution point (strong control model, high-quality service, and low frequency of use). Therefore, the government should strengthen regulation during the development of transport infrastructure to promote the provision of high-quality services by operational service providers, while providing users with a diversity of transport options.

5. A BP Neural Network-Based Prediction Model for Sustainable Operation of Transportation Infrastructure

In order to investigate whether the above results based on game theory are in line with reality and whether the conclusions can serve as a guide to reality, based on the simulation using MATLAB, the paper uses BP neural network to

+ $C_3 = 10$
— $C_3 = 14$
--- $C_3 = 18$
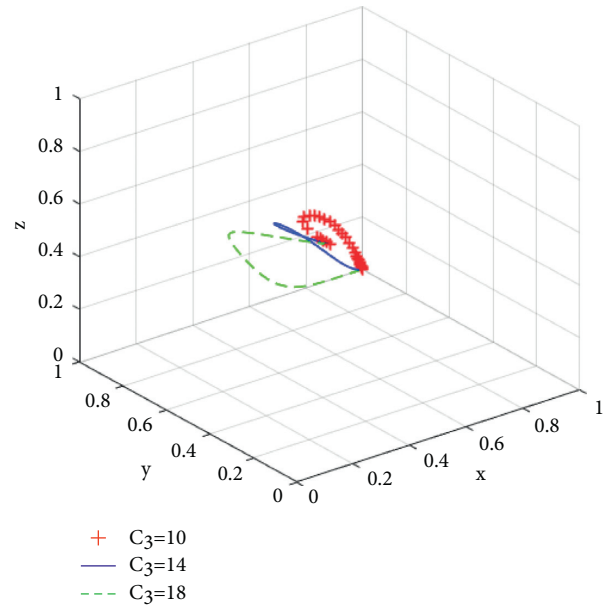
FIGURE 2: Impact of regulatory costs.

build a prediction model based on the results of game theory by collecting actual operation cases of transportation infrastructure across the country and investigates whether the conclusions obtained from the evolutionary game are correct and whether the results of the evolutionary game can be used to further guide the future construction. Based on the results of the game and the sensitivity of each parameter, the index parameters in the BP neural network construction process are further selected. The indicators selected in the model construction process are based on the indicators used in the evolutionary game process, including government subsidies,
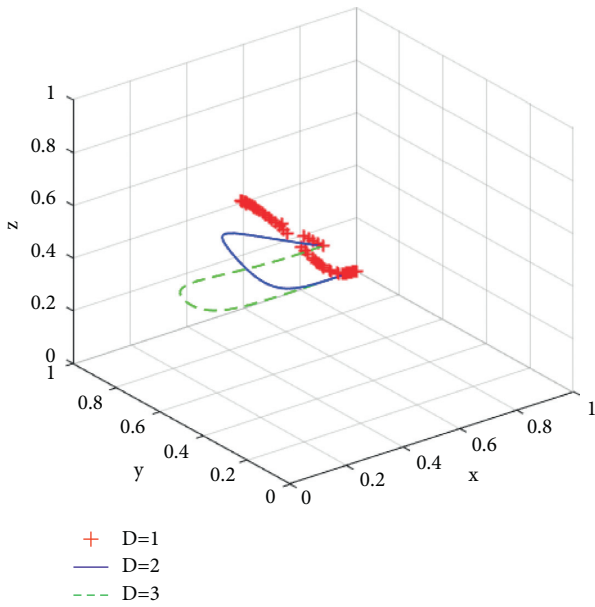
Figure 3: Impact of risk aversion factors.
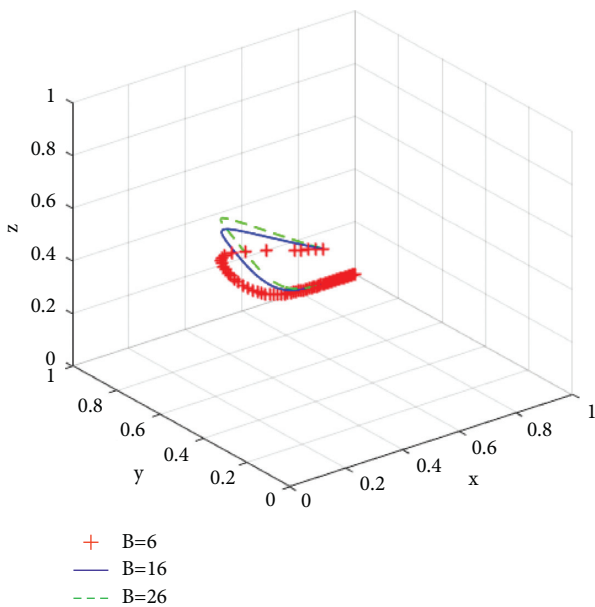
Figure 4: Impact of incentive grants.

Figure 5: The results of 50 evolutions of situation 1.
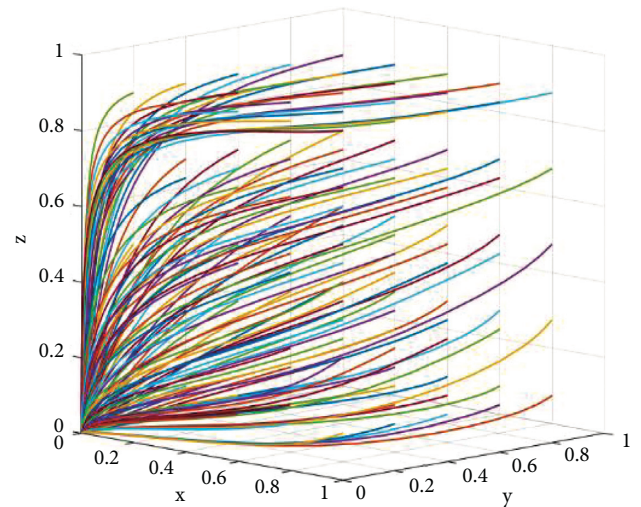
Figure 6: The results of 50 evolutions of situation 2.
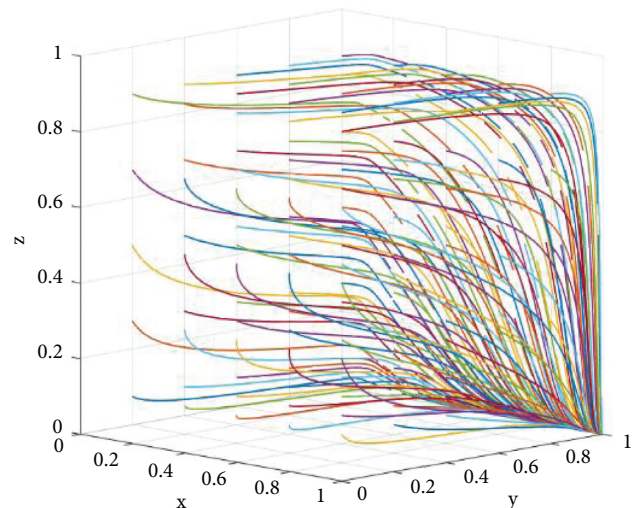
Figure 7: The results of 50 evolutions of situation 3.

government support, GDP, operating income, and other indicators, and supplemented with data indicators that reflect the scale of the project. The selected evaluation result is the final revenue of the project.

The relevant data of the paper are obtained from China National Statistical Yearbook, local economic operation reports, and other official public data.

The mean square error of the model is shown in Figure 8, and the model fit data are shown in Figure 9.

From the fitted data (Figure 9), it can be found that the outcome of the transportation infrastructure operation process can be predicted better using the indicators determined by the evolutionary game and sensitivity analysis of

Figure 8: Mean square error.
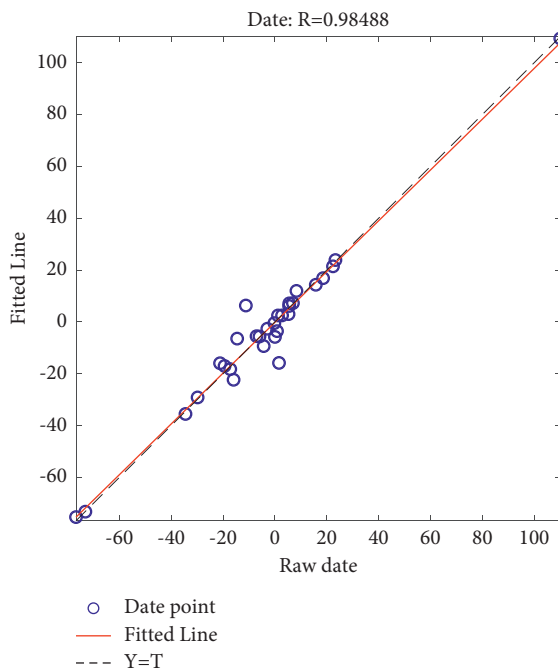
Figure 9: Sample regression plot.

indicators. The mean square error of the original data and the fitted data are small, and the BP neural network can be used to predict and analyze the operation process of the transportation infrastructure. At the same time, by adjusting the corresponding indexes in the prediction set according to the results of game theory analysis, we can find that the error of the model constructed by the BP neural network is smaller than that of the conclusion obtained by using the evolutionary game. Therefore, the combination of the deep learning method and game theory in the traditional operation process can better assist the government and the operator to predict and analyze the possible situation of the project. The use of the BP neural network can better predict the operation results of transportation infrastructure by external influencing factors; therefore, relevant deep learning algorithms can be added to analyze and evaluate the feasibility of the project in the process of relevant policy formulation and project evaluation.

6. Conclusions

This paper introduces prospect theory to the sustainable operation of transport infrastructure projects under government regulation, establishes an evolutionary game model with the participation of three parties: government, operator, and user, and analyzes it using relevant theories to obtain the following conclusions.

(1) The growth rate of the three parties in the game to participate in the construction of transport infrastructure by adopting positive strategies is not the same, and the size of the government's incentive for the construction operator to adopt positive strategies affects the government's willingness to participate. Excessive rewards are not only more burdensome for the government but also do not encourage operators to improve their service quality

(2) The government's willingness to participate is influenced by the cost of monitoring the low-quality of services provided by the operator. The government should dynamically adjust the level of supervision according to the operator's own positive level of operation and the actual operating results and control the supervision expenses while ensuring the operating results. Strong government control does not necessarily result in high-quality service from the service provider. In some cases, it has led to low-quality services being provided. Therefore, the interests of the operators need to be taken care of in the control process as well

(3) In all cases of evolutionary stability, users tend to choose low-frequency strategies, and the government should provide users with diverse transport options to meet their needs. The government should provide users with a wide range of transport options to meet their needs, and in the construction of the city, it should balance the development of each region as far as possible and build up distinctive industries in each region to avoid the emergence of "sleeping cities"

(4) In many cases, operators evolve a balanced strategy of "choosing low-quality services" in order to maximize their own profitability. This is a reflection of the fact that operators are currently rewarded for providing high-quality services and are not sufficiently penalized for providing low-quality services, resulting in operators choosing to adopt low-quality services in order to avoid loss of profit during the operation process, which is not conducive to good project operation

(5) In the process of research for infrastructure operations, the use of deep learning algorithms can better simplify the original research process on the basis of complex calculations to obtain similar conclusions and at the same time can realize the prediction of the future based on the underlying data. Therefore, in the process of future research and use, we can try to

combine deep learning algorithms with traditional analysis and evaluation methods to further optimize the operation process

Data Availability

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Conflicts of Interest

The authors declare that they have no conflicts of interest regarding this work.

Acknowledgments

References

[1] R. Gao, "Research on the success criteria of PPP projects," *Management Modernization*, vol. 38, no. 02, pp. 96–100, 2018.

[2] T. Besley and M. Ghatak, "Government versus private ownership of public goods," *Quarterly Journal of Economics*, vol. 116, no. 4, pp. 1343–1372, 2001.

[3] C. Xu, "Comparative study on subsidy models of government and social capital cooperation (PPP) projects--based on the practice of urban rail transit PPP projects," *Macroeconomic Research*, vol. 94, no. 05, 106+165 pages, 2018.

[4] F. Medda, "A game theory approach for the allocation of risks in transport public private partnerships," *International Journal of Project Management*, vol. 25, no. 3, pp. 213–218, 2007.

[5] P. Xia and F. Gao, "Research on social regulation strategy of infrastructure operation based on credit game," *Management Modernization*, vol. 38, no. 05, pp. 66–71, 2018.

[6] C. Jiang, "Regulatory mechanism of PPP mode operation from the perspective of game theory," *Gansu Finance*, vol. 20, no. 01, pp. 58–62, 2017.

[7] W. He, "Research on government supervision of urban infrastructure under PPP mode," *China Business Theory*, vol. 26, no. 33, pp. 152–154, 2017.

[8] N. M. El-Gohary, H. Osman, and E. E. D. Tamer, "Stakeholder management for public private partnerships," *International Journal of Project Management*, vol. 24, no. 7, 2006.

[9] Y. Wu, "Research on the introduction of PPP model for infrastructure stock assets," *Macroeconomic Research*, vol. 40, no. 02, pp. 84–91, 2020.

[10] W. Sun, "Empirical laws and strategy optimization of PPP mode financing for infrastructure construction--an analysis based on two typical cases," *Economic Vertical*, vol. 41, no. 07, pp. 120–128, 2019.

[11] Y. Yang and M. Chen, "On the structural model and tax policy of REITs--Taking infrastructure public REITs as an example," *Finance and Accounting Monthly*, vol. 29, no. 13, pp. 133–141, 2021.

[12] Z. Zhang, M. Jia, and W. D. Fang, "The optimal control allocation of PPP from the perspective of incomplete contract and relationship contract," *Foreign Economics and Management*, vol. 24, no. 08, 29+44 pages, 2007.

[13] E. T. Ma and X. Li, "Research on the cooperative relationship between project participants under PPP model--an analytical framework based on social network theory," *Finance and Trade Economics*, vol. 38, no. 07, pp. 49–63+77, 2017.

[14] Y. He and H. Xiong, "Research on government supervision of PPP mode from the perspective of game theory," *Journal of Engineering Management*, vol. 32, no. 06, pp. 64–68, 2018.

[15] W. Liu, J. N. Lv, and Q. Zou, "Decision model for concession period of transportation BOT projects under revenue uncertainty," *Systems Engineering*, vol. 30, no. 12, pp. 51–56, 2012.

[16] B. Liu and W. Liu, "Game analysis of private supply mechanism of quasi-public goods: the case of transportation infrastructure investment in China," *China Soft Science*, vol. 22, no. 08, pp. 145–151, 2007.

[17] Z. G. Shao, C. Han, L. Meng, and Q. Wu, "A logistic-based model for the evolution of regional transportation infrastructure ecosystem," *Systems Engineering Theory and Practice*, vol. 38, no. 11, pp. 2918–2928, 2018.

[18] D. Yanchao, Z. Hengyu, Y. Yongbo, and L. Xiaoxue, "Explore knowledge-sharing strategy and evolutionary mechanism for integrated project team based on evolutionary game model," *Advances in Civil Engineering*, vol. 2019, Article ID 4365358, 23 pages, 2019.

[19] H. A. Mesa, K. R. Molenaar, and L. F. Alarcón, "Exploring performance of the integrated project delivery process on complex building projects," *International Journal of Project Management*, vol. 34, no. 7, pp. 1089–1101, 2016.

[20] D. R. Hale, P. P. Shrestha, G. E. Gibson, and G. C. Migliaccio, "Empirical comparison of design/build and design/bid/build project delivery methods," *Journal of Construction Engineering and Management*, vol. 135, no. 7, pp. 579–587, 2009.

[21] Q. Shi, J. Zhu, L. Qian, and P. D. Lellis, "Cooperative Evolutionary Game and Applications in Construction Supplier Tendency," *Complexity*, vol. 2018, Article ID 8401813, 13 pages, 2018.

[22] T. Feng, S. Tai, C. Sun, and Q. Man, "Study on cooperative mechanism of prefabricated producers based on evolutionary game theory," *Mathematical Problems in Engineering*, vol. 2017, Article ID 1676045, 6 pages, 2017.

[23] D. Friedman, "Evolutionary games in economics," *Econometrica*, vol. 59, no. 3, pp. 637–666, 1991.

[24] J. Liu, R. Gao, C. Y. J. Cheah, and J. Luo, "Evolutionary game of investors' opportunistic behaviour during the operational period in PPP projects," *Construction Management & Economics*, vol. 35, no. 3, pp. 137–153, 2017.

[25] Y. Lei, "A tripartite evolutionary game analysis of the Co-construction of an international transport corridor," *Discrete Dynamics in Nature and Society*, vol. 12, no. 2021, pp. 1–11, 2021.

[26] X. Sun and Y. Lei, "Research on financial early warning of mining listed companies based on BP neural network model," *Resources Policy*, vol. 73, Article ID 102223, 2021.

[27] L. Bai, K. Zheng, Z. Wang, and J. Liu, "Service provider portfolio selection for project management using a BP neural network," *Annals of Operations Research*, pp. 1–22, 2021.

Research Article

Human Action Recognition Technology in Dance Video Image

Lei Qiao[1] and QiuHao Shen (ID)[2]

[1]*Jiaozuo University Art College Dance Teaching and Research Section, Jiaozuo 454000, China*
[2]*School of Humanities and Law, Zhejiang A&F University, Zhejiang 311300, China*

Correspondence should be addressed to QiuHao Shen; 2019012389@zwu.edu.cn

In order to effectively improve the recognition rate of human action in dance video image, shorten the recognition time of human action, and ensure the recognition effect of dance motion, this study proposes a human motion recognition method of dance video image. This recognition method uses neural network theory to transform and process the human action posture in the dance video image, constructs the hybrid model of human motion feature pixels according to the feature points of human action in the image coordinate system, and extracts the human motion features in dance video image. This study uses the background probability model of human action image to sum the variance of human action feature function and update the human action feature function. It can also use Kalman filter to detect human action in dance video image. In the research process, it gets the human multiposture action image features according to the linear combination of human action features. Combined with the feature distribution matrix, it processes the human action features through pose transformation and obtains the human action feature model in the dance video image to accurately identify the human action in the dance video image. The experimental results show that the dance motion recognition effect of the proposed method is good, which can effectively improve the recognition rate of human action in dance video image and shorten the recognition time.

1. Introduction

The advent of the intelligent era promotes the rapid development of computer image processing technology. Human action recognition technology has always been a hot topic in the field of computer vision [1]. The purpose of studying human action recognition technology is to effectively extract the human motion features of video image, analyze human action features in video through effective preprocessing of video image, extract human action features from video image and classify them, and finally realize effective recognition of image features. This recognition technology has been widely used [2, 3]. It is of great significance to analyze the human action in the dance video image by using the characteristics of computer vision, realize the recognition of human action, and correct the wrong action in time to achieve high-quality development.

At present, scholars in related fields have conducted research on video image recognition technology and achieved some research results. Guo et al. [4] proposed a human action feature recognition technology based on image feature similarity. Firstly, they use the image analysis method to make action recognition of the video image, then carry out the dimension reduction, and obtain a group of new human action representation models at the same time, use the image feature similarity technology for secondary recognition analysis, and calculate the similarity of the same two images through the adaptive analysis of the image, and finally obtain the recognition results of human action features by the weighted processing method. This method can effectively improve the recognition accuracy of similar actions in human action video. Yu and Min [5] proposed a human action recognition algorithm based on improved time network. Firstly, they extract human action features based on improved time network, construct human recognition model through neural network, use CNN framework for grid fusion, and analyze the characteristic of neural grid. Then, they use the same structure as spatial network for

weighted summation, obtain a set of new feature vectors, and iterate the processing results. Finally, they get a new set of human action features and obtain the classification recognition results through the recognition of the two groups of human motion features. This method has good recognition effect. Based on the above analysis, this paper proposes a method of human action recognition in dance video image, which provides a certain reference for further improving the recognition rate of human action in dance video image and shortening the recognition time of human action.

2. Design of Human Motion Recognition Technology in Dance Video Image

2.1. Extraction of Human Action Features from Dance Video Image. In the design process of this technology, the human action features in the dance video image are extracted by using the neural grid theory, and the feature points of the extracted feature image are classified and processed. Combined with the above theory, the following research carries out the posture transformation of the human action target in the dance video image, which can be expressed as

$$A_{(i,j)}(k,l) = B(i,j) \cdot C(k,l). \qquad (1)$$

In the above formula, $A_{(i,j)}(k,l)$ represents the filtering result of the human action image in the dance video image after processing, $B(i,j)$ represents the human dance target action image, and $C(k,l)$ represents the filtering combination result. By processing the above results, the definition of $p + q$ order matrix human action image $D(x,y)$ in dance video image is obtained, which is expressed as

$$E_{pq} = \sum_x \sum_y (x - \overline{x})^p (y - \overline{y})^q D(x,y). \qquad (2)$$

In formula (2), the center coordinate of the dance video image is $(\overline{x}, \overline{y})$.

Through the iterative processing of human action image in dance video image [6–8], the feature point (i,j) in the coordinate system of human dance action image is obtained, and the pixel probability within the moment t is

$$p(G_{ij}^t) = \sum_{s=K,L} p_s(G_{ij}^t | \theta_{ij,s}^t). \qquad (3)$$

In this formula, G_{ij}^t represents the pixel value of the human action in the dance video image at time t. It is assumed that the actual probability of the pixel (i,j) of the human motion image is $p_s(G_{ij}^t | \theta_{ij,s}^t)$ at any time t. The mixed model of human action feature pixel $p_s(G_{ij}^t | \theta_{ij,s}^t)$ in the dance video image at time t is as follows:

$$p_b(G_{ij}^t) = \sum_k w_{ij,b,k}^t (\mathbf{G}_{ij}^t, \mu_{ij,b,k}^t). \qquad (4)$$

In the above formula, k represents the number of recognition models of human action image in dance video image. The weight of feature vector of human action image in dance video image is $w_{ij,b,k}^t$ at time k. When $\sum_k w_{ij,b,k}^t = 1$, the characteristic vector value of human action image is

$\mu_{ij,b,k}^t$, and the variance matrix of human action features in dance video image is $\mathbf{G}$ at time t[9–11].

Through weight analysis, the fitness value of individual action image in the dance video image can be obtained, and suitable human action feature distribution model can be found in the dance video image. At any time t, in the dance video image, the probability value of the action image pixel (i,j) of the human action is expressed as follows:

$$\eta(Y,\mu,\mathbf{v}) = \frac{1}{(2\pi)^{\pi/2}|\mathbf{v}|^{1/2}} \exp\left(-\frac{1}{2}(Y - \mu)^T \mathbf{v}\right). \qquad (5)$$

In this formula, Y and η represent the feature vectors of the coordinate point (i,j) of the human action image in the dance video image, in which the probability value of the coordinate point is μ, and v represents the variance matrix of the human action image in the dance video image.

According to the pixel information of the human action image in the obtained dance video image, the judgment pixel can be obtained, which is as follows:

$$p(H_B) = \frac{\lambda_a}{5},$$
$$p(H_F) = 1 - p(H_B). \qquad (6)$$

In formula (6), $p(H_B)$ represents the characteristic distribution function of the pixel value HB of the human action image, $p(H_F)$ represents the characteristic distribution function of the pixel value HF of the body action image. When $p(H_B) \geq p(H_F)$, the human action image feature can be used as foreground pixel. When $p(H_B) < p(H_F)$, the human action image can be used as the background pixel [12–14].

After obtaining the background of the human action image in the dance video image, comparing the human action image in the dance video image with the standard motion image, the features of the human action in the dance video image can be obtained, which can be expressed as

$$p_u(y) = I_h \sum_{i=1}^n k\left(\left\|\frac{y - x_i}{h}\right\|^2\right) \delta[b(x_i) - u]. \qquad (7)$$

In the above formula, I_h represents the normalization processing result of the human action image in the dance video image, u represents the characteristic value of the human action image, y represents the center pixel (i,j) of the human action image in the dance video image. The feature of the human action in the dance video image can be obtained by formula (7), so as to realize the extraction of the human action features.

2.2. Detection of Human Action in Dance Video Image. Assuming that the variance of the gray value distribution of the human movements in the dance video image is σ, all the gray values meet the expected value μ of the human action image and the Gaussian distribution of the human action in dance image at this time [15–17]. Then, the background probability model of human action image in dance video image can be obtained by using the following formula:

$$p(x|w_i) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right). \tag{8}$$

In formula (8), $p(x|w_i)$ represents the normal distribution of human action image features, and w_i represents the feature vector of gray value of human action. Assuming that $K = 2, 3, \ldots L$, the variance summation of the human action feature function in the dance video image can be expressed as

$$\mu'_{K+1} = \frac{K-1}{K}\mu'_K + \frac{1}{K}\mu_K,$$
$$\sigma^{2'}_{K+1} = \frac{K-1}{K}\sigma^{2'}_K + \frac{1}{K}\sigma^2_K. \tag{9}$$

Then, in frame $K = L + 1$, if updated the variance of the human action feature function in the dance video, it can be expressed as

$$\mu'_{K+1} = (1-\rho)\mu'_K + \rho\mu_K,$$
$$\sigma^{2'}_{K+1} = (1-\rho)\sigma^{2'}_K + \rho\sigma^2_K. \tag{10}$$

In the above formula, $\mu_K = 1/N_K \sum_{j=1}^{N_K} I_j$ represents the statistic mean value of human action image in the dance video image of frame K, $\sigma^2_K = 1/N_K \sum_{j=1}^{N_K} (I_j - \mu_K)^2$ represents the variance of human motion feature vector, μ'_K represents the Gaussian model feature function, $\sigma^{2'}_K$ represents the Gaussian model variance of human action, N_K represents the sample number of human action, I_j represents the gray value of human action feature points, and ρ represents the feature update rate.

In the analysis of human action in dance video image, in order to detect the relationship between human action features and basic feature points, this study carries out the target comparison of human action image in dance video image by using Kalman filter [18]. Assuming that the action amplitude of the human action image in the first dance video image is X_1, then the feature of detecting the human action in the dance video image can be expressed as

$$\hat{X}_{K+1} = MX_K. \tag{11}$$

In formula (11), M represents the structure parameter in the dance video image, and X_K represents the detection variance value of human action in the K-th dance video image. Combined with the above formula, the detection of human action in dance video image can be completed.

2.3. Recognition of Human Action in Dance Video Image.
In order to recognize human action in dance video image, the mathematical model is established as follows:

By connecting the human action samples in the dance video image, the recognition model of the human action in the dance video image can be obtained. The coordinate point of this model is p, and $T = (Y_1, Z_1, A_1, \ldots, Y_p, Z_p, A_p)^U \in S^{3p}$. The number of human action samples in this model is n, and the human action component in the dance video image is T_j, and then $j = 1, 2, \ldots, n$. T represents the behavior classification of human action. It can

be obtained through normalization processing of all human actions in dance video image [19, 20] and also can be used to describe the linear combination of human action features, which is as follows:

$$T = \sum_{j=1}^{n} x_j T_j. \tag{12}$$

Using the above analysis method, both the human action matrix of the dance video image and the human action image features can be obtained:

$$U = \overline{U} + Tu. \tag{13}$$

In the above formula, $\overline{U}$ represents the mean value for classification of human action behavior in dance video image, and u represents the human action matrix in dance video image. According to the human action feature vector [21], the multiposture features of the human body can be obtained:

$$U^s = \overline{U}^s + \mathbf{G}^S U. \tag{14}$$

In formula (14), $\mathbf{G}^S$ represents the offset matrix of human action in the dance video image; U^s and $\overline{U}^s$ are the state variables in the changing process of human dance posture. Based on the feature distribution matrix K, the human action feature model of dance video image can be obtained by carrying out pose transformation of human dance action features. The formula is as follows:

$$\left(U^l\right)^s = \mathbf{K}\left[\left(\overline{U}^l\right) + \left(G^l\right)U^s\right]. \tag{15}$$

In the above formula, U^l represents the feature points of human action in dance video image [22]. By extracting the state variables of human dance action image, the feature vector of human dance action and the feature components $\overline{U}^l$ and G^l of human dance action can be obtained. Through the above steps, the recognition of human action in dance video image can be realized.

3. Experimental Analysis

In order to verify the effectiveness of human action recognition technology in dance video image in practical application, this experiment chooses the AIST++ dance video dataset and uses MATLAB simulation software as the experimental platform to carry out simulation verification.

In the AIST++ dance video dataset, the most common dance actions mainly include belly dance action, Latin dance action, ballet dance action, national dance action, and folk dance action. This experiment respectively uses the method proposed in this paper and the methods in literature [4] and literature [5] for recognition and compares the recognition effects of dance action with different methods, as shown in Figure 1.

According to Figure 1, for the belly dance action, the dance action recognized by the methods in literature [4] and literature [5] cannot accurately identify the image details, while the method proposed in this paper can accurately

(A) Original image   (B) The method proposed in this paper   (C) Reference [4] method   (D) Reference [5] method

(a)

(A) Original image   (B) The method proposed in this paper   (C) Reference [4] method   (D) Reference [5] method

(b)

(A) Original image   (B) The method proposed in this paper   (C) Reference [4] method   (D) Reference [5] method

(c)

Figure 1: Continued.

(A) Original image   (B) The method proposed in this paper   (C) Reference [4] method   (D) Reference [5] method

(d)

(A) Original image   (B) The method proposed in this paper   (C) Reference [4] method   (D) Reference [5] method

(e)

Figure 1: Recognition effect of different methods on dance action. (a) Belly dance action. (b) Latin dance action. (c) Ballet dance action. (d) National dance action. (e) Folk dance action.

identify the image contour. For Latin dance action, the dance action contour edges recognized by the methods in literature [4] and literature [5] are missing, while the dance action contour edges of the method proposed in this paper are complete. For the ballet dance action, the hand and foot areas of dance action recognized by the methods in literature [4] and literature [5] are deformed, and the method proposed in this paper can effectively restore the original image shape. For the national dance action, the methods in literature [4] and literature [5] still cannot effectively identify the details of the hand area of dance action, but the method proposed in this paper can accurately restore the original shape of the original image on the hand area. For the folk dance action, both the methods in literature [4] and literature [5] have deviations from the original image, while the method proposed in this paper avoids the image recognition deviations and accurately identifies the image edges. Based on the above analysis, it can be seen that the recognition effect of the proposed method is better, because the proposed method uses Kalman filter to detect human action in dance video image, which can effectively remove image noise, so as to ensure the recognition effect of dance action.

During the experiment, the neural network method is to extract the dance video image in the AIST++ dance video dataset frame by frame, and the descriptor matrix is taken as the training sample. According to the dance action samples of the collected dance video image, the experiment aims at five dance action types in AIST++ dance video dataset, and it also tests the recognition rate of the dance action type by the method proposed in this paper. The results are shown in Figure 2.

According to Figure 2, in the AIST++ dance video dataset, when the method proposed in this paper is applied to identify dance action types, the recognition rates of five dance actions are high. In addition, among them, the recognition rate of belly dance action is the highest, which can reach 92%, and the recognition rate of folk dance action is the lowest, which is 85%. Then, the calculation shows that the average recognition rate of the proposed method is 88.7% when identifying five dance action types in AIST++ dance video dataset. Therefore, the recognition rate of dance action types by the proposed method is high.

In order to further verify the recognition effect of the proposed method on human action in dance video image, this experiment selects the data in 400 MB AIST++ dance video dataset, uses the proposed method for recognition, and obtains the recognition rate of human action in dance video image by the proposed method, as shown in Figure 3.
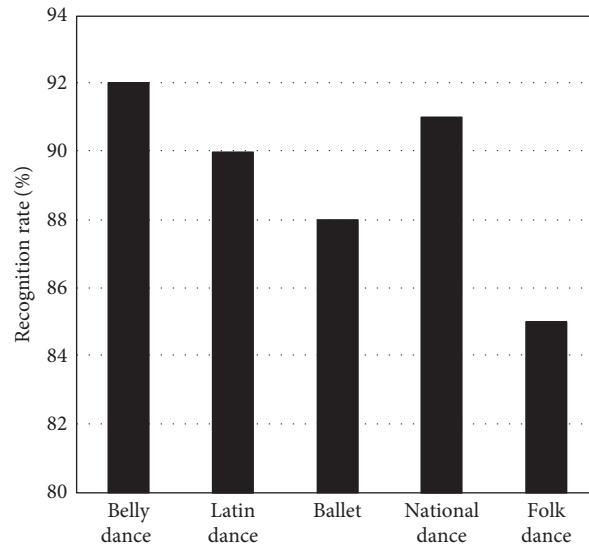
FIGURE 2: The recognition rate of dance action type by the proposed method.
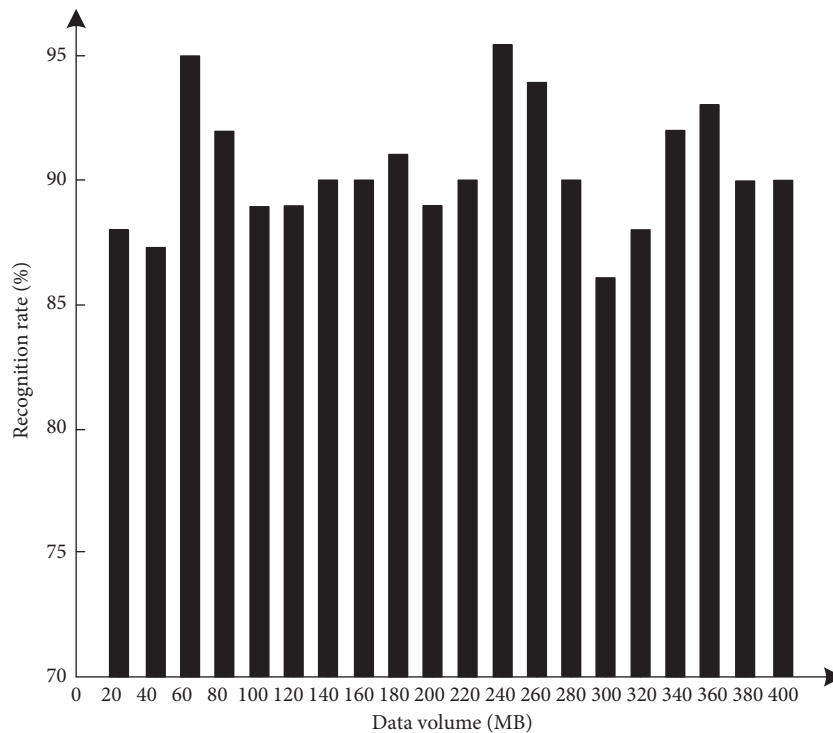
FIGURE 3: The recognition rate of human action in dance video image by the proposed method.

As it can be seen from the results in Figure 3, when the proposed method is used to identify human action in the AIST++ dance video dataset, the recognition rate is high. Among them, when the data amount of AIST++ dance video dataset is 240 MB, the highest recognition rate reaches 95.5%. However, when the data amount in the AIST++ dance video dataset is 300 MB, the lowest recognition rate is 86%. According to the calculation, when the data volume in AIST++ dance video dataset is 400 MB, the average recognition rate of human action in dance video image by the proposed method is 90.4%. Therefore, the proposed method

can effectively improve the recognition rate of human action in dance video image.

On the above basis, the recognition time of human movements in dance video image is further verified. The 400 MB AIST++ dance video data are selected and the recognition time of human movements in dance video image is obtained by using the method proposed in this paper, as shown in Figure 4.

According to the data in Figure 4, the recognition time of human action in the dance video image by the proposed method also increases with the increase of the data amount
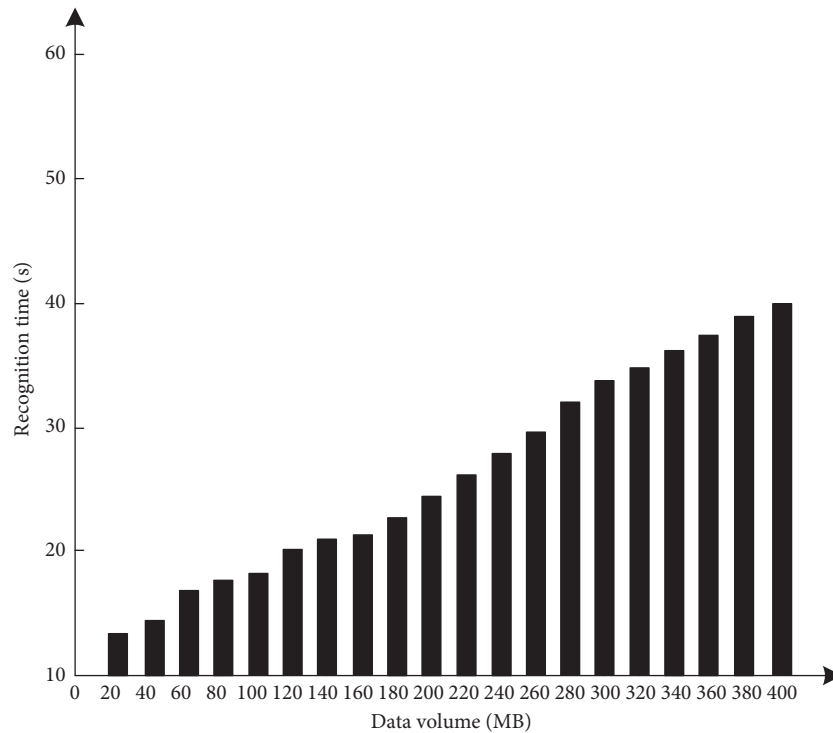
FIGURE 4: The recognition time of human action in dance video image by the proposed method.

in the AIST++ dance video dataset. When the data amount in the AIST++ dance video dataset is 400 MB, the recognition time of human action in dance video image by the proposed method is only 40 s. Therefore, the proposed method can effectively shorten the recognition time of human action in dance video image.

4. Conclusion

This paper studies the human action recognition technology in dance video image, detects the human action in the dance video image by extracting the features of human action, and finally realizes the human action recognition in the dance video image according to the principle of human action recognition in the dance video image. What is more, the research results show that the human action recognition technology designed in this paper has high recognition rate, and it can effectively shorten the recognition time, which further verifies the practicality of this technology [23–26].

Data Availability

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Conflicts of Interest

The authors declared that they have no conflicts of interest regarding this work.

References

[1] H. Wu and Z. Cheng, "Action recognition algorithm based on complexity measure and multi-scale motion coding," *Optical Technique*, vol. 44, no. 04, pp. 427–434, 2018.

[2] H. Li, "Research on motion recognition method in dance video image," *Video Engineering*, vol. 42, no. 07, pp. 34–37+52, 2018.

[3] Z. Wu and Z. Zheng, "Motion recognition algorithm based on deep learning and motion information," *Computer Engineering and Design*, vol. 39, no. 8, pp. 2668–2674, 2018.

[4] Z. Guo, X. Cao, and Y. Hu, "Human motion recognition algorithm based on feature optimization and image similarity," *Science Technology and Engineering*, vol. 19, no. 18, pp. 228–233, 2019.

[5] H. Yu and Z. Min, "Human motion recognition based on improved CNN framework," *Computer Engineering and Design*, vol. 40, no. 07, pp. 2071–2075, 2019.

[6] W. Ding, K. Liu, F. Tang, and X. Fu, "Application of linear dynamic system inversion model in human behavior recognition," *Journal of Image and Graphics*, vol. 24, no. 9, pp. 1450–1457, 2019.

[7] T. Exner, C. A. Beretta, Q. Gao et al., "Lipid droplet quantification based on iterative image processing," *Journal of Lipid Research*, vol. 60, no. 7, pp. 1333–1344, 2019.

[8] F. Kokkinos and S. Lefkimmiatis, "Iterative joint image demosaicking and denoising using a residual denoising network," *IEEE Transactions on Image Processing*, vol. 28, no. 8, pp. 4177–4188, 2019.

[9] T. L. Gomes, R. Martins, J. Ferreira, R. Azevedo, G. Torres, and E. R. Nascimento, "A shape-aware retargeting approach to transfer human motion and appearance in monocular

videos," *International Journal of Computer Vision*, vol. 129, no. 2, pp. 2057–207, 2021.

[10] P. B. Zhang and Y. S. Hung, "Articulated deformable structure approach to human motion segmentation and shape recovery from an image sequence," *IET Computer Vision*, vol. 13, no. 3, pp. 267–276, 2019.

[11] G. Singh, M. Chowdhary, A. Kumar, and R. Bahl, "A personalized classifier for human motion activities with semi-supervised learning," *IEEE Transactions on Consumer Electronics*, vol. 66, no. 4, pp. 346–355, 2020.

[12] S. Jiang, E. Chen, and M. Zheng, "Human motion recognition based on ResNeXt," *Journal of Engineering Graphics*, vol. 41, no. 2, pp. 277–282, 2020.

[13] A. Kushwaha, A. Khare, O. Prakash, and M. Khare, "Dense optical flow based background subtraction technique for object segmentation in moving camera environment," *IET Image Processing*, vol. 14, no. 5, pp. 3393–3404, 2020.

[14] Z. Lin, H. Qin, and S. C. Chan, "A new probabilistic representation of color image pixels and its applications," *IEEE Transactions on Image Processing*, vol. 28, no. 4, pp. 2037–2050, 2019.

[15] H. Luo, K. Tong, and F. Kong, "The progress of human action recognition in videos based on deep learning: a review," *Acta Electronica Sinica*, vol. 47, no. 5, pp. 1162–1173, 2019.

[16] Y. Xinbo, H. Wei, L. Yanan et al., "Bayesian estimation of human impedance and motion intention for human-robot collaboration," *IEEE transactions on cybernetics*, vol. 51, no. 4, pp. 1822–1834, 2019.

[17] J. S. Park, C. Park, and D. Manocha, "I-planner: intention-aware motion planning using learning-based human motion prediction," *The International Journal of Robotics Research*, vol. 38, no. 1, pp. 23–39, 2019.

[18] Y. Liu, S. Qiu, and L. Sun, "Human motion recognition method based on multi-view self-paced learning," *Computer Engineering*, vol. 44, no. 2, pp. 257–263, 2018.

[19] D. Hu, H. Ke, and W. Zhang, "Research on human body attitude recognition based on kinect and ROS," *Chinese High Technology Letters*, vol. 30, no. 2, pp. 177–184, 2020.

[20] S. Chen, W. Wei, B. he, S. Chen, and J. Li, "Action recognition based on improved deep convolutional neural network," *Application Research of Computers*, vol. 36, no. 3, pp. 945–949+953, 2019.

[21] X. Shen, S. Yu, and Y. Dong, "Human motion recognition method based on deep learning," *Computer Engineering and Design*, vol. 41, no. 4, pp. 1153–1157, 2020.

[22] R. Zhang, Q. Li, and J. Chu, "Human action recognition algorithm based on 3D convolutional neural network," *Computer Engineering*, vol. 45, no. 1, pp. 259–263, 2019.

[23] Y. Zhu, X. Huang, and J. Huang, "Research on human motion recognition based on 3D CNN," *Modern electronic technology*, vol. 43, no. 18, pp. 150–152+156, 2020.

[24] B. He, W. Wei, and B. Zhang, "Improved deep convolutional neural network for human action recognition," *Application Research of Computers*, vol. 36, no. 10, pp. 3107–3111, 2019.

[25] S. Liu, X. Bai, M. Fang, L. Li, and C. C. Hung, "Mixed graph convolution and residual transformation network for skeleton-based action recognition," *Applied Intelligence*, pp. 1–12, 2021.

[26] X. Ma, K. Uu Rbil, and X. Wu, "Denoise magnitude diffusion magnetic resonance images via variance-stabilizing transformation and optimal singular-value manipulation[J]," *NeuroImage*, vol. 215, Article ID 116852, 2020.

Research Article

Research on Mathematical Model of Smart Service for the Elderly in Small- and Medium-Sized Cities Based on Image Processing

Chunmei Feng [ORCID]

School of Economics and Management, Chuzhou University, Chuzhou 239000, China

Correspondence should be addressed to Chunmei Feng; mqy1232021@chzu.edu.cn

Image processing technology is to use computer, camera, and other technologies to calculate and process images and make the image clearer and convenient for quick extraction of information. Image processing technology has entered an all-round development stage. It also plays a great role in the components of the intelligent service model for the aged. Now many countries in the world have entered the aging stage, but old-age equipment is relatively backward and personnel management is not standardized. Based on these problems, this paper studies the intelligent model of old-age care in small- and medium-sized cities by using the image recognition method. Based on the analysis of the present situation of intelligent old-age care, an intelligent system is proposed, which solves the problems of defects in old-age care facilities and insufficient comprehensive management of medical staff in some small- and medium-sized cities. This system has RTID positioning system and APP client, which can ensure the privacy of the elderly. Through real-time identification of images in the elderly service, the rationality and layout optimization of existing old-age facilities are analyzed. The mathematical model is used to detect the regularity of participants' daily activities. The image experiment results show that the prediction accuracy is over 90%, and the optimal prediction effect is obtained. In addition, a questionnaire survey was conducted among many elderly people over 50 years old to investigate their willingness to use smart old-age products.

1. Introduction

With the development of computers, image technology has also been applied to all aspects of human life. In the construction of smart models for the elderly in certain small- and medium-sized cities, image processing technology has also played a great role. Literature [1] introduces the development of image processing technology and the connection with the smart service model for the elderly. We analyzed the results of previous research and related data. Because of the physical reasons of the elderly, their social and victory functions will decline with age, so they will encounter various difficulties in daily life. The article is based on the elderly. The daily activity habit of creating scenes provides a specific solution for the activities of the elderly at home. Literature [2] discusses the views and suggestions of alternatives to restraint and

isolation for service users and their families, as well as the conditions that can be used in adults in short-term psychiatric care and residents in long-term care. Set up 3 nursing institutions related to middle-aged and elderly family members and 5 focus groups related to service quality. These small focus groups discussed and analyzed records and found that more companionship and caring for the elderly are the most effective way to solve the mental disorders of the elderly. Listen to the inner thoughts of the elderly and give them more love and company. Immature old-age care policies in big cities result in the high poverty rate of the elderly and a series of old-age care problems. Literature [3] studies the elderly to provide location-based, customized job search services to actively support the elderly to participate in economic activities, provide customized positions for the elderly according to their residence, physical condition, and

working conditions, so that the elderly can choose suitable positions according to their interests and living conditions, and contribute to the expansion of the employment market for the elderly. Literature [4] aims to develop smart healthcare glove system (SHGS) and transcutaneous electrical nerve stimulator (TENS) based on electronic textiles as healthcare equipment for elderly hypertension. Using image processing technology, the patient's pulse and blood pressure generated pictures can be transmitted to the computer for better observation. Observing the blood pressure and pulse of multiple patients wearing SHGS, the results show that SHGS can not only be used for the elderly to lower blood pressure and improve the irregular blood circulation but also be used for hypertensive patients of any age. The goal of literature [5] is to deploy an intelligent nursing service integration agent to provide individualization and integration for each elderly person. Every elderly person's needs and preferences are different. In time, a robot with strong learning ability also needs to spend a lot of time learning everyone's preferences and needs. In order to provide basic nursing tasks, we have deployed nursing templates in the cloud. There are different scenes for each character. Literature [6] provides a theoretical basis for combining biophilic and smart home technology and provides a framework for smart home services to ensure that elderly residents can have a biophilic experience. Smart home technology is not a mechanized achievement. It simulates nature and creates a high quality of life, which can support the physical and mental health of the elderly. It not only provides effective information for the elderly smart home industry but also makes a huge contribution to the trend of smart home services. Literature [7] studies the contradiction between the design of smart home products under the medical care model and the daily needs of the elderly and conducts case analysis of smart home products for the elderly. The research results show that smart homes can maximize the satisfaction of the elderly. Literature [8] introduces WITSCare, which is a research prototype of a web-based IoT smart home system. It is a smart home system designed using the intelligence of the Internet of Things, which can help the elderly live safely and independently at home. Literature [9] proposes an application architecture suitable for such online service network platforms and designed it with the most advanced concepts such as server-side JavaScript, NoSQL databases, and machine learning. The rapid increase in the number of elderly people has brought great pressure to our country's medical treatment and society. It has become an increasingly popular topic to provide the elderly with a high-quality living environment. Literature [10] proposes a conceptual model of an integrated and personalized system to solve this problem. We interviewed three healthcare experts to better explore the needs of the elderly and customized a conceptual model for the elderly to live independently, centering on the concepts of comfort, safety, and environmental protection and providing good protection for the elderly to live alone. The purpose of literature [11] is to propose services based on behavioral patterns. The range of activities for the elderly is limited,

and the living room is an important area. We have proposed a method of developing a smart living room specifically designed for the elderly. We put forward a number of behavior patterns of the elderly, including preventing falls and injuries, and explain safety issues and health issues to provide a comfortable living environment in all aspects. There are many hidden safety hazards for the elderly at home alone. To ensure the safety of the elderly at home, we must consider the hazards such as gas leaks in the kitchen, fire, and falls in the bathroom. Literature [12] deals with the hazards and intelligent services in the kitchen according to the behavior of the elderly. According to multiple accidents such as fires and gas leaks, a smart service has been set up. The smart service can not only detect safety problems at home but also has functions such as automatic alarms and automatic ventilation to further protect the safety of the elderly at home. Studies have shown that the most common disease in the elderly is cardiovascular disease, which also poses a great threat to the health of the elderly. Literature [13] develops smart clothes to record 3-lead electrocardiogram (ECG). Using the image elementary technology, the analyzed data are quickly generated into pictures. The system is composed of fiber clothes with electrodes which have the function of acquiring physiological signals and can analyze health data. Experimental results show that the accuracy of ECG is as high as 86.82%. Starting from the two ways of family pension and institutional pension, family pension is the pension mode chosen by most people. The deficiency of traditional family pension lies in that it is generally difficult for the elderly to get professional and meticulous care, medical care, and spiritual and cultural services in the family. Under the background that "421 structure" families have become the mainstream of urban society, social competition has intensified and the pace of life has accelerated. The social labor cost and people's work burden are generally increasing, family members and children cannot have enough energy to take care of the elderly at home, family pension is facing severe challenges, the traditional family pension model is increasingly difficult to maintain and play its social function and role, and the urban family pension is gradually weakened and socialized. Nursing homes charge higher fees, so most families cannot bear the economic pressure, and they have higher requirements for medical staff. Based on these problems, this paper creates an intelligent mathematical model to help solve the problem of providing care for the aged and uses image processing technology to create an intelligent system. The system has many functions, such as detecting health problems of the elderly and real-time positioning. By analyzing the present situation of old-age wisdom in China, this paper finds some problems in old-age care in China, such as backward old-age care equipment and irregular personnel management. This paper uses image processing technology to build an old-age wisdom service model, which solves the problems of old-age care facility defects and insufficient comprehensive management of medical staff in some small- and medium-sized cities. Using this model, the problem of providing care for the

aged in China has been greatly solved. Through the test, it can be found that the system has a very high accuracy for the health detection of the elderly.

1.1. Overview and Development of Digital Image Processing Technology. With the rise of the computer industry, people gradually began to pay attention to digital image processing technology. In 1964, the US Jet Propulsion Laboratory used a computer to process a large number of lunar photos sent by the "Prowler" spacecraft. The results are very satisfactory. The digital image technology has also become an emerging subject. Until the 1990s, digital images have developed rapidly. Today, various industries have put forward high requirements on image processing technology, which has also promoted the better development of image processing technology. Common methods of processing technology include image transformation, image enhancement and restoration, and image segmentation. Image processing technology has the following characteristics: the diversity of image processing, the sharpness of the processed image is getting higher and higher, and the amount of data processing is large.

1.2. Research Background. The problem of population aging in our country is becoming more and more serious. The issue of elderly care is one of the problems that our country must face in its development. According to the report of the civil affairs department, the elderly in our country now account for 17.17% of the total population, and the problem of population aging is aggravated. The imperfections of old-age facilities show that our country is not fully prepared for the problem of population aging again. The number of elderly people living alone has increased rapidly in recent years. Because with the acceleration of social development and the pace of life, young people in many big cities have chosen to develop in big cities, so they cannot accompany these old people for a long time to take care of their own old people. The elderly care in modern society is mainly divided into two types: one is home care and the other is "material home." Because of the early implementation of the family planning policy in our country, a pair of husbands may have a single child every day. This traditional family management model is called "4-2-1" by us. It refers to a special family that requires 4 independent elderly people, a husband and a wife, and a child. This is for a couple of two people. In other words, it is necessary to take care of 4 independent old people and one child at the same time. Because of the livelihoods of working outside, there is no way to accompany the old people at all times. The traditional family care model is difficult to achieve. When the elderly live alone, there are also many hidden dangers. When the elderly suddenly falls ill and cannot seek medical treatment in time, they cannot deal with the sudden accident in time. These hidden dangers make children of many ages gravitate towards nursing care institutions. The elderly care institutions in our country are showing a trend of two levels of differentiation. Senior nursing homes are equipped with advanced equipment and high level of management personnel, but the fees are too high. Most families cannot afford the high expenses. The general nursing homes have aging equipment and limited service quality and cannot provide high-quality services for the elderly.

1.3. Research Significance. With the increase of age, the elderly may also have some problems in life, such as falling hair and teeth, shaking legs and feet, and memory loss. They tend to have a strong sense of self-dependence and fear of loneliness. They pay more attention to their own life and health problems and their dependence on family and affection. As modern Chinese elderly people decline in their cognitive abilities and active actions, their ability to accept new things has also declined. Most nursing homes in our country have imperfect medical service facilities and imperfect management, which will cause many retired elderly people to have to enjoy high-quality medical care. At present, most nursing homes in our country monitor and care for the elderly mainly through videos. Although video monitoring can directly monitor the situation and behavior of the elderly in the life of the home and nursing homes, there are dead corners in this kind of monitoring and cannot provide full coverage to the elderly. This kind of monitoring easily reveals the personal privacy of the left-behind elderly. Therefore, this article proposes an intelligent elderly care service system. On the premise of ensuring that the privacy of the elderly in the family is not leaked and will not affect the normal life of the elderly in the family, the RFID technology equipment can monitor the elderly in the family in real time. The RFID technology equipment can monitor the situation of medical sanatorium in real time for the elderly at home.

1.4. Status Quo of Research on Smart Elderly Care. Population aging is an inevitable phenomenon. Many developed countries have entered a new era of population aging earlier than our Chinese nation and have accumulated a lot of advanced experience. Therefore, when dealing with these pension problems, our country should appropriately learn from the practices of developed countries. From the successful examples of the elderly care service systems in these developed countries, we can find some more excellent high-quality smart elderly care products, which not only integrate advanced technology but also incorporate humanistic care that meets the characteristics of the elderly, so that the elderly can enjoy high tech while bringing convenience and fun and feel the humanistic care that meets your requirements. With the development of computer technology and the Internet of Things, the convenience brought by automation to the elderly care problem is of great significance, and it provides effective experience for the development of our country's smart elderly care industry.

2. Related Technical and Theoretical Research

2.1. Positioning Technology. Positioning technology mainly includes two major components: indoor positioning and outdoor positioning. Outdoor positioning technology has been widely used in various scenarios. At present and

internationally, outdoor positioning systems that can achieve commercial operation and normal operation mainly include the United States' global positioning system, Europe's Galileo satellite positioning system, and Russia's global positioning system. The navigation satellite system and China's BeiDou satellite positioning system are based on satellite signals, but when a satellite signal is transmitted indoors, the signal strength will be severely degraded and the error is large, making it impossible for anyone to receive. Therefore, it cannot be suitable for indoor positioning. However, the area of daily life of the elderly is basically indoors, so the indoor automatic positioning system can greatly promote the daily life of the elderly.

2.2. Frequency Radio Identification.

2.2. Frequency Radio Identification. RFID wireless radio frequency technology is a technology that uses radio frequency signals to achieve information interaction through a magnetic field. The basic structure diagram is shown in Figure 1.

The development process of RFID technology is shown in Table 1.

There are 3 categories of RFID, and their main characteristics are shown in Table 2.

The comparison chart of RFID technology under different carrier frequencies is shown in Table 3.

3. Image Preprocessing

3.1. Image Binarization. Image binarization is a common image segmentation method. The grayscale processed image is binarized again. Assuming that an image has L gray levels and T is the binarization value, the entire image can be divided into two districts, namely, C_0 and C_1. The number of sharing speed points of gray level i is n_i, and N is the total number of pixels in the image. Then,

$$N = \sum_{i=0}^{t-1} n_i. \tag{1}$$

The probability of occurrence of i is

$$p_i = \frac{n_i}{N}. \tag{2}$$

The probability of the pixels in the two regions appearing in the image is

$$\omega_0 = \sum_{i=0}^{T} P_i,$$
$$\omega_1 = \sum_{i=T+1}^{L-1} P_i = 1 - \omega_0. \tag{3}$$

The average grayscale of the two regions is

$$\mu_1 = \frac{1}{\omega_0} \sum_{i=T+1}^{L-1} i p_i, \tag{4}$$

$$\mu_0 = \frac{1}{\omega_0} \sum_{i=0}^{T} i p_i. \tag{5}$$

Through formulas (4) and (5), the average gray value of the entire image is obtained, and according to the pixel points, the color image is divided into three components: R, G, and B, which show various colors such as red, green, and blue, respectively. Grayscale is the process of making the R, G, and B components of color equal. The pixels with large gray value are brighter (the maximum pixel value is 255, which is white), and the opposite is darker (the lowest pixel is 0, which is black).

The background and the interclass variance formula of the target can be obtained as follows:

$$\mu = \sum_{i=0}^{L-1} i P_i = \sum_{i=T+1}^{T} i P_i = \omega_0 \mu_0 + \omega_1 \mu_1,$$
$$\sigma^2(T) = \omega_0(\mu_0 - \mu)^2 + \omega_1(\mu_1 - \mu)^2 = \omega_0 \omega_1 (\mu_0 - \mu_1)^2. \tag{6}$$

3.2. Image Morphological Filtering. Image morphological filtering is widely used in the process of image processing, and its common arithmetic methods are divided into the following types.

(1) *Expansion Algorithm.* The principle of the expansion algorithm is to assume that A is an image, B is a structural element, and $\oplus$ is an operator in Figure 2, which is defined as

$$A \oplus B = \left\{ Z \mid (\widehat{B})_Z \cap A \neq \varnothing \right\}. \tag{7}$$

(2) *Corrosion Algorithm.* The basic principle of image expansion is assuming that A is an image, B is a structural element, and $\odot$ is an erosion mathematical operator in Figure 3, which is defined as

$$A \odot B = \left\{ Z \mid (B)_z \subseteq A \right\}. \tag{8}$$

3.3. Use Image Processing Data

3.3.1. Extreme Learning Machine. When training an extreme learning machine to deal with classification problems, a set $(x_i, t_i) \in R^d \times R^m$ is given, where x_i is the input vector and t_i represents the category to which it belongs. Then,
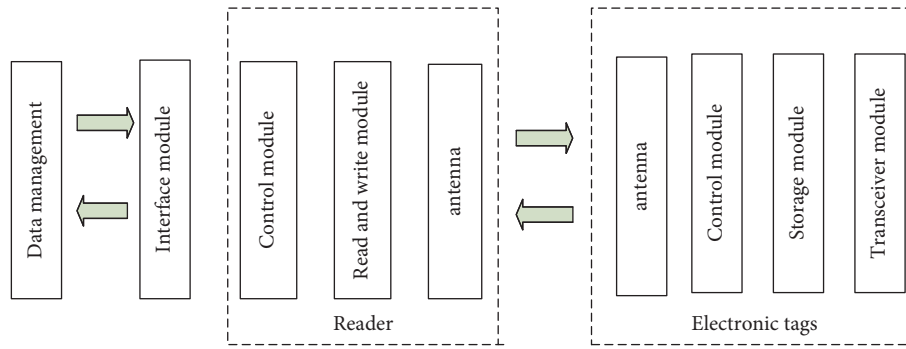
Figure 1: RFID structure diagram.

Table 1: RFID technology development process.

Year	Development process
1941–1950	RFID technology is separated from radar technology and appears in front of people as an independent technology
1951–1960	RFID technology is separated from radar technology and appears in front of people as an independent technology
1961–1970	The first RFID-related paper was published, and the successful application of EAS for electronic article monitoring marks the further development of RFID technology
1971–1980	A large number of RFID patents appeared, and RFID technology appeared in commodity applications for the first time
1981–1990	RFID has been officially used in commercial production, and various large-scale applications have begun to appear
1991–2000	The standardization of RFID technology is getting more and more attention, RFID products are widely used, and RFID products have gradually become a part of people's daily lives
2000 later	RFID product types are more abundant, the production level is continuously improved, the cost of electronic tags is continuously reduced, and the scale of application industries is expanded

Table 2: Comparison of characteristics of different types of RFID.

Types of RFID	Passive RFID	Semiactive RFID	Active RFID
Label power supply	Without battery	Built-in battery	Part of the built-in battery
Range of action	Limited	Farther	General
Service life	Longer	Shorter	General
Label cost	Lower	Higher	General
Adapt to harsh environments	Suitable	Inappropriate	General

Table 3: Technical comparison table under different frequencies.

Frequency	Low frequency	High frequency	UHF
Carrier frequency	<125 kHz	13.56 MHz	>433 Hz
General characteristics	High price, affected by the environment	Low price, suitable for short-distance and multiple target recognition applications	Advanced IC technology makes the cost the lowest, suitable for multiple target recognition
Data transfer rate	Low (8kbit/s)	High (64kbit/s)	High (64kbit/s)
Recognition speed	Low (<1 m/s)	Medium (<5 m/s)	High (<50 m/s)
Label structure	Coil	Printed coil	Dipole antenna
Directionality	None	None	Part
Humid environment	No effect	No effect	Greater impact
Market share	74%	17%	9%
Transmission performance	Penetrable conductor	Penetrable conductor	Linear propagation
Imitation impact performance	Limited	Good	Good
Existing standards	ISO11784 ISO11785	ISO18000-3 ISO14443	EPC G2 ISO18000-6
Recognition distance	<60 cm	0.1–1 m	1–6 m
Scope of application	Access control, fixed equipment, natural gas	Library, product tracking, transportation	Shelves, truck tracking, containers
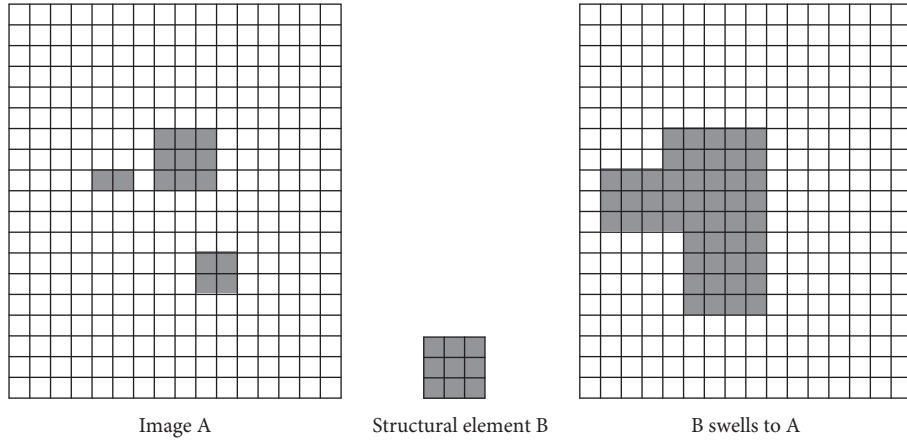
FIGURE 2: Schematic diagram of expansion algorithm.
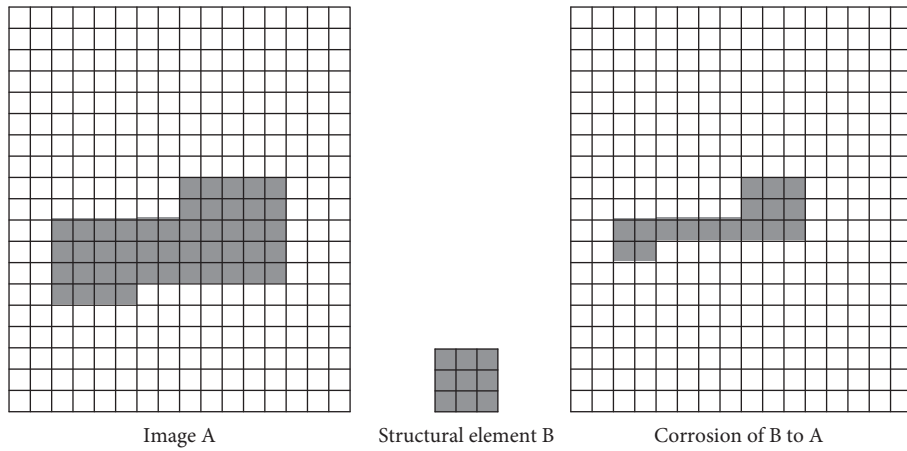
FIGURE 3: Schematic diagram of corrosion algorithm.

$$t_j = \sum_{i=1}^{L} \beta_i G_i(w_i \cdot x_i + b_i), \quad j = 1, \ldots, N. \tag{9}$$

Among them, $G(w, b, x)$ is the activation function. Equation (10) can be simplified to

$$H\beta = T. \tag{10}$$

In equation (10),

$$H = \begin{bmatrix} G(w_1 \cdot x_1 + b_1) \cdots G(w_L \cdot x_1 + b_L) \\ \cdots \\ G(w_1 \cdot x_N + b_1) \cdots G(w_L \cdot x_N + b_L) \end{bmatrix}_{N \times L},$$

$$\beta = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_N^T \end{bmatrix}_{L \times M}, \tag{11}$$

$$T = \begin{bmatrix} t_1^T \\ \vdots \\ t_N^T \end{bmatrix}_{L \times M}.$$

According to the permission matrix of the learning machine, it can be set as

$$\widehat{\beta} = (H^T H + \lambda I)^{-1} H^T T, \quad \lambda > 0. \tag{12}$$

When entering an unknown sample of $\widetilde{x}$,

$$\widetilde{t} = \arg \max (\widetilde{h} \widehat{\beta}), \tag{13}$$

where

$$\widetilde{h} = [G(W_1 \cdot \widetilde{X} + b_1) \cdots G(w_L \cdot \widetilde{x} + b_L)], \tag{14}$$

where $\widetilde{t}$ is the predicted value of unknown sample $\widetilde{x}$.

3.3.2. Support Vector Machine. The training process of the support vector machine model is to solve a minimum problem with constraints:

$$\begin{cases} \{y_i[(w_i \cdot x_i) + b] \geq 1, & i = 1, 2, \ldots m, \\ \min\left\{J(w) = \dfrac{\|w\|^2}{2}\right\}. \end{cases} \tag{15}$$

Solving by Lagrangian function,

$$L(w, b, \alpha) = \frac{1}{2}(w \cdot w) - \sum_{i=1}^{m} \alpha\{[(x_i \cdot w) + b]y_i - 1\}, \quad (16)$$

Take the partial derivative of w and b, and then take the value of 0. Then, we obtain

$$\frac{\partial L}{\partial W} = W - \sum_{i-1}^{m} \alpha_i y_i x_i = 0,$$

$$\frac{\partial L}{\partial b} = \sum_{i=1}^{m} a_i y_i = 0. \quad (17)$$

The optimal hyperplane needs to meet

$$\alpha_i\{[x_i \cdot w + b]\}y_i - 1 = 0, \quad (18)$$

which is transformed into

$$\min W(\alpha) = \frac{1}{2}\sum_{i=1}^{m}\sum_{i=1}^{m} y_i y_j \alpha_i \alpha_j (x_i \cdot x_j) - \sum_{j=1}^{m} a_j,$$

$$\text{s.t} \begin{cases} a_i \geq 0, & i = 1, \ldots, m, \\ \sum_{i=1}^{m} \alpha_i y_i = 0. \end{cases} \quad (19)$$

The value of α^* is the optimal solution, and there are

$$\|w^*\| = 2w(\alpha^*) = \sum sv\alpha_i^* \alpha_j^* (x_i \cdot x_j) y_i y_j. \quad (20)$$

The optimal decomposition function is

$$f(x) = \text{sgn}\left(\sum_{sv}^{m} y_i \alpha_i^* (x_i \cdot x_j) + b^*\right). \quad (21)$$

For the linear inseparable case, the constraint conditions are transformed into

$$y_i[(w \cdot x_i) + b] - 1 + \varepsilon_i \geq 0, \quad i = 1, \ldots, m. \quad (22)$$

The objective function is transformed into

$$J(w, \varepsilon) = \frac{1}{2}(w \cdot w) + C\left(\sum_{i=1}^{m} \varepsilon_i\right). \quad (23)$$

In formula (15), C is the penalty factor, and the larger the value of $C > 0$, the greater the loss caused by the outlier to the objective function.

4. System Design

The smart elderly care system is a smart elderly care system designed using image processing technology, remote monitoring, and information service platforms, combined with the life needs of the elderly, and the overall structure is shown in Figure 4.

The system has many functions. When the elderly wear the ECG monitoring bracelet, the heart rate and heartbeat of the elderly will be generated and transmitted to the mobile

phones of family members and medical staff. So that they can pay more attention to the health problems of the elderly and spend the day of the elderly. The pictures generated by the system will also be sent to the family members' mobile phones. Based on the actual management and service needs of nursing homes, a nursing home personnel positioning management system is developed, which can realize daily basic information management, real-time positioning and tracking of the elderly, vital signs monitoring of the elderly, one-button alarm of dangerous situations of the elderly, and other functions. When the elderly have special circumstances, they can respond as soon as possible, and the system can truly realize the significance and requirements of intelligent old-age care and be responsible for the daily life and health status of the elderly. It is no longer the traditional way of providing care for the aged, and there is a shortage of nursing staff. There is an independent nursing home in this system.

4.1. Simulation Experiment Results.
We investigated, learned, and collected the daily activity trajectories of many elderly people, as shown in Figure 5.

The collected images are counted, and the statistics chart can view the real-time location of each person, as shown in Figure 6.

You can also query the activity track of a user in a day, as shown in Figures 7 and 8.

There are many recreational facilities in nursing homes. These recreational facilities bring great fun to the elderly and enrich the daily life of the elderly; some facilities are used by a large number of people, and the resource allocation is not reasonable. Nursing homes should strengthen the construction of the project and make elderly care services better.

4.2. Obtaining Experimental Results.
In order to check the accuracy of the system, we arranged the RFID equipment and collected the daily data of 4 students for a month. The ReaderIDs of the two places are represented by 1 and 2, respectively. The specific data are shown in Table 4.

The smart elderly care system generates many pieces of location data every day to prevent data memory from being confused. We should process these data to facilitate more accurate results. The data of each experimenter are accumulated in one place. Because there are 5 active areas, the experimental data are divided into 5 groups.

4.3. Analysis of Simulation Experiment Results.
The experiment process lasted for one month. We used the data of 4 students within 25 days as the training set and used the data of the last 5 days to test the robustness of the prediction model and the support vector machine prediction model. Figure 9 shows the prediction results. The results show that, in the five-day life of four people, there were only two errors, and the accuracy rate was as high as 90%. The specific results are shown in Figures 9 and 10.
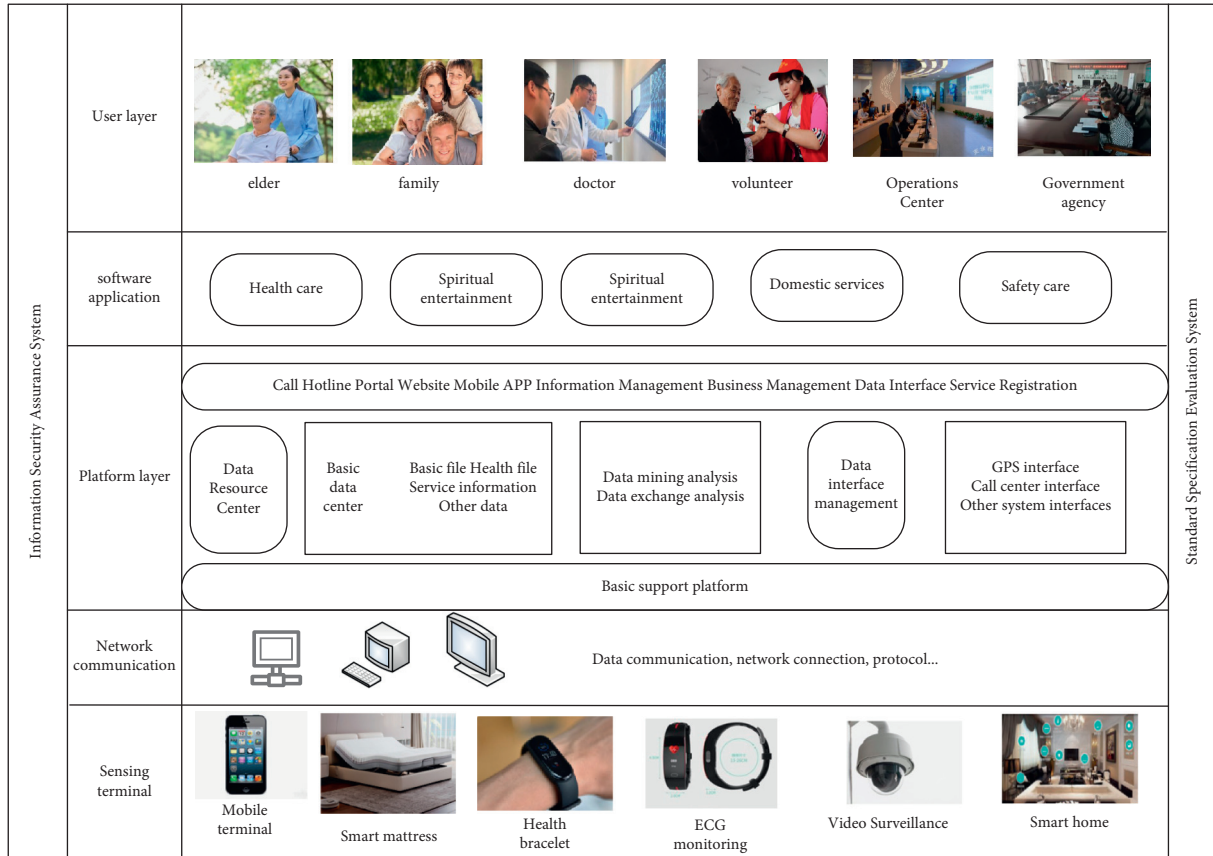
FIGURE 4: The overall structure of the smart elderly care system.

FIGURE 5: Activity trajectory diagram of the elderly.
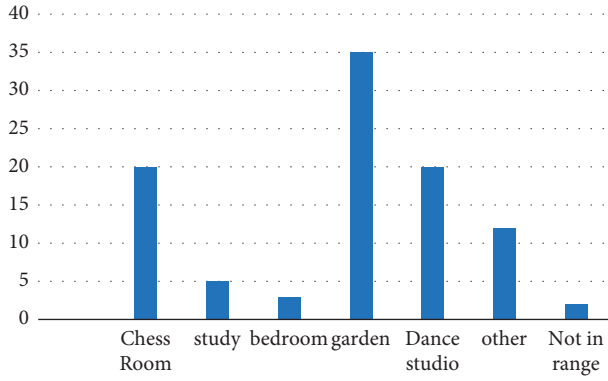
FIGURE 6: Statistics of the scope of activity.
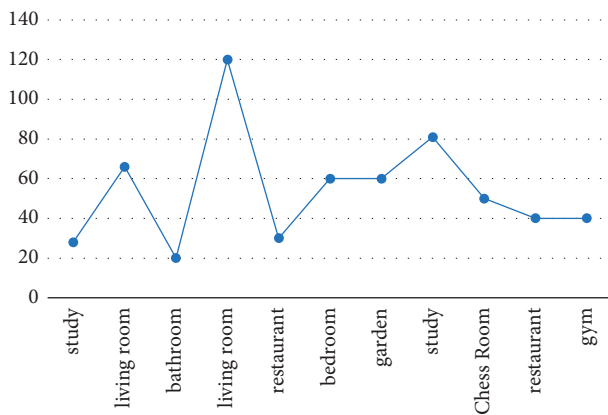
FIGURE 7: Statistics of the range of activities of the elderly.
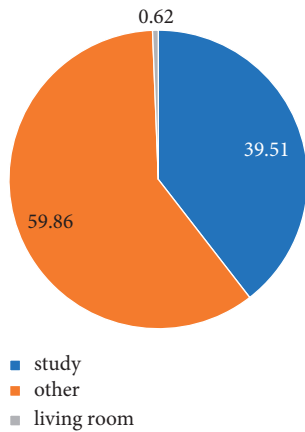
FIGURE 8: Statistics of the scope of activities for the elderly.

TABLE 4: Stored location data.

TagID	Time	Date	ReaderID	AntID	ReaderRssi	AntRssi
82027158	10:35:00	2020-09-16	2	2	−12	−7
82028058	10:35:00	2020-09-16	2	2	−16	−17
82027159	10:35:00	2020-09-16	1	1	−23	−20
82032425	10:35:00	2020-09-16	2	2	−21	−14
82027158	10:36:00	2020-09-16	2	2	−12	−7
82028058	10:36:00	2020-09-16	2	2	−16	−17
82027159	10:36:00	2020-09-16	1	1	−23	−20
82032425	10:36:00	2020-09-16	2	2	−21	−14
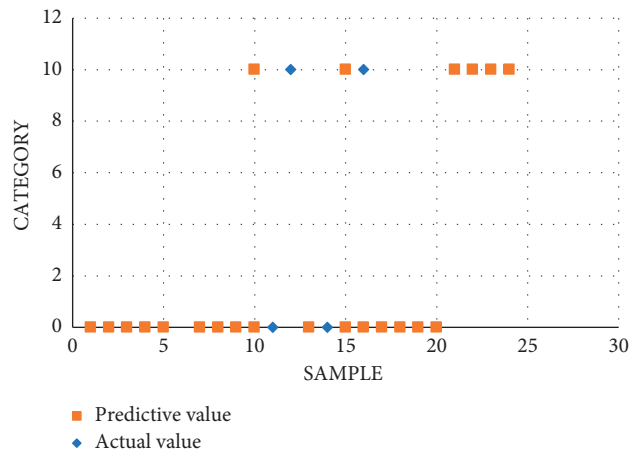
FIGURE 9: Prediction results of the extreme learning machine.

The predicted value and actual value of the extreme learning machine are 70 and 60, respectively, and the accuracy rate is as high as 90%.

4.4. Research Methods. We started a round of questionnaire surveys with 50 pairs of elderly people and those older than 50 years as the survey subjects and distributed questionnaires to elderly people in different regions, different ages, and educational backgrounds. After removing invalid questionnaires, there were 254 valid questionnaires remaining. The distribution results are shown in Table 5.

The willingness to use smart aged care products is shown in Figure 6.

According to the experimental results obtained in Figures 9 and 10, we can clearly observe the accuracy of prediction by drawing images with image processing technology. For the questionnaire survey obtained in Table 6, through image analysis, we can also clearly see the different percentages of statistical characteristics of different samples and other conditions.
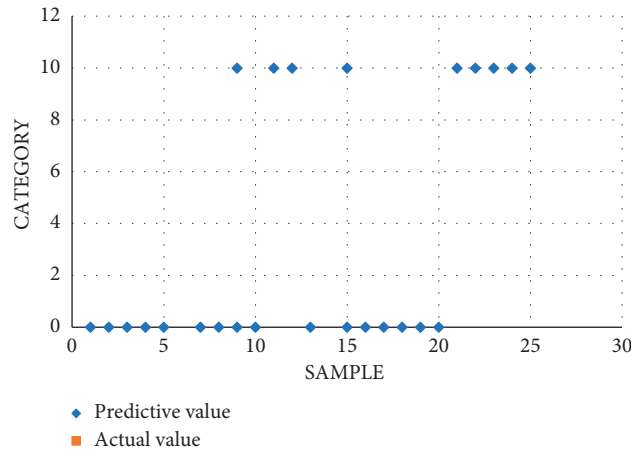
FIGURE 10: Prediction results of the support vector machine.

TABLE 5: Distribution results.

Sample statistical characteristics	Category	Frequency	Percentage%	Cumulative percentage%
Gender	Male	113	39.8	39.8
	Female	171	60.2	100.0
Age	50–69 years old	172	60...6	60.06
	70 years old and above	112	39.4	100.0
Education	Elementary school and below	144	50.7	50.7
	Junior high school	60	21.1	71.8
	High school	39	13.7	85.6
	Bachelor degree and above	41	14.4	100.0
Monthly income	<1000 yuan	85	30.0	30.0
	1000–2000 yuan	80	28.3	58.3
	2000–3000 yuan	53	18.7	77.0
	>3000 yuan	65	23.0	100.0

TABLE 6: Regression analysis table.

D V	I V	Nonstandardized coefficient		Standard coefficient	t	Sig.	VIF	Durbin–Watson	R^2
		B	Standard error						
U M	Constant	0.424	0.151		2.801	0.005			
		0.293**	0.048	0.291	6.104	0.000	1.583		
		0.402**	0.046	0.433	80848	0.000	1.663	1.758	0.597
		0.210**	0.043	0.211	4.879	0.000	1.299		

Through the results of the questionnaire survey, we can find that the degree of education, education level, and salary of the elderly are related to the willingness to use smart pension products. In general, the elderly with high wages also have a higher education level and they will use their own knowledge reserve ability to quickly learn how to use smart products. Therefore, in the process of product development and promotion, we should focus on the elderly who are slow to accept new things.

5. Conclusion

In view of the difficulty of providing care for the aged in cities, this paper puts forward an RTID positioning system and APP client, which can ensure the privacy of the elderly.

Through image processing technology, it analyzes the old-age situation in each activity center and the living conditions of the elderly in different environments. It can reasonably plan urban old-age facilities and activity centers, so as to improve the satisfaction of elderly services.

Data Availability

The experimental data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest regarding this work.

Acknowledgments

References

[1] S.-H. Lee, "A scenario-based approach in smart agriculture services," *Journal of the Korea Institute of Information and Communication Engineering*, vol. 19, no. 7, pp. 1705–1710, 2015.

[2] M.-P. Gagnon, M. Desmartis, M. T. Dipankui, J. Gagnon, and M. St-Pierre, "Alternatives to seclusion and restraint in psychiatry and in long-term care facilities for the elderly: perspectives of service users and family members," *The Patient - Patient-Centered Outcomes Research*, vol. 6, no. 4, pp. 269–280, 2013.

[3] M. Kim and D. Seo, "Smart senior job search: the elderly-oriented services for job searching with the spatial information," *Journal of Korea Multimedia Society*, vol. 19, no. 8, pp. 1433–1443, 2016.

[4] R.-H. Kim and G. Cho, "Effectiveness of the smart healthcare glove system for elderly persons with hypertension," *Human Factors and Ergonomics in Manufacturing & Service Industries*, vol. 23, no. 3, pp. 198–212, 2013.

[5] S. Tokunaga, H. Horiuchi, K. Tamamizu, S. Saiki, M. Nakamura, and K. Yasuda, "Deploying service integration agent for personalized smart elderly care," in *Proceedings of the IEEE/ACIS 15th International Conference on Computer and Information Science (ICIS)*, pp. 1–6, Okayama, Japan, June 2016.

[6] E. J. Lee and S. J. Park, "A framework of smart-home service for elderly biophilic experience," *Sustainability*, vol. 12, no. 12, pp. 1–10, 2020.

[7] M. Yang, H. Huang, H. Yuan, and Q. Sun, "Interaction design of products for the elderly in smart home under the mode of medical care and pension," Edited by J. Zhou and G. Salvendy, Eds., in *Proceedings of the Human Aspects of IT for the Aged Population. Healthy and Active Aging. ITAP 2016*, vol. 9755, pp. 145–156, Toronto, Canada, July 2016.

[8] L. Yao, B. Benatallah, X. Wang, N Khoi, and Q Lu, "Context as a service: realizing Internet of things-aware processes for the independent living of the elderly," *Springer International Publishing*, vol. 12, no. 5, pp. 22–52, 2016.

[9] P. Brune and R. Rockmann, "Towards an application architecture for A smart online service network platform for the elderly," *Procedia Computer Science*, vol. 20, no. 12, pp. 22–52, 2017.

[10] P. Pirzada, N. White, and A. Wilde, "Sensors in Smart Homes for Independent Living of elderly," in *Proceedings of the 2018 5th International Multi-Topic ICT Conference (IMTIC)*, pp. 1–8, Jamshoro, Pakistan, April 27 2018.

[11] H. Lee and S. Lee, "Smart services of the livingroom based on behavior for the habitability of the elderly," *Design Convergence Study*, vol. 12, no. 3, pp. 45–85, 2013.

[12] H. Lim and H. Lee, "The smart services of the hazard management in the kitchen for the elderly," *Journal of the Architectural Institite of Korea Planning & Design*, vol. 29, no. 6, pp. 139–146, 2013.

[13] Y. H. Wang, C. G. Chung, C. C. Lin, and C. M. Lin, "The study of the electrocardiography monitoring for the elderly based on smart clothes," in *Proceedings of the 2018 Eighth International Conference on Information Science and Technology (ICIST)*, pp. 478–482, IEEE, Cordoba, Spain, March 2018.

Research Article

Subway Tunnel Construction Settlement Analysis Based on the Combination of Numerical Simulation and Neural Network

Qiangqiang Ma [iD], Wentao Li, and Yongjun Zhang [iD]

Qingdao University of Technology, Qingdao 266033, Shandong, China

Correspondence should be addressed to Qiangqiang Ma; 230218294@seu.edu.cn

Under the background of rapid economic development and the continuous expansion of city size, people's travel is deeply troubled by the problem of urban traffic congestion. Subway is an effective way to relieve traffic pressure and plays an important role in its use. In the process of building the city subway, the excavation of the underground tunnel is the most critical. However, the excavation of the tunnel will inevitably disturb the nearby soil, change the stratum stress conditions, and make the stress distribution uneven. If the surface settlement is too large, it will not only affect the normal construction of the tunnel but also cause damage to the surrounding buildings (structures), roads, underground pipelines, and so on, resulting in very serious malignant consequences. In this study, Cuobuling Station is taken as a case study. First, the construction status of the subway tunnel in the station is analyzed, and then the monitoring results are analyzed. According to the cross-section settlement law, the numerical simulation and neural network are used to build a model, calculate the numerical simulation results, and carry out a risk assessment of regression model. Finally, combined with the tunnel construction situation of the station, according to the risk assessment results, the concrete measures are put forward to deal with the subway tunnel construction settlement problem.

1. Introduction

As an effective way to relieve traffic pressure, the subway plays an important role. It is foreseeable that the construction of urban subways in the future will usher in new development opportunities. The well-developed shield construction will still disturb the soil layer, and different levels of ground displacement and surface subsidence will occur [1]. Once this phenomenon exceeds a certain range, the safety of adjacent buildings and underground pipelines will not be guaranteed, and geotechnical environmental engineering problems will be caused one after another [2]. In order to reduce the disturbance of surrounding soil and reduce the adverse impact on adjacent buildings (structures), many experts and scholars are committed to finding a method to control surface settlement to a great extent [3]. Therefore, it is very important to study the evolution law and control method of surface subsidence, which is one of the most critical research issues [4, 5]. Based on the Cuobuling Station project of Qingdao Metro Line 4, the evolution law of ground subsidence caused by subway shield construction was analyzed and studied in this paper. And a series of measures are summarized to control the ground subsidence, so as to ensure the safety of Qingdao metro shield construction, but also to provide a reference for the ground subsidence caused by the construction of similar areas in the future.

2. Project Summary

Cuobuling Station is a nine-seat station of Qingdao Metro Line 4, which interchanges with the Cuobuling Station of Line 3. The strata involved in the temporary construction of the shaft and the cross passage are mainly the artificial filling soil of the Quaternary-Holocene, alluvial and diluvial deposits of the Quaternary Holocene, and late Yanshanian granite. Controlled by the Cathaysia tectonic system, the regional main tectonic trajectory in the Qingdao area is a

group of NE-trending faults distributed diagonally, with the characteristics of high inclination, overall inclination to the northwest, and local counterinclination. From west to east, there are the Cangkou fault, the Pishikou-Fushan fault, and the Wanggezhuang-Shandongtou fault. The main types of groundwater in the field are Quaternary pore water and bedrock fissure water. The Quaternary pore water mainly occurs in the filling soil and the ninth layer of coarse sand, which is rich in water and belongs to medium to strong permeable layer. The bedrock fissure water occurs in layers and bands in the strong wind zone of the bedrock and the densely developed zone of fissures. Due to the uneven development of fissures, its water richness is uneven.

3. Monitoring Results

3.1. Testing Arrangement. In Cuobuling Station, a total of seven survey lines are arranged on the road surface, including five-line survey and two-point survey. The surface detection line layout diagram is shown in Figure 1:

Within the detection range, the electromagnetic wave is distributed regionally, and the distribution uniformity of the energy group is poor. There are obvious diffraction and refraction phenomena in the local electromagnetic wave. There is a fault in the in-phase axis, and the reflected wave energy in local areas is strong, and the amplitude is high. In view of the complex structure of the overlying soil layer and numerous pipelines under the road surface, there is a great interference to the identification of radar Atlas. It is expected that the local area rock strata are broken, and there are weak structures and bedding structural planes. Surface detection results map is shown in Figure 2.

3.2. Cross-Section Settlement Law. Three sections are selected, and five groups of monitoring data are taken from each section [6]. During the whole process of the shield machine passing through each section (including before, during, and after the shield machine passing through), the field surface subsidence monitoring is carried out. Because the number of rings advanced by the shield machine is different in different monitoring periods, the interpolation method is adopted locally for data processing when the monitoring data is counted.

3.3. Analysis Method. With the development of computer technology, numerical analysis methods have been realized; the application of numerical analysis methods in geotechnical engineering began in the 1960s [7], among which the most widely used methods are the finite element method (FEM) and finite difference method (FDM).

3.3.1. Finite Element Method (FEM). The finite element method establishes the relationship between force and displacement based on the basic mechanics principle of continuum and calculates the displacement and stress of unknown nodes according to the known stress and strain of each division element [8].
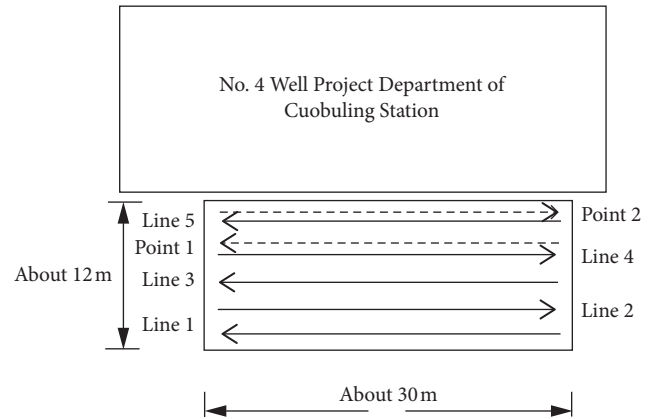
Figure 1: Surface detection line layout diagram.

3.3.2. Finite Difference Method (FDM). The solution of a given initial and boundary value problem is the finite difference method, an earlier numerical analysis method. With the rapid development of computer technology, the finite difference method has shown certain advantages in the numerical analysis due to its unique calculation format and calculation flow [9].

3.4. Numerical Simulation Software. At present, the common numerical analysis software includes ANSYS, MIDAS/GTS, ABAQUS, PLAXIS, and so on. Different numerical simulation software have their own advantages and disadvantages and are suitable for different research fields.

In the civil construction industry, ANSYS simulation software can accurately deal with and simulate the situation of stressed nodes. Because of its own characteristics, ANSYS software can be used for physical simulation analysis in many disciplines, such as fluid mechanics, geotechnical engineering, structural engineering, and other disciplines.

The MIDAS/GTS provides a reference for the research and calculation of underground engineering. It has advanced representation, and it can provide good solutions in tunnel construction or other more complex environments [10]. The ABAQUS is not only well used in simple linear domain simulations, but also for relatively difficult nonlinear simulations. The simulation in the linear domain can also be solved a lot.

In this paper, PLAXIS 3D numerical analysis software is used to carry out numerical simulations of the shield construction process. Now the PLAXIS software is briefly introduced, with emphasis on its calculation and analysis function and modeling process. PLAXIS has powerful calculation and analysis functions, stable calculation, and accurate results [11]. The software CCP includes three modules: main module, seepage module, and dynamic module, which can perform numerical analysis of various conditions such as consolidation, fluid-solid coupling, plasticity, and dynamics.

4. Model Building

4.1. Shield Excavation Model. The lining of shield tunnels is usually composed of precast concrete segments, and the precast concrete segments are connected by bolts during
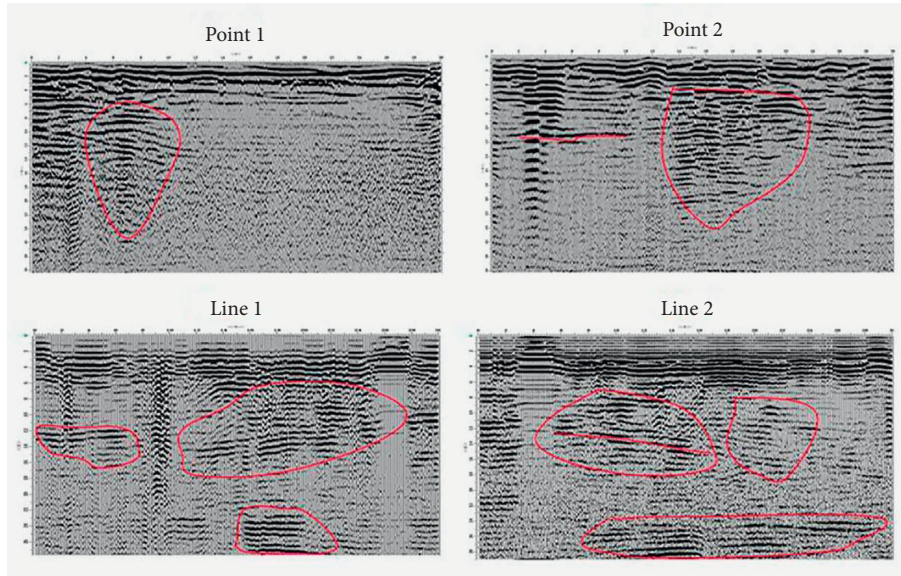
FIGURE 2: Surface detection results map.

shield tunneling to form the tunnel lining [12]. The simulated construction process is divided into several construction stages according to the length of each ring lining segment, which is generally about 1.5 m long. Considering that the overlaying soil layer above the tunnel is different at different sections, numerical simulation analysis is carried out on the three selected representative sections to calculate the settlement results. The detailed modeling process is as follows.

4.1.1. Geometric Model Building and Parameter Input. Shield construction is mainly a construction method in which the shield machine is used as a tunnel excavation tool. In order to prevent the collapse of the tunnel surrounding the rock, the shield machine is pushed forward while laying tiles and lining. Before excavation, the soil is excavated with a cutting device. The excavated machine is out of the hole, the jack is pressed on the top, and precast concrete segments are assembled to form a tunnel structure. In the numerical simulation of the tunnel excavation construction process, the project is often simplified appropriately. According to the geological investigation data and the study of boundary conditions, two adjacent borehole intervals are selected as the study area of the geometric model [13]. The length of the model is 80 m (in the direction of the X-axis); the width is 20 m (in the direction of the Y-axis); and the thickness is 30 m (in the direction of the Z-axis). The buried depth of the tunnel is 18 m; the outer diameter of the shield is 6.3 m; the inner diameter of the shield is 5.7 m; each segment ring is 1.5 m; and the wall thickness is 0.3 m. The physical and mechanical parameters such as cohesion, friction angle, elastic modulus, and Poisson's ratio of the overburden soil layer are simulated by solid elements. Reference table of section soil material is shown in Table 1.

Model plate unit material properties are shown in Table 2.

Numerical simulation of three-dimensional model meshing is shown in Figure 3.

4.1.2. Defining Structural Units and Creating Advanced Excavation Segments. Shield excavation, support, and other relevant contents are defined under the "structural mode," and the tunnel shield machine (TBM) is used to carry out the excavation [14]. The length of the shield is 9 m, and the diameter is 6.3 m. The tunnel inner and outer diameter profile is established, and the inner and outer diameter profile surface of the left half tunnel lining is established to form the left half tunnel entity. In order to simulate the shield machine's propulsion process, it is divided into five stages, including the initial position of the shield and four stages of propulsion, with each stage advancing 10 m and each advancing 1.5 m, divided into seven times. The initial position of the section shield is shown in Figure 4.

(1) Creating the Tunnel Face to Balance the Pressure and Jack Thrust. The equilibrium pressure of the tunnel face is bentonite pressure, which increases linearly with depth. At the same time, in the process of tunnel excavation, the shield needs to push itself forward and separate from the completed lining. This process is realized by hydraulic jacks. Therefore, the force exerted by the jacks on the tunnel lining should be considered in the model.

(2) Defining Shield, Surface Shrinkage, Grouting Pressure, and Soil-Structure Interaction. The "Create Plate" of the outer cylinder of the model is selected to specify the plate element for the surface representing the shield machine. Each stage is only 9 m long in the active state to simulate the shield machine [15]. Once again, the "Creation Surface Shrink" of the outer cylinder is selected, and the distribution type and value of the contraction will be defined in the calculation stage. The grouting pressure was defined, and the "original surface load" of the outer cylinder was selected to simulate the grouting pressure. Finally, the interaction between soil and structure is simulated around the tunnel.

TABLE 1: Reference table of section soil material.

Tab	Parameter	Symbol	Concrete	Marl layer	Clay layer	Plain fill layer	Unit
General	Material model	—	Linear elasticity	Moore-Cullen	Moore-Cullen	Moore-Cullen	—
	Drainage type	—	Nonporous	Drain	Drain	Drain	—
	Natural severity	γ_{unsat}	27	17	16	17	kN/m³
	Saturation severity	γ_{sat}	—	20	18	20	kN/m³
Parameter	Elastic modulus	E_{ref}	3.10E7	7.50E4	1.00E4	1.30E4	kN/m³
	Poisson's ratio	V	—	0.10	0.30	0.35	0.30
	Cohesion	C'	—	35	15	10	kN/m³
	Friction angle	ρ	—	31	25	21	(°)
	Dilatancy angle	ψ	—	0	0	0	(°)
Interface	Interface strength	R_{inter}	Rigidity	Rigidity	Rigidity	Rigidity	—
Initial conditions	Initial horizontal stress coefficient	K_0	Automatic	Automatic	Automatic	Automatic	—

TABLE 2: Model plate unit material properties.

Name of the material	Natural density (g/cm³)	Compression modulus (MPa)	Angle of internal friction (°)	Cohesion (kPa)	Poisson's ratio
Shield shell (70 mm thickness)	7.72	210,000	50.00	20,000	0.28
C50 concrete segment (300 mm thickness)	2.49	35,000	62.63	4,241	0.16
Shield tail void medium (grouting material)	1.63	0.94	10.00	1.12	0.45
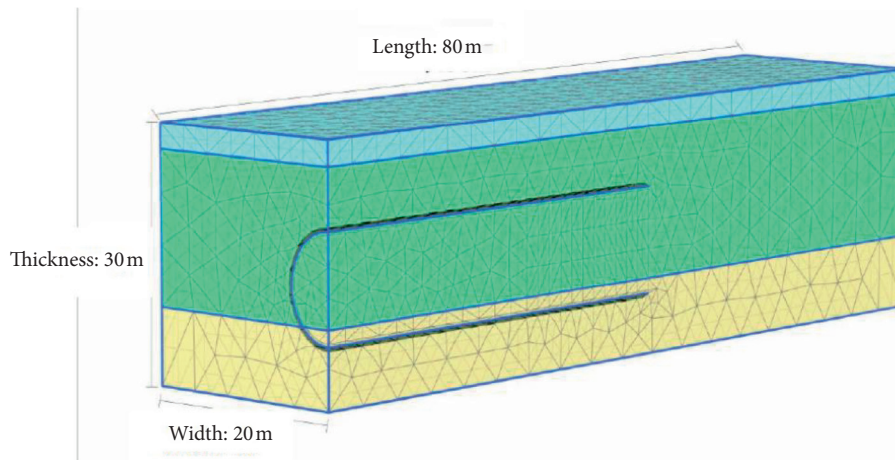
FIGURE 3: Numerical simulation of three-dimensional model meshing.

4.2. Numerical Results. The width of each ring of segments is 1.5 m. For the convenience of numerical simulation calculation, each shield tunneling is carried out in a group of seven-ring segments with 1.5 m of each advance. When the distance of the shield tunneling machine from the monitoring surface is −10 m, 0 m, 10 m, 20 m, and 30 m, the research and analysis are carried out. The settlement amount of the study section is counted, and the cloud map of a complete calculation stage is shown as follows.

(1) The Shield Machine Advances to a Position of −10 m from the Monitoring Surface. The shield machine advances to a position of −10 m from the monitoring surface, which is equivalent to the part of the soil before "freezing." The excavation releases Earth pressure, and a small amount of settlement is generated on the monitoring section. The

maximum stratum uplift value is 2.814 mm, which occurs at the grouting behind the bottom wall of the tunnel. Vertical cumulative subsidence cloud image at −10 m from section YDK13 + 001 is shown in Figure 5.

4.2.1. Simulating Ring Wall Grouting for Excavated Soil. In the second stage, due to soil excavation, a small part of the monitoring section has subsided, so it is necessary to simulate ring wall grouting in real-time. Cloud image of ring wall grouting settlement before −10 m from section YDK13 + 001 is shown in Figure 6.

4.2.2. Shield Machine Advances One Step Forward. A new stage is added to hide the surrounding soil of the tunnel, making it easy to select TBM, lining, surface load, and
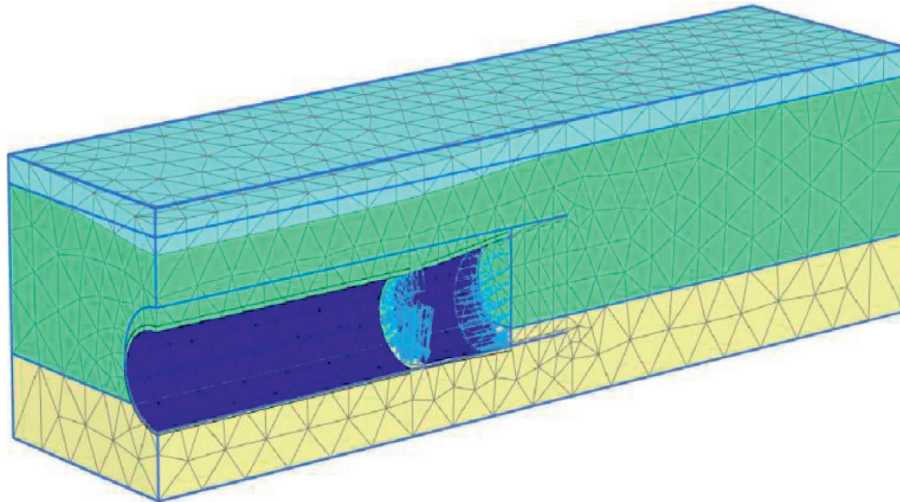
FIGURE 4: Initial position of section shield.
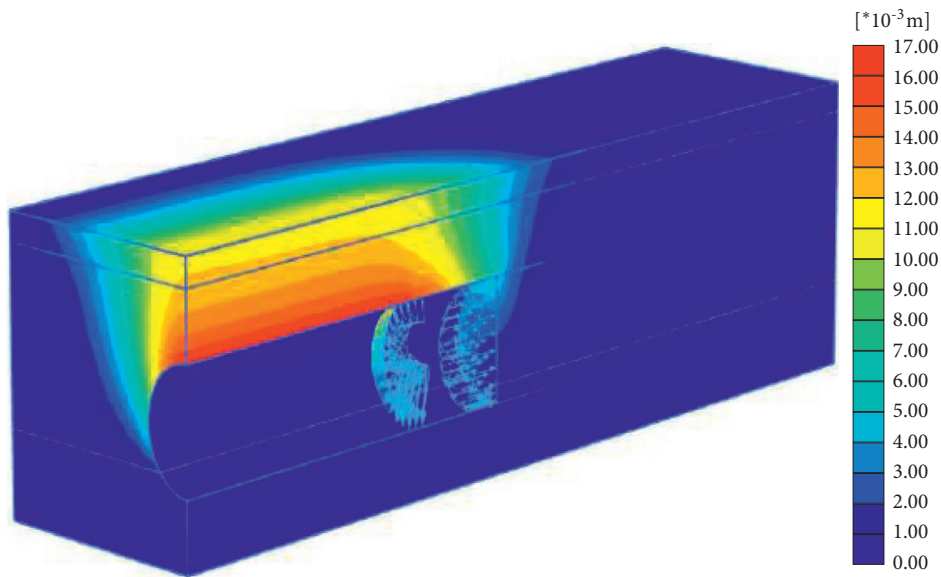
Total displacement | u |
Maximum=0.007 m (Unit 1896 is at node 1224)

FIGURE 5: Vertical cumulative subsidence cloud image at −10 m from section YDK13 + 001.

surface shrinkage from the outside and inside of the tunnel. The load between −10 m and −8.5 m is frozen; its forward interface is activated; and the TBM is used to advance 1.5 m. Sedimentation cloud map at −8.5 m from section YDK13 + 001 is shown in Figure 7.

4.2.3. The Shield Propulsion Process from −8.5 m to −6 m Is Simulated. In the fourth stage, TBM was advanced from −8.5 m to −6 m, and the simulation process is basically the same as that in the previous stage, except that the whole process is advanced along the *X*-axis by 1.5 m. Grouting pressure from −10 m to −8.5 m remains unchanged; the jack thrust at −10 m is removed; the grouting pressure from −10 m to −8.5 m needs to be changed to the lining; and the

jack thrust is applied; Only 6 m–8.5 m is required from the tail of TBM to the grouting pressure Settlement cloud diagram at −6.5 m from YDK13 + 001 section is shown in Figure 8

4.2.4. The Shield Reaches the Monitoring Section. The shield machine advances cyclically, advancing forward 1.5 m each time, and When the TBM shield machine advances to the ydk13 + 001 section, the numerical simulation cloud diagram is shown in Figure 9.

4.3. Regression Model Risk Assessment. The artificial neural network evaluation model for large-section tunnels uses a

Total displacement | u |
Maximum=0.00830 m (Unit 1532 is at node 5873)
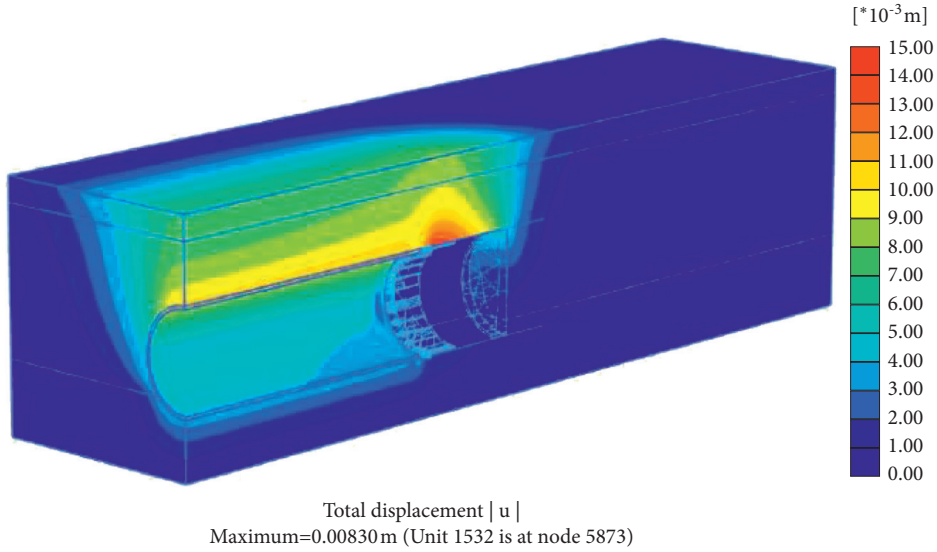
FIGURE 6: Cloud image of ring wall grouting settlement before −10 m from section YDK13 + 001.

Total displacement | u |
Maximum=0.00870 m (Unit 1475 is at node 5764)
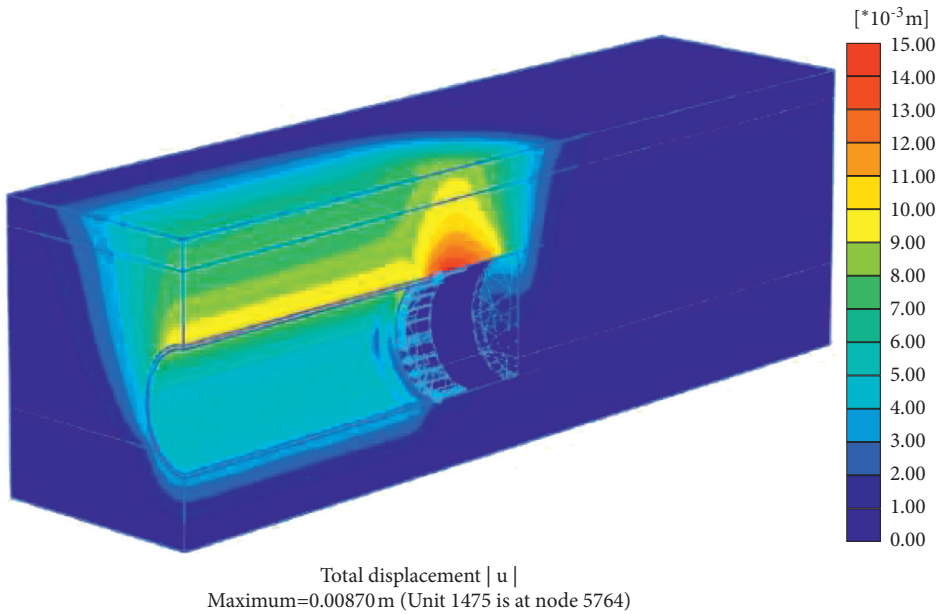
FIGURE 7: Sedimentation cloud map at −8.5 m from section YDK13 + 001.

three-layer neural network. The geological conditions and construction parameters of the surrounding rock of the tunnel are used as the input layer, and the evaluation indicators of the dome subsidence, ground settlement, structural force, and the extent of the plastic zone are used as the output layer [16]. Suppose the input neuron is h, the hidden layer neuron is i, and the output layer neuron is j using nh, ni, and nj to denote the number of nodes in the input layer, hidden layer, and output layer, θ_i and θ_j are thresholds of hidden layer node and output layer node, respectively. The input sample is as follows:

$$\left\{X_{k,h} | k = 1, 2, ..., nk; h = 1, 2, ..., nk; \right\}. \tag{1}$$

The output sample is as follows:

$$\left\{d_{k,j} | k = 1, 2, ..., nk; h = 1, 2, ..., nj; \right\}. \tag{2}$$

In this evaluation, the median value nk is the number of nodes in the input layer, which is 75, and the number of nodes in the output layer nh is 14. The larger the number of hidden layer nodes ni, the higher the fitting accuracy of the network, but the lower the generalization ability of the network and the longer the training time. Therefore, under the condition of satisfying accuracy, the value ni should be as small as possible. It is taken as 15 in this assessment.

Functions used:

$$y = (x - \min)(\max - \min). \tag{3}$$
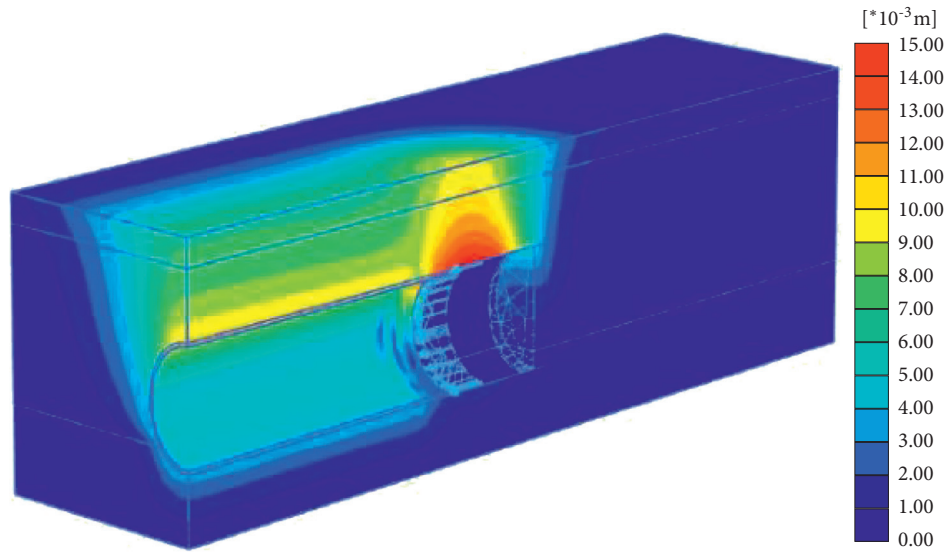
Total displacement | u |
Maximum=0.00960 m (Unit 1872 is at node 6549)

FIGURE 8: Settlement cloud diagram at −6.5 m from YDK13 + 001 section.
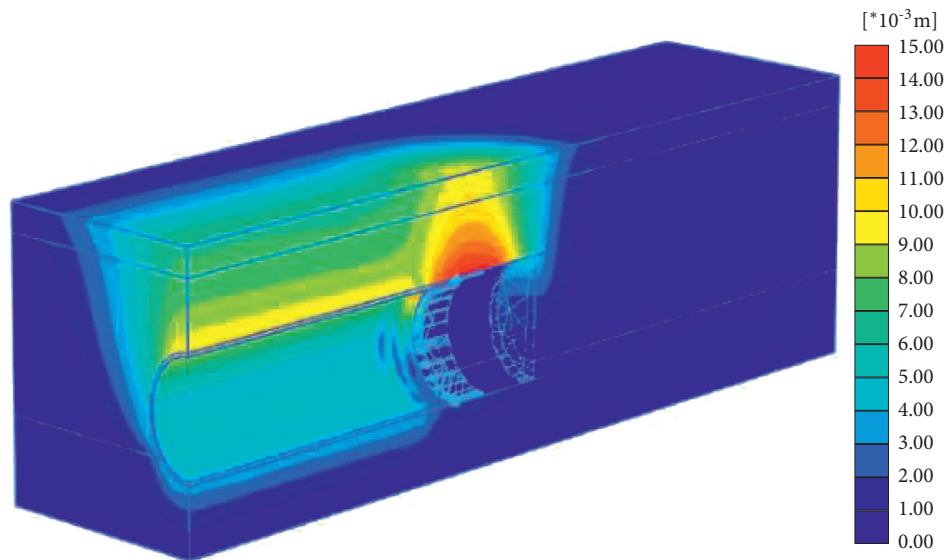
Total displacement | u |
Maximum=0.00960 m (Unit 1872 is at node 6549)

FIGURE 9: Deformation cloud image when TBM shield reaches YDK13 + 001 section.

The input and output samples are normalized, where x represents the original data and represents the normalized data. The 1 in the normalization process is processed into 0.99, and the 0 in the normalization process is processed into 0.11.

In order to reduce the learning error and make the evaluation result more accurate, the four evaluation indexes of the output layer are trained separately. All the samples corresponding to the four evaluation indicators are learned successively, and their weights and thresholds are obtained. The evaluation result is substituted into the input layer of the artificial neural network model. After calculation, the result of the output layer is the evaluation index value of the current tunnel. The evaluation results are shown in Table 3.

5. Responses

5.1. Management Process Optimization.
Based on a large amount of engineering practice experience in the past, combined with many engineering accident lessons, and consulting a large number of literature materials, according to the specific engineering characteristics of Cuobuling

TABLE 3: The evaluation results.

Projects	Risk degree (%)
Water burst	71.21
Surrounding rock deformation	88.96
Plastic zone	99.12
Landslide	83.42
Supporting structure	82.47

Station, the management process optimization measures are formulated.

5.1.1. Advanced Geological Forecast.

The survey of subway tunnel is very difficult. At the same time, due to the equipment and technical reasons, the geological conditions considered in the design are often different from reality [17]. Therefore, it is necessary to use an advanced geological forecast system and geological radar to forecast the geological condition of the tunnel face.

5.1.2. Monitoring Measurement.

The tunnel vault subsidence, tunnel circumferential convergence and deformation of the support structure, and water inflow are monitored and measured. According to the disclosed surrounding rock information, a geological sketch of the palm surface is done [18]. The monitoring data and construction information are analyzed in a timely manner and fed back to the technical leader and site supervisor in time.

5.1.3. Dynamic Risk Management.

In the process of excavation, a dynamic information management system with the advanced geological forecast, risk assessment, and monitoring measurement as the core should be developed to timely deal with advanced geological forecast results and feed back to the risk assessment system. And the revised construction method should be reassessed according to whether the risk is reduced. The dynamic management diagram is shown in Figure 10.

5.2. Grouting Process Control.

Generally speaking, there are two aspects to ensure the effect of field grouting. Firstly, it should strictly control the grouting speed. The filling speed of the slurry should be consistent with the excavation speed of the shield machine as far as possible, so as to minimize the generation of shield tail voids, so as to avoid the settlement caused by shield tail voids. When the grouting speed is too fast, it is necessary to pay attention to the leakage phenomenon of shield tail and damage the construction effect. Therefore, the specific grouting speed should be determined according to the actual construction situation. Secondly, the grouting time is accurately grasped, through the analysis of the above surface subsidence reasons; one of the important reasons for the subsidence is not timely grouting. In the construction of a shield, it is necessary to grasp the timing of grouting. Generally, the best time for grouting is the time when the segment is off the tail of the shield.
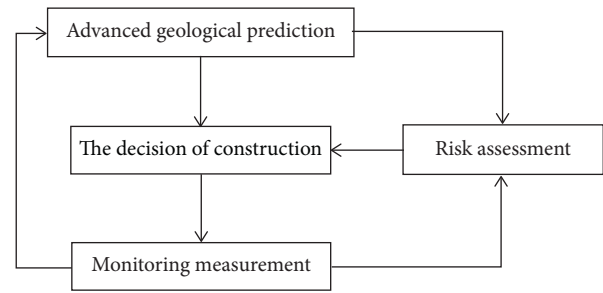
FIGURE 10: Dynamic management diagram.

5.3. Setting of Heading Parameters.

The three parameters of formation earth pressure, groundwater pressure, and preliminary pressure are the earth pressure parameters that need to be controlled during shield construction. Because there are too many factors that cause ground settlement, it is impossible to accurately calculate the specific value of the settlement, so the shield parameters should be adjusted while digging. There are many factors to consider when calculating the earth pressure. The construction of shallow tunnels should follow their respective formulas and make selections based on past experience.

6. Conclusion

By increasing the grouting amount, the soil deformation caused by shield construction can be well controlled. The influence range of grouting pressure is 10 M smaller than that of the tail, and it has an obvious effect on reducing the instantaneous settlement of the tail. In addition, in the process of shield propulsion, the influence of shield thrust force and cutter head torque on soil deformation is small, and the change of cutter head torque has no obvious relationship with the change of vertical displacement. The influence of working face pressure is mainly concentrated around the shield incision and has little influence on the shield tail, which is the main factor affecting the soil deformation in front of the shield. It should be pointed out that the soil deformation caused by shield tunneling is a comprehensive expression of the interaction of various influencing factors, and the relationship between vertical displacement and each influencing factor cannot be fully reflected by the relationship between them. According to the measured settlement data and numerical simulation analysis, the influence factors of shield tunnel construction on ground settlement are obtained, which provides a reference for further study of settlement control of shield tunnel construction. In order to deal with settlement caused by metro tunnel construction, the management process should be optimized; grouting technology should be controlled; and tunnel parameters should be set. Taking the ground settlement caused by the shield construction of Cuobuling Station as an example, the field settlement is monitored and analyzed, and the influence of different strata is considered. In the process of tunnel excavation, the shield machine damages the earth pressure balance of the original soil, which leads to the settlement of the top part of the tunnel

and the cracking of the bottom of the tunnel. PLAXIS 3D software is used to construct the shielding model, and the calculated results are close to the measured data, indicating that the ground subsidence and subsidence trend caused by the shielding constant can be accurately predicted by using PLAXIS 3D software for finite element analysis.

Data Availability

The experimental data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest regarding this work.

References

[1] R. L. Sousa and H. H. Einstein, "Risk analysis during tunnel construction using Bayesian Networks: Porto Metro case study," *Tunnelling and Underground Space Technology incorporating Trenchless Technology Research*, vol. 27, no. 1, pp. 86–100, 2011.

[2] G. Atzl, G. Ullmann, and M. Schmidt, "Planning of tunnel excavation for the Vienna underground - construction Lot U1/9/Planung der Tunnelvortriebe für die Wiener U-Bahn - b," *Geomechanics and Tunnelling*, vol. 8, no. 1, pp. 74–82, 2015.

[3] Z. Zhao, "Study on differential grey neural net-AR model based on wavelet decomposition and its application in subway tunnel settlement prediction," *Bulletin of Surveying and Mapping*, vol. 34, no. 1, pp. 104–108, 2020.

[4] K. Pan, L. Tian, and C. Li, "Analysis on the prediction effect of dynamic grey time-series neural network combined model in subway operation period," *Site Investigation Science and Technology*, vol. 100, no. 2, pp. 49–52, 2019.

[5] Y.-H. Kim, "Progress in civil, architectural and hydraulic engineering IV," *American Journal of Anatomy*, vol. 10, no. 1201, pp. 123–131, 2015.

[6] X. Liu and F. Sun, "Comparison and selection of tunnel construction schemes based on numerical simulation," *Journal of Civil Engineering and Management*, vol. 35, no. 3, pp. 1–6, 2018.

[7] C. Di Perna, A. Carbonari, R. Ansuini, and M. Casals, "Empirical approach for real-time estimation of air flow rates in a subway station," *Tunnelling and Underground Space Technology*, vol. 42, no. 5, pp. 25–39, 2014.

[8] H. Liu and D. Wang, "Study on subsidence change law of underground tunnel underpass existing buildings," *Research on Urban Rail Transit*, vol. 22, no. 6, pp. 47–51, 2019.

[9] A. Scheller, "The second S-Bahn trunk line in Munich/Die zweite S-Bahn-Stammstrecke München," *Geomechanics and Tunnelling*, vol. 8, no. 2, pp. 115–128, 2015.

[10] G. Papastamos, S. Stiros, V. Saltogianni, and V. Kontogianni, "3-D strong tilting observed in tall, isolated brick chimneys during the excavation of the Athens Metro," *Applied Geomatics*, vol. 7, no. 2, pp. 1–7, 2014.

[11] A. L. Costa, R. L. Sousa, and H. H. Einstein, "Probabilistic 3D alignment optimization of underground transport infrastructure integrating GIS-based subsurface characterization," *Tunnelling and Underground Space Technology*, vol. 72, no. 2, pp. 233–241, 2017.

[12] K. Rabensteiner and S. Notarianni, "Cityringen - extension of the Copenhagen metro/Cityringen - Ausbau der Metro Kopenhagen," *Geomechanics and Tunnelling*, vol. 6, no. 5, pp. 542–550, 2013.

[13] S. Chen, T. Ren, and P. Deng, "Tunnel settlement prediction based on artificial bee colony optimization wavelet neural network," *Modern Tunneling Technology*, vol. 52, no. 4, pp. 56–61, 2019.

[14] P. Qi, "Analysis and study on the influence of additional load on the construction of transfer structure foundation of upper cross metro tunnel," *Modern Tunneling Technology*, vol. 390, no. 1, pp. 105–110, 2020.

[15] H. Chakeri, Y. Ozcelik, and B. Unver, "Investigation of ground surface settlement in twin tunnels driven with EPBM in urban area," *Arabian Journal of Geosciences*, vol. 8, no. 9, pp. 7655–7666, 2015.

[16] S. Kim and J. Heo, "Development of 3D underground cadastral data model in Korea: based on land administration domain model," *Land Use Policy*, vol. 60, pp. 123–138, 2017.

[17] X. Wang, H. Jiang, and K. Zhu, "Ground subsidence law of urban metro tunnel construction process based on earth pressure shield," *Earth Science*, vol. 44, no. 12, pp. 4293–4298, 2019.

[18] B. D. Doelder, R. Berkelaar, C. Luijten, and V. Thumann, "Practical integration of risk management and monitoring during rebuilding of subway station CS Rotterdam," *Iabse Symposium Report*, vol. 99, no. 3, pp. 606-607, 2013.

Research Article

Target Tracking Algorithm of Basketball Video Based on Improved Grey Neural Network

You Jun Wang [1,2] **and Guo Huang** ⓘ [1,2]

[1] *Jiaozuo University, Department of Taichi and Martial Arts, Jiaozuo 454002, China*
[2] *Zhejiang Wanli University, Ningbo 315100, Zhejiang, China*

Correspondence should be addressed to Guo Huang; 2002900057@zwu.edu.cn

Considering the limitation of tracking range on sports video target tracking in basketball games, there are some problems, such as poor tracking effect, low accuracy, low anti-interference ability, and being time-consuming. Therefore, this study proposes a target tracking algorithm of basketball video based on improved grey neural network. According to the pixel grey difference of the target image in basketball video, this study applies the adaptive threshold algorithm in order to segment the target image of basketball video and obtain the target area of basketball video. This algorithm can normalize the grey level of the target area, build the generating sequence of the target area, and collect the target data of the basketball video. It obtains the feature output matrix of basketball video target based on the geometric dispersion of the target image and extracts the key feature points of basketball video target by single frame visual difference analysis. In addition, it makes use of the improved grey neural network to track and locate the feature points of basketball video target and reconstruct the basketball video target image with superresolution to realize basketball video target tracking. The experimental results show that the proposed algorithm has good target tracking effect of basketball video, can effectively improve the target tracking accuracy and anti-interference ability, and can shorten the target tracking time.

1. Introduction

As it is known, the development of computer network technology promotes the update of image processing technology, and the video processing technology is also applied in people's daily lives [1, 2]. At present, the image accuracy of basketball video is high, but due to the variable target trajectory of basketball video, the similar moving target, and complex background, it is more difficult to track the target of basketball video. In addition, in the actual basketball game, it is difficult for the target tracking effect to achieve the ideal state due to the influence of various factors such as illumination, occlusion, and the definition of image acquisition equipment [3]. Therefore, the research on target tracking algorithm of basketball video has become a hot issue.

Yang and Wang [4] proposed a target tracking algorithm of basketball video based on DSP and FPGA. This study constructed a target index of the basketball video on the basis of collecting the basketball video, initialized the video, and determined the basic color map of the target index. Then according to the input basketball image, it used FPGA technology to obtain a new target index image and processed the new target image and the original image to get a new target position, so as to realize the target tracking of basketball video. The algorithm in this study can effectively track moving targets in basketball game video, but the tracking accuracy of moving targets is low. Wang and Hou [5] proposed a target tracking algorithm of sports video based on optimized particle filter. This study firstly detects the sports video, removes the background color histogram, then tracks the color target by using the particle filter technology, combines it with the color distribution, and finally realizes the target tracking of sports video by optimizing the tracked target model. The algorithm has high efficiency, but poor performance in target tracking of

moving video. Aiming at solving the above problems, this study proposes a target tracking algorithm of basketball video based on improved grey neural network, which provides a certain reference for accurately tracking basketball video target.

2. Design of Target Tracking Algorithm in Basketball Video

2.1. Obtaining the Target Area of Basketball Video. In order to obtain the target area of basketball video, this study assumes that the pixel point of basketball video image is (x, y) and sets the grey neural network as $I(x, y, t)$, and the time is t. At time $t + 1$, $I(x, y, t + 1)$ is used to represent the grey neural network. The pixel grey difference at time t and time $t + 1$ is calculated by the following formula, which is expressed as

$$\Delta I(x, y) = \{I(x, y, t + 1) - I(x, y, t)\}. \tag{1}$$

By iterative processing of formula (1), it can obtain the segmentation threshold T of basketball video that is processed by improved grey neural network and play the advantages of visual technology [6–8] to obtain the possible areas of the moving target in the basketball video:

$$I(x, y) = \begin{cases} 1, & \Delta I(x, y) > T, \\ 0, & \Delta I(x, y) \le T. \end{cases} \tag{2}$$

Particle filter is used to detect the target of the processed basketball video [9–11]. In order to increase the accuracy, it is assumed that when the basketball video is at time t, the pixel point (x, y) of the image is represented by grey neural network $I(x, y, t)$. After simulation of the motion of basketball video, if this pixel point remains unchanged, when $dI/dt = 0$, it can obtain the following formula:

$$uI_x + vI_y + I_t = 0. \tag{3}$$

In formula (3), $u = dx/dt$, $v = dy/dt$, and the differentiation of the target area images x, y, and t in the basketball video can be represented by I_x, I_y, and I_t.

It can obtain the solution of (u, v) by using the iterative algorithm:

$$u_{x,y}^{k+1} = u_{x,y}^{-k} - \frac{Ix\left(Ixu_{x,y}^{-k} + Iyv_{x,y}^{-k} + It\right)}{\Delta I(x, y)^2 + Ix^2 + Iy^2}, \tag{4}$$

$$v_{x,y}^{k+1} = v_{x,y}^{-k} - \frac{Iy\left(Ixu_{x,y}^{-k} + Iyv_{x,y}^{-k} + It\right)}{a^2 + Ix^2 + Iy^2}. \tag{5}$$

Here, a represents the iterative coefficient.

Referring to formulas (4) and (5), it can segment the basketball video image by using the threshold after iterative processing:

$$\text{Energyd}(I_i, C_j) = \left((u)^2 + (v)^2\right)^{1/2}. \tag{6}$$

Here, I_i represents the area where the target may appear and C_j represents the area where the target actually appears.

Taking clustering analysis of the segmented image by the image processing technology [12–14], it can obtain the target area of the basketball video, which can be shown in the following formula:

$$j = \min\left\{j/d\left(I_i, C_j\right)\right\} \le \theta_1. \tag{7}$$

In formula (7), the distance between the two areas is $d(I_i, C_j)$, the maximum distance threshold of the two target areas of basketball video is θ_1, and the nonzero pixel in the basketball video is I_i.

2.2. Acquisition of Target Data in Basketball Video. The algorithm in this study fuses the feature vectors by matching technology, monitors three-dimensional fusion target in the target area of basketball video, reconstructs the video that is not in the target area [15–17], and takes resolution reconstruction and spatial exchange matching of target data in the target area of basketball video. The calculation formula is as follows:

$$S(\omega) = \frac{1}{K} \sum_{k=0}^{K-1} S_k(\omega). \tag{8}$$

In formula (8), K represents the reconstruction coefficient and $S_k(\omega)$ represents the target data.

Assuming that the tracking targets in the target area of the basketball video are distributed linearly, the linear tracking target model in the public area of the basketball video is as follows:

$$x(n) = \frac{1}{\sqrt{N}} A \sum_{k=0}^{N-1} X(k)\exp\frac{j2\pi kn}{N}, \quad n = 0, 1, \ldots, N - 1. \tag{9}$$

Here, j represents the number of image pixel node, k represents the image pixel node, A represents the amplitude of image reconstruction in basketball video, N represents the sensor node in the collected basketball video, and $X(k)$ represents the target image pixel of the basketball video. Assuming that $A = \{a_{i,j}, 0 < i, j < N\}$, the basketball video image can be decomposed into multimodule color map. Through the feature classification processing of the target data in the video, it can form the target pixel matrix $M \times N$ of basketball video.

Suppose the feature vector of the i-th pixel in the basketball video is $B = \{b_j(k)\}$; then when $0 < j < N, 0 < k < M$, and the initial pixel of the target data distribution in the basketball motion video is π, the probability distribution is $\pi = \{\pi_i, i = 1, 2, \ldots, n\}$ at this point. Using the grey normalization processing of grey neural network [18–20] and combining with the basketball video after template matching, it can obtain the single frame pixel feature set of the basketball video, which is as follows:

$$s(k) = [s_1(k), s_2(k), \ldots, s_m(k)]^T. \tag{10}$$

In the target area of basketball video, this algorithm matches the target image template of the block area and

constructs the generation sequence of the target area in basketball video, which is as follows:

$$\mathrm{Dif}\,(C_1, C_2) = N \min w(v_i, v_j). \tag{11}$$

In formula (11), $N = 1, 2, \ldots, T$ represents the feature collection sample of the target area in basketball video, $w(v_i, v_j)$ represents the pixel width of the basketball video, and the feature vector is composed of the eigenvalue of the target image in the target area of basketball video, thus achieving the acquisition of the target data of basketball video.

2.3. Extraction of Key Feature Points of Basketball Video Target.

The following research is mainly to collect the target data of basketball video and extract the key feature points of the basketball video target [21]. Firstly, it is necessary to use the image degradation model to detect the feature of the target in the basketball video, combined with the improved grey neural network to optimally match the target image in the basketball video. If the threshold of the main feature sequence $h(j)$ of the target image in the basketball video is set as $h(j)$, the acquisition model of the target data in basketball video can be obtained. The formula is as follows:

$$x(n) = \sum_{j=0}^{\infty} h(j)w(n - j). \tag{12}$$

Secondly, according to the motion curve of basketball video target image in three-dimensional space [22, 23], it can obtain the combined dispersion of target image in basketball video. The formula is as follows:

$$\psi_{a,b}(t) = [U(a, b)\psi(t)] = \frac{1}{\sqrt{|a|}} \psi\left(\frac{t - b}{a}\right). \tag{13}$$

Here, $U(a, b)$ represents the similarity of pixels, and $\psi(t)$ represents the discrete features of the image.

The characteristic output matrix of basketball video target is as follows:

$$X' = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ x_{(n-1)1} & x_{(n-1)2} & \cdots & x_{(n-1)n} \end{bmatrix}. \tag{14}$$

Thirdly, after obtaining the feature vector of the basketball video, it can detect the moving image in the target area of the basketball video by template matching method [24] and get the fusion image, which is as follows:

$$X_t = AX_{t-1} + t. \tag{15}$$

In formula (15), A represents the matching amplitude, t represents the image frame, the coordinate point of the key feature point of the target in the basketball video is $X = [x_t, y_t]^T$. Assuming that the two most boundary target

feature points in the target area of basketball video are PE_1 and PE_2, it can obtain the average matching pixel of the feature points.

$$A = s \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix},$$
$$t = \begin{bmatrix} t_x \\ t_y \end{bmatrix}. \tag{16}$$

Finally, it also needs to match each feature point obtained above and geometrically describe the sparse items of the image based on the improved grey neural network and extract the motion feature of the target image in the basketball video by using the single frame visual difference analysis method.

$$\frac{\partial u(x, y; t)}{\partial t} = M\Delta_s u(x, y; t) + N\Delta_t u(x, y; d, t). \tag{17}$$

Here, M and N, respectively, represent the common sparse item set of the image, and Δ_t represents the image area. Through the above operations, the key feature points of the basketball video target can be extracted.

2.4. Design of Target Tracking Algorithm of Basketball Video.

Based on the improved grey neural network, it can reconstruct the target data of the target area in the basketball video, extract the key points of the motion video in the target area, and obtain the feature vectors of the basketball video. According to the improved grey neural network, the image feature points of the basketball video can be tracked and positioned; the tracking and positioning information can also be obtained, which is as follows:

$$\theta(t) = -2\pi K \ln\left(1 - \frac{t}{at_0}\right) + \theta_0. \tag{18}$$

Here, it can obtain $\theta_0 = -2\pi K \ln(1 + T/2t_0)$ after the amplitude segmentation of image in the target area of the basketball video, in which K represents the reconstruction coefficient. Then it is necessary to fuse the target image of basketball video through the fusion technology [25] and get the fused n subimage blocks, so as to obtain the permutation matrix of the dynamic feature points of the basketball motion image. The permutation matrix is as follows:

$$e_{mk}^{ij} = \begin{cases} -p_k \log(p_k), & p_k \neq 0, \\ 0, & p_k = 0. \end{cases} \tag{19}$$

In formula (19), the probability value of dynamic characteristic distribution of basketball video is p_k, and then $m = 1, 2, \ldots, N$.

After extracting the implicit features of the target image in the target area of the basketball video and reconstructing the image superresolution of basketball video, it can obtain the output:

$$p_{i,j}(A) = \begin{cases} \dfrac{w_{i,j}}{w_i}, & \text{if } i \neq j \text{ and } e_{i,j} \in A, \\\\ 0, & \text{if } i \neq j \text{ and } e_{i,j} \notin A, \\\\ 1 - \dfrac{\sum\limits_{j: e_{i,\varepsilon} \in} w_{i,j}}{w_i}, & \text{if } i = j. \end{cases}$$

(20)

Here, $w_{i,j}$ represents the weight of basketball video target image, w_i represents the weight of pixel i, and $e_{i,j}$ represents the matching value of pixel.

The 3D feature reconstruction model of target area in basketball video is as follows:

$$\text{NLM}[g](i) = \sum_{j \in \Omega} w(i, j)g(j).$$

(21)

In formula (21), $w(i, j)$ represents the weight of image pixels i and j, and $g(j)$ represents the sequence of pixel j. Assuming that the three-dimensional features of the target image of basketball video achieve uniform distribution [26], it can calculate the features of the target tracking image of basketball video.

$$\mu_{pq} = \sum_{m=1}^{M} \sum_{n=1}^{N} (x - \dot{x})^p (y - \dot{y})^q f(x, y).$$

(22)

Here, $f(x, y)$ represents the scale information of target features, $(x - \dot{x})$ represents the feature quantity of tracking target x, and $(y - \dot{y})$ represents the feature quantity of tracking target y. According to the result of pixel reconstruction of basketball video, it can realize the reconstruction of dynamic feature in the target area of basketball video and achieve the design effect of target tracking algorithm of basketball video.

3. Experimental Analysis

In order to verify the effectiveness of the target tracking algorithm of basketball video based on the improved grey neural network in target tracking, this experiment tracks two video sequences with different video frame number and tracking area size. In the VS platform under Windows environment, the experiment uses C++ language for programming, sets the experimental error threshold as 0.55, and chooses three algorithms for comparison experiment, namely, the target tracking algorithm designed in this paper and the algorithm in [4] and in [5]. Three groups of basketball video sequences are selected for the test. And the image sequences of frames 23, 38, and 56 of the three groups of basketball video in a group of videos are collected as the sample sequences of the experimental test. This experiment mainly tests the basketball video target tracking effect of different algorithms under different basketball video frames, as shown in Figure 1.

According to Figure 1, when tracking the 23rd frame image of basketball video, the proposed algorithm, the algorithm in [4], and that in [5] can accurately track the basketball video target. When tracking the 38th frame image of basketball video, both the proposed algorithm and the algorithm in [4] can accurately track the basketball video target, while the algorithm in [5] has lost the basketball video target. When tracking the 56th frame image of the basketball video, both the proposed algorithm and the algorithm in [5] can accurately track the basketball video target, while the algorithm in [4] has errors in tracking the basketball motion video target. It can be seen that the proposed algorithm has a good effect on basketball video target tracking.

According to the image sequence of the basketball video, when tracking the image sequence of the 23rd, 38th, and 56th frames, this experiment tests the accuracy of the target tracking algorithm of basketball video based on the improved grey neural network. The results are shown in Figure 2.

As can be seen from the results in Figure 2, in the test of the target tracking algorithm of basketball video based on the improved grey neural network, when tracking the 23rd frame image of basketball video, the tracking accuracy is 80%~85%. When tracking the 38th frame of the basketball video, the tracking accuracy is between 84% and 91%. When tracking the 56th frame of the basketball video, the tracking accuracy is between 85% and 98%. Therefore, the target tracking algorithm designed in this paper has high accuracy in tracking the basketball video target. The target tracking algorithm designed in this paper can determine the actual position of the basketball video target by obtaining the target area of the basketball video, which can improve the tracking accuracy of the basketball video target.

In order to further verify the performance of the target tracking algorithm designed in this paper in the target tracking, this experiment collects three target images with different tracking areas of $600 * 400$, $480 * 320$, and $420 * 280$ and tests the target tracking effects of different algorithms in different tracking areas. The experimental results are shown in Figure 3.

According to Figure 3, it is known that when the tracking area of the target image is $600 * 400$, the proposed algorithm, the algorithm in [4], and that in [5] can accurately track the basketball video target. When the tracking area of the target image is $480 * 320$, the proposed algorithm can accurately track the basketball video target, while the algorithm in [4] has lost the basketball video target, and the algorithm in [5] has errors in tracking the basketball video target. When the tracking area of the target image is $420 * 280$, the proposed algorithm can still accurately track the basketball video target, while the algorithm in [4] and that in [5] have errors in tracking the basketball video target. Therefore, the target tracking effect of the proposed algorithm is good.

Under the condition of tracking target image sample sequences with different tracking area sizes, this experiment takes three target images with different tracking area sizes collected in basketball video as sample data and tests the tracking accuracy of the target tracking algorithm of basketball video based on improved grey neural network. The results are shown in Figure 4.

Frame 23 basketball video image sequence

Frame 38 basketball video image sequence

Frame 56 basketball video image sequence

——— The proposed algorithm

——— Reference [4] algorithm

——— Reference [5] algorithm
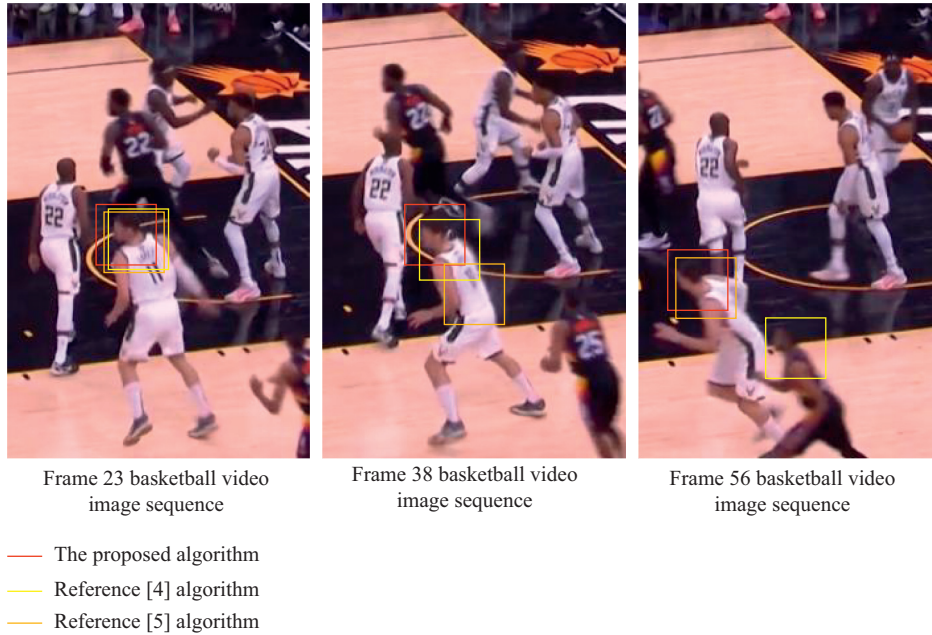
FIGURE 1: Target tracking effect of the algorithm under different basketball video frames.

····· Frame 56 image sequence

–·–· Frame 38 image sequence

——— Frame 23 image sequence
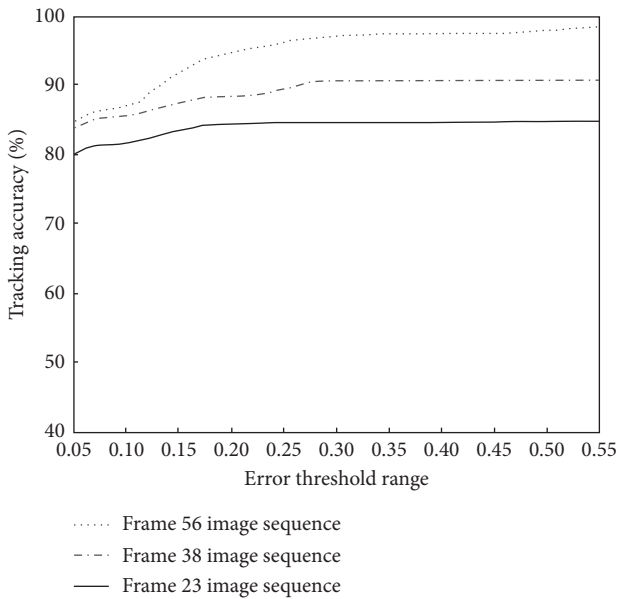
FIGURE 2: Test results of target tracking accuracy of basketball video with different frames.

As can be seen from the results in Figure 4, with the expansion of the error threshold range, the tracking accuracy of target image in different tracking areas gradually increases. When the tracking area of the target image is $600 * 400$, the target tracking accuracy of basketball video is between 86% and 98%; when the tracking area of the target image is $480 * 320$, the target tracking accuracy of basketball video is between 83% and 95%; when the tracking area of the target image is $420 * 280$, the target tracking accuracy of basketball video is between 80% and 90%. Therefore, the target tracking algorithm designed in this paper has high

accuracy in tracking the basketball video target. Among the three tracking areas with different sizes, this algorithm with the tracking area of $600 * 400$ has a higher target tracking accuracy of basketball video, which ensures the performance of the basketball video target tracking.

In order to further verify the anti-interference ability of the target tracking algorithm designed in this paper, the basketball video can be simulated by using 5.3dmax 3D simulation software in simulation environment, and the noise size of the image is set to $-0.003{\sim}0.003$ m, as shown in Figure 5.

In order to further verify the anti-interference ability of the target tracking algorithm designed in this paper, this experiment sets the noise with the size of $0.003{\sim}0.003$ m in the simulation environment, compares the target tracking algorithm designed in this paper with the algorithm in [4] and that in [5], and obtains the target recognition rates of basketball video with different algorithms under noise interference, as shown in Table 1.

According to the data in Table 1, in the case of noise interference, when the amount of basketball video data is 500 MB, the average recognition rate of basketball video target of the algorithm in [4] is 74.8%, and that of the algorithm in [5] is 65.2%, while the average target recognition rate of basketball video target of the algorithm designed in this paper is as high as 90.6%. Therefore, in the case of noise interference, the algorithm designed in this paper has a high target recognition rate and strong anti-interference ability. Because the algorithm designed in this paper processes the target area through grey level normalization, establishes the target area generation sequence, and collects the target data of the basketball video, it can effectively remove the image noise and redundant data, thus effectively improving the anti-interference ability of the algorithm designed in this paper.

600∗400   480∗320   420∗280

——— The proposed algorithm
——— Reference [4] algorithm
——— Reference [5] algorithm
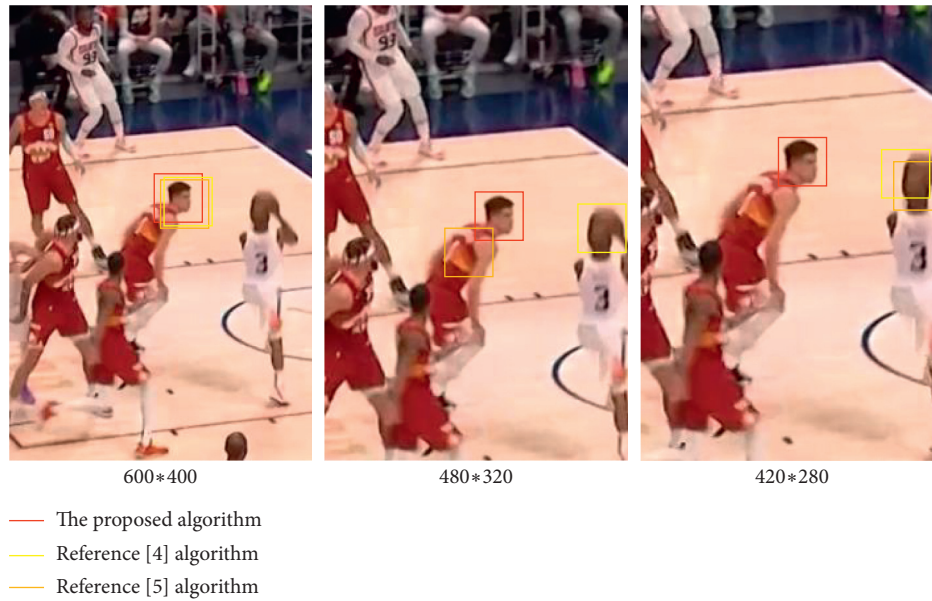
FIGURE 3: Target tracking effect of the algorithm in different tracking areas.

······▶ 600∗400 tracking area
———▶ 480∗320 tracking area
—·—▶ 420∗280 tracking area
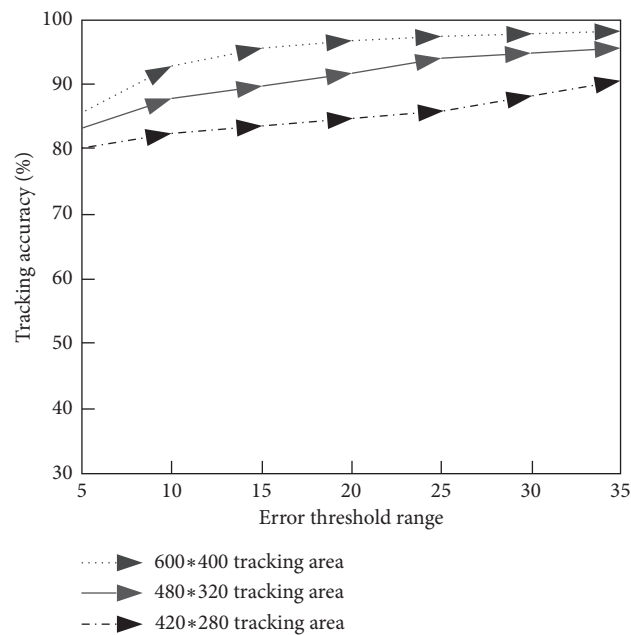
FIGURE 4: Test results of target tracking accuracy of basketball video in different areas.

On this basis, in order to verify the basketball video target tracking time of the target tracking algorithm designed in this paper, this experiment selects 500 MB basketball video data and compares the basketball video target tracking time of the target tracking algorithm designed in this paper with that of the algorithm in [4] and that of the algorithm in [5]. The test results of basketball video target tracking time of different algorithms are shown in Table 2.

According to the data in Table 2, as the amount of basketball video data increases, the target tracking time of basketball video with different algorithms increases. When the amount of basketball video data is 500 MB, the basketball video target tracking time of the algorithm in [4] is 34 s, and the basketball video target tracking time of the algorithm in [5] is 39 s, while the basketball video target tracking time of the algorithm designed in this paper is only 22 s. Therefore, the basketball video target tracking time of the algorithm designed in this paper is relatively short. The algorithm designed in this paper uses single frame visual difference analysis to extract the key feature points of the basketball video target and combines it with the improved grey neural

FIGURE 5: Simulation video of basketball.

TABLE 1: Target recognition rate of basketball video with different algorithms under noise interference.

Data size of basketball video (MB)	The designed algorithm (%)	The algorithm in [4] (%)	The algorithm in [5] (%)
100	89	78	69
200	92	72	65
300	95	75	63
400	87	79	67
500	90	70	62

TABLE 2: Test results of target tracking time of basketball video with different methods.

Data size of basketball video (MB)	The designed algorithm (s)	The algorithm in [4] (s)	The algorithm in [5] (s)
100	5	8	14
200	9	17	20
300	14	23	26
400	19	29	31
500	22	34	39

network to track and locate the feature points of the basketball video target, thus effectively shortening the tracking time of the basketball video target.

This paper extracts the key feature points of basketball video target and tracks and locates the feature points of basketball video target combined with the improved grey neural network, thus effectively shortening the tracking time of basketball video target.

4. Conclusion

(1) For the target tracking of video images of basketball, different frames of video images are selected to complete the target tracking test. The results show that the proposed method can achieve high-precision target tracking of different images, and the tracking accuracy can be maintained between 80% and 98%, which has application advantages compared with the literature algorithm.

(2) When the tracking region of the target image is different, the target tracking results of different algorithms are quite different. When the target image tracking area is large, the target tracking is not difficult, and both the literature algorithm and the proposed algorithm can complete the experimental task.

When the tracking area of the target image keeps shrinking, the recognition of target feature becomes more difficult, and the target tracking error of the literature algorithm appears obviously. But the proposed algorithm can still achieve the target tracking with high precision, and the accuracy can be maintained between 80% and 98%.

(3) In the case of noise interference, the target recognition rate of basketball video designed in this paper is high.

When the amount of basketball video data is large, the target tracking time of basketball video designed in this paper is only 22 s. It can be seen that the target tracking time of basketball video with the algorithm designed in this paper is short.

Based on improved grey neural network proposed in this paper, the target tracking algorithm of basketball video takes advantage of improved grey neural network. It reconstructs the basketball video target image by obtaining the basketball video target area, collecting the basketball video target data, and extracting the key feature points of the basketball video target, and it finally realizes the basketball video target tracking. Therefore, the target tracking algorithm designed in this paper has good tracking

effect, which can effectively improve the tracking accuracy and ensure the performance of the algorithm.

Data Availability

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Conflicts of Interest

The authors declare that they have no conflicts of interest regarding this work.

Acknowledgments

References

[1] Li Chen, "Research on target tracking algorithm of fast human moving in video image," *Modern Electronic Technique*, vol. 42, no. 3, pp. 49–51, 2019.

[2] J. Mao, "A video target tracking algorithm based on wireless sensor networks," *Control Engineering*, vol. 25, no. 1, pp. 71–75, 2018.

[3] Y. Li, J. Lu, H. Li, W. Zhang, and Y. Zhang, "Research on target tracking algorithm from fisheye camera based on compressive sensing," *Journal of Electronics and Information Technology*, vol. 40, no. 5, pp. 1242–1249, 2018.

[4] B. Yang and T. Wang, "Implementation of target tracking algorithm of embedded basketball video based on DSP-FPGA," *Natural Science Journal of Xiangtan University*, vol. 40, no. 6, pp. 104–108, 2018.

[5] J. Wang and X. Hou, "Design of target tracking algorithm of sports video based on optimized particle filter," *Machine Tool & Hydraulics*, vol. 46, no. 6, pp. 164–169, 2018.

[6] D. Wang, H. Fang, Y. Liu, S. Wu, Y. Xie, and H. Song, "A target tracking algorithm of panoramic video based on improved RT-MDNet," *Journal of Harbin Institute of Technology*, vol. 52, no. 10, pp. 152–160+174, 2020.

[7] S. Javed, A. Mahmood, J. Dias, and N. Werghi, "Robust structural low-rank tracking," *IEEE Transactions on Image Processing*, vol. 29, pp. 4390–4405, 2020.

[8] C. Fang, J. Huang, and K. Cuan, "Comparative study on poultry target tracking algorithms based on a deep regression network," *Biosystems Engineering*, vol. 190, pp. 176–183, 2020.

[9] P. Zheng, H. Bai, Z. Li, and G. Hongwei, "Design of accurate detection and tracking algorithm of moving target under jitter interference," *Chinese Journal of Scientific Instrument*, vol. 40, no. 11, pp. 90–98, 2019.

[10] T. Wang, "Unsupervised video multi-target tracking based on fast resampling particle filter," *The Journal of Supercomputing*, vol. 76, no. 2, pp. 1293–1304, 2020.

[11] Z Cai and H Feng, "Target tracking based on improved square root cubature particle filter via underwater wireless sensor networks," *IET Communications*, vol. 13, 2019.

[12] Yi Liu, "Research on improved particle filter model technology for target tracking of moving video," *Modern Electronics Technique*, vol. 42, no. 3, pp. 65–67+72, 2019.

[13] A. Silva Ramos, C. Hora Fontes, A. Magdiel Ferreira et al., "Somatic cell count in buffalo milk using fuzzy clustering and image processing techniques," *Journal of Dairy Research*, vol. 88, no. 1, pp. 69–72, 2021.

[14] Y. Kim and G. Dodbiba, "A novel method for simultaneous evaluation of particle geometry by using image processing analysis," *Powder Technology*, vol. 393, pp. 60–73, 2021.

[15] Y. Qi and M. Wang, "Tracking algorithm of moving target based on LBP and kernel correlation filter," *Infrared Technology*, vol. 41, no. 6, pp. 572–576, 2019.

[16] S. Unar, X. Wang, C. Wang, and Y. Wang, "A decisive content based image retrieval approach for feature fusion in visual and textual images," *Knowledge-Based Systems*, vol. 179, no. SEP.1, pp. 8–20, 2019.

[17] F. Yuan, G. Li, X. Xia, B. Lei, and J. Shi, "Fusing texture, edge and line features for smoke recognition," *IET Image Processing*, vol. 13, no. 14, pp. 2805–2812, 2019.

[18] W. Ou, H. Yin, H. Xu, and Z. Liu, "A multi-odject tracking algorithm based on trajectory reconstruction on multi-egocentric video," *CAAI Transactions on Intelligent Systems*, vol. 14, no. 2, pp. 246–253, 2019.

[19] K. Z. Rajab, B. Wu, P. Alizadeh, and A. Alomainy, "Multi-target tracking and activity classification with millimeter-wave radar," *Applied Physics Letters*, vol. 119, no. 3, Article ID 034101, 2021.

[20] D. Yao, C. Dou, D. Yue, Z Nan, and Z Tingjun, "Adaptive neural network consensus tracking control for uncertain multi-agent systems with predefined accuracy," *Nonlinear Dynamics*, vol. 101, no. 1, pp. 2249–2262, 2020.

[21] X. Yu and S. Li, "Research on tracking algorithm of moving target based on multi-step prediction and Mean-Shift," *Infrared Technology*, vol. 40, no. 12, pp. 1182–1187, 2018.

[22] J. Li, Z. Shang, and H. Liu, "Target tracking algorithm of correlation filtering moving based on multi-layer convolution feature," *Computer Science*, vol. 046, no. 7, pp. 252–257, 2019.

[23] H. Zhang, P. Rao, H. Xia, D. Weng, X. Chen, and Y. Li, "Modeling and analysis of infrared radiation dynamic characterisitcs for space micromotion target recognition," *Infrared Physics & Technology*, vol. 116, no. 3, Article ID 103795, 2021.

[24] C. Chao, "Target tracking algorithm involving the Gaussian mixture model and weighted likelihood computer engineering and applications," *Computer Engineering and Applications*, vol. 55, no. 12, pp. 124–131, 2019.

[25] P. Jia and W. Jia, "Research on coal mine multi-target trajectory tracking algorithm," *Computer Engineering and Applications*, vol. 54, no. 2, pp. 227–232, 2018.

[26] B. Wei, P. Ge, H. Wu, G. Wang, and W. Han, "Object tracking algorithm based on static-adaptive appearance model correction," *Computer Applications*, vol. 038, no. 4, pp. 1170–1175, 2018.

Research Article

Research on Spoken Language Understanding Based on Deep Learning

Hui Yanli (ID)

Faculty of Foreign Languages and Business, Jiaozuo Normal College, Jiaozuo 454001, China

Correspondence should be addressed to Hui Yanli; 1295004004@jzsz.edu.cn

Aiming at solving the problem that the recognition effect of rare slot values in spoken language is poor, which affects the accuracy of oral understanding task, a spoken language understanding method is designed based on deep learning. The local features of semantic text are extracted and classified to make the classification results match the dialogue task. An intention recognition algorithm is designed for the classification results. Each datum has a corresponding intention label to complete the task of semantic slot filling. The attention mechanism is applied to the recognition of rare slot value information, the weight of hidden state and corresponding slot characteristics are obtained, and the updated slot value is used to represent the tracking state. An auxiliary gate unit is constructed between the upper and lower slots of historical dialogue, and the word vector is trained based on deep learning to complete the task of spoken language understanding. The simulation results show that the proposed method can realize multiple rounds of man-machine spoken language. Compared with the spoken language understanding methods based on cyclic network, context information, and label decomposition, it has higher accuracy and F1 value and has higher practical application value.

1. Introduction

In our society, with the development of science and informatization, more tasks have been applied to the field of artificial intelligence, and it has become an irreversible trend. With the rapid development of artificial intelligence, there are countless tasks to process serialized data in our society. For example, speech recognition, natural language understanding, and time series data all need to process serialized data. The spoken language system integrating speech recognition and speech synthesis is the core technology of human-computer interaction, and oral understanding is the core of spoken language system [1]. Therefore, the research on oral comprehension can enable people to apply it more accurately to spoken language system, which is more convenient for people's life and work. It has always been explored and studied to enable machines to communicate with people without barriers. Oral English is used in human communication. If the computer can understand spoken language and make correct answers and can correctly

complete various operation tasks required by people according to people's instructions, it can use robots to complete operations in many occasions, especially in some dangerous fields, which can save a lot of resources and reduce the risk. It can be seen that spoken language system is widely used in life and is of great significance. Spoken language system is such that people express their ideas in natural language to communicate with a certain field of computer [2]. This way can make the computer understand human requirements more efficiently and quickly, so as to complete people's demands according to the corresponding model processing. Oral understanding is to convert the natural speech input by people into text through speech recognition, convert the text into corresponding word vector or sentence level vector in the oral understanding system, then send the vector through the encoder or directly into the built model, and finally decode or directly output the sentence. For the whole spoken language system, the key part is oral understanding and dialogue management. If oral comprehension cannot be performed correctly or the

performance of oral comprehension is poor, such a situation will lead to the error of query results of subsequent dialogue management, resulting in the poor performance of the whole spoken language system, which cannot complete various operations and tasks required by people [3]. It can be seen that oral understanding plays a key role in spoken language system. The performance of oral comprehension directly determines the performance of spoken language system. Therefore, the study of oral comprehension is of great value and significance.

The academic community has carried out extensive research on spoken language understanding. Zhang and others improved the effectiveness of information feature extraction and oral comprehension performance by adding and storing historical state information [4]. Yang and others constructed the initial representation of the current round of text and context text in combination with the phonetic features and used the context semantic information to assist the intention detection of the current round of text, which improved the detection effect [5]. Xu and Huang transformed label classification into independent classification and introduced external word vector to improve the classification performance of the model [6]. The above research results have improved the performance of spoken language understanding methods, but they still have the problem of poor slot value recognition, which affects the accuracy of dialogue tasks. With the successful application of deep learning, deep neural network has made remarkable achievements [7]. The development of deep neural network has gone through a very long time. Now remarkable achievements have been made in the fields of speech recognition, image processing, text processing, computer vision, and natural language processing. Deep neural network is a successful application in many fields such as natural language processing, compared with the traditional oral comprehension model. The biggest feature of deep neural network is to train a large amount of data and then extract the characteristic information. The characteristic information obtained through the network structure can achieve good results in oral comprehension tasks. Therefore, based on deep learning, this paper proposes a spoken language understanding method to improve the effect of oral understanding tasks, promote the development of human-computer interaction technology, and better meet the needs of practical application scenarios.

2. Spoken Language Understanding Method Based on Deep Learning

2.1. Semantic Text Classification. In man-machine dialogue system, oral comprehension is mainly used to understand what users say and extract important information. Oral comprehension model includes three tasks: domain recognition, intention recognition, and semantic slot filling. Domain recognition task and intention recognition task are used to understand what users say. Both tasks belong to text classification. Semantic slot filling task is to extract important information, which belongs to sequence annotation problem. Domain recognition and intention recognition are

to classify the dialogue text entered by the user, and the classified label is the domain involved in the dialogue text or the user's intention. Therefore, this paper first designs a semantic text classification model, uses the algorithm to extract the local features of the text, and matches it with the current task to better classify the text. The model consists of text convolution neural network and bidirectional long-term and short-term memory neural network combined with attention mechanism. The multiconvolution kernel mechanism of text convolution neural network model can better extract the n-gram features of text data, so that the text convolution neural network model can learn more abundant local features [8].

The bidirectional long-term and short-term memory neural network model can effectively extract the context information of semantic text and concentrate the more important information of the current task [9]. There are 160 neurons in the hidden layer of bidirectional long-term and short-term memory neural network model, the size of four-layer convolution nuclei is 2–5, the number of convolution nuclei is 32, and the size of attention mechanism is 128. The two-channel neural network model is used to extract text information at the same time, which makes the model more efficient. The first channel is the word vector of the word. First, the text is segmented, converted into word vectors, and input into the first channel. Unsupervised stacked deconvolution neural network is used to learn from the word vector of the text to obtain the feature mapping matrix. The feature mapping matrix is used as the convolution kernel of the deep convolution neural network to convolute and pool the word vector layer by layer [10]. Then, the important information of text context is extracted through the two-way long-term and short-term memory model. The second channel is the input word vector. Convert a single Chinese character in the text into a word vector. The local semantic features of the text are extracted through the convolution model of multiconvolution kernel. The hierarchical attention mechanism is used to select the important sentences in the text at the network layer, and the network is extracted layer by layer to obtain the text feature vector [11]. Finally, the left and right channel outputs are spliced into vectors, and then text classification is carried out through the maximum layer of fully connected neural network. In text classification, each output is related to the context of the input and context at that time [12, 13]. If the output vector is added each time and the average value is taken directly, the contribution of each output to text classification is the same, but this is not the case. Keywords in the classification should have greater weight. Therefore, when outputting vectors, we want to focus on vectors that are more important to the current task, so we introduce the attention mechanism. Note that the mechanism model can be expressed as

$$\alpha = \tanh(w\beta + \varepsilon). \tag{1}$$

In formula (1), α represents the output value of the attention mechanism; tanh represents hyperbolic tangent function; w and ε represent the weight and bias of attention mechanism; β represents the splicing of outputs in

both positive and negative directions at each time. The outputs of forward and reverse are spliced to obtain the randomly initialized attention mechanism column vector. Finally, a normalized attention mechanism weight is obtained by softmax operation. The calculated output weights at each time are weighted and summed to obtain the output of the model; that is, the text classification is completed. Figure 1 shows the semantic text classification model.
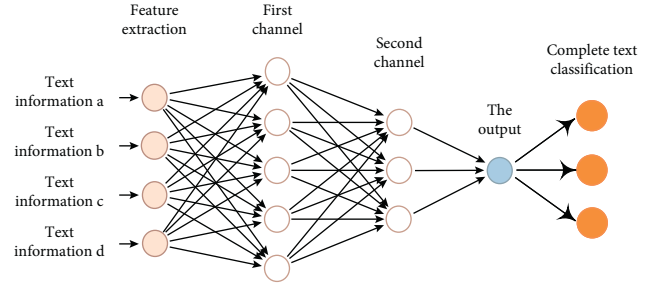
FIGURE 1: Semantic text classification model.

2.2. Design Intent Recognition Algorithm.

In oral comprehension task, intention recognition task is a classification problem, which is used to extract users' specific intention [14]. Semantic slot filling task is a sequence labeling problem, that is, each word in a given sentence is labeled respectively. There is a certain relationship between intention recognition task and semantic slot filling task. Therefore, these two tasks can be completed in the same model [15]. In this paper, Albert pretraining model and convolutional neural network are used to complete the task of intention recognition, and Albert pretraining model and conditional random field are used to complete the task of semantic slot filling. Albert model backbone network adopts transformer encoder framework and GELU nonlinear activation function. The dimension of the input embedded vector dimension is smaller than the output vector. The word level embedding vector has no context dependence, but the output of the hidden layer includes not only the meaning of the word itself, but also some context information. Therefore, the expression of the hidden layer contains more information [16]. In natural language processing tasks, the word embedding matrix is usually large. Due to the large number of parameters and the process of back propagation, the updated content is also sparse. Combined with the above two points, Albert model adopts a factorization method to reduce the amount of parameters. Firstly, the unique heat code is mapped to a low dimensional space and then mapped to a high-dimensional space. Through this decomposition method, the amount of word embedding parameters can be reduced [17]. On this basis, the parameters of the full-connection layer and the attention mechanism layer are shared; that is, all the parameters inside the encoder are shared. In order to retain only consistency tasks and remove the impact of subject prediction, the positive samples of SOP are obtained in the same way, and the negative samples reverse the order of positive samples. That is, SOP only focuses on the order of sentences and has no influence on the subject [18]. Albert model input needs to add [CLS] at the beginning of the text, and the output corresponds to the input [CLS] vector containing the information coding of the whole sentence, which can be used for text classification tasks [19]. The remaining eigenvectors are used for the sequence annotation task. Use [CLS] vector and feature vector to identify semantic intention. Conditional random field is a probability graph model and belongs to discriminant model. In the field of natural language understanding, linear chain conditional random fields are often used to solve the problem of sequence annotation [20]. In the linear chain random field, the sequence and label meet the following conditions:

$$\theta(B_i \mid A, B_1, B_2, \cdots, B_n) = \theta(B_i \mid A, B_{i-1}, B_{i+1}). \quad (2)$$

In formula (2), θ represents probability; A represents sequence; B_n represents the tag sequence; n indicates the number of labels; i indicates the marked serial number. The parametric form of linear chain conditional random field is as follows:

$$\theta(B_i \mid A) = \frac{1}{\delta} e^{(\lambda_1 + \lambda_2)}. \quad (3)$$

In formula (3), $\theta(B_i|A)$ represents the parametric form; δ represents normalization factor; e is the natural constant; λ_1 and λ_2 are local characteristic function and node characteristic function, respectively. λ_1 is only related to the current node and the previous node, and λ_2 is only related to the current node. λ_1 and λ_2 values can only be 0 or 1.

Each network layer of Albert model has two subnetwork layers: the first layer is multihead self-attention mechanism layer. The second layer is the common feedforward network layer, which is used to integrate the position information of words. In addition, each subnetwork layer contains an add label layer, which is used to add and normalize the input and output of this layer, and then the residual connection is used between the two subnetwork layers [21].

Let L be the additional weight matrix to compress the spliced matrix dimension into the sequence length, Q, K, V is the vector of each corresponding label in the input sequence, and Q^{ij}, K^{ij}, V^{ij} is the weight matrix of Q, K, V; DK represents the vector dimension of each label, and Dm is the normalized activation function [22]. σ is vector point multiplication; r is the hidden layer of the network layer. The calculation formula of dynamic word vector W is as follows:

$$\text{MultiHead}(Q, K, V) = L\sigma(Q^{ij}, K^{ij}, V^{ij}),$$

$$Dm = \frac{DK(Q^{ij}, K^{ij}, V^{ij})}{\sigma}, \quad (4)$$

$$W = \frac{Dm}{\sigma r}(Q, K, V).$$

The algorithm uses the Chinese Albert pretraining model to obtain the dynamic word vector with context and then uses the conditional random field model which can effectively deal with the sequence annotation problem to

complete the idiom meaning slot filling task [23]. At the same time, the multicore convolutional neural network is used for training, so that each datum has a corresponding intention label, so as to complete the intention recognition task [24]. Figure 2 shows the structure of Albert model.

2.3. Conversation State Tracking Based on Slot Feature.

In each round of conversation, the user input is used as an important information source for conversation status tracking, which directly contains the slot or slot value pair related to the user's needs. In the process of interaction, users are allowed to modify or improve their needs at any time. Therefore, the spoken language system needs to update the dialogue status according to the user's current round input [25]. In order to improve the recognition accuracy of new slot values and rare slot values, a multiround dialogue state tracking model based on local slot features is proposed, including coding module and state evaluation module. Since each state in state tracking consists of multiple sets of slot value pairs and the dataset of state tracking is often small [26], many slot value pairs rarely appear in the dataset, so wrong inference of rare slot value pairs often leads to poor state tracking results. The slot information word vector is spliced into the word vector of the text to be encoded to strengthen the connection between the slot information and the text sequence [27]. At the same time, the attention mechanism is applied to the slot information to obtain the weight of the hidden state, so as to obtain the feature representation of the whole text for the slot. The encoder is composed of a forward LSTM (the input is the original sequence input) and a backward LSTM (the input is the reverse sequence input) [28]. It can be expressed as

$$y_t = L(c_t). \tag{5}$$

In formula (5), y_t represents the bidirectional feature vector corresponding to the word in the input text; L represents bidirectional LSTM structure; c_t represents the input sequence; t indicates time. For the input sequence and slot set, in order to calculate the sequence representation of any slot, the coding vector of the slot is spliced with the sequence, and finally the feature vector corresponding to the word in the input text is generated, which can be expressed as

$$R = L[\omega(c_t, u) + \kappa]. \tag{6}$$

In formula (6), R represents the feature vector corresponding to the word; ω represents hidden state weight; u represents the slot set; κ represents the slot value vector. The feature vector is the set of word feature vectors in the input text and the semantic vector of the whole sentence under the condition of different slot information. In the new round of state generation, the purpose is to predict the actual intention of users in this round. Intuitively, judge whether the slot value pair in the candidate is expressed in the user's statement. There are three main contribution sources in the dialogue information, namely, the user's current round of input, the previous round of system reply, and the previous round of system action. This means that the status evaluation
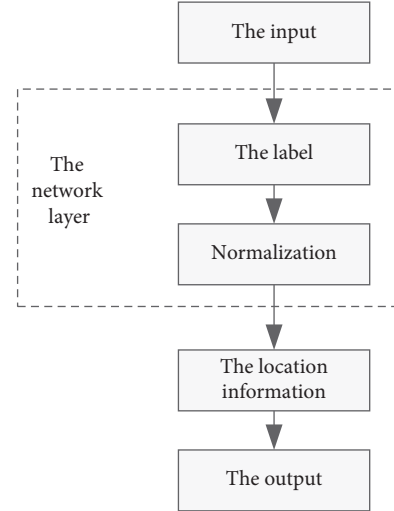
FIGURE 2: Albert model structure.

module will use these three information contribution sources to score the slot value pairs in each candidate [29]. As the most important information source, the user's current input can directly state the goal or request. Score a specific slot value set based on the current user's words. The calculation formula is

$$f = hg + v. \tag{7}$$

In formula (7), f represents the degree of expressing the specific slot value under consideration in the user's discourse; h represents the discourse of the current user; g represents the context representation generated by the attention mechanism; v represents a specific set of slot values. Considering that the last round of system reply can enrich or enhance the text information input by the user in detail, the user input text and the last round of system reply are jointly modeled in the evaluation module, so as to score the candidate slot value. The calculation formula is

$$f' = \sum_j h_j g_j. \tag{8}$$

In formula (8), j represents all candidate slot value sets; f' indicates the degree of expressing the candidate slot value under consideration in the user's discourse. After the score weighted sum is obtained by the scoring module, the score is mapped to the range of [0, 1] by using the activation function. The result is used as the basis for selecting the current candidate slot value, and the threshold is set. When the score exceeds the threshold, it indicates that the candidate slot value is characterized in the user requirements, and the slot value is used to update the dialog status.

2.4. Establishing Spoken Language Understanding Model Based on Deep Learning.

In multiround and multitask oral comprehension, because the user conversation may switch multiple times in different tasks, not all historical information will be useful. It is very important to select the historical information related to the current conversation.

When the current conversation encounters task jump; that is, all historical information talks about other tasks, its historical information has a negative impact on the intention identification and slot filling of the current conversation, resulting in the historical information not correctly helping to understand the user's current conversation [30]. In order to alleviate this problem, this model introduces current dialogue into historical coding and combines historical dialogue with current dialogue to reduce the negative impact of irrelevant historical context. That is, an auxiliary gate unit is constructed to learn the relationship between historical context and slot position. The input of the auxiliary gate unit is the output state and history information coding of the decoding layer LSTM, which are jointly input into the auxiliary gate unit for the following calculation:

$$z = \text{sigmoid}\left(\vartheta_1 o + \vartheta_2 m\right). \tag{9}$$

In formula (9), z represents a weight feature of the historical information and the output state of each step; ϑ_1 and ϑ_2 represent weights; o represents the output state of the decoding layer LSTM; m represents the historical dialogue information code; sigmoid is the activation function. Learning the relationship between the historical information and the output state of each step of the current dialogue, we can better use the historical information to find the key information in the current dialogue. The intention of user statement can affect the generation of slot. Introducing the context vector and intention representation vector of slot into a gate structure at the same time can improve the performance of slot filling task [31]. The function of the input layer is to convert the text in the dialog box into a pattern that can be understood by the computer. In the neural network channel, word vector and text vector are used to represent the classified text information. Through the expression method of information distribution, the transformed text information is mapped into high-dimensional space by using the method based on deep learning, and the semantic relationship can be inferred by using distance [32]. In this paper, Word2Vec toolkit is used to train on the prediction set. Finally, the word vector obtained by training is used as the word coding of the model. Assuming that a training sample is a given conversation, the conditional probabilities of intention label and slot label predicted by the model can be expressed as

$$\begin{cases} p_1 = \max\left(M_1\right), \\ p_2 = \prod_1^v \max\left(M_2^v\right). \end{cases} \tag{10}$$

In formula (10), p_1 and p_2 represent the conditional probability of intention label and slot label; M_2^v represents the probability vector of softmax transport slot label corresponding to each word; v represents the number of corresponding slot labels; M_1 represents the intention probability vector output by softmax. The probability of the corresponding category in each dimension and the maximum probability are taken as the intention category predicted by the sample [33]. Intent recognition and slot filling

share the same encoder. In the process of training the model, the two loss functions are added, and the parameters of the joint model are updated through back propagation [34]. In this model, the cross direct function is used as the loss function of the model. The model proposed in this paper introduces the current dialogue and attention mechanism into the historical information coding, learns the relationship between the historical information and the slot through the auxiliary gate unit mechanism, and effectively uses the historical dialogue information to improve the effect of the model.

3. Simulation Experiment

3.1. Dialogue Effect Test. In order to verify the performance of this method, the following simulation experiments are carried out. The experimental platform is MATLAB simulation platform. The computer used is Windows 10 system, equipped with i7 processor and running memory of 16 G. In this simulation experiment, the API of iFLYTEK is adopted for speech recognition and speech synthesis and transplanted to the ROS operating system. The spoken language understanding part is the content of this paper. Each module is encapsulated and organized as a node of ROS. The speech recognition node collects the user's audio and publishes the recognition results in the form of topic after speech recognition. Using spoken language understanding node, and after the question preprocessing, question and answer module, and dialogue management module based on slot features discussed above, the processing results are released again in the form of topic. Finally, the speech synthesis node subscribes to the topics published by the semantic analysis node and feeds back to the user in the form of audio after synthesis to complete a round of interaction. In practical application, the spoken language system runs on the ROS system and the microphone collects the user's speech and finally feeds back to the user with audio. Generally, there is no graphical interface. For demonstration, this section uses flask as the background framework to realize the API of semantic recognition and Apache as the static server of web resources, and the front end uses HTML, JavaScript, and CSS to realize a web application. Taking weather query as an example, the multiround interaction of the spoken language understanding method is shown in Figure 3.

According to the demonstration results in Figure 3, the spoken language understanding method proposed in this paper can ask and answer common questions and has the effect of real-time interaction of multiple rounds of dialogue. On this basis, the performance of the proposed method is further tested.

3.2. Experimental Environment. The KVRET dataset used in this paper is from Stanford Natural Language Processing Group. Task-oriented dialogue focuses on participating in the dialogue on specific topics initiated by users. Generally speaking, if researchers want to do task-oriented dialogue and the training model dataset is not large and diverse enough, the next work is likely to be blocked. To help

(a)   (b)   (c)
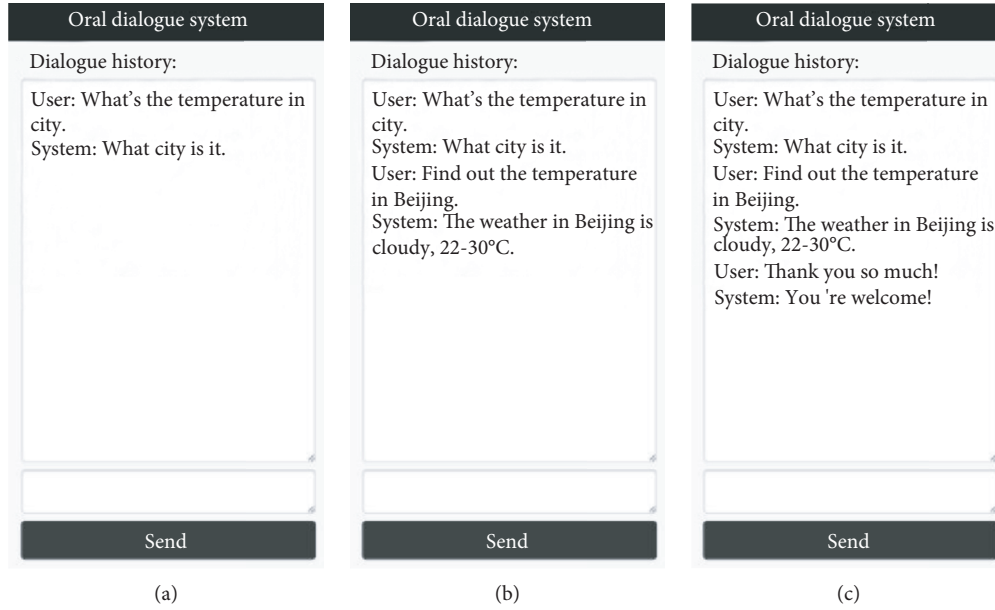
Figure 3: Schematic diagram of multiwheel interaction. (a) First round of dialogue. (b) Second round of dialogue. (c) Third round of dialogue.

alleviate this problem, the Stanford natural language processing group published a corpus. This dataset contains more than 3000 rounds of conversation data, mainly distributed in schedule, weather retrieval, and navigation. Because there is only one task in each conversation in KVRET dataset, in order to fit the reality, this paper reorganizes the KVRET dataset and obtains the conversation dataset containing multiple tasks. The reorganization method is as follows: two dialogue paragraphs of schedule, weather retrieval, and navigation are randomly selected for cross splicing, so that the spliced dialogue paragraphs contain two different tasks. The learning rate of shallow neural network is set to 0.064, the size of context window is set to 8, the dimension of word vector is set to 150, and the number of hidden layer neurons is 120. The number of training steps is set to 10 and the number of iterations is 100. In order to compare the performance of this method, it is compared with the spoken language understanding methods based on circular network, context information, and label decomposition. The experimental evaluation criteria are the accuracy and F1 value, which are widely used at present.

3.2.1. Experiment of Measuring Accuracy.

According to the experimental environment, taking 1000 rounds of training as an example, the accuracy of the four methods is calculated. The precision experiment comparison diagram of Figure 4 is obtained.

As can be seen from Figure 4, the test accuracy of these four methods is basically relatively stable and the fluctuation is small. The test accuracy of the three methods based on cyclic network, context information, and oral understanding of label decomposition is always higher than 90%, and the test accuracy is relatively stable. The test accuracy of the method studied in this paper remains above 94%, up to 97%,
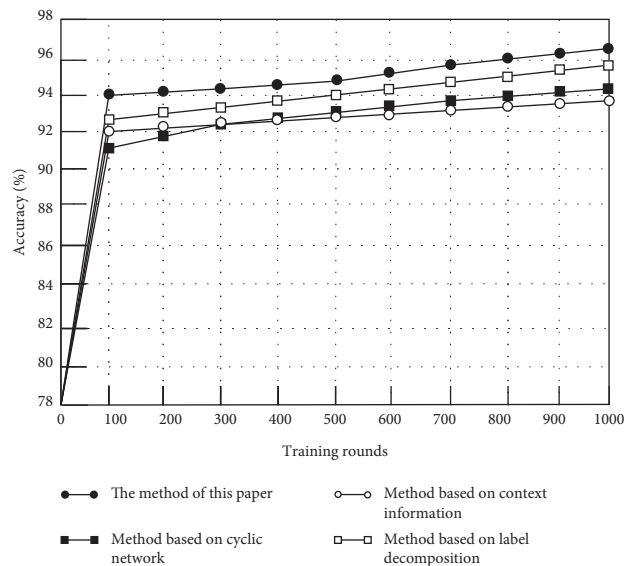
Figure 4: Comparison results of accuracy test.

which is always higher than the other three methods, indicating that the performance of this method is better. According to the accuracy comparison results in Figure 4, the comparison diagrams of the highest accuracy and the lowest accuracy of different methods can be drawn, as shown in Figure 5.

As shown in Figure 5, the highest accuracy of the method based on cyclic network is 93%, and the lowest accuracy is 91%. The highest accuracy of context information method is 93.5%, and the lowest accuracy is 92%. The highest accuracy of label decomposition method was 96%, and the lowest accuracy was 92.5%. The highest accuracy of this method is 96% and the lowest is 94%. Compared with the oral

FIGURE 5: Comparison of maximum and minimum accuracy of different methods.

FIGURE 6: Comparison results of F1 value test.

FIGURE 7: Comparison of the highest and lowest F1 values of different methods.

It can be seen from Figure 6 that the test F1 value of the method studied in this paper is higher than that of the oral comprehension method based on circular network, context information, and label decomposition. Basically, the F1 values of the four methods are relatively stable and have little fluctuation. According to the F1 value comparison results in Figure 6, the comparison diagrams of the highest F1 value and the lowest F1 value of different methods can be drawn, as shown in Figure 7.

As can be seen from Figure 7, the highest F1 value is 92% and the lowest F1 value is 91% based on the cyclic network method. The highest F1 value of context information method is 94%, and the lowest F1 value is 89%. The highest F1 value of label decomposition method is 92%, and the lowest F1 value is 90%. The highest F1 value is 95%, and the lowest F1 value is 91%. The highest F1 value is 3%, 1%. and 3% higher than the oral comprehension methods based on circular network, context information. and label decomposition, respectively. F1 value is an effective evaluation standard for comprehensive accuracy and recall, which can comprehensively reflect the performance of this method. Experiments can prove that this method has good application performance and has certain advantages.

Through experiments, it can be concluded that the method proposed in this paper has the highest accuracy of 97%, the highest accuracy of 96%, and the highest F1 value of 95%, which can realize man-machine oral English understanding and has a good application prospect.

4. Conclusion

Man-machine dialogue system is the concentrated embodiment of the level of artificial intelligence. As the core part of man-machine dialogue system, oral comprehension model is the focus and difficulty of research. This paper proposes an spoken language understanding method based on deep learning. The test results show that this method can

comprehension method based on circular network, context information, and label decomposition, the accuracy is 3%, 2.5%, and 3.5% higher, respectively. It shows that this method has high accuracy and good practical application performance.

3.2.2. Test of F1 Value. In order to better test the actual performance of the method in this paper, the F1 value of the four methods is tested. F1 value is an index used to measure the accuracy of binary classification model in statistics. Its maximum value is 1 and its minimum value is 0. The F1 value is tested in the form of percentage, and the comparison test results are shown in Figure 6.

significantly improve the accuracy and F1 value and has high practical value. The application research of oral comprehension is a complex and far-reaching topic, and there are still deficiencies in this paper. Due to the limitation of hardware equipment, too many rounds will lead to too large model parameters and failure to run. However, in actual situations, such as the communication between online customer service and users, there may be dozens or even hundreds of rounds of dialogue between them. How to solve the difficulty of multiple rounds of dialogue needs further research. The data of oral comprehension in this paper focus on the fact that each sentence contains only one intention. However, in real life, a sentence may contain multiple intentions, which needs further exploration.

Data Availability

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Conflicts of Interest

The authors declared that they have no conflicts of interest regarding this work.

References

[1] L. Hou, Y. Li, and C. Li, "Review of research on task-oriented spoken language understanding," *Computer Engineering and Applications*, vol. 55, no. 11, pp. 7–15, 2019.

[2] P. Qiao, "Automatic language generation simulation of two-way interactive robot," *Computer Simulation*, vol. 36, no. 4, pp. 310–314, 2019.

[3] A. Vanzo, D. Croce, E. Bastianelli, R. Basili, and D. Nardi, "Grounded language interpretation of robotic commands through structured learning," *Artificial Intelligence*, vol. 278, pp. 1–21, 2020.

[4] J. Zhang, H. Huang, Y. Hu, and Y. Wushour, "Modified recurrent neural networks in spoken language understanding," *Computer Engineering and Applications*, vol. 55, no. 18, pp. 155–160, 2019.

[5] X. Yang, J. Wang, L. Qi-yuan, and L. Shou-shan, "Intention detection in spoken language based on context information," *Computer Science*, vol. 47, no. 1, pp. 205–211, 2020.

[6] Y. Xu and H. Huang, "Spoken language understanding model based on label decomposition," *Computer Engineering*, vol. 45, no. 7, pp. 237–241, 2019.

[7] Q. Dong, H. Li, G. Cao, and L. Xia, "An exploratory posts detecting method for MOOC forums based on deep learning," *Library and Information Service*, vol. 63, no. 5, pp. 92–99, 2019.

[8] G Wang and X Huang, "Convolution neural network text classification model based on word2vec and improved TF-IDF," *Journal of Chinese Computer Systems*, vol. 40, no. 5, pp. 1120–1126, 2019.

[9] D Yang, Y Wu, and F. Chun-xiao, "Chinese short text keyphrase extraction model based on attention," *Computer Science*, vol. 47, no. 1, pp. 193–198, 2020.

[10] E. Lv, X. Wang, and Y. Cheng, "Deep convolution neural network learning based on deconvolution feature extraction," *Control and Decision*, vol. 33, no. 3, pp. 447–454, 2018.

[11] Q. Zhao, X. Cai, and L. Bo, "Text feature extraction method based on LSTM-Attention neural network," *Modern Electronics Technique*, vol. 41, no. 8, pp. 167–170, 2018.

[12] T. Xue, Y. Wang, and M Nan, "Convolutional neural network based on word sense disambiguation for text classification," *Application Research of Computers*, vol. 35, no. 10, pp. 2898–2903, 2018.

[13] H. Wei and L. Fan, "Feature extraction of binary documents using neural network," *Communications Technology*, vol. 12, pp. 2881–2887, 2019.

[14] J. Liu, Y. Li, and M. Lin, "Review of intent detection methods in human-machine dialogue system," *Computer Engineering and Applications*, vol. 55, no. 12, pp. 1–7, 2019.

[15] L. Hou, Y. Li, and M. Lin, "Joint recognition of intent and semantic slot filling combining multiple constraints," *Journal of Frontiers of Computer Science & Technology*, vol. 14, no. 9, pp. 1545–1553, 2020.

[16] S. C. Akkaladevi, M. Plasch, M. Hofmann, and A. Pichler, "Semantic knowledge based reasoning framework for human robot collaboration," *Procedia CIRP*, vol. 97, no. 5, pp. 373–378, 2021.

[17] A. K. Bosen and E. Buss, "Short-term audibility is a better predictor of vocoded speech-in-speech recognition than long-term target-to-masker ratio," *Journal of the Acoustical Society of America*, vol. 148, no. 4, p. 2465, 2020.

[18] Q Zhou and Z. Li, "BERT based improved model and tuning techniques for natural language understanding in task-oriented dialog System," *Journal of Chinese Information Processing*, vol. 34, no. 5, pp. 82–90, 2020.

[19] V. Kadyan, M. Dua, and P. Dhiman, "Enhancing accuracy of long contextual dependencies for Punjabi speech recognition system using deep LSTM," *International Journal of Speech Technology*, vol. 24, no. 2, pp. 517–527, 2021.

[20] M. Wang, D. Yu, Y. Rui, H Wenpeng, and Z. Dongyan, "Chinese multi-turn dialogue tasks based on HERD model," *Journal of Chinese Information Processing*, vol. 34, no. 8, pp. 78–85, 2020.

[21] Y. Xu and H. Huang, "Spoken language understanding method based on recurrent neural network with persistent memory," *Computer Engineering and Applications*, vol. 12, pp. 145–148, 2019.

[22] Z. Matteo, B. Luca, S. Ivan, and G. Alfonso, "Evaluating different natural language understanding services in a real business case for the Italian language," *Procedia Computer Science*, vol. 176, pp. 995–1004, 2020.

[23] M. Mcshane and S. Nirenburg, "Context for language understanding by intelligent agents," *Applied Ontology*, vol. 14, no. 4, pp. 1–34, 2019.

[24] Z. Yang, L. Wang, and Y. Wang, "Application research of deep learning algorithm in question intention classification," *Computer Engineering and Applications*, vol. 55, no. 10, pp. 154–160, 2019.

[25] J. Hu and H. Tao, "Design and implementation of domain question answering system based on deep learning," *Journal of Chengdu University Of Information Technology*, vol. 34, no. 3, pp. 232–237, 2019.

[26] N. Zhu, J. Dong, and Z. Zhang, "A human-computer dialogue model for digital reference consultation in library," *Library and Information Service*, vol. 63, no. 6, pp. 5–11, 2019.

[27] H. Akay and S.-G. Kim, "Measuring functional independence in design with deep-learning language representation models," *Procedia CIRP*, vol. 91, pp. 528–533, 2020.

[28] Z. Chen and H. Yang, "Yi language speech recognition using deep learning methods," in *Proceedings of the 2020 IEEE 4th*

Information Technology, Networking, Electronic and Automation Control Conference (ITNEC), IEEE, Chongqing, China, June 2020.

[29] M. Ali, M. L. Rahman, J. Chaki, N. Dey, and K. Santosh, "Machine translation using deep learning for universal networking language based on their structure," *International Journal of Machine Learning and Cybernetics*, vol. 3, pp. 1–12, 2021.

[30] A. H. Zadeh, Z. Poulos, and A. Moshovos, "Deep learning language modeling workloads: where time goes on graphics processors," in *Proceedings of the 2019 IEEE International Symposium on Workload Characterization (IISWC)*, November 2019.

[31] R. W. Filice, "Deep-learning language-modeling approach for automated, personalized, and iterative radiology-pathology correlation," *Journal of the American College of Radiology: JACR*, vol. 16, no. 9, pp. 1286–1291, 2019.

[32] L. Wu, D. Han, and Z. Du, "Assembly language and assembler for deep learning accelerators," *High Technology Letters*, vol. 25, no. 4, pp. 42–50, 2019.

[33] Y. Kim, J. H. Lee, S. Choi et al., "Validation of deep learning natural language processing algorithm for keyword extraction from pathology reports in electronic health records," *Scientific Reports*, vol. 10, no. 1, Article ID 20265, 2020.

[34] V. Sorin, Y. Barash, E. Konen, and E. Klang, "Deep learning for natural language processing in radiology-fundamentals and a systematic review," *Journal of the American College of Radiology*, vol. 17, no. 5, pp. 639–648, 2020.

Research Article

Design of Process Products Based on Image Processing Multimode Interaction

Wei Sun[1] **and Lijun Li** [2]

[1]*School of Digital Creation and Animation, Shenzhen Polytechnic, Shenzhen 518000, Guangdong, China*
[2]*Evergreen Institute of Elderly Education, Shenzhen Polytechnic, Shenzhen 518000, Guangdong, China*

Correspondence should be addressed to Lijun Li; muzi7728@szpt.edu.cn

With the development of modern science and technology and more and more image processing systems, related technologies are becoming more and more complex. The application of image processing technology can be seen in various fields of society, such as medical field, aerospace field and life, and entertainment field. Due to the increasing amount of information on the picture, the requirements for the speed and clarity of image processing are also increasing. The existence of various external factors will lead to the production of image products and objects between the error and distortion problems. In order to make the process product design more authentic and reliable, this paper studies the process product design based on image processing multimode interaction. It uses radiometric correction and geometric correction to process distorted images and uses GPU parallel computing technology to accelerate the correction process. In this paper, this technology is applied to the visual recognition of welding robot, and the experiment shows that the product produced by the image processed by this module can obviously reduce the error.

1. Introduction

Image refers to the general information carrier existing on paper, photo, TV, or computer screen. Image processing technology refers to the process of image collection, processing, processing, and storage. Most images are stored in computer in digital mode, and image processing is mainly aimed at digital image processing. With the rapid development of science and technology, the amount of information accepted by human beings has expanded rapidly, and the research and application of image processing have also increased. In the 1960s, the United States used image processing technology to process nearly 100,000 photos of spacecraft returning to Earth, which laid the foundation for the lunar landing plan of the United States. In the 1970s, Britain applied image processing technology to human medicine and obtained sectional images of various parts of human body. Now, in all fields of social production, such as space remote sensing, agricultural production, scientific research, and other fields, we can see the application of

image processing technology, and image processing technology is also concerned as a whole. With the increase of image information, the requirements of image processing speed and resolution are getting higher and higher. In many fields, image imaging is affected by many factors. External factors are weather factors, illumination angle, exposure, and so on. There are internal factors, such as low precision of camera sensors and image transmission errors. These are errors and distortions between the produced image products and the actual target objects. These errors and distortions are not conducive to users' feelings and evaluation of image products. In order to eliminate the image distortion, this paper uses efficient image correction technology to study the method of dealing with the image distortion and reducing the error between the image and the real target and devotes itself to displaying practical and reliable image data to users [1].

Multimodal interaction refers to the comprehensive interaction between products and users [2]. However, the innovation of products has changed from technology type to

user perception and human-computer interaction. Exactly, multimodal interaction can reflect this perception and experience. At present, we propose a design method based on human-computer interaction and TRIZ cooperation. It includes the whole process from requirement analysis to scheme formation and provides a user experience view. After verification, it shows its effectiveness and helps product innovation. Internet of things devices have been integrated into the public physical environment. It is subject to the application and voice assistant as the interface [3], which makes its expansion performance poor. On this basis, the researchers decompose the user's IOT commands into two components, selection and interaction in the design space, and then clarify the possibility that they go beyond the current way. The design is transformed into a perception platform, showing a novel interactive scene of the Internet of things and solving the fuzzy direction of messy devices. Nowadays, a wide range of computing and communication resources can support us to think about sound in a positive way [4]. Acoustic interaction design belongs to a new field and is in a cross field. Sid is the abbreviation of acoustic interactive design. Its application has the artistic significance from alarm to music creation. It emphasizes the interactive system of auditory mode and is realized by the integration of computing, communication, and interactive technology. Background augmented reality classroom has become meaningful [5], and its limitations are mainly in two aspects: first, the inconvenience caused by using cards to operate experiments and second, single-mode interaction lacking the sense of real interaction. The proposed multimodal interaction algorithm based on augmented reality (argev) and the virtual reality fusion interaction tool kit (vrfits) enhance the visual and tactile feedback. By comparing the availability of single-mode and multimode interaction and existing TV remote control applications [6], we find out the appropriate interaction mode. Through their tests, they prefer to adopt unimodal interaction design in TV remote control applications, but due to interface problems, favorable interaction design involving brush gestures will replace this design. The handicraft design in the traditional design is manual training to achieve a certain technical height, but there are skilled operators; handicraft design defective rate is high, resulting in the cost of handicraft design. It affects the large-scale production of products, and the process of the same batch of products is different. This paper puts forward a design method of handicrafts based on image processing technology, thus improving the design accuracy of handicrafts.

2. Introduction of Related Theories and Technologies

In digital image processing, generally, the image information collected by image sensors is transmitted to computers or other devices, and a series of processing is carried out on this image by means of software or hardware, so as to highlight the interesting feature information in the image. Digital image processing technology can be seen everywhere in our lives, such as aerospace, consumer payment, artificial intelligence, intelligent transportation, and medical imaging, which plays a great role in promoting the progress of human civilization.

This section introduces digital image processing and FPGA technology in detail before using image processing algorithm to identify PCB defects.

2.1. Representation Method of Related Images. $F(X, J)$ is often used to represent a gray image, where $F(X, Y)$ is a two-dimensional mathematical matrix, (X, Y) represents the pixels in this gray image, and the value of a single $F(X, Y)$ represents the gray value at the position (X, Y). Most signals in nature are analog quantities, and digital image processing is a process of processing digital quantities, so it is generally necessary to transform analog quantities into digital quantities, and a gray image is usually represented by matrix and binary values.

2.1.1. Array Representation of Gray Image. Let the analog image become a digital image $F(X, Y)$ after a series of processing, which is usually expressed in the matrix form of $M * N$ size and can also be expressed in the matrix size of $N * N$), which is expressed by

$$f(x, y) = \begin{bmatrix} f(0,0) & \cdots & f(0, N-1) \\ \vdots & \ddots & \vdots \\ f(M-1, 0) & \cdots & f(M-1, N-1) \end{bmatrix}. \quad (1)$$

In a digital image, a pixel is the value of each element in the matrix $f(x, y)$.

2.1.2. Binary Image Representation. In digital images, the values of matrix $F(X, Y)$ generally have only two values, such as logic 0 and logic 1 in logic circuits, and there are only two possibilities. For digital image processing, gray images are often changed into binary images for image processing, which can greatly reduce the amount of computation. The characteristic of binary image is that it does not occupy too much memory and is convenient to store. The disadvantage is that when describing complex things, it can only outline its rough outline and cannot describe it in detail [7].

2.2. Basic Algorithms of Digital Image Processing. In the external natural environment, when the image sensor is used to collect the image to be detected, due to external interference, making the images we collected, more or less, it will not be so clear. There may be certain noise, low image contrast, and other characteristics. These phenomena will cause great interference to our image recognition. In order to get the feature information of the image to be detected better, it is necessary to carry out a series of image preprocessing on these collected images first, so as to enhance the feature information of the image to be detected to detect the related defects of the image to be detected more quickly and accurately [8].

2.2.1. Image Enhancement Algorithm. In digital image processing, image contrast enhancement is usually carried out in spatial domain and frequency domain. Histogram equalization is a digital image processing algorithm in spatial domain. Its main implementation process is to change the overall gray value of the image by changing the distribution of the gray histogram. Histogram equalization algorithm is a common digital image processing algorithm [9], and its basic principle is as follows.

Let an image have L gray levels and n pixels, and n_k is the number of pixels of gray value f in image $f(x, y)$; then the probability of gray level k is

$$p(f_k) = \frac{n_k}{n}. \tag{2}$$

The realization of histogram equalization is to change the gray value of the original image through mapping function. If T represents its mapping function, histogram equalization can be expressed as follows:

$$S_k = T(f_k) = \frac{L-1}{n} \sum_{i=0}^{k} n_i \quad 0 \le k \le L-1, \tag{3}$$

where f_k is the gray level of the original image, S_k is the equalized gray level, N is the total number of pixels in the image, and L is the total gray level in the image.

2.2.2. Median Filtering Algorithm. In the external natural environment, when the image sensor is used to collect the image to be detected, because of the external interference, the images we collected will be more or less clear and may have certain noise, low image contrast, and other characteristics. These phenomena will cause great interference to our image recognition, so we must first filter the collected images. Median filtering is an image processing algorithm used to remove noise. It can carry out a series of processing procedures on images containing noise to suppress noise. Generally, the median value in the neighborhood space of the template is used to replace the gray value at the center of the template [10], and its mathematical formula is shown in

$$f(x, y) = median\{g(x-i, y-i)\}(i, j) \in W, \tag{4}$$

where g(x, y) is the gray value of the image before processing (x, y) and $f(x, y)$ is the gray value of the image after processing (x, y). For the specific implementation of median filtering, sliding window is very important for median filtering. The size of sliding window is usually 3×3 or 5×5. As shown in Figure 1, it is a specific filtering process of 3×3 sliding window.

2.2.3. Edge Detection Algorithm. In digital image processing, the types of image edges mainly include three types as shown in Figure 2.

The types of image edges and the variation law of derivatives are shown in Figure 2. In digital image processing, in order to describe the edges of images, derivatives in mathematics are often used to express them, which are
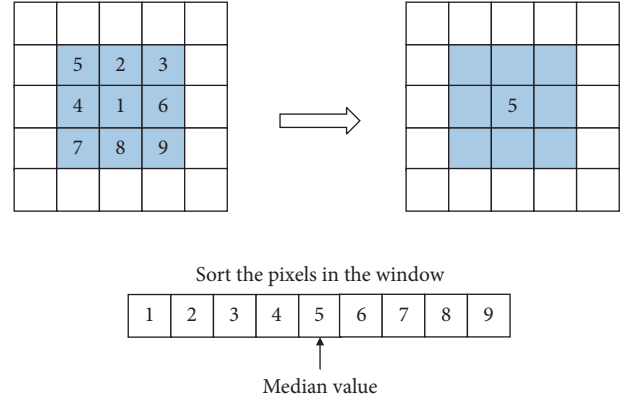
Figure 1: Implementation framework of median filtering algorithm.

generally obtained by calculating the first and second derivatives of image matrix functions. In digital image processing, edge detection is equivalent to derivation of two-dimensional function. In mathematics, derivation of discrete function is equivalent to derivation of differential operator. Therefore, the derivative gradient vector in digital image processing can be defined as follows:

$$G(x, y) = \begin{pmatrix} G_x \\ G_y \end{pmatrix} = \frac{\partial f / \partial x}{\partial f / \partial y}. \tag{5}$$

The direction and amplitude of the gradient are expressed as

$$\alpha(x, y) = \arctan\left(\frac{G_y}{G_x}\right),$$

$$|G(x, y)| = \sqrt{G_x^2 + G_y^2}. \tag{6}$$

In general, for digital image processing, in order to facilitate calculation, the formula of gradient amplitude is simplified:

$$|G(x, y)| = |G_x| + |G_y|. \tag{7}$$

However, in the actual image processing, some directional template operators are usually constructed, and it is easy to obtain derivatives through these directional template operators. Generally, Sx and Sy are used to represent horizontal and vertical gradient operators. According to the size of template and the value of element, Sobel operator is the most commonly used edge detection operator. In digital image processing, Sobel edge detection operator often uses convolution operation in mathematics to realize edge detection algorithm. Generally, it includes template operators in horizontal and vertical directions. The gradient operator templates in these two directions are shown in Figure 3.

G and G are used to represent the edge detection gradient in horizontal direction and vertical direction. In mathematics, the formula can be expressed as in Figure 4:
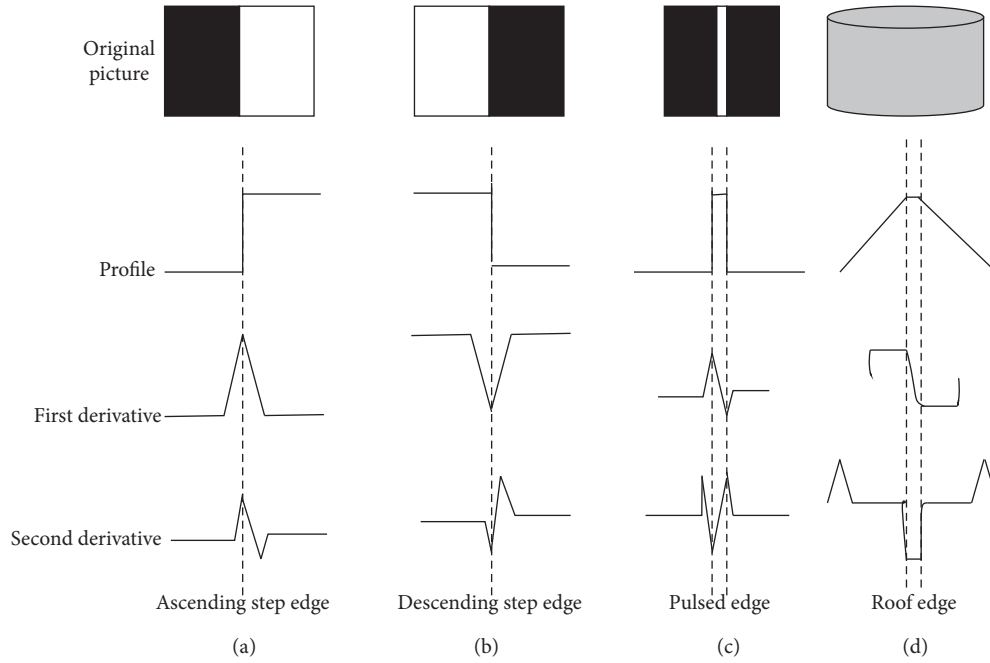
FIGURE 2: Example of image edge type and derivative curve law.

FIGURE 3: Horizontal and vertical templates of Sobel operator.

FIGURE 4: Schematic diagram of 3×3 data block of original image.

$$G_x = (P3 + 2P6 + P9) - (P1 + 2P4 + P7),$$
$$G_x = (P1 + 2P2 + P3) - (P7 + 2P8 + P9). \tag{8}$$

2.2.4. Algorithm of Target Recognition. There are three commonly used target defect detection methods, one is reference method, the other is nonreference method, and the third is mixed method. The most commonly used method is the reference method [11]. The reference method generally compares the pixel values of the image to be tested with those of the standard image, so as to detect the defects of the image to be tested. This method is relatively simple, but the detection speed is fast and easy to complete; however, the nonreference method is to detect interesting things through the feature size of objects or some unique design criteria, so it is difficult to realize this method, and it can only identify objects through some feature information, which requires more complex detection algorithms. Mixed method is a combination of reference method and nonreference method, which can achieve twice the result with half the effort. But for the specific actual situation, the most commonly used method is the reference method.

Background subtraction algorithm is a commonly used target recognition algorithm to realize reference method, which is widely used in the related fields of image detection and recognition. Background subtraction method is to subtract the current image and the background image point by point based on pixels, take their absolute values, and binarize them with thresholds, so that the interested objects in the current image can be determined. It can also be used for defect detection in practice in Figure 5.

The concrete implementation of the background subtraction algorithm is shown in Figure 5. $B(X, Y)$ is a two-dimensional matrix function processed in advance, FN (X, Y) is a two-dimensional matrix function of the current image, and a two-dimensional matrix function DN can be obtained by processing the two-dimensional matrix function of the background image and the two-dimensional matrix function of the current image by

$$D_n(x, y) = |f_n(x, y) - B(x, y)|. \tag{9}$$

A two-dimensional matrix function R can be obtained by an appropriate threshold T and corresponding operation according to the following formula:

$$R_n = \begin{cases} 255, & D_n(x, y) \geq T \\ 0, & \text{others} \end{cases}, \tag{10}$$
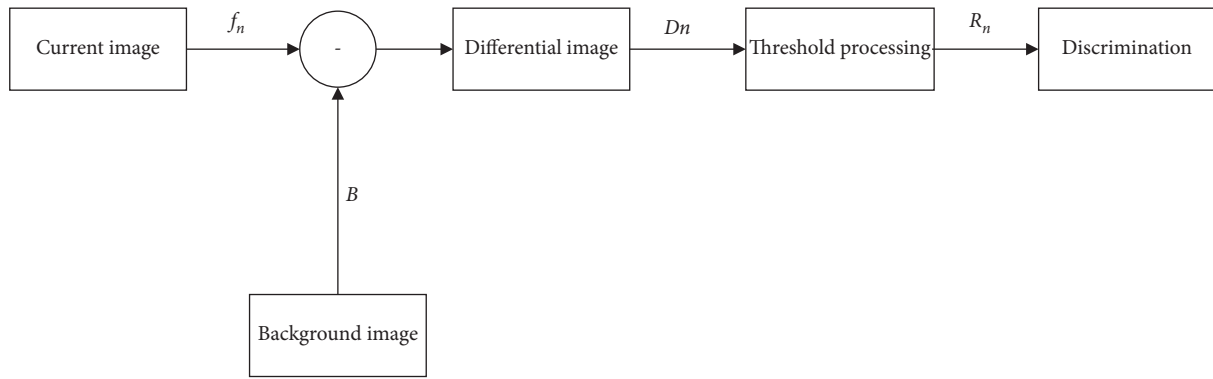
Figure 5: Operation process of background subtraction algorithm.

where 255 stands for the target and 0 stands for the background, the interesting target area in the image can be obtained finally. There are four processes in target recognition using background subtraction: obtaining background image, updating background image, obtaining target area, and extracting feature information.

3. Welding Image Processing

3.1. Welding Image Preprocessing. In the process of generation and transmission, the pretreated weld image will be affected by imaging system and transmission medium, which will lead to the degradation of image quality, resulting in image blurring, distortion, noise, and so on. At this time, it is necessary to introduce image processing technology to preprocess the weld image, so as to make the weld image clearer.

3.1.1. Fuzzy Welding Image Processing. When the camera takes pictures, the front end of the welding gun moves faster than the weldment, which is easy to cause motion blur, which brings a lot of inconvenience to the acquisition of weld image. Using motion blur image restoration technology, recognizable weld image can be obtained.

3.1.2. Improving the Contrast of Weld Image. After obtaining the weld image, the weld profile may not be obvious, which is not conducive to extracting the important information of the weld. In order to make the contour of weld image obvious and discriminate, it is necessary to improve the image contrast. Commonly used methods are gamma value adjustment and stretchlim function method. For this experimental environment, gamma value adjustment method has better effect.

3.1.3. Gray Scale Processing. The weld image obtained by industrial camera is a RGB three-channel color image, which requires a long calculation time. Gray processing can change the color image of RGB three-channel data into a single-channel data image with only brightness information but no color information. After gray-scale processing, the main contour information of the image is not affected, which can improve the speed of image processing and reduce the amount of data processed.

3.1.4. Image Noise. In gray image processing, image noise should be considered firstly, which is caused by spatter, smoke, and welding electrical noise in the welding process. In addition, noise will also be generated when the weld image is stored, processed and transmitted. Noise is the pixel value of the obtained image, which cannot reflect the real scene brightness and will affect the accurate understanding of image information. There are two typical noises in digital image processing: Gaussian noise and salt and pepper noise [12].

(1) Gaussian Noise. Gaussian noise is based on the normal distribution of probability density function caused by poor illumination and high temperature. Its amplitude follows Gaussian distribution, but its power spectral density is also evenly distributed.

(2) Salt and Pepper Noise. Salt and pepper noise is caused by interference of image signal, which is random white dot or black dot and may be white dot noise or white image on black image.

3.1.5. Image Denoising. The existence of noise has an important feature that makes the image flooded. Usually, the image obtained by using image emphasis technology and image filtering technology is processed by noise reduction, which reduces the influence of multiple interference signals and improves the image quality after processing. Image noise removal mainly refers to suppressing or removing the noise of the target image while using image filtering technology to keep the details of the image as much as possible. General image noise removal methods include frequency domain filtering and spatial domain filtering. Frequency domain filtering enhancement processing does not process the input image directly, but processes the converted image, which has a large amount of computation and poor real-time performance and is rarely used in beacon tracking system using machine vision system. Therefore, spatial region filtering is usually selected to process images, and the most

common processing methods are average filtering and median filtering.

(1) Mean Filtering. Typical random noise is caused by sharp changes in image gray level. Mean filtering can reduce sharp changes in image gray level by replacing pixel values with neighborhood mean value, thus reducing image noise. $3 * 3$ template smoothing filter, $5 * 5$ template smoothing filter, $7 * 7$ template smoothing filter, and $9 * 9$ template smoothing filter are used for average filtering.

It is found that $5 * 5$ template smoothing filter has a good effect on removing Gaussian noise and can retain more details. For salt and pepper noise, its amplitude is basically the same, but the position of noise points is random, and the average value of noise in statistical sense is not 0, so the filtering effect of average filter on salt and pepper noise is not ideal [13].

(2) Median Filtering. Median filter is a nonlinear signal processing method, which is insensitive to outliers and can reduce the influence of outliers without reducing image contrast. The template of median filtering is usually a square adjacent area, and the median point is the target pixel point. When calculating, the gray value of the surrounding points is sorted, and the median value obtained becomes the filtering result of the central point. By this method, the median filtering of the whole image is realized for each pixel point, and better smoothing processing is obtained.

The research shows that if the value filter of $7 * 7$ template is selected, the filtering of pepper noise is the most effective, and the important information of image edge can be kept to the maximum extent.

3.1.6. Gray Scale Transformation Enhancement.
The contrast of the image becomes worse, and the distribution of the histogram of the image becomes uneven. The main pixels are concentrated in the vicinity of several pixel values. Gray scale linear transformation every pixel stretches linearly, gray scale value is transformed into new gray scale value by certain mathematical transformation formula, and the dynamic range of gray scale changes. If the contrast of gray image is enlarged, the display effect of the image will be improved, and the image quality will be improved. If the gray scale range of the original image $f(x, y)$ is set to $[a, b]$ and the gray scale range of the converted image $g(x, y)$ is linearly expanded to $[c, d]$, the gray scale value $f(x, y)$ of any point in the image is converted to $g(x, y)$. The formula is as follows:

$$g(x, y) = \frac{d - c}{b - a} \times [f(x, y) - a] + c. \tag{11}$$

If the gray levels of most pixels in the image are distributed in the interval $[a, b]$, the maximum gray level of the original image is maxf, and only a small part of the gray levels exceed this interval. In order to improve the image enhancement effect, you can make

$$g(x, y) = \begin{cases} c & 0 \le f(x, y) \le a \\ \dfrac{d - c}{b - a} \times [f(x, y) - a] & a \le f(x, y) \le b \\ d & b \le f(x, y) \le \max f \end{cases} \tag{12}$$

There are images after direct gray processing. The gray scale in [0, 1] is stretched to [0.1, 0.5] by local stretching of the image. The image is stretched locally, and the gray scale in [0, 1] is stretched to [0.3, 0.5]. There are images after direct gray processing and images after stretching gray range.

Compared with the images with linear gray-scale changes, it is found that the images with direct gray-scale processing are not obvious in light and shade, and the contrast is not high. After the gray level changes linearly, the contrast between light and shade of the image is enhanced, which makes the weld clear and the contour obvious, effectively improves the visual effect of the image, and brings convenience for the subsequent processing.

3.2. Image Binary Processing and Scatter Removal

3.2.1. Image Binarization Processing.
The image after gray-scale processing is binary, which can make the weld image only black and white, so that it is convenient to identify the edge features of the weld. Image binarization processing mainly sets the gray value of the image and divides the image through a certain gray value. The expression of binarization processing is

$$g(x, y) = \begin{cases} 255, & if \ f(x, y) < T \\ 0, & if \ f(x, y) > T \end{cases} \tag{13}$$

The algorithm of this expression is as follows: given a threshold T, when the gray value is greater than T, it is white, and when it is less than T, it is black. After processing, the two-dimensional matrix is only composed of 0 and 1, and the gray-scale image will become a special gray-scale image with only black and white.

In the image after binary processing, the contour of weld edge is discontinuous and the image feature is not obvious, so the bimodal method is used for image processing. This method considers that the image is composed of foreground and background, and the gray distribution of the image can be considered as the superposition of two normal distribution functions, so there will be two separated peaks in the histogram of the image, and the trough between the two peaks is the image reading value.

3.2.2. Scattered Removal.
After image denoising, gray transformation enhancement, and binarization, the weld image basically meets the requirements of edge detection, but there are some problems in the welding process, such as rough welding surface, welding slag splashing, and so on, which will make the collected weld image appear. The light spots with large area are not related and distributed

irregularly, which are usually called scatter spots. The area of scatter points is much larger than that of noise points. But it is much smaller than the area of continuous weldments on both sides of the weld. At this time, it is necessary to use image algorithm to remove these large scatter points, that is, calculating the area of each bright spot area in the image and then judging whether the area is large enough. If the area is small, it can be considered as scatter points and removed directly, so that the final output image is only the weld and weldment.

After the scatter points are removed, the image can be reversed, and the black part can be changed into white, while the white part can be changed into black, which does not affect the effect of the weld image and is convenient to compare with the original weld image and extract the edge contour of the weld.

After the image is inverted, it can be seen that the white part in the middle is the weld and the black part is the workpiece to be welded on both sides. Compared with the original weld image, we can see that the weld area is basically the same as the original weld area, which proves that the image processing flow is feasible and the method is correct.

3.3. Edge Detection and Centerline Extraction.

In the image, the region boundary with sharp change of gray value is called edge, and the local features of edge image are discontinuous, and there are sudden variation of gray level and sudden change of texture structure. Edge detection is an important research direction in the field of machine vision and image processing, and it is an important feature extraction method. The commonly used edge detection is Roberts operator detection, Sobel operator detection, Prewitt operator detection, Cany operator detection, and Logo operator detection.

Comparing the results of edge detection, it is found that Prewitt operator edge detection and Sobel operator edge detection cannot meet the requirements, the edge contour is not clear, and the complete weld cannot be extracted. It is found that the contour of the weld seam is consistent between the image processed by edge detection and the original image collected, which proves that the flow of weld seam image processing can be carried out.

The ultimate goal of the seam tracking system is to control the welding gun to move along the seam centerline, so that the welding has real time and high accuracy, so it is particularly important to accurately extract the seam centerline. In this paper, the image skeleton method is used to extract the center line, which takes the center of the target as the benchmark, refines the target, and extracts the center line of the part of the figure, thus removing unnecessary information and facilitating recognition.

4. Experiment

4.1. Trajectory Extraction Experiment.

Through the seam edge detection algorithm based on directional wavelet transform in Section 3, the seam edge image is obtained, and the seam trajectory image is extracted by morphological image processing method. Seam trajectory extraction experiments are carried out on different shapes of welds such as straight lines, broken lines, and curves, and the experimental results are as follows:

(1) The weld trajectory extraction experiment of straight weld is carried out on the weld trajectory extraction experiment of straight weld, and the experimental results are shown in Figure 6.

(2) Experiment on extraction of weld trajectory of 90 broken line weld: the weld trajectory extraction experiment is carried out on the 90 broken line weld, and the experimental results are shown in Figure 7.

(3) Experiment of extracting weld trajectory of broken line weld greater than 90: the weld trajectory extraction experiment is carried out on the broken line welds greater than 90, and the experimental results are shown in Figure 8.

(4) Experiment of extracting weld trajectory of broken line weld less than 90: the weld trajectory extraction experiment is carried out on the broken line welds less than 90, and the experimental results are shown in Figure 9.

(5) Experiment of weld trajectory extraction of regular arc weld: the weld trajectory extraction experiment is carried out on regular arc weld, and the experimental results are shown in Figure 10.

(6) Experiment of weld trajectory extraction of irregular arc curve weld: the weld trajectory extraction experiment is carried out on the irregular arc curve weld, and the experimental results are shown in Figure 11.

4.2. Experiment and Error Analysis of Center Line Extraction.

Through the above experiments, the weld trajectory image is obtained, and the weld trajectory image is further processed. The weld trajectory centerline is extracted by thinning method, and the weld trajectory centerline image and coordinates are obtained. The weld trajectory centerline extraction experiments are carried out for different shapes of welds, such as straight line, broken line, and curve. In order to verify the accuracy of the algorithm for extracting the centerline of weld trajectory, the error analysis of the extracted centerline of weld trajectory with different shapes is carried out. The actual weld width measured by microscope is about 0.25 mm. The industrial camera selected by the platform is 2 million pixels, its resolution is 1624×1240 pixels, and the corresponding actual working area is 80 mm x 60 mm. Therefore, the actual distance between every two pixels can be calculated to be about 0.05 mm, which is the design accuracy of extracting the center line of weld trajectory, which is far less than the actual weld width of 0.25 mm, and can meet the design requirements. From the above analysis, it can be concluded that the extraction accuracy of weld trajectory centerline is related to the resolution of the selected industrial camera. Improving the resolution of industrial camera can improve the extraction

(a)　　　　　　　　　　(b)　　　　　　　　　　(c)
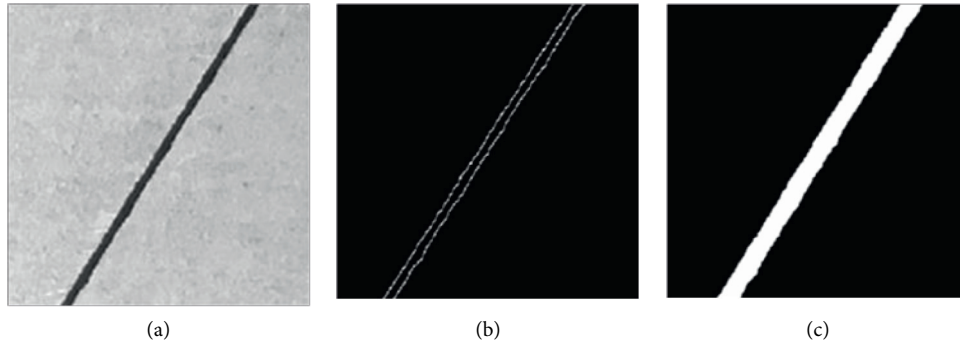
FIGURE 6: Direct weld. (a) Original image of weld. (b) Image of weld edge. (c) Image of weld trajectory.

(a)　　　　　　　　　　(b)　　　　　　　　　　(c)
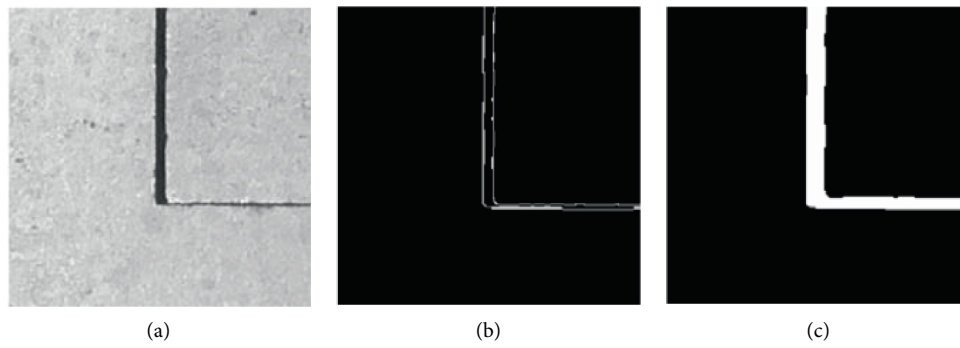
FIGURE 7: 90-degree broken line weld. (a) Original image of weld. (b) Image of weld edge. (c) Image of weld trajectory.

(a)　　　　　　　　　　(b)　　　　　　　　　　(c)
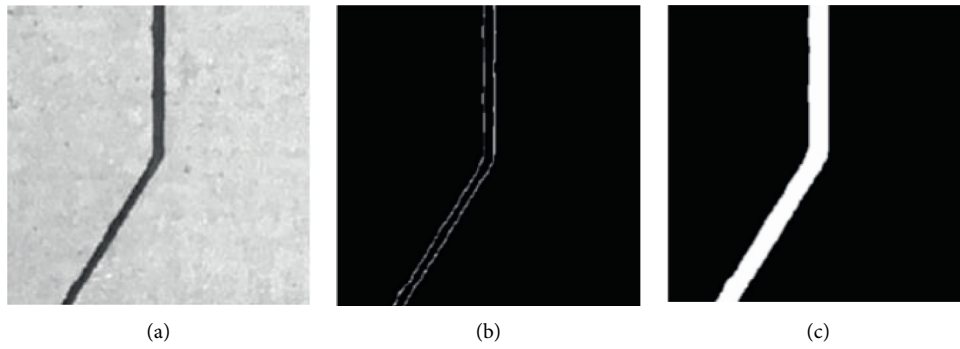
FIGURE 8: Broken line weld greater than 90°. (a) Original image of weld. (b) Image of weld edge. (c) Image of weld trajectory.

(a)　　　　　　　　　　(b)　　　　　　　　　　(c)
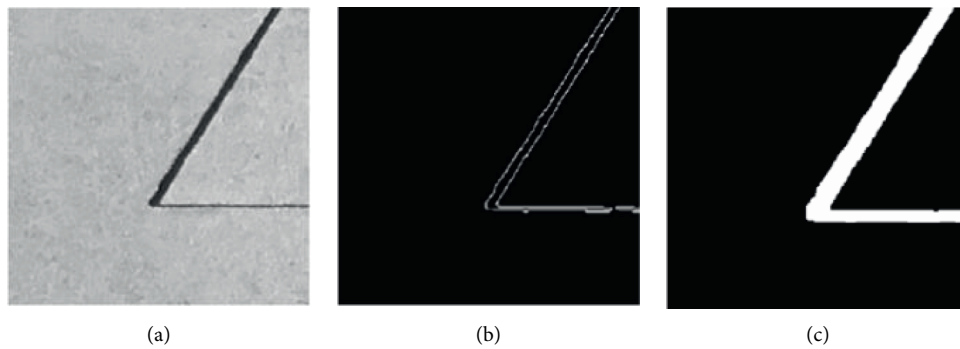
FIGURE 9: Broken line weld less than 90°. (a) Original image of weld. (b) Image of weld edge. (c) Image of weld trajectory.

(a)   (b)   (c)
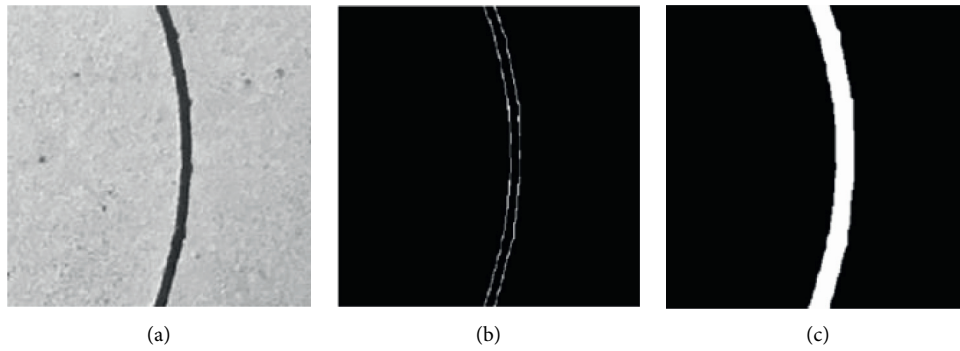
FIGURE 10: Regular arc weld. (a) Original image of weld. (b) Image of weld edge. (c) Image of weld trajectory.

(a)   (b)   (c)
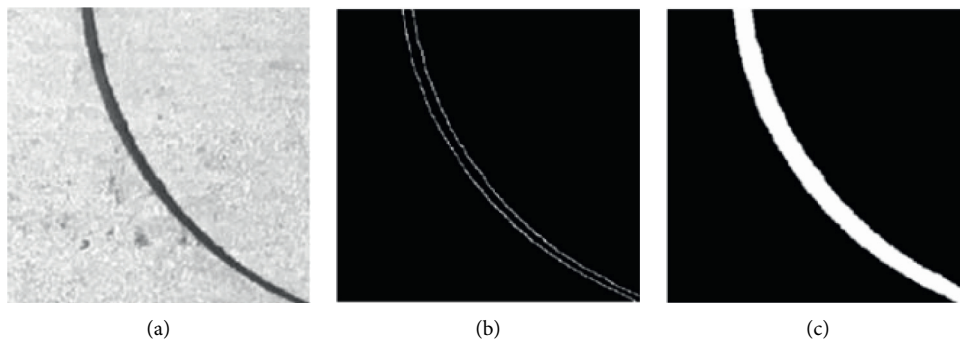
FIGURE 11: Irregular arc weld. (a) Original image of weld. (b) Image of weld edge. (c) Image of weld trajectory.
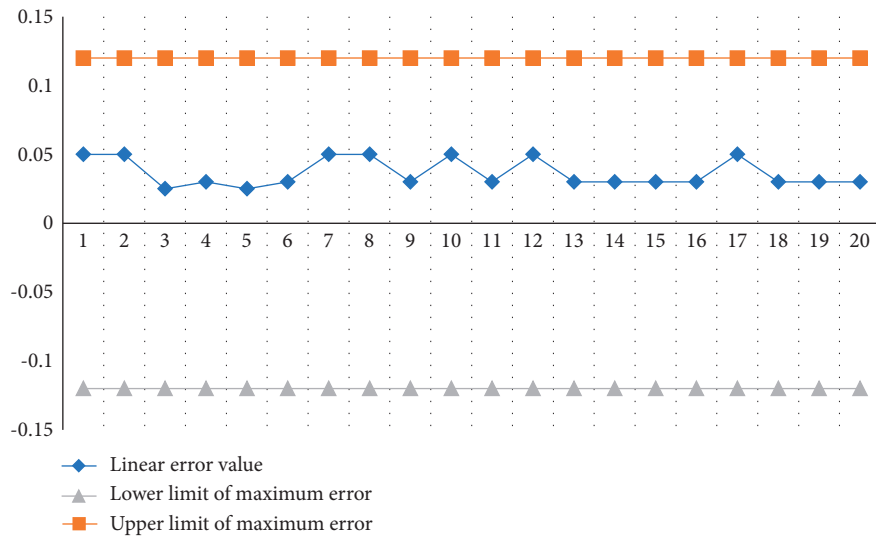
FIGURE 12: Center error of direct welding trajectory.

accuracy of weld trajectory centerline. Take 20 coordinate points on the center line of each weld track. Error measurement is carried out with the weld center of the original weld image. The extraction results of the weld trajectory center line and the measured error curves are shown in Figures 12–17. The red line in the error curve indicates that the maximum error value range is 0.12 mm, and the blue

curve indicates the error curves of the weld trajectory center lines with different shapes.

It can be seen from the error graphs of the center line of weld trajectory with different shapes.

For the weld image with cusp, such as right angle, more than 90° broken line and less than 90° broken line, there is a large error in the center line of the weld track at the cusp. The
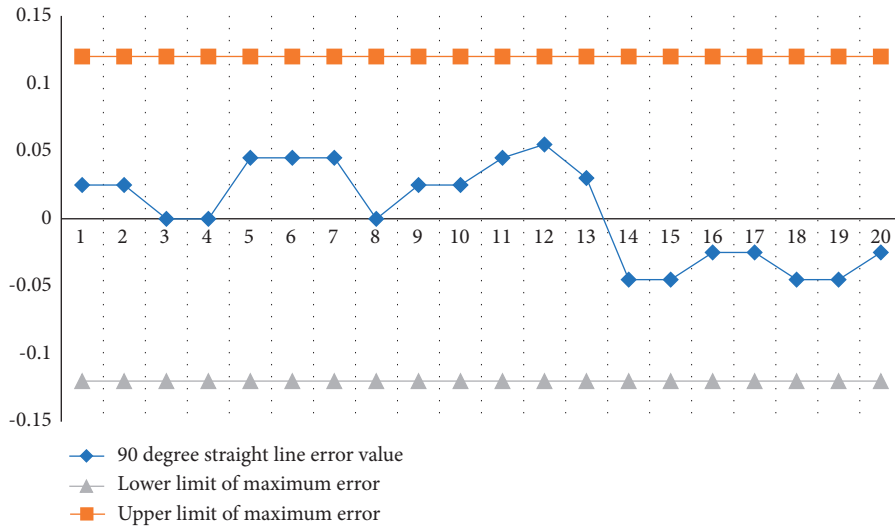
Figure 13: 90-degree welding trajectory center error.
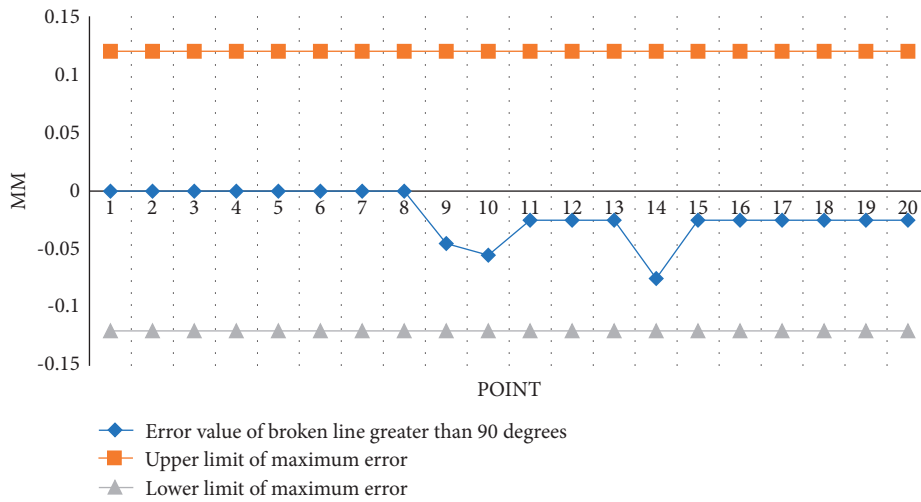
Figure 14: Center error of welding trajectory greater than 90°.
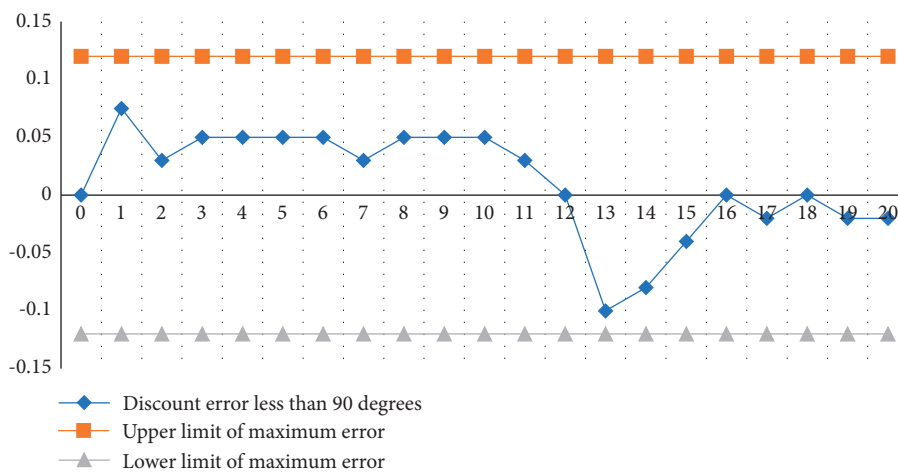
Figure 15: Center error of welding trajectory less than 90°.
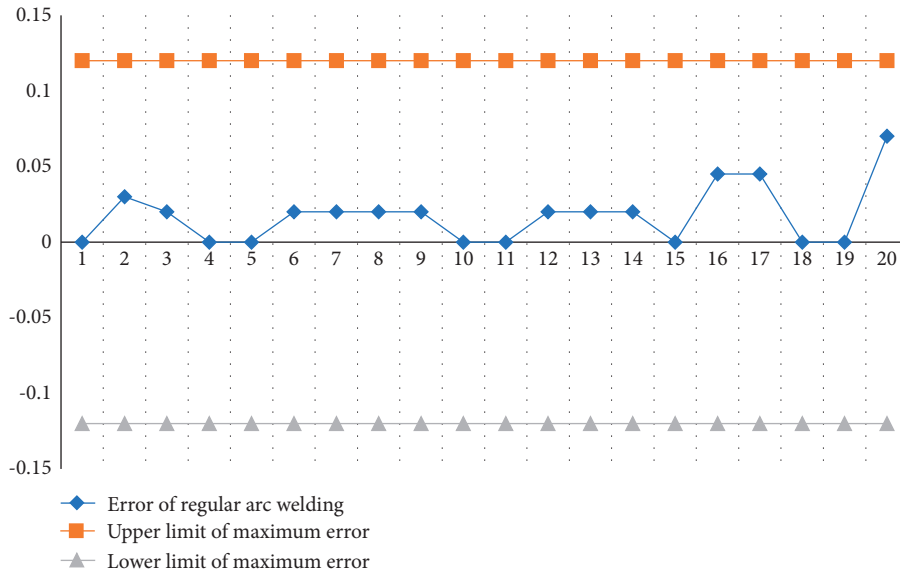
FIGURE 16: Center error of regular arc welding trajectory.
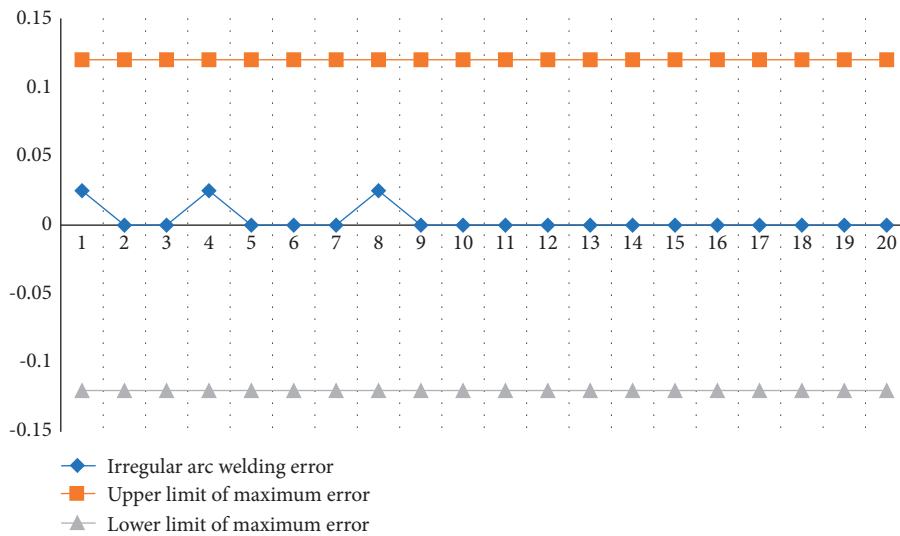
FIGURE 17: Center error of irregular arc welding trajectory.

maximum error value is 0.105 mm, which is less than the maximum allowable error value of 0.12 mm, which is generated from the center line of the weld track of less than 90°broken line. For weld images without cusps, such as straight and curved welds, the error fluctuation range is small, which can be controlled within the range of +0.07 mm. Through the error analysis of weld trajectory centerline with different shapes, it can be verified that the errors produced by the extraction algorithm of weld trajectory centerline studied in this paper are all less than the maximum allowable error, which meets the system accuracy requirements.

Finally, in order to prove that the welding robot using this image enhancement technology has a lower error rate, we have done two groups of experiments, which are the center line error in straight and curved welds in Figures 18 and 19.

Through comparative experiments, it is found that the accuracy of process products based on image processing has been greatly improved, especially the weld error before enhancement has exceeded the standard value in several points, which is not allowed for process products, which further shows that our technology has brought a breakthrough in the design of process products. At the end of the article, some conclusions are given without in-depth analysis. Image enhancement technology can improve the accuracy of products in process product design. The improvement of accuracy brings a new direction to the development of process technology, continuously integrates into high-precision equipment, does not rely on imports,
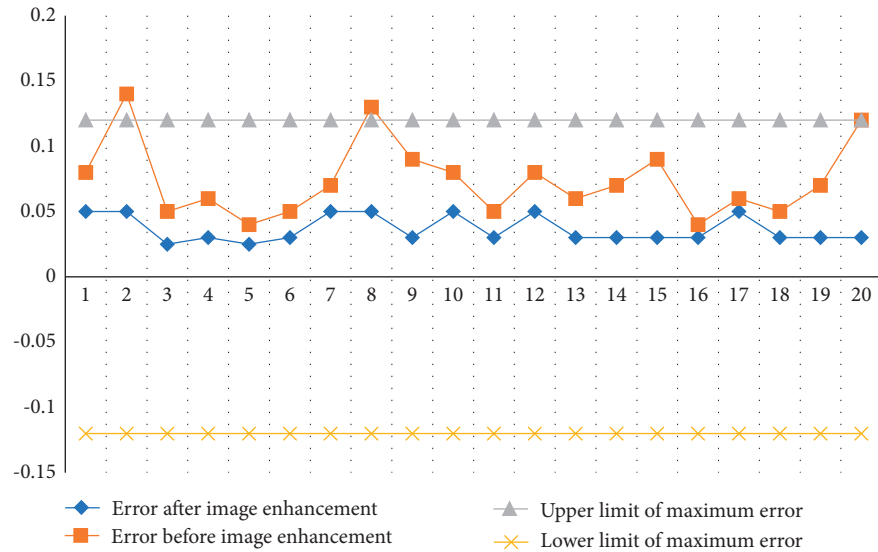
FIGURE 18: Contrast chart of linear welding error.
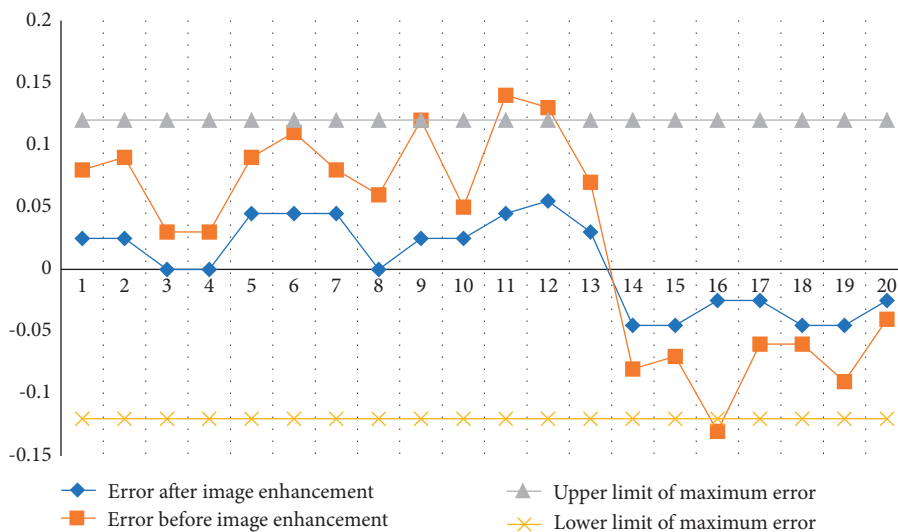
FIGURE 19: Comparison chart of curve welding error.

and realizes technological breakthroughs and opens up new ideas.

5. Conclusion

Due to the increasing amount of information on images, the requirements for the speed and clarity of image processing are also increasing. The existence of various external factors will lead to errors and distortions between the produced image products and the target objects. In this paper, the process product design based on multimode interaction of image processing is studied. It uses radiometric correction and geometric correction to process distorted images, and GPU parallel computing technology is used to accelerate the correction process. In order to validate the effect of image enhancement technology in this paper on improving process product design, it is applied to

the visual recognition of welding robot. As the most important visual recognition of welding robot in welding is the first process, the image quality of recognition is directly related to the follow-up process. Finally, through experiments, it is found that the error value of the center line extracted by welding robot through image enhancement is less than the industry requirements. It shows that the image enhancement technology based on image processing studied in this paper can improve the design of process products and make the design of process products more authentic and reliable.

Data Availability

The experimental data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] F. M. Bublitz, A. Oetomo, K. S. Sahu et al., "Disruptive technologies for environment and health research: an overview of artificial intelligence, blockchain, and Internet of things," *International Journal of Environmental Research and Public Health*, vol. 16, no. 20, pp. 1–24, 2019.

[2] S. Chen, K. M. Kamarudin, and S. Yan, "Product innovation: a multimodal interaction design method based on HCI and TRIZ," *Journal of Physics: Conference Series*, vol. 1875, no. 1, Article ID 012012, 2021.

[3] R. Kang, A. Guo, G. Laput, Y. Li, and X. A. Chen, "Minuet: multimodal interaction with an Internet of things," in *Proceedings of the Symposium on Spatial User Interaction*, New Orleans, LA, USA, October 2019.

[4] D. Rocchesso, S. Serafin, F. Behrendt et al., "Keywords sound design, auditory display, multimodal interaction, product design, interactive arts and music, sound perception and cognition," in *Proceedings of the Sound Modelling. CHI 2008 Proceedings Workshops*, Florence, Italy, April 2008.

[5] M. Xiao, Z. Feng, X. Yang, T. Xu, and Q. Guo, "Multimodal interaction design and application in augmented reality for chemical experiment," *Virtual Reality & Intelligent Hardware*, vol. 2, no. 4, pp. 291–304, 2020.

[6] H. Li, S. Zhou, F. Zhang, and C. Yang, "Immersive VR theater with multi-device interaction and efficient multi-user collaboration," *Procedia Computer Science*, vol. 147, pp. 468–472, 2019.

[7] J. Jia, J. Y. Lai, G. H. Zhang, and H. Ling, "The design of image acquisition and preprocess system based on FPGA," *Applied Mechanics and Materials*, vol. 71-78, pp. 4269–4273, 2011.

[8] L. J. Zhong and W. W. Li, "A method of online color-difference detecting based on image processing and its application," *Applied Mechanics and Materials*, vol. 37-38, pp. 14–17, 2010.

[9] R. G. Wang, J. Zhu, W. T. Yang et al., "An improved local multi-scale retinex algorithm based on illuminance image segmentation," *Acta Electronica Sinica*, vol. 38, no. 5, pp. 1181–1186, 2010.

[10] X. Ren, W. Wang, J. Ren, X. Mao, and M. Zhang, "Research and application of label defect detection method based on machine vision," *Journal of Physics: Conference Series*, vol. 1453, no. 1, Article ID 012084, 2020.

[11] J. Acuna, J. Walter, and I. Kallfass, "Very fast short circuit protection for gallium-nitride power transistors based on printed circuit board integrated current sensor," in *Proceedings of the 20th European Conference on Power Electronics and Applications (EPE'18 ECCE Europe)*, pp. 1–10, IEEE, Riga, Latvia, 2018.

[12] J. A. Guo, Z. B. Zhu, B. B. Sun, and T. Zhang, "A novel field box girder welding robot and realization of all-position welding process based on visual servoing," *Journal of Manufacturing Processes*, vol. 63, pp. 70–79, 2020.

[13] W. Cai, X. Wen, Q. Tu, and X. Guo, "Research on image processing of intelligent building environment based on pattern recognition technology," *Journal of Visual Communication and Image Representation*, vol. 61, pp. 141–148, May 2019.

Research Article

Research on Music Emotion Recognition Model of Deep Learning Based on Musical Stage Effect

Cuiqing Huang and Qiang Zhang [ORCID]

School of Music and Dance, China-ASEAN College of Arts, Chengdu University, Chengdu, Sichuan 610106, China

Correspondence should be addressed to Qiang Zhang; zhangqiang01@cdu.edu.cn

The change of life style of the times has also prompted the reform of many art forms (including musicals). Nowadays, the audience can not only enjoy the wonderful performances of offline musicals but also feel the charm of musicals online. However, how to bring the emotional integrity of musicals to the audience is a technical problem. In this paper, the deep learning music emotion recognition model based on musical stage effect is studied. Firstly, there is little difference between the emotional results identified by the CRNN model test and the actual feelings of people, and the coincidence degree of emotional responses is as high as 95.68%. Secondly, the final recognition rate of the model is 98.33%, and the final average accuracy rate is as high as 93.22%. Finally, compared with other methods on CASIA emotion set, the CRNN-AttGRU has only 71.77% and 71.60% of WAR and UAR, and only this model has the highest recognition degree. This model also needs to update iteration and use other learning methods to learn at different levels so as to make this model widely used and bring more perfect enjoyment to the audience.

1. Introduction

With the development of the times and technology, people can easily get digital music, drama, film, and television on mobile phones, iPad, computers, and other electronic devices, while tapes, CDs, records, videos, and so on gradually disappear into people's daily life. The stage appeal of musicals is very strong, but simply the video recording technology is poor, and the audience cannot be there. It is difficult for the stage effect to have an effect on the audience online. If the musical is to be moved online well, so that the audience can be infected by the musical without being on the scene, it is necessary to study how to completely restore the musical scene. This paper is one of the most important step, using the existing technology to make a musical emotional recognition model, which can perfectly identify the emotions contained in the musical. Literature [1] used the LEO model to process tree structure format with Gabor feature representation in order to recognize human facial emotion. Literature [2] proposed a layered coding cascade optimization model for the facial expression recognition system, which optimizes direct similarity and Pareto-based function

optimization. Literature [3] easily created an entropy-based maximum emotion recognition model using individual average differences of emotion signals. Literature [4] proved that small-world network is the most suitable model to capture the cognitive basis of facial emotions. Reference [5] designed an improved wave physics model based on depth wave field inference in speech emotion recognition. Literature [6] used the Gaussian mixture model fitting method to design neutral profile dictionary to solve the baseline problem. Literature [7] collected emotional physiological data sets under four induced emotions, and the group-based IRS model improved the performance of emotion recognition. Literature [8] introduced traditional acoustic features and new vector functions to represent speech signals abstractly. In Literature [9], cyclic neural network is effectively replaced by convolution network and self-attention, which is close to transformer performance. Literature [10] proposed a new multidimensional cyclic convolution algorithm to achieve the least number of multiplications in theory. Based on CRT and Vinograd's minimum multiple complexity theorem, a simplified cyclic convolution formula is obtained in literature [11]. Literature [12] used the music emotion

recognition model based on deep learning to solve the problem of low accuracy of emotion recognition. Literature [13] computed audio function ideas, capturing music form, texture, and expression elements to advance music emotion recognition. Literature [14] constructed a balanced music video emotional data set and integrated multimodal transport information based on deep learning. Literature [15] created a model based on deep neural network to identify and classify music genres and Chinese traditional musical instruments. In order to learn to restore the important human emotion nodes in music, this paper focuses on designing the emotional recognition model of musical based onstage effect, introduces the related theoretical basis, respectively, and then designs a hybrid model based on CNN and RNN according to several points that need to be discussed emphatically in emotional recognition (signal preprocessing, emotional data set, recognition algorithm, and evaluation). Then, the model is simulated and tested.

2. Theoretical Basis

2.1. Emotional Description Model under Stage Effect. The charm of musicals is always infectious, which most people cannot stop. Different from the performance of simple music and drama dance, musical is a mixture of music elements, dance elements, and drama elements and gradually develops and changes into a unique artistic expression in the torrent of years. It brings wonderful enjoyment in visual, auditory, and other senses. The stage effect cooperates with the music to immerse the audience in the performance of the plot. In this study, we want to establish an emotion recognition model [16]. We want to build an emotion recognition model [16] to define and classify various emotions.

(1) A large number of scholars [17] have studied the definition of relatively basic emotion [18] as shown in Table 1.

Emotion is a reflection of the relationship between objective things and subject needs. However, emotion is an extremely complex psychological process, which is a complex body with multidimensions, multiforms, and multifunctions. Each individual is influenced by his own environment, social environment, and his own experience and cognition, and everyone has different definitions of emotion. Furthermore, human research on psychology is not thorough, and there are many blank fields. Therefore, there are too many reasons for the differences in definitions of different scholars, involving a wide range of fields, which need to be systematically studied and discussed by later generations.

(2) Because human emotions are complex and changeable, it is difficult to express them with simple basic emotional definitions. Therefore, in this study, we tend to combine a more comprehensive and complex two-dimensional emotion description model with a simple emotion definition as shown in Figure 1.

Table 1: Definition of emotion by scholars.

Scholar	Emotion
Mowrer	Pain, joy
Panksepp	Expectation, fear, anger, panic
James	Fear, sadness, love, anger
Pultchick	Acceptance, anger, expectation, disgust, joy
Gray	Fear, sadness, surprise
Tomkins	Anger, fear, anxiety, joy
Ekman, Freesen	Anger, interest, contempt, disgust, pain
Weiner, Graham	Fear, joy, shame, surprise
Frijida	Anger, disgust, fear, joy, sadness, surprise

2.2. Deep Learning Method. Machine learning [19] has a method called deep learning [20]. This method is very mature in the field of speech recognition, so we can use it to recognize musical emotion.

2.2.1. Deep Confidence Networks. Deep confidence network (DBN) [21] is a probability generation model [22], and it is also a special neural network [23]. It consists of a variety of "constrained Boltzmann machines" as shown in Figure 2.

2.2.2. Parameter Pretraining. Constrained Boltzmann machine [24] is widely used. Set n, m, v, and h as visible layer nodes, hidden layer nodes, visible units, and hidden units, respectively. The energy of the system [25] is defined as follows:

$$E(v, h \mid \theta) = -\sum_{i=1}^{n} a_i v_i - \sum_{j=1}^{m} b_j h_j - \sum_{i=1}^{n} \sum_{j=1}^{m} v_i W_{ij} h_j. \quad (1)$$

When the model state is constant, the joint probability distribution is as follows:

$$P(v, h \mid \theta) = \frac{e^{-E(v,h|\theta)}}{Z(\theta)},$$
$$Z(\theta) = \sum_{v,h} e^{-E(v,h|\theta)}. \quad (2)$$

The activation probability of hidden unit is as follows:

$$P(h_j = 1, \mid v, \theta) = \sigma\left(b_j + \sum_j v_i W_{ij}\right). \quad (3)$$

Visible cell activation probability is as follows:

$$P(v_i = 1, \mid h, \theta) = \sigma\left(a_i + \sum_i h_j W_{ij}\right). \quad (4)$$

The sigmoid activation function is as follows:

$$\sigma(x) = \frac{1}{1 + \exp(-x)}. \quad (5)$$

2.2.3. Parameter Tuning. The output of the activation function hidden layer node is as follows:
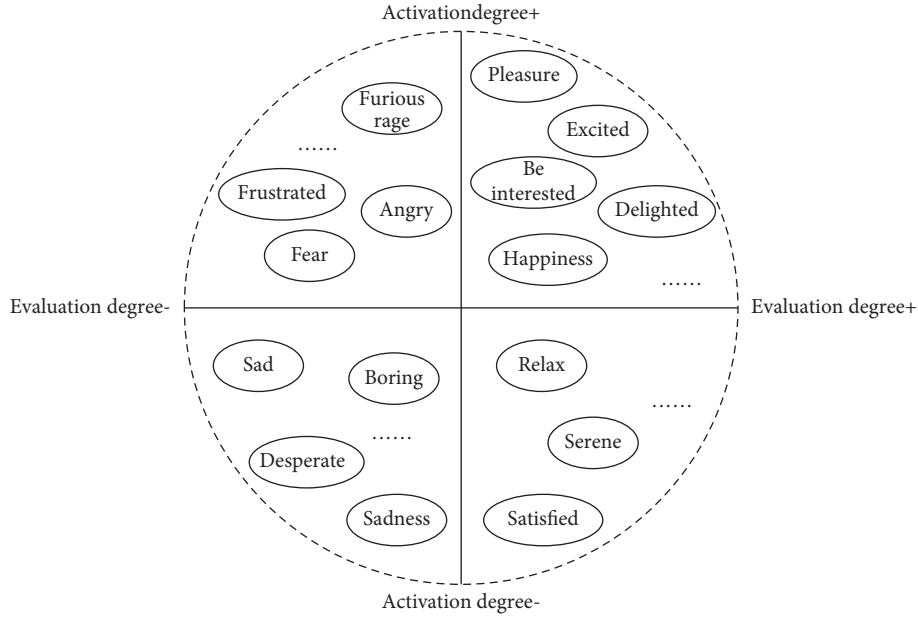
Figure 1: Two-dimensional (activation and evaluation) emotion model.
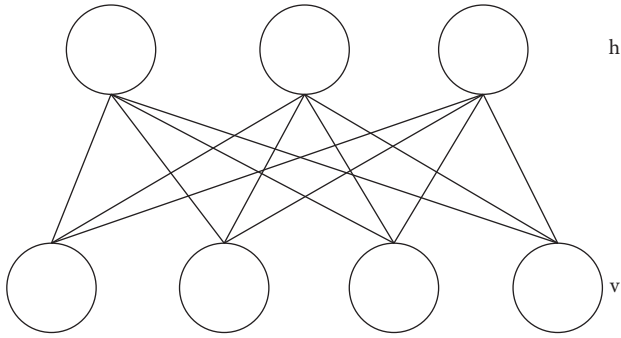
Figure 2: Constrained Boltzmann machine.

$$v^{k+1} = \omega^k h^k + b^k,$$
$$h^k = \sigma(v^k). \tag{6}$$

Softmax function is as follows:

$$p_s = \frac{\exp(v_s^{N+1})}{\sum_j \exp(v_s^{N+1})}. \tag{7}$$

When $d_s = 0$, the cross entropy function is as follows:

$$L = -\sum_s d_s \log p_s. \tag{8}$$

2.2.4. CNN. "CNN" is the abbreviation of convolution neural network. CNN offers supervised and unsupervised learning. It is usually used for processes corresponding to natural access and language. Generally, three-dimensional CNN has two operations: convolution and pooling. Important formulas of convolution layer are as follows:

$$Z^{l+1}(i, j) = [Z^l \otimes \omega^{l+1}](i, j) + b,$$

$$Z^{l+1}(i, j) = \sum_{k=1}^{K_l} \sum_{x=1}^{f} \sum_{y=1}^{f} [Z_k^l(s_0 i + x, s_0 j + y)\omega_k^{l+1}(x, y)] + b,$$

$$L_{l+1} = \frac{L_l + 2p - f}{s_0} + 1\,(i, j) \in \{0, 1, \dots L_{l+1}\},$$

$$Z^{l+1} = \sum_{k=1}^{K_l} \sum_{i=1}^{L} \sum_{j=1}^{L} (Z_{i,j,k}^l \omega_k^{l+1}) + b = \omega_{l+1}^T Z_{l+1} + b,$$

$$L^{l+1} = L. \tag{9}$$

Excitation function to help express complex characteristics is as follows:

$$A_{i,j,k}^l = f(Z_{i,j,k}^l). \tag{10}$$

Manifestations of *Lp* pooling are as follows:

$$A_k^l(i, j) = \left[\sum_{x=1}^{f} \sum_{y=1}^{f} A_k^l(s_0 i + x, s_0 j + y)^p\right]^{1/p}. \tag{11}$$

Hybrid pooled linear combination is as follows:

$$A_k^l = \lambda L_1(A_k^l) + L_\infty(A_k^l), \lambda \in [0, 1]. \tag{12}$$

2.2.5. RNN. "RNN" is short for cyclic neural network. The information of the sequence can be better handled. The details are as follows:

$$O_t = g(V \cdot S_t),$$
$$S_t = f(U \cdot X_t + W \cdot S_{t-1}). \tag{13}$$

Simply RNN is shown in Figure 3.

3. Research on Emotion Recognition

3.1. Music Signal Preprocessing.
When we want to identify the influence of music in musicals on the audience, the preprocessing of music signals is the first step of all work. We try our best to extract many features of music, then identify their categories, and build models to identify different music. This is an extremely important step, and all subsequent work is based on this step. Music signals will be uniformly converted into good formats, which is convenient for data management and high-quality operation. The flowchart is shown in Figure 4.

Preprocessing is to read audio information for feature extraction. We use SciPY speech processing tool and Python for speech information reading and feature extraction, librosa for speech processing, and MATLAB for pre-filtering. Advantages can eliminate the influence of aliasing, high frequency, and other factors brought by equipment on the quality of voice signals, ensure that the processed music signals are more uniform and smooth, provide high-quality parameters, and improve the processing quality.

The preprocessing parameter data are shown in Table 2.

The parameter settings in Table 2 are basically reasonable. The input of the model is the digital amplitude mel spectrum, and the audio frequency of 22 s is input, and the fast Fourier transform STFT is performed (FFT is the content of decomposing the whole time domain into countless small processes with equal length). Hop size is the overlapping area between two windows. Mel filter is used to obtain the logarithmic amplitude mel spectrum; the preprocessing process is carried out in Librosa, and the output size is array. Zero padding is the last step in the preprocessing process. By adding 75 frames on the time axis to inject more data with the same information, the input signal has better frequency resolution.

3.2. Music Emotional Data Set.
For many years, researchers have been studying databases that contain many kinds of human emotions, and these databases contain various forms of data. Music emotion database can collect signals through audience's reaction to different music materials. In this paper, we quote EMO-DB, CASIA, SAVEE, DEAP, and MAHNOB-HCI.

3.3. Emotion Recognition Algorithm

(1) Bayesian network is shown in Figure 5.
(2) Hidden Markov model (HMM) is shown in Figure 6.
(3) The Gaussian mixture model (GMM) is shown by the following formula:
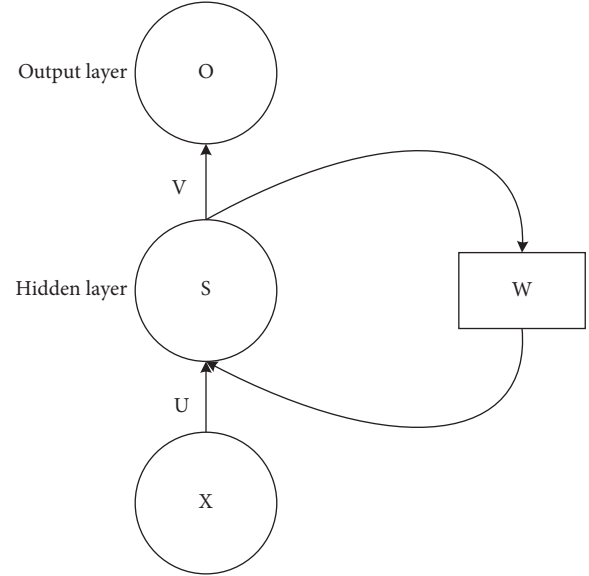
FIGURE 3: Simple cyclic neural network.

$$P(X \mid \theta) = \sum_{k=1}^{M} \pi_k N\left(X; \mu_k; \sum_k\right). \tag{14}$$

3.4. Evaluation Aspects.
When we apply the musical emotion recognition model, we will find that there will be an important problem in the process of construction: overfitting and under-fitting. In order to solve this problem well, we choose to evaluate the model again and again and use various evaluation indexes to test it. In this study, the K-fold cross-validation method and UA, WA, SD, SI, Precision, Recall, and other evaluation indicators were used. Here are some of the formulas:

$$UA = \frac{TP + TN}{TP + FP + TN + FN}, \tag{15}$$

$$WA = \frac{\sum_i^n a_i}{n}, \tag{16}$$

$$Precision = \frac{TP}{TP + FP}, \tag{17}$$

$$Recall = \frac{TP}{TP + FN}. \tag{18}$$

Formula 19 is used to evaluate the accuracy of the model. TP means positive case prediction is positive case. FN means positive case prediction is negative case. FP means negative case is predicted as positive case. TN means negative case predicts negative case.

3.5. Emotion Recognition Based on Hybrid Model.
Convolution neural network and cyclic neural network can not meet the research needs. Because both theories and methods have their own advantages and disadvantages, we
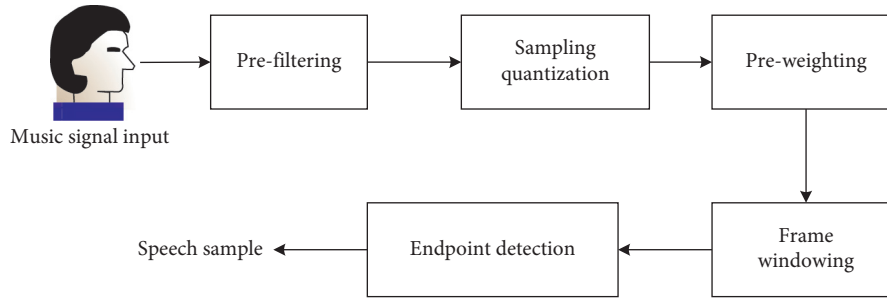
FIGURE 4: The flowchart of feature extraction.

TABLE 2: Preprocessing parameters of music signal.

Duration	n
FFT frame size	512
Hop size	255
Number of mel filters	96
Mel-time matrix size	$1 * 95 * 1300$
Extended mel-time matrix	$1 * 95 * 1356$
Extended boundary number	75

FIGURE 5: Basic structure of Bayesian network.

take the essence and discard the dross, so we improve these two kinds of networks and combine them into a more complex neural network to meet the needs of this study, that is, convolutional recurrent neural network (CRNN). We have modified many excellent performances for this model. It can have all kinds of advantages of CNN and RNN at the same time.

After the improvement of this model, the operation of variable length input can be carried out, and the interference of filling values to model data can be avoided. It can also ensure that the accuracy of the model will not be lost.

$$S_{\text{conv}} = \text{Conv}(S) \cdot \text{Mask}(S). \qquad (19)$$

LSTM is an extension of RNN. It is a long-term and short-term memory network. We add this extension to the CRNN model, which can solve the problem of processing sequence change data. Then, we transport the data to the Softmax loss and quantification loss layer for processing and finally the optimization goal. The CRNN model is shown in Figure 7.

4. Music Emotion Recognition Model Test

In this experiment, the music emotion model is tested. The main tasks of the test model are as follows: accept music data samples of musicals for certain processing; after analyzing the sample, speculate and identify what emotions the stage effect will cause the audience (such as happiness, sadness, sadness, and excitement), the emotion description model should be applied here, and finally the audience's emotions should be collected, and the accuracy of the opinion collection should be compared with the results of our experimental model test, so as to evaluate whether the established CRNN model is available. If there is too much difference between the two results, it means that our music emotion recognition model is unqualified,
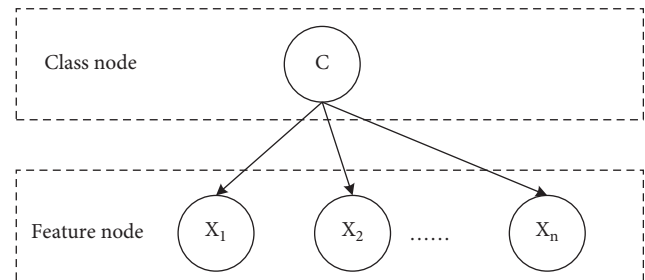
and more modifications and adjustments are needed to carry out technical innovation. If the difference accuracy meets the requirements, it proves that the deep learning music emotion recognition model based on musical stage effect can be used.

4.1. Experimental Environment. In the process of music performance, various unpredictable external interference factors affect the audience, so it is necessary to deal with the noise. We chose to conduct the experiment in a theater where the surrounding environment is quiet and there is no performance task for the time being. In this experiment, we invited 30 volunteers (native speakers of Chinese) who were in good mental state and did not stay up late, drink alcohol, or get sleepy to watch a Chinese musical together. While watching the experiment, the volunteers were unable to do anything else. After watching the musical, the emotional feelings of these 30 people were collected. The music emotion recognition model will be tested by this musical music sample, and finally the test results will be compared with the feelings of 30 voluntary participants.

4.2. Experimental Settings. The experimental test samples are Chinese musicals, and Chinese emotional data sets are used. From the matching selection in the database to the data suitable for this musical, it is divided into pretest, development set, and test set. *Note.* The main purpose of pretest is to detect the interference caused by sudden factors such as garbled code and disordered data, and all the extracted data should be tested by pretest. As shown in Table 3, it is the specific data situation.
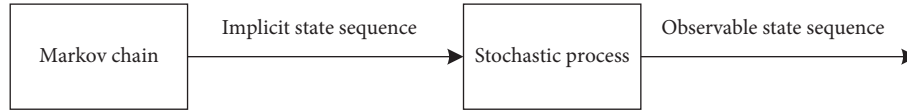
FIGURE 6: Schematic diagram of hidden Markov process.
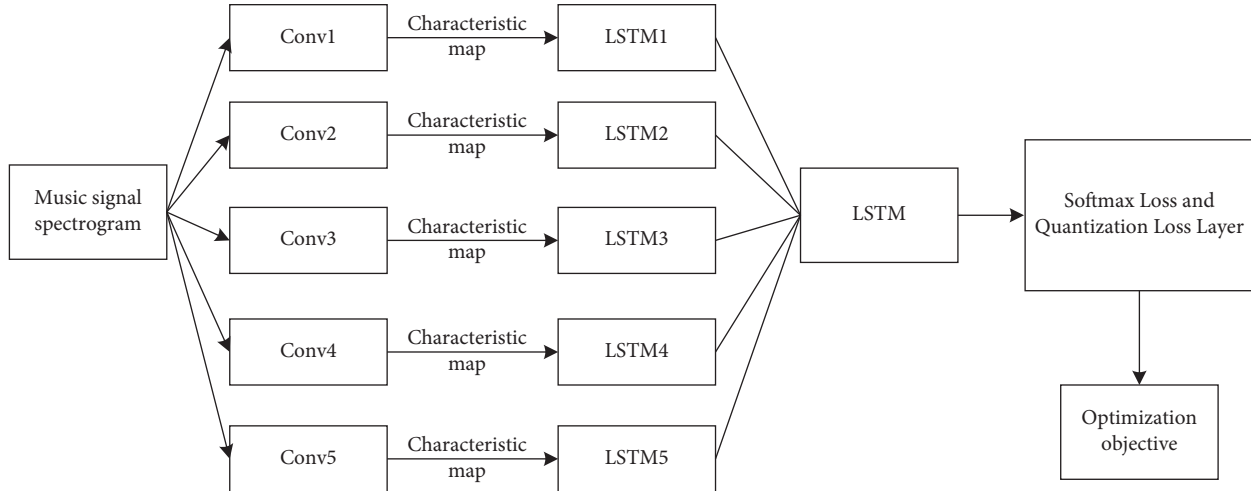
FIGURE 7: Flow chart of CRNN music recognition framework.

TABLE 3: Pretest, development set, and test set dataset settings.

	Pretest	Development set	Test set
Is there any situation	No abnormality	No abnormality	No abnormality
Total number of sentences (sentences)	60	120	360

4.3. Parameter Optimization. In the experiment, parameters are the part that needs special attention. Without adjustment and optimization, the final result may not achieve the best experimental effect. This will cause trouble to the experimental results. The relevant parameters are shown in Table 4.

We carry out 20,000 iterations on music samples. When the learning rate is very small (such as 0.001), the learning process is extremely slow and the recognition is unstable; high learning rate (such as 0.1) leads to unstable conditions and even reduces performance as shown in Figure 8.

Figure 8 is a discussion of recognition rate for different iteration times of learning rate in experimental parameters, and the most suitable learning rate interval range is selected. The reason of performance deviation is that the learning rate affects the recognition stability of recognition rate.

Momentum coefficient can speed up the learning process. The deviation of coefficient will cause oscillation in the initial stage, and the fluctuation will cause performance degradation, as shown in Figure 9.

However, if the weight attenuation coefficient is too large (such as 0.005), the stability of the learning process will be destroyed. On the contrary, a relatively small weight attenuation coefficient will be more stable and safe, as shown in Figure 10.

4.4. Key Elements of Musical Emotion. There are some fragments in the music. It is useful for experiments, including emotional parts, and some fragments are not important, which may cause suspension, transition, and other effects. Let us take time as an example and look at the emotional information carried by each sentence as a characteristic, one is unimportant and the other is important.

As shown in Figure 11, the music sentence part is divided into five parts, and the emotional information is mainly distributed in the first, second, and third parts, which takes more time to test.

TABLE 4: Parameters related to experiments.

Parameter	Learning rate	Batch size	Momentum coefficient	Weight attenuation coefficient	Dropout coefficient
Value	0.01	16	0.9	0.0005	0.5
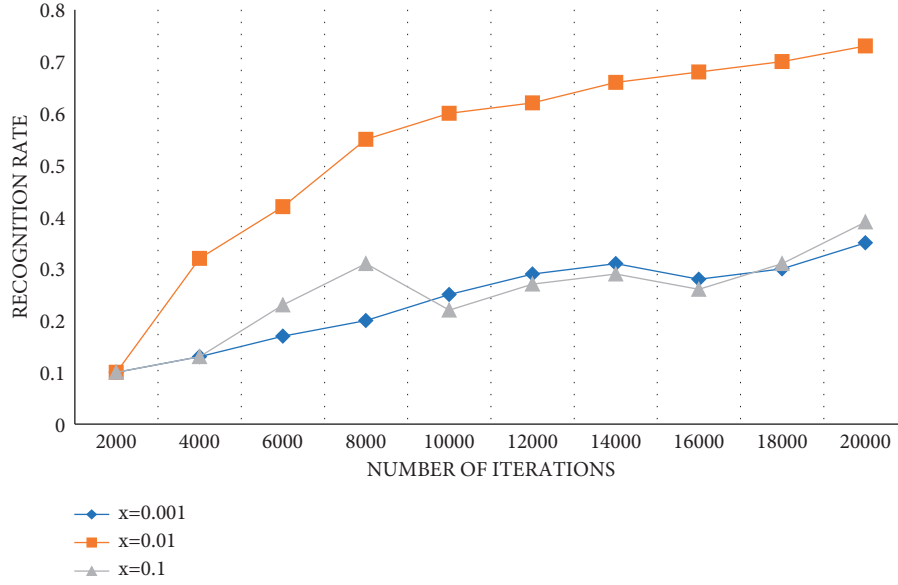
FIGURE 8: Learning rate-related effects.
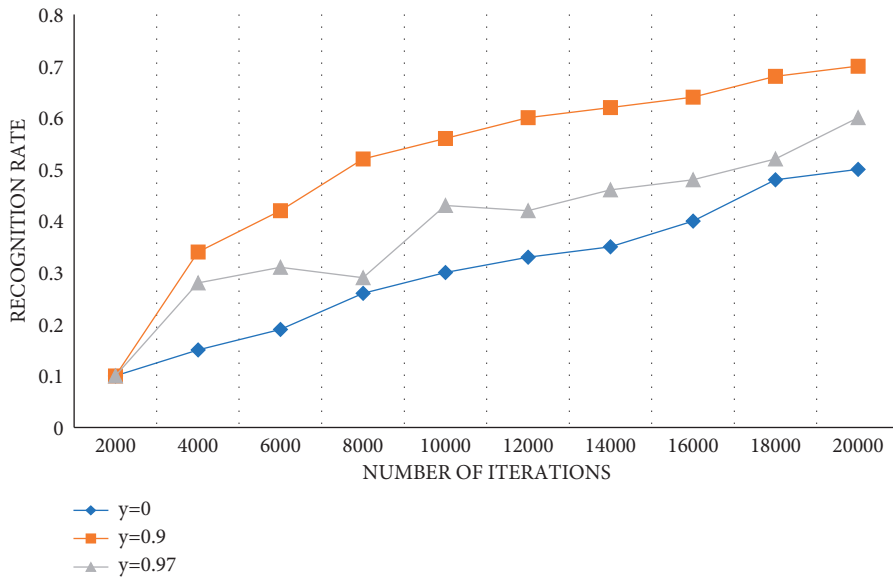
FIGURE 9: Correlation effect of momentum coefficient.

4.5. Recognition Performance. The performance calculation formula of music recognition is as follows:

$$\text{rejection rate} = \frac{\text{number of successfully recognized music sample fragments}}{\text{total number of sample fragments of test music}}. \tag{20}$$
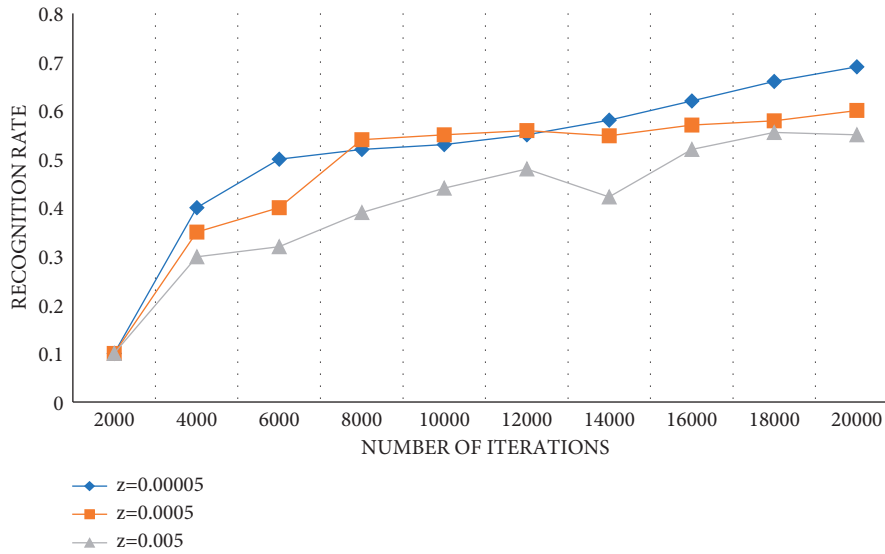
Figure 10: Correlation effect of weight attenuation coefficient.
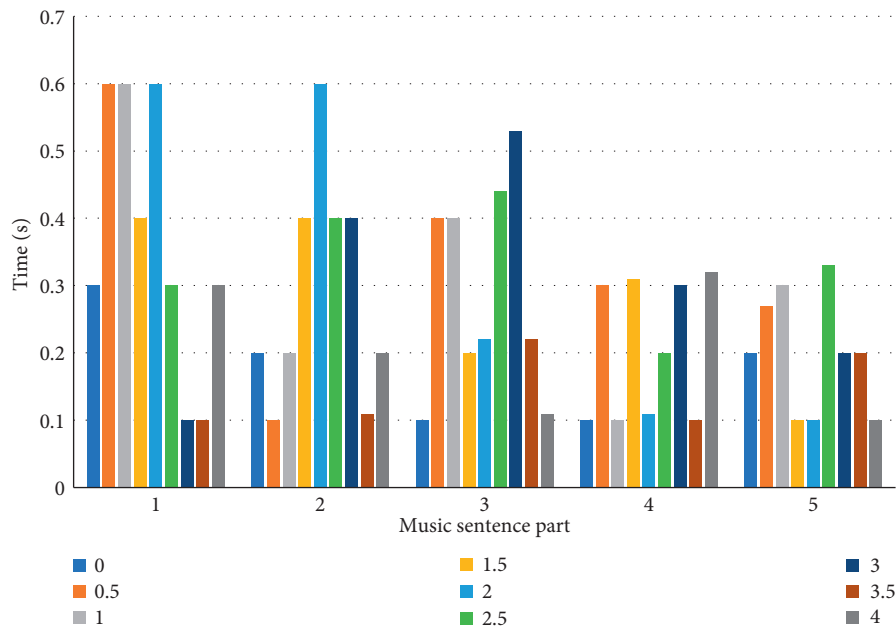
Figure 11: Visual emotional information distribution map.

The specific identification is shown in Table 5.

From Table 5, we can see that the recognition performance of this experiment is very good, the final recognition rate is 98.33%, and the final average accuracy rate is as high as 93.22%.

4.6. Experimental Test Results. Because the content of this musical is sad as a whole, the emotional reaction of positive energy is excluded. The specific feelings of 30 voluntary participants are collected as shown in Table 6.

The results tested by the music emotion model are shown in Table 7.

After calculating and comparing the experimental results, it can be found that the coincidence degree of emotional response with 30 voluntary participants is as high as 95.68%.

4.7. Comparison with Other Methods. Compare WAR and UAR on Chinese emotion database (which can be changed in other tests). The results show that only the model created by us has the highest recognition performance and the most accurate accuracy. Among the three methods in the table, even CRNN-AttGRU with the highest recognition performance has only 71.77% and 71.60% recognition rates, respectively, as shown in Table 8.

TABLE 5: Table of recognition performance.

Numbering	1 (%)	2	3	4 (%)	5 (%)	6 (%)	7 (%)	...
False acceptance	0	0%	0%	0	1.19	0	0	...
False rejection rate	0	20%	2.14%	0	1.9	0	0	...

TABLE 6: Emotional feelings of volunteers.

Voluntary participant number	Emotion
1	Pain, fear
2	Expectation, fear, anger, panic
3	Fear, sadness, anger
4	Acceptance, anger, expectation, disgust
5	Fear, sorrow
6	Anger, fear, anxiety
7	Interest, disgust, pain, fear
8	Anxiety, surprise
9	Anger, disgust, fear, sadness, surprise
...	Boredom, fear, sadness
30	Surprise, sadness, fear

TABLE 7: Test results of the music emotion model.

Name	Emotion
CRNN model	Sadness, sadness, pain, disgust, fear, fear, surprise, anger, surprise, surprise, anxiety, and so on

TABLE 8: Comparison of WAR and UAR on CASIA by different methods.

Model	WAR (%)	UAR (%)
CRNN-CTC	70.42	69.75
CRNN-GRU	60.48	61.72
CRNN-AttGRU	71.77	71.60

5. Conclusion

In this paper, based on the changes of the times, electronic instruments are of great significance for musicals to be perfectly moved to mobile devices. The stage effect provided by musicals for the audience leads the audience to immerse themselves in the emotional changes in music. This study needs to restore people's emotional feelings brought by music in musicals with technology. Therefore, the CRNN model is designed for testing. The recognition accuracy of the model for emotion is more accurate than that of other methods. The results show that (1) there is little difference between the emotional results identified by the CRNN model test and the actual feelings of people, and the coincidence degree of emotional responses is as high as 95.68%; (2) the final recognition rate of the model is 98.33%, and the final average accuracy rate is as high as 93.22%; and (3) compared with other methods on the CASIA emotion set, the CRNN-AttGRU has only 71.77% and 71.60% of WAR and UAR, and only this model has the highest recognition degree. Although this paper has some results in music emotion recognition, it still needs further exploration by later workers. This model also needs to update iteration and use other learning methods to learn at different levels, so as to make this model widely used and bring more perfect enjoyment to the audience.

Data Availability

The experimental data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest regarding this work.

Acknowledgments

References

[1] J.-J. Wong and S.-Y. Cho, "A local experts organization model with application to face emotion recognition," *Expert Systems with Applications*, vol. 36, no. 1, pp. 804–819, 2009.

[2] S. C. Neoh, L. Zhang, K. Mistry et al., "Intelligent facial emotion recognition using a layered encoding cascade optimization model," *Applied Soft Computing*, vol. 34, pp. 72–93, 2015.

[3] S.-Y. Park, D.-K. Kim, and M.-C. Whang, "Maximum entropy-based emotion recognition model using individual average difference," *The Journal of the Korean Institute of Information and Communication Engineering*, vol. 14, no. 7, pp. 1557–1564, 2010.

[4] T. Takehara, F. Ochiai, and N. Suzuki, "A small-world network model of facial emotion recognition," *Quarterly Journal of Experimental Psychology*, vol. 69, no. 8, pp. 1508–1529, 2016.

[5] C. Zheng, C. Wang, and N. Jia, "Emotion recognition model based on multimodal decision fusion," *Journal of Physics: Conference Series*, vol. 1873, no. 1, Article ID 012092, 2021.

[6] S. Ulukaya and C. E. Erdem, "Gaussian mixture model based estimation of the neutral face shape for emotion recognition," *Digital Signal Processing*, vol. 32, pp. 11–23, 2014.

[7] C. Li, C. Xu, and Z. Feng, "Analysis of physiological for emotion recognition with the IRS model," *Neurocomputing*, vol. 178, no. 20, pp. 103–111, 2016.

[8] T. Zhang and J. Wu, "Speech emotion recognition with i-vector feature and RNN model," in *Proceedings of the 2015 IEEE China Summit and International Conference on Signal and Information Processing (ChinaSIP)*, pp. 524–528, IEEE, Chengdu China, July 2015.

[9] M. Chen, Y. Li, and R. Li, "Research on neural machine translation model," *Journal of Physics: Conference Series*, vol. 1237, Article ID 052020, 2019.

[10] A. H. Diaz-Perez and D. Rodriguez, "One dimensional cyclic convolution algorithms with minimal multiplicative complexity," in *Proceedings of the 2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings*, Toulouse, France, May 2006.

[11] A. Rodriguez, "Cyclic convolution algorithm formulations using polynomial transform theory," *Journal of Computers*, vol. 2, no. 7, pp. 40–48, 2007.

[12] R. Sarkar, S. Choudhury, S. Dutta, A. Roy, and S. K. Saha, "Recognition of emotion in music based on deep convolutional neural network," *Multimedia Tools and Applications*, vol. 79, no. 10, pp. 765–783, 2020.

[13] R. Panda, R. M. Malheiro, and R. P. Paiva, "Audio features for music emotion recognition: a survey," *IEEE Transactions on Affective Computing*, vol. 99, p. 1, 2020.

[14] Y. R. Pandeya and J. Lee, "Deep learning-based late fusion of multimodal information for emotion classification of music video[J]," *Multimedia Tools and Applications*, vol. 80, no. 38, pp. 1–19, 2021.

[15] K. Xu, "Recognition and classification model of music genres and Chinese traditional musical instruments based on deep neural networks," *Scientific Programming*, vol. 2021, Article ID 2348494, 8 pages, 2021.

[16] M. Fredrikson and R. Gunnarsson, "Psychobiology of stage fright: the effect of public performance on neuroendocrine, cardiovascular and subjective reactions," *Biological Psychology*, vol. 33, no. 1, pp. 51–61, 1992.

[17] R. Studer, P. Gomez, H. Hildebrandt, M. Arial, and B. Danuser, "Stage fright: its experience as a problem and coping with it," *International Archives of Occupational and Environmental Health*, vol. 84, no. 7, pp. 761–771, 2011.

[18] Y. Lyu, "Research on the influence of music educational psychology on saxophone players' mental state and stage performance," *Journal of Intelligent and Fuzzy Systems*, vol. 2, no. 2, pp. 1–12, 2021.

[19] G. Litjens, T. Kooi, B. E. Bejnordi et al., "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, no. 9, pp. 60–88, 2017.

[20] S. Levine, P. Pastor, A. Krizhevsky, J. Ibarz, and D. Qullien, "Learning hand-eye coordination for robotic Grasping with deep learning and large-scale data collection," pp. 421–436, 2016, https://arxiv.org/abs/1603.02199.

[21] Z. Jian and O. G. Troyanskaya, "Predicting effects of noncoding variants with deep learning-based sequence model," *[J]. Nature Methods*, vol. 12, no. 10, pp. 931–934, 2015.

[22] K. Warburton, "Deep learning and education for sustainability," *International Journal of Sustainability in Higher Education*, vol. 4, no. 1, pp. 44–56, 2003.

[23] J. Kim and E. Andre, "Emotion recognition based on physiological changes in music listening," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 12, pp. 2067–2083, 2008.

[24] Y. P. Lin, C. H. Wang, T. P. Jung et al., "EEG-based emotion recognition in music listening," *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 7, pp. 1798–1806, 2010.

[25] A. D. Patel, "A neurobiological strategy for exploring links between emotion recognition in music and speech," *Behavioral and Brain Sciences*, vol. 31, no. 5, pp. 589-590, 2008.

Retraction

Retracted: Apparel Design and Development Based on 3D Scanning Technology

Scientific Programming

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope

(2) Discrepancies in the description of the research reported

(3) Discrepancies between the availability of data and the research described

(4) Inappropriate citations

(5) Incoherent, meaningless and/or irrelevant content included in the article

(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] G. Zhang, "Apparel Design and Development Based on 3D Scanning Technology," *Scientific Programming*, vol. 2021, Article ID 7933206, 7 pages, 2021.

Research Article

Apparel Design and Development Based on 3D Scanning Technology

Guodong Zhang (ORCID)

Hubei Institute of Fine Arts, Wuhan, Hubei 430205, China

Correspondence should be addressed to Guodong Zhang; 20181850@hifa.edu.cn

With the development of computer science, especially the application of 3D scanning technology in garment design, intelligent modeling is realized, which is impossible to achieve in traditional design methods. In this paper, we propose the 3D model construction of human garments based on the motion recovery structure method. The eigenmatrix is obtained from the camera parameters, and the transformation matrix is calculated by matching the image feature points with the help of scale-invariant feature conversion algorithm to realize the 3D reconstruction technology of human garments based on multiview image sequences. The effectiveness of this method is verified through experiments, and it has good robustness and accuracy. Through the form of style modeling, the design thinking and method can be extended to form a more reasonable garment structure and guide the innovation of garment production mode.

1. Introduction

As computer vision technology is widely used in various fields such as virtual reality and digital cities, more convenient and efficient acquisition and construction of 3D models of human clothing have become a current research hotspot, which in turn promotes the improvement of computer vision algorithms and applications [1–3]. The use of relevant algorithms to construct 3D models of features from multiview image sequences captured with ordinary digital cameras has the advantages of wide use and convenient application, and they are most widely used in the field of 3D model construction [4–7].

Currently, 3D technology in China has received strong support from the government. China's manufacturing industry must rely closely on deepening reform and innovation to promote its transformation from bits to strong, thus promoting the development of 3D technology in China [5]. This paper discusses the design and development of garments based on 3D scanning technology, taking women's shirts as an example. The purpose is to change the traditional design made through a new creative thinking in the 3D era [6].

With the advancement of virtual reality, computer vision, and computer graphics, interactive apparel design has attracted attention as an innovation in apparel technology and a future trend in apparel design. In interactive apparel design, 3D virtual human models based on 3D scanning technology are used to instantly display the garments in the design phase and are a crucial part of the entire workflow [7, 8].

The new interaction design development process includes 5 objects: human body, virtual mannequin, 3D garment, 2D paper pattern, and physical garment. First, the point cloud data of the human body are captured using a modern 3D scanning system, which is used as the basic input for building the 3D virtual model. Then, according to the initial design sketch, the paper pattern of the garment is drawn on it using the relevant design software [9]. Then, the latest 3D fitting software CLO3D can be used to generate the 2D paper pattern directly into a 3D virtual garment. After the final 3D garment effect is determined, the 3D garment can be unfolded into a 2D garment paper pattern using CATIA (Computer Aided Tri-Dimensional Interface Application) software. Finally, the 2D paper pattern can be put into industrial production for mass production. Throughout

the process, there are 8 interactions and 3 traditional interaction techniques that are used for the virtual mannequin, namely, garment fitting/fabric simulation/texturing, data scanning, and garment unfolding. Production and physical inspection interactions are more traditional practices in the apparel design industry (Figure 1).

In this paper, we propose a 3D model construction of human clothing based on the motion recovery structure method. We obtain the eigenmatrix from the camera parameters, match the image feature points with the help of scale-invariant feature conversion algorithm, calculate the transformation matrix, and realize the 3D reconstruction technology of human clothing based on multiview image sequences [10]. The effectiveness of this method is verified through experiments, and it has good robustness and accuracy. Through the form of style modeling, the design thinking and method can be extended to form a more reasonable garment structure and guide the innovation of garment production mode.

2. Related Work

The current research on the algorithm for constructing 3D models of human clothing based on multiview image sequences has been focused by scholars on two aspects: image sources and algorithm improvement. In terms of image sources, the study in [11] used images acquired using digital cameras and associated linkage maps to achieve measurements and 3D reconstruction of human clothing. The work in [12] used handheld cameras to simulate drone shots and useful features around objects to obtain camera calibration parameters for each image using image pairs to complete 3D modeling. A method for building high-resolution digital elevation models using captured images from consumer-grade digital cameras was proposed in [13]. These studies are based on a small amount of image data obtained with a specific instrument and are not adapted to the large amount of image data obtained using different instruments. For this reason, Mao et al. [14] used a large collection of unstructured images downloaded from the Internet to successfully construct a 3D model of the target. High-resolution, multi-channel image sources have a fundamental impact on the application area and accuracy of 3D model construction.

For the same image data, the improvement of the algorithm is a necessary way to improve the accuracy of 3D model construction. Lamb et al. [15] introduced the basic principle and process of Structure from Motion (SfM) technique to show the simplicity and effectiveness of the technique to obtain high-precision 3D terrain data, which are suitable for areas with sparse vegetation.

Roberts et al. [16] addressed the efficient modeling of large, unordered, highly redundant, and irregularly sampled photo sets by SfM. Morris et al. [17] implemented an image-based 3D model reconstruction process using the global SfM method instead of iterative SfM based on the opensource code. Brkic et al. [18] proposed a new incremental motion structure recovery algorithm (SfM-Y) to solve the problems of poor algorithm robustness, low efficiency, and computational redundancy.
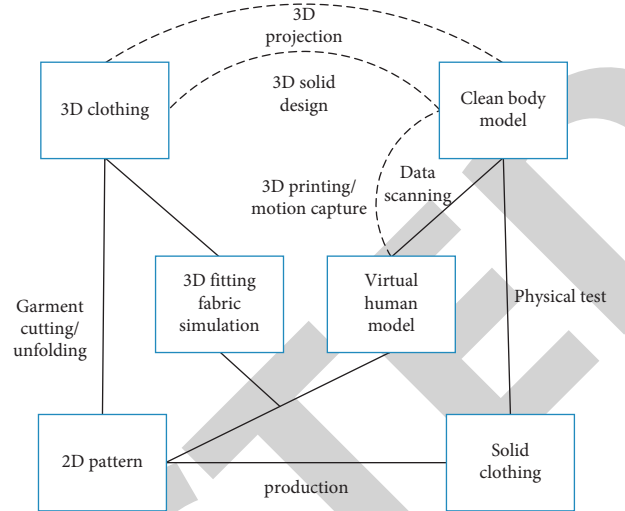
FIGURE 1: Development process of interactive fashion design.

This paper constructs 3D models of human garments based on SfM and uses multiview image sequences and computer vision techniques to extract, match, and reconstruct features of the image sequences in 3D.

3. Extraction and Matching of Feature Points

3.1. Extraction of Feature Points. The key step of 3D modeling using multiview image sequences is to fuse images taken from different viewpoints into the same coordinate system, so as to display 3D scenes in all directions. The key of image matching is to extract the corresponding points with similar features in the adjacent images, and this paper uses the scale-invariant feature transform (SIFT) [19] algorithm to extract feature points. The algorithm is based on Gaussian fuzzy, and the Gaussian pyramid and Gaussian differential pyramid are established by using different standard deviation values of feature points and down sampling, and then, the corresponding feature points are extracted and the edge points are rejected.

Gaussian blurring is used to extract high-precision feature points on images of different scales, and Gaussian blurring is the basis of the algorithm [20]. Gaussian blurring uses a Gaussian function as a template to convolve the image to make it smooth. In this paper, let the blurring degree, i.e., the standard deviation, be the size of the template $a \times b$ and the coordinates of the pixel point being processed be (x, y). Then, the Gaussian function is shown in the following equation:

$$G(x, y) = \frac{1}{2\pi\sigma^2} e^{\left(-\left((x-a/2)^2+(y-b/2)^2/2\sigma^2\right)\right)}. \tag{1}$$

In this paper, a 5×5 Gaussian template with $\sigma = 0.6$ is used to blur the image.

This paper invokes the scale space theory to establish a Gaussian differential pyramid. The original image is first processed with different Gaussian blurring to obtain a set of images with different scales, then the image is downsampled, and the downsampled image is processed in the same way to

obtain a set of downsampled images with different scales, and so on, to place the original image group at the bottom and stack the downsampled image groups in turn to build a Gaussian pyramid [21].

The number of groups of the Gaussian pyramid is calculated using the empirical value formula given by Zhang et al. [22], with different parameter sets according to the specific need. The number of groups is calculated as shown in the following equation:

$$0 = [\log_2(\min(M, N))] - 3. \tag{2}$$

where 0 is the number of groups and M, N is the image size.

After the Gaussian pyramid is built, the Gaussian differential pyramid is obtained by differencing the adjacent images of each group in turn.

3.2. Matching of Feature Points.

3.2. Matching of Feature Points. After the feature points are extracted, each feature point has information such as position, feature vector, and scale; in order to accurately represent the feature points, descriptors are constructed for the feature points and ensure that they are rotationally invariant for matching. In this paper, we use the descriptor construction method recommended by Kellomaeki et al. [23], which takes 4*4 windows around the key points, and each subwindow has gradient information in 8 directions, east, south, west, north, southeast, northeast, southwest, and northwest and a total of 128 gradient vectors constitute the descriptors of the feature points, as shown in Figure 2.

The determination of descriptor edge length requires interpolation, which is calculated by bilinear interpolation in this paper. And, the direction needs to be rotated to the main direction of the descriptor before the descriptor calculation to ensure the rotation invariance, as shown in Figure 3.

After the descriptors are obtained, the descriptors between matching image pairs are found and the one with the most similar descriptors is the matching feature point. The previously extracted feature points are used to match, and the feature points that do not have corresponding points are eliminated [24].

4. Construction of the 3D Model of Human Clothing Based on Multiview Image Sequences

4.1. 3D Model Construction of Human Clothing with Dual-View Image Sequences. Before constructing the 3D model of the dual-view image sequence, the relationship between the target point locations in the camera image plane of two consecutive shots should be determined first [25], that is, to determine the relationship between the key points of the overlapping parts of the two images when taking consecutive shots, as shown in Figure 4.

In Figure 4, the target point is point X, the image coordinates are x_1 and x_2, and the distances between point X and the two image planes are z_2 and z_1. The internal reference matrix of the camera is K, and the external reference matrix of the two shots is T_1R_1 and T_2R_2.

Let $x_1' = K^{-1}x_1$ and $x_2 = K^{-1}x_2$. The eigenmatrix between the two shots is calculated from equation (6) by matching points (at least 5 pairs) between the two images $\widehat{T}_2R_2$.

$$x_2'\widehat{T}_2R_2x_1' = 0. \tag{3}$$

Using the eigenmatrix and matching the relationship between the two image planes, the image points are processed using the triangulate method to obtain the world coordinate system of the target points. The point cloud image of the target feature is formed by displaying all the points of the world coordinate system, and the 3D model construction based on the dual-view image sequence is completed [26, 27]. In this paper, the results of constructing 3D models based on dual-view images are shown in Figure 5.

4.2. 3D Model Construction of Human Clothing with Multiview Image Sequence. The 3D model construction based on the dual-view image sequence is the basis for the 3D model construction of the multiview image sequence. However, if only the previous matching method is followed, matching the previous image slice with the next two images in sequence, only the rotation matrix between adjacent images can be obtained and their displacement vectors cannot be obtained, because the displacement vectors calculated by the above function are unit vectors. The result of such a method is that when the difference between the images is large, there will be few matching points to use and the accumulated errors will lead to poor matching results.

In this paper, we use the Perspective-n-Point Projection (PNP) method [28] to derive the shooting position of the image from the real 3D coordinates of the target point and its image point coordinates, so as to know the displacement vector between the images. The PNP problem is shown in Figure 6.

Firstly, the initial point cloud is established; i.e., the first two images are matched using the dual-view image sequence reconstruction method, the external reference matrix is solved, and the coordinates of the matched points are triangulated. After getting part of the world coordinates and image point coordinates, we use the PNP method to calculate the camera pose of each image in turn to get the displacement vector, then obtain the rotation matrix by triangulation reconstruction, and match the images in turn to get the 3D point cloud of the target feature for incremental update. After that, we use software such as Geomagic Warp to thicken [29], encapsulate, and add textures to the generated point cloud data and finally get the results of the 3D model of human clothing based on multiview images.

4.3. Example Applications. The experimental data in this section are obtained from the human body photographs taken by the Canon 5D3 camera, and the experimental analysis is performed using the multiview-based image sequence 3D reconstruction method [30, 31]. The image sequence data are shown in Figure 7.
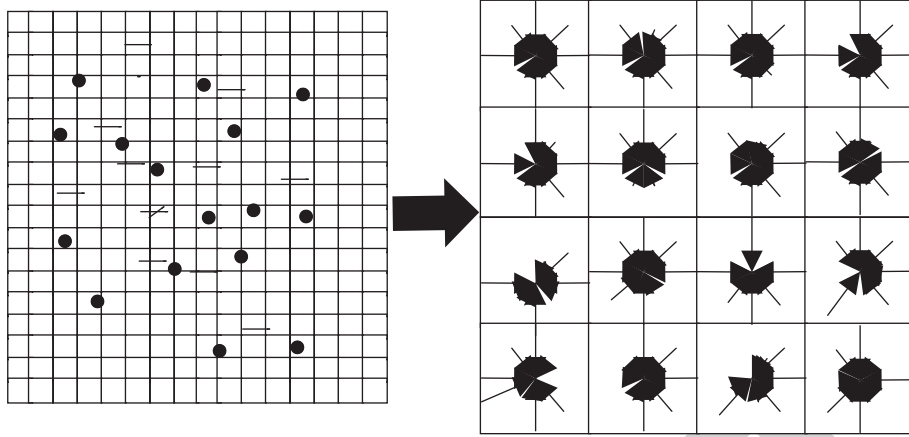
Figure 2: Descriptor gradient histogram.
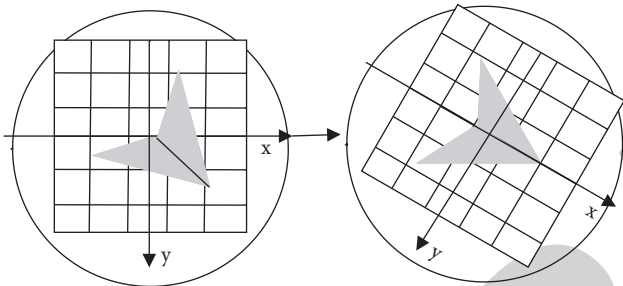
Figure 3: Schematic diagram of rotation direction.
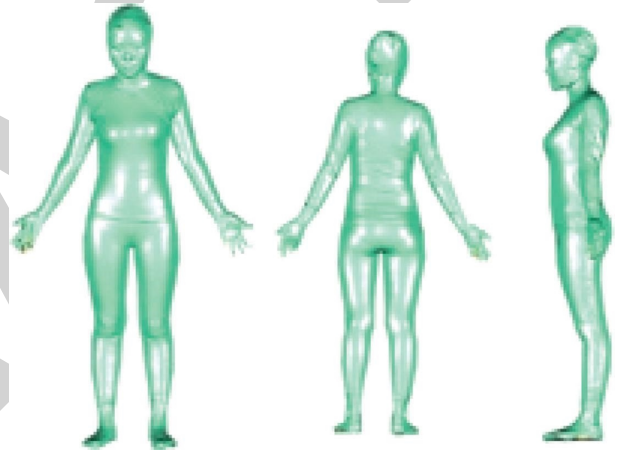
Figure 5: Construction results of 3D models of dual-view image sequences.
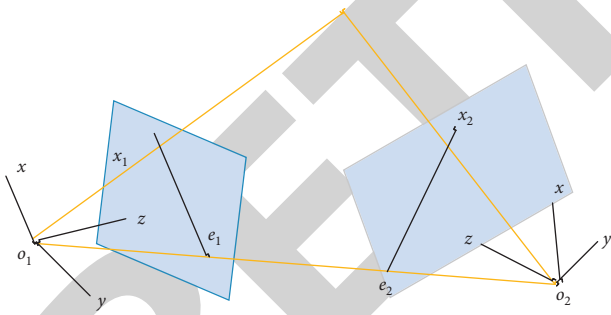
Figure 4: Results of constructing 3D models of dual-view image sequences.

On the basis of the 3D virtual human model, the upper body surface of the human model is intercepted and combined with both automatic extraction and manual definition to obtain human feature information from the PHBC human model [32, 33], to realize the extraction of key points and lines of the human body, and to release the key parts of the human body, such as chest circumference, waist circumference, and hip circumference, to ensure the fit of the garment. Finally, we adjust the B-sample curve for the virtual design of the basic shirt and generate the basic shirt model (Figure 8).
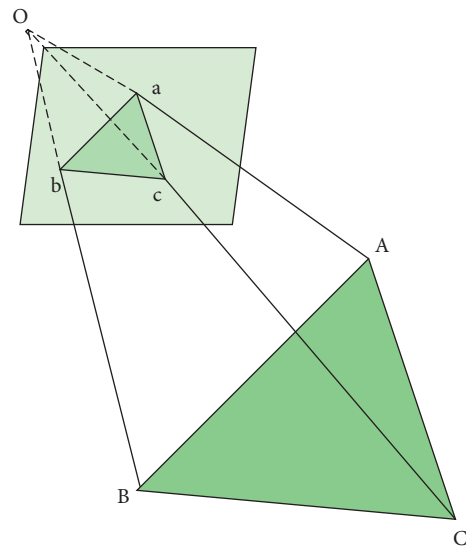
Figure 6: Diagram of the PNP problem.
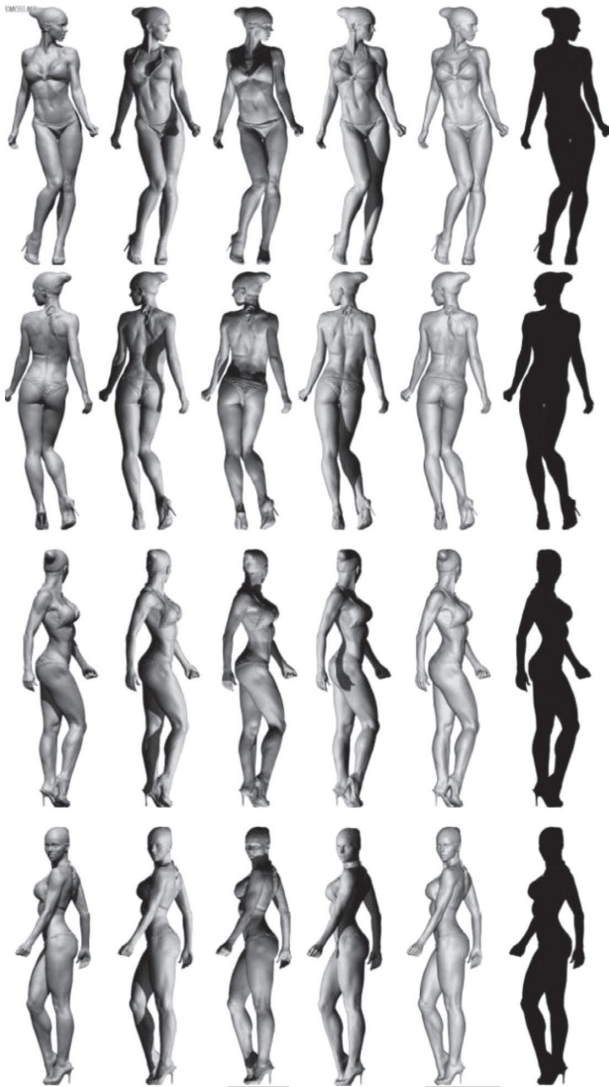
Figure 7: Human multiview image sequence.

Figure 8: Basic shirt style.

Figure 9: Paper sample unfolding.

Due to the irregularity of the human body, a fitted shirt must have certain split lines and the location and shape of the split lines are determined by the design style. First, we imported the base shirt surface into CATIA software [34], selected the free style design platform, and drew the split lines on the base shirt surface. After drawing, we switch to the wireframe and surface design platform, select the unfolded surface from the unfolded shape, select the surface to be unfolded, and set the reference point and unfold direction.

In addition, in order to have the surfaces distributed on the same plane, the same plane should be selected in the unfolding interface, and in this experiment, the ZX plane is selected for both the front and back pieces and the unfolding direction is the Z-axis direction [35]. After unfolding, it is necessary to adjust the position uniformly to form a complete paper sample. After these processes, the unfolded paper pattern of the front and rear surfaces of the net body is obtained (Figure 9).

Figure 10 shows one of the results of edge extraction using the cable operator. From the extraction results, it can be seen that the cable operator has a good effect on the simple image and the clothing edges are successfully extracted, but there are still some background spurious points in the image, so it can be seen that the edge extraction using only the cable operator cannot achieve the expected effect [36].

Figure 10 shows the results of the connected-domain clutter removal, which, when compared, shows that the results are more satisfactory, with the background clutter outside the target being largely removed. Figure 10 shows the new binary image sequence obtained by the preprocessing operation. The size of the images used in the experiments is 589*800 pixels. In one set of experiments, the processing time of 10 images is 64.417 s and the average time per image is only 6.44 s, which are a good image processing result. By comparing the processing results before and after the process, the processing process can remove the background clutter better for the input 3D reconstructed sequence images and has a better processing effect.

Figure 10: Example of processing results of clothing sequence images used in the experiment.

5. Conclusions

In this paper, the 3D model of human clothing is constructed based on multiview image sequences. The eigenmatrix is obtained from the camera parameters, the scale-invariant feature transformation algorithm is used to match the image feature points, the transformation matrix is calculated, and the problems arising in the modeling process are analyzed and solved to realize the 3D reconstruction technology of human clothing based on multiview image sequences. Due to the existing technical conditions, how to weaken the perspective error and realize the fine model construction needs further in-depth research.

Data Availability

The datasets used in this paper are available from the corresponding author upon request.

Conflicts of Interest

The author declares no conflicts of interest regarding this work.

Acknowledgments

References

[1] S. Paquette, "3D scanning in apparel design and human engineering," *IEEE Computer Graphics and Applications*, vol. 16, no. 5, pp. 11–15, 1996.

[2] J. Jian Li, S. K. Zhou, and R. Chellappa, "Appearance modeling using a geometric transform," *IEEE Transactions on Image Processing*, vol. 18, no. 4, pp. 889–902, 2009.

[3] C. Zhang, T. Xie, K. Yang et al., "Positioning optimisation based on particle quality prediction in wireless sensor networks," *IET Networks*, vol. 8, no. 2, pp. 107–113, 2019.

[4] J. Jo, M. Suh, T. Oh et al., "Automatic human body segmentation based on feature extraction," *International Journal of Clothing Science & Technology*, vol. 26, no. 1, pp. 4–24, 2014.

[5] J. M. Nassar, J. P. Rojas, A. M. Hussain, and M. M. Hussain, "From stretchable to reconfigurable inorganic electronics," *Extreme Mechanics Letters*, vol. 9, pp. 245–268, 2016.

[6] A. Haleem and M. Javaid, "3D scanning applications in medical field: a literature-based review," *Clinical Epidemiology and Global Health*, vol. 7, no. 2, pp. 199–210, 2019.

[7] A. Becker, Y. Donchin, and N. Bitterman, "Operating room clothing: design and ergonomic concepts," *Journal of Clinical Anesthesia*, vol. 21, no. 6, pp. 459–461, 2009.

[8] G. Doretto, T. Sebastian, P. Tu, and J. Rittscher, "Appearance-based person reidentification in camera networks: problem overview and current approaches," *Journal of Ambient Intelligence and Humanized Computing*, vol. 2, no. 2, pp. 127–151, 2011.

[9] Q. Ramadan and M. A. M. Gijs, "In vitro micro-physiological models for translational immunology," *Lab on a Chip*, vol. 15, no. 3, pp. 614–636, 2015.

[10] H. Cui and Y. Yan, "The optimization design of uniform's hanging production line," *International Journal of Clothing Science & Technology*, vol. 27, no. 3, pp. 370–389, 2015.

[11] Z. Wei, W. Cong, and Y. Li, "Model-based human gait recognition via deterministic learning," *Cognitive Computation*, vol. 6, no. 2, pp. 218–229, 2014.

[12] D. Wu, C. Zhang, L. Ji, R. Ran, H. Wu, and Y. Xu, "Forest fire recognition based on feature extraction from multi-view images," *Traitement du Signal*, vol. 38, no. 3, pp. 775–783, 2021.

[13] J. S. Lee and C. Jirousek, "The development of design ideas in the early apparel design process: a pilot study," *International Journal of Fashion Design, Technology and Education*, vol. 8, no. 2, pp. 151–161, 2015.

[14] A. Mao, J. Luo, Y. Li, X. Luo, and R. Wang, "A multi-disciplinary strategy for computer-aided clothing thermal engineering design," *Computer-Aided Design*, vol. 43, no. 12, pp. 1854–1869, 2011.

[15] J. M. Lamb and M. J. Kallal, "A conceptual framework for apparel design," *Clothing and Textiles Research Journal*, vol. 10, no. 2, pp. 42–47, 1992.

[16] T. J. Roberts, S. J. Mckenna, and I. W. Ricketts, "Human tracking using 3D surface colour distributions," *Image and Vision Computing*, vol. 24, no. 12, pp. 1332–1342, 2006.

Hindawi

Retraction

Retracted: Application of Thematic Context-Based Deep Learning in Foreign Language Teaching

Scientific Programming

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope

(2) Discrepancies in the description of the research reported

(3) Discrepancies between the availability of data and the research described

(4) Inappropriate citations

(5) Incoherent, meaningless and/or irrelevant content included in the article

(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] Q. Zhao, "Application of Thematic Context-Based Deep Learning in Foreign Language Teaching," *Scientific Programming*, vol. 2021, Article ID 8664219, 11 pages, 2021.

Research Article

Application of Thematic Context-Based Deep Learning in Foreign Language Teaching

Qinhua Zhao [ID] [1,2]

[1]*Department of Foreign Languages, Guangxi University for Nationalities, Nanning 530000, Guangxi, China*
[2]*Euro-Languages College, Zhejiang Yuexiu University, Shaoxing 312400, Zhejiang, China*

Correspondence should be addressed to Qinhua Zhao; 20052036@zyufl.edu.cn

Foreign language teaching is not simply the transfer of knowledge, but rather the placement of students in contexts to explore and discover problems. The thematic contexts do not exist in isolation. Teachers should adopt certain teaching strategies based on thematic contexts, rely on relevant discourse, study the discourse text, and use rich learning and activities as the driving force to highlight students' active experience and emotional experience. In this paper, we propose an ELT (English Language Teaching) affective analysis method based on contextual classification and genetic algorithms. The method first constructs ELT topic sets and ELT topic word sets using the LDA (latent Dirichlet allocation) model, then applies genetic algorithms to each ELT topic word set one by one using ELT label data to automatically iterate the sentiment values of words in the word sets, and finally calculates the sentiment polarity of ELT texts using the sentiment values of words in the word sets. The experimental results show that the accuracy of this method improves 3.12% compared with LDA, the recall rate reaches 87.32%, and F1 reaches 73.79%, which can obtain ELT sentiment information from contextual and nonfeatured sentiment words and effectively improve the accuracy of sentiment classification.

1. Introduction

As society develops, education should also keep pace with the time. Based on the fundamental task of establishing moral education, foreign languages curriculum standards have updated the content of the curriculum and emphasized the thematic contexts [1]. In foreign languages teaching, teaching design based on thematic contexts can guide students to integrate the development of language ability, cultural awareness, ideological quality and learning ability, and shape students' core foreign languages subject literacy [2, 3].

Context is the language environment, including the natural language environment and the classroom language environment. In the process of teaching foreign languages, due to the lack of a natural language environment, students mainly rely on the classroom language environment, in which they learn foreign languages by retelling, remembering, or imagining some scenes in their minds. The latest

"2017 Curriculum" has established three major thematic contexts: "Man and Self, Man and Society, and Man and Nature," and each of these three thematic contexts is divided into more subthematic groups, which are interconnected and inseparable from each other [4, 5]. The thematic context of "Man and Self" advocates a correct and healthy lifestyle, which is conducive to students' better understanding, enriching and perfecting themselves and cultivating a correct human attitude; the thematic context of "Man and Society" is conducive to students' forming good social interaction and establishing good interpersonal relationships [6]. The theme context of "People and Society" is conducive to the formation of good social interaction, the establishment of good interpersonal relationships, the formation of good literacy among students, the cultivation of the innovative spirit of developing information technology, and the better integration of students into social life; the theme context of "People and Nature" advocates understanding nature, knowing nature, caring nature, cultivating students'

curiosity to explore the natural world, and enhancing the ecological concept of people and nature [7, 8].

Sentiment tendency analysis of ELT texts is one of the hot elements of current ELT data mining research. Mining the sentimental tendency of ELT texts can obtain information related to students' liking for foreign languages [9], policy support [10], hot topic tendency [11], and position [12], which are important references for issues such as ELT improvement.

Text sentimental tendency analysis belongs to the category of natural language research, and the diversity of natural language descriptive viewpoints is one of the main factors affecting the accuracy of text sentiment tendency analysis. Compared with media with a good content classification such as news, forums, and postings, ELT [13] has broad content and poor classification. Currently, there are two main methods of text sentiment analysis, based on sentiment dictionaries and based on machine learning, both of which perform text sentiment polarity calculation through some algorithm based on the subcategorization of texts. A large number of research results show that the accuracy of sentiment analysis by these two research methods is constrained by the relevance of the text content domain. Since the same word may show different sentiment polarity in different contexts, it is difficult to guarantee the accuracy of sentiment analysis of ELT texts without differentiating word contexts. The LDA extended model is one of the most important methods for text sentiment analysis, but the current research fails to consider the difference in sentiment polarity of the same word in different contexts and the influence of non-featured sentiment words on the sentiment polarity of ELT texts. Therefore, this paper proposes an ELT sentiment analysis method based on contextual classification and genetic algorithms.

2. The Role of Thematic Contexts in Reading Instruction

2.1. Thematic Contexts Can Enhance Students' Interest in Reading.
Compared with foreign languages teaching in college, foreign language teaching in college has more characteristics. For example, there are more diverse genres, more lengths, more complex sentences, and much more difficult in the discourse. For this reason, students need to enhance their comprehension of textual content and improve their language skills. Teachers also need to adopt active teaching strategies to meet the challenges. College students are often confronted with boring and tasteless topics in reading materials and feel that they do not match their actual level, which leads to students' reluctance to read the materials or even their inability to read them and their low interest in reading. In college foreign language reading teaching, teachers combine the three categories of contextual themes with reading materials, which can increase the connection between reading materials and real life; let students experience different cultures in learning, feel real life, learn knowledge, and understand language in real contexts, which can largely motivate students to actively participate in learning, give full play to their own initiative, help students apply what they learn, and increase reading interest [14, 15].

2.2. Thematic Contexts Facilitate Students' Better Understanding of Texts.
When reading texts, students often focus their attention on heavy words or understanding long and difficult sentences, neglecting to grasp the overall meaning of the text and lacking knowledge of the logical relationships between small sentences in the paragraphs of the text, thus failing to correctly access information in the text, discover the thematic meaning of the text accurately, and read and understand reading materials on common topics. Teaching college foreign languages reading under the guidance of thematic contexts requires teachers to first study the text in depth and then guide students to read the text and analyze it, grasp the culture and meaning embodied in the text, and discover the thematic meaning, which facilitates students to use a variety of methods to obtain information creatively. Teachers are expected to use thematic reading text materials to provide appropriate instruction to students and promote students' initiative to read and think actively and take the best out of them [16]. By creating authentic thematic contexts, teachers can guide students to relate to the context and combine thematic contexts to grasp the content of the text, understand the deeper meanings of key sentences, avoid reading misunderstandings caused by biased generalizations and words that do not make sense and ultimately improve students' reading comprehension skills [17].

2.3. Thematic Contexts Can Improve Students' Foreign Languages Application Skills.
Themed foreign language reading teaching in college is a process of active discovery and learning with students as the main body, and it is also a learning process that follows students' cognitive rules from the surface to the deeper level gradually. Due to the limitations of China's traditional examination-based education system, schools currently focus too much on students' test scores in college foreign language teaching, ignoring students' ability to apply foreign languages in real-life situations, resulting in "dumb foreign languages" that students can learn but cannot use [18]. In order to change the current situation of college foreign languages reading teaching, teachers should change the traditional teaching concept, design the whole teaching based on the theme context, and create a context that is closely related to the meaning of the theme and students' real life. In the process of in-depth study of the text, teachers should aim at solving problems, focus on learning foreign languages language knowledge and language skills, and closely connect the theme of the text with students' lives [19, 20]. At the same time, teachers should cultivate students' logical and critical thinking and diverse cultural perspectives by comparing Chinese and foreign cultures, adopt thematic contextual teaching, make a profound analysis of texts, reasonably design thematic contexts that help students improve their comprehension as well as their application skills, and adopt creative teaching methods

and approaches to actively improve students' practical foreign languages application skills.

3. Related Work

Currently, there are two main methods of text sentiment analysis based on sentiment dictionary and based on machine learning. The method based on sentiment dictionary is to first extract the sentiment feature words of the text [21], then compare the sentiment feature words with the words in the sentiment dictionary, and use the sentiment polarity of the marked words in the sentiment dictionary to calculate the foreign languages teaching sentiment tendency. The classification accuracy of this method depends on the sentiment dictionary, and the goodness of the sentiment dictionary directly affects the results of sentiment tendency calculation. The machine learning-based method extracts text features first and then applies some algorithms to the features for classification [22] to get the text sentimental tendency. Machine learning-based methods are divided into three types of methods: strongly supervised, weakly supervised, and unsupervised. The main strongly supervised methods are support vector machines [23]. The accuracy of plain Bayesian and decision trees [24, 25] depend on the accuracy of the labeled data. Weakly supervised methods mainly include long- and short-term memory networks [26], convolutional neural networks [27], and so on. These methods require massive labeled data to train the model to ensure accuracy. Unsupervised methods mainly include LDA [28], K-nearest neighbor algorithm [14], random forest [5], and so on. Compared with supervised methods, unsupervised methods do not depend on labeled data and are less affected by the size of data.

To address the above problems, in order to further improve the accuracy of text sentiment polarity analysis using the LDA model extension method, this paper proposes an ELT sentiment analysis method based on context classification and genetic algorithm. The method first classifies ELT into contextual topics using the LDA model and divides ELT words into different contextual topics to form ELT topic sets and ELT topic word sets; then for each topic of ELT and topic word sets, a genetic algorithm is used to calculate the sentiment values of all words (including sentiment feature words and nonsentiment feature words), and finally, the sentiment values of words are used to calculate ELT sentiment tendency.

4. ELT Theme Analysis Method

4.1. Overall Process. The overall process of ELT sentiment classification method based on contextual classification and genetic algorithm is as follows: (1) ELT data preprocessing, screening, and word separation of ELT data; (2) LDA ELT topic contextual word set construction, using LDA to classify ELT in topic context and construct ELT topic word set; and (3) genetic algorithm based on topic ELT sentiment tendency calculation. The overall process is shown in Figure 1.

5. ELT Data Preprocessing

The ELT platform is aimed at the mass population, and some students post information with unclear purpose, and a considerable number of these sentences do not carry an opinion tendency. Therefore, nonopinion sentences are removed first, and only sentences with emotional tendency are kept before word separation. Main Chinese word separation tools are Jieba, SnowNLP, THULAC, NL-PIR, PKU-SEG, and so on [29]. Since the content of foreign languages teaching is relatively brief, PKU-SEG can maintain the original word formation relationship of sentences better.

5.1. LDA ELT Theme Context Word Set Construction. ELT is a relatively open and free media; compared with news, forums, and other media with good thematic classification performance, its content range is broader and more arbitrary, without a strict classification structure, so there are quite a lot of words in the ELT text set showing different sentiment tendencies in different contexts. LDA is a document topic generation probability model, which is able to obtain "document—topic," "topic—word," and "topic—word." This paper applies the LDA model to categorize ELT document sets and their words by topic context and constructs ELT topic sets and ELT topic context word sets based on topic context division.

5.2. LDA ELT Topic Context Classification. The LDA ELT topic model is shown in Figure 2. k topics are set manually in the LDA ELT topic model, and the preprocessed corpus D has m ELTs, which is denoted as $D = \{d_1, d_2, \ldots, d_m\}$, and the number of words that are deemphasized after splitting ELTs is c, and the word set is denoted as $W = \{\text{word}_1, \text{word}_2, \ldots, \text{word}_c\}$. Topic conditional distribution of ELT i is denoted as $p(\overrightarrow{z}_i | \overrightarrow{\alpha}), i \in (1, 2, \ldots, m)$, and the topic conditional distribution of all documents can be obtained by using the LDA ELT topic model, which is normalized as shown in the following equation:

$$p(\overrightarrow{z} | \overrightarrow{\alpha}) = \prod_{i=1}^{m} p(\overrightarrow{z}_i | \overrightarrow{\alpha}) = \prod_{i=1}^{m} \frac{\Delta(\overrightarrow{n}_i + \overrightarrow{\alpha})}{\Delta(\overrightarrow{\alpha})}, \quad (1)$$

where $\overrightarrow{n}_i$ denotes the number of words distributed under k topics in the i-th ELT, and α is a k-dimensional hypercalcemia variable.

Similarly, the distribution of subjective conditions for word word_t can be obtained as $p(\overrightarrow{\text{word}} | \overrightarrow{z}, \overrightarrow{\beta})$, $t \in (1, 2, \ldots, c)$ and normalized as shown in the following equation:

$$p(\overrightarrow{\text{word}} | \overrightarrow{z}, \overrightarrow{\beta}) = \prod_{j=1}^{k} p(\overrightarrow{\text{word}}_t | \overrightarrow{z}, \overrightarrow{\beta}) = \prod_{j=1}^{k} \frac{\Delta(\overrightarrow{n}_j + \overrightarrow{\beta})}{\Delta(\overrightarrow{\beta})}, \quad (2)$$

where $\overrightarrow{n}_j$ denotes the number of words under the j-th topic and β is a c-dimensional hypernatremia variable.

Microblog data
preprocessing

Construction of
LDA subject
word set

Analysis of emotional
tendency based on
genetic algorithm

NPCC
dataset

Da topic model

Genetic
algorithm

Preprocessing
(opinion
sentence
selection, word
segmentation)

Weibo topic
probability
distribution

Emotional
value of
microblog
subject word set

Corpus

Microblog
theme set and
microblog
theme word set

Microblog
affective
tendency
calculation

FIGURE 1: Overall process.
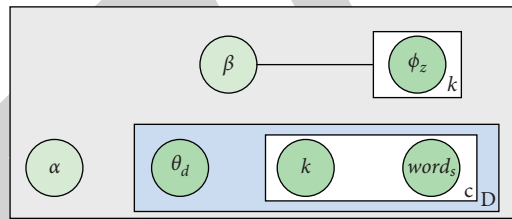
FIGURE 2: LDA ELT theme model.

Combining equations (1) and (2), the joint distribution of topics and words can be obtained as shown in the following equation:

$$p(\overrightarrow{\text{word}}|\overrightarrow{z}) \propto p(\overrightarrow{\text{word}}, \overrightarrow{z}|\overrightarrow{\alpha}, \overrightarrow{\beta}) = p(\overrightarrow{z}|\overrightarrow{\alpha})p(\overrightarrow{\text{word}}|\overrightarrow{z}, \overrightarrow{\beta}) = \prod_{i=1}^{m}\frac{\Delta(\overrightarrow{n_i} + \overrightarrow{\alpha})}{\Delta(\overrightarrow{\alpha})}\prod_{j=1}^{k}\frac{\Delta(\overrightarrow{n_j} + \overrightarrow{\beta})}{\Delta(\overrightarrow{\beta})}. \tag{3}$$

The distribution of topics after removing the t-th word is represented by (t). The conditional probability of the topic for the t-th word is shown in the following equation using Gibbs' sampling method:

$$p\left(z_i = j|\overrightarrow{\text{word}}, \overrightarrow{z}_{(t)}\right) \propto p\left(z_i = j, \text{word} = t|\overrightarrow{\text{word}}_{(t)}, \overrightarrow{z}_{(t)}\right) = \frac{n_{d,(t)}^{j} + \alpha_j}{\sum_{s=1}^{k} n_{d,(t)}^{s} + \alpha_s \sum_{f=1}^{t} n_{j,(t)}^{t} + \beta_f}\sum_{j,(t)}^{f} + \beta_f. \tag{4}$$

The probability distribution of all the words in ELT d was summed to obtain the probability distribution of tweet

under k topics $A_d = \left\{ p_{d_1}, p_{d_2}, \ldots, p_{d_i} \right\}$, and the value with the highest probability was selected as the topic context of the d-th tweet $p_{d\text{Max}} = \max\left\{ p_{d_1}, p_{d_2}, \ldots, p_{d_i} \right\}$.

5.3. ELT Theme Context Word Set Construction.

According to the above maximum probability division method, the subject context classification of the ELT set is completed to form the ELT topic set $T = \{T_1, T_2, \ldots, T_k\}$, where $T_j = \left\{ d_{jl}, d_{j2}: \ldots, d_{jy} \right\}$, $j \in (1, 2, \cdots\cdots. k)$, y denote the number of ELTs of topic j. The word set $Z_j, Z_j = (v_{j1}, v_{j2}, \ldots, v_{jn})$ of topic j is obtained by dividing and deduplicating the tweets in T_j, and n denotes the number of deduplicated words of the j-th topic. The word set of all k topics constitutes the LDA ELT topic word set $Z = \{Z_1, Z_2, \ldots, Z_k\}$, and the pseudocode of the LDA ELT topic word set construction algorithm is shown in Algorithm 1.

5.4. Genetic Algorithm-Based Calculation of Affective Disposition for Teaching Foreign Languages as a Foreign Language.

Considering the influence of nonfeatured sentiment words on the sentiment tendency of ELT texts, this paper calculates the sentiment values of all words in each topic context separately after calculating the LDA ELT topic word sets, which include nonfeatured sentiment words and feature sentiment words, and finally calculates the ELT sentiment tendency using the sentiment values of words. Sentiment values of words in each topic word set are obtained automatically by genetic algorithm calculation using ELT (labeled data) with manually labeled sentimental tendencies. The sentiment value calculation method first assigns a random initial sentiment value to the words within a predefined range; then the optimal sentiment value is obtained by designing the objective function and fitness function associated with the label data to self-optimize the sentiment value of the words; and finally, the optimal sentiment value of the topic words is used to calculate the optimal sentiment tendency value of ELT [30]. Topic context word set Z_j is used as the individual in the genetic algorithm, and individual Chrom_x corresponds to the sentimental value of all words in Z_j, which is $\text{Chrom}_x = \{w_{x1}, w_{x2}, \ldots, w_{x3}\}$. w_{xt} is the sentimental value of the t-th word in individual x. The sentimental value of each word corresponds to the chromosome code of the individual. The population is composed of M individuals, denoted as $P = \{\text{Chrom}_1, \text{Chrom}_2, \ldots, \text{Chrom}_m\}$, and the initial word sentimental value of an individual is a random value of $[-10, 10]$, as shown in Figure 3. The population is iteratively optimized in the genetic algorithm, and the individual word sentimental value calculated when the number of iterations reaches a predefined value is the optimal sentimental value of all words in the topic context.

5.5. Genetic Algorithm Objective Optimization Function.

In fact, some words do not have the same sentiment tendency in different contexts, so corresponding to this

situation, this paper sets the same word to have different sentiment values in different individuals, and classifying words into different thematic contexts is to consider this variability. In order to make the sentiment tendency of ELT in individuals close to the sentiment tendency of labeled data, that is, in order to apply labeled data to automatically obtain the sentiment tendency of words, the objective function of genetic algorithm is designed in this paper to achieve word sentiment value optimization [13]. The sentimental value of the s-th ELT under topic j is calculated using the following equation:

$$\text{Senti}\left(Z_j, T_j, \text{Chrom}_x, d_{js}\right) = \sum_{word \in d_k} w_{jt}\left(\text{Chrom}_x\left(\text{word}_t\right)\right), \tag{5}$$

where word_t is the word in ELT d_{js}, $w_{jt}\left(\text{Chrom}_x\left(\text{word}_t\right)\right)$ indicates the sentiment value of word word_t in individualChrom$_x$. When $\text{Senti}\left(Z_j, T_j, \text{Chrom}_x, d_{js}\right)$ is greater than or equal to 0, it is positive, and vice versa, it is negative. $\text{class}\left(d_{js}\right)$ indicates the affective tendency of ELT d_{js} as shown in the following equation:

$$\text{class}\left(d_{js}\right) = \begin{cases} \text{positive}, \text{Senti}\left(Z_j, T_j, \text{Chrom}_x, d_{js}\right) \geq 0, \\ \text{negative}, \text{Senti}\left(Z_j, T_j, \text{Chrom}_x, d_{js}\right) < 0. \end{cases} \tag{6}$$

This paper sets $\text{Acc}\left(\text{Chrom}_x, T_j\right)$ as the degree of difference between the affective tendency of ELT in individual Chrom_x and the affective tendency of the labeled data under topic j, as shown in the following equation; the smaller the value, the closer the affective tendency of ELT in individual Chrom_x is to the affective tendency of the labeled data.

$$\text{Acc}\left(\text{Chrom}_x, T_j\right) = 1 - \frac{\text{num}_{1 \leq s \leq y, d_k \in T_j} \text{class}\left(d_{js}\right)}{y}, \tag{7}$$

where T_j is the set of topics d_{js}. The number of ELT affective tendencies that are consistent with the labeled ELT data affective tendencies is calculated in individual 33 by equation (6). The minimum difference degree individuals are determined by setting the objective optimization function according to equation (7) as shown in the following equation:

$$\text{Acc}_{\min} = \min_{1 \leq i \leq M, \text{Chrom}_x \in P}\left(\text{Accuracy}\left(\text{Chrom}_x, T_j\right)\right). \tag{8}$$

5.6. Genetic-Algorithm-Based Word Sentiment Value Calculation.

In order to make the probability of individuals with the smaller variance being retained higher, the adaptation function is set in this paper as shown in the following equation:

Input: $D = \{d_1, d_2, \ldots, d_m\}$, $A_d = \{p_{d1}, p_{d2}, \ldots p_{dk}\}$
//ELT corpus and ELT topic probability distributions
Output: $Z = \{Z_1, Z_2, \ldots, Z_k\}$//LDA ELT subject headings
(1) for $i = 1$ to m do//traverse ELT D
(2) //Get the topic T to which the i-th ELT topic probability maximum belongs
(3) $T_a \leftarrow \text{Get}(P_{i\text{Max}})$
(4) //put the i article of ELT into the T_ath topic
(5) Topic.append (T_a, i)
(6) end for
(7) for $j = 1$ to k do//traverse k topics
(8) //ELT splitting and deduplication in the j-th topic
(9) lexical $\leftarrow$ set(seg(Topic[j]))
(10) Z_j append(lexical)//constitute a set of subject words Z_j.
(11) end for
(12) return (Z)//return to LDA ELT subject word set

ALGORITHM 1: LDA ELT topic word set construction algorithm.

FIGURE 3: Some of the initial word sentiment scores correspond to.
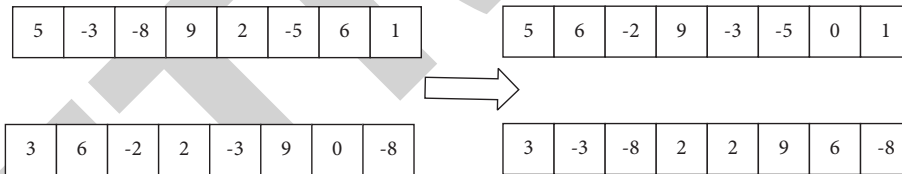
FIGURE 4: Multipoint crossover.
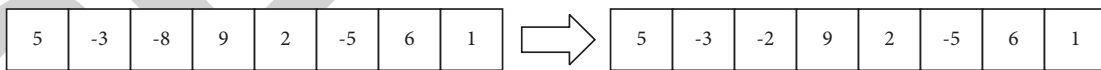
FIGURE 5: Genetic variation.

$$\text{gather edfit}\left(\text{Chrom}_x, T_j\right) = \frac{y}{\text{Acc}\left(\text{Chrom}_x, T_j\right)} \times (-1) + \left(y - \frac{y}{\text{Acc}\left(\text{Chrom}_x, T_j\right)}\right), \tag{9}$$

where $y/\text{Acc}(\text{Chrom}_x, T_j)$ is the number of ELTs in which the emotional disposition was judged incorrectly in individual Chrom_x and $(y - y/\text{Acc}(\text{Chrom}_x, T_j))$ indicates the number of ELTs in which the emotional disposition was judged correctly in individual Chrom_x. The roulette wheel method was used for individual selection, using equation (9) so that the greater the fitness (smaller the variance), the greater the probability of individuals being selected for retention. The selected individuals are then subjected to

crossover and mutation operations to produce new individuals. Individual selection, crossover, and mutation operations are repeatedly performed in the genetic algorithm to optimize the population of individuals iteratively until a predetermined number of iterations is reached and the calculation is stopped, and finally, the individual with the smallest variance is selected as the word sentiment value using the objective optimization function. The pseudocode of word sentimental value calculation based on a genetic

```
Input: D = {d₁, d₂, ..., dₘ}, T = {T₁, T₂, ..., Tₖ}, Z = {Z₁, Z₂, ..., Zₖ}
//ELT corpus, ELT topic collection, ELT topic word collection
Output: Chromₓ = [wₓ₁, wₓ₂, ..., wₓₙ]
/minimum variance individual (optimal sentiment value of topic word set words)
 (1) P = [Chrom₁, Chrom₂, ..., Chrom_M]//Randomly generated populations P
 (2) for j = 1 to MAX do//MAX is the maximum number of iterations
 (3)  for ch = 1 to M do//traverse the population
 (4)   fit ← fitness (ch)//calculate individual fitness
 (5)   fitness.append (fit)//save fitness
 (6)  end for
 (7)  new_p ← select(P, fitness)//select operation
 (8)  new_p ← cross(new_p)//cross operation
 (9)  new_p ← mutation (new_p)//mutation operation
(10)  p ← new_p//new population P
(11) end for
(12) Chrom_min ← Acc_min //Minimum variance individual
(13) return (Chrom_min)//Return the minimum variance individual
```

ALGORITHM 2: Genetic algorithm-based word sentiment value calculation.

algorithm is shown in Algorithm 2, and the individual crossover and variation operations are shown in Figures 4 and 5.

5.7. Calculation of Emotional Disposition in ELT.

The minimum variance individual 11 was obtained by the genetic algorithm, and the codes of chromosomes in the individual corresponded to the optimal sentiment values of words in the subject word set. In the minimum variance individual 22, the sentiment values of all words in ELT are first summed up by equation (5) to obtain the sentiment value of ELT and then judged by equation (6) [31], if the sentiment value is greater than or equal to 0, it means that the ELT has a positive sentiment tendency, and the opposite means that it has a negative sentiment tendency. The calculation example is shown in Figure 6.

6. Experimental Results and Analysis

6.1. Experimental Data Set.

The datum was obtained from the 2012–2014 NLPCC public data set [22], with 17,253 ELTs. There were 7,188 ELTs after removing nonopinion sentences, and this data was used as a corpus for the calculation of affective tendencies for ELTs, of which 3,314 were positive and 3,874 were negative, and the tenfold cross-validation was used to train and test the method of this paper.

The NLPCC data set has eight labels: none, happiness, like, sadness, disgust, anger, fear, and surprise. The eight labels are simplified into two types of labels: positive and negative, as shown in Table 1.

After the labels were categorized, the ELT content was stored in a uniform format, and the data format is shown in Table 2, with the polarity 1 for ELT indicating positive and −1 for negative.

6.2. Experimental Procedure and Results.

In this paper, the effects of all words on the effective polarity of ELT were involved in the calculation, and all words in ELT were retained after the splitting of ELT. Sample results of PKU-SEG splitting are shown in Table 3.

After ELT word classification, the LDA ELT topic model is used to construct the topic word sets. K values of the number of topics in the LDA ELT topic model need to be set in advance, and the selection of suitable k values is beneficial to topic classification. In this paper, the k value is set to 5 (the collected data are divided into 5 topics), and the sample ELT topic context classification is shown in Table 4.

The results of the ELT theme context classification are: theme one is related to foreign languages and foreign languages commentary; theme two is related to personal emotional expressions; theme three is social status commentary and event descriptions; theme four is dynamic ELT commentary about socially prominent people; theme five is more colloquial popular Internet phrases, and the classification results are consistent with reality. After completing the ELT topic context classification, the ELT topic word set is constructed, as shown in Table 5.

After obtaining the topic word sets, the word sentiment values were calculated using a genetic algorithm-based method for calculating sentimental tendencies in ELT. The population P is randomly generated in the algorithm, and the population size is set to 1,000, and the selection, crossover, and mutation operations are performed on the individuals. In this paper, the number of words after deweighting the current topic ELT words is used as the individual coding length, and the chromosome coding is coded with real integers in the [−10, 10] interval, and the initial word sentiment values are shown in Table 6. The initial word sentiment values are shown in Table 6. The genetic algorithm is run with the labeled data, objective function and fitness function for each word set in each topic, so that the sentiment values are optimized iteratively until the predetermined number of iterations is reached and the calculation stops (the iteration threshold is set to 2,000 in this paper), and the results of word sentiment value optimization are shown in Table 7.

This is	One	Good	Bad	late	Company	go to work	Good morning
0	-1	8	-9	-2	1	-6	3

Emotional value table of words

Test data set → This is a good unit → 0-1+8+1=+8 → +8>,positive electrode
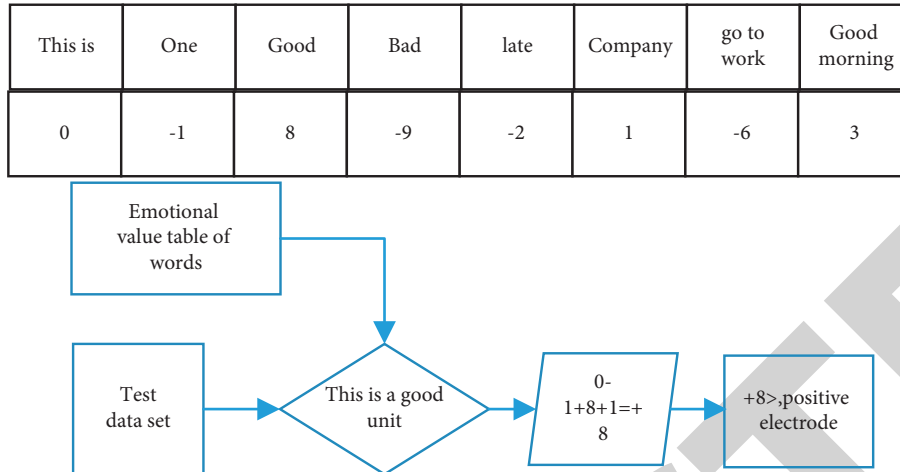
FIGURE 6: Example of emotional disposition calculation.

TABLE 1: Label classification.

Positive	Negative
None	Sadness
Happiness	Disgust
Like	Fear
	Surprise

TABLE 2: Sample ELT texts.

Serial number	Microblog content	Polarity
1	I feel strange and uncomfortable.	−1
2	The film is really moving; especially with the development of the plot, it is melodious and beautiful.	1
3	Suddenly, I found that Liu Xin's song names are so powerful	1
4	It's too late.	−1
5	These days have been wonderful, all kinds of wonderful.	1
6	Superman is everywhere, and Kobe responds only to roars and mistakes.	−1
7	I don't know if the performance of iPad3 is good. If it is good, I can consider buying one.	1
8	Our unit is a good unit. Although we get off work late, we go to work early!	−1
9	After reading a lot about the laptop dock of MTO ATRIX, I like it more and more.	1
10	Personality explosion	1

TABLE 3: Sample ELT PKU-SEG word separation results.

Serial number	Word segmentation result
1	I feel strange and uncomfortable.
2	I've had a wonderful time these days.
3	Suddenly, I found that Liu Xin's song names are so powerful.
4	Personality explosion

TABLE 4: Sample ELT topic context classification.

Topic 1	I don't know whether the performance of iPad3 is good or not. If it is good, I can consider buying a laptop dock that has read a lot about MTO ATRIX. I like it more and more.
Topic 2	The past few days have been wonderful. I feel strange and uncomfortable in my heart. Our unit is a good unit. Although we get off work late, we go to work early.
Topic 3	The film is really moving, especially with the development of the plot; the melodious songs are really sensational.
Topic 4	Suddenly, I found that Liu Xin's songs were so powerful. They were superman everywhere. Kobe responded only with roars and mistakes.
Topic 5	Burst out of character. It's so sad.

TABLE 5: Selected subject headings.

Theme	Partial word set				
Subject word set 1	iPad3	Performance	Screen	Keyboard	Product
Subject word set 2	Make complaints	Speechless	Night	Company	Good morning!
Subject word set 3	Pain	Development	Labour	Reactionary	Challenge
Subject word set 4	Fire	Sing	Night shanghai	Vocal concert	Baidu
Subject word set 5	Yes	Ah	Burst	Vulnerable	Alas

TABLE 6: Initial word sentiment values.

Individual number	Make complaints	Speechless	Happiness	Awesome	Fond dream	Life
1	−8	5	−3	8	9	2
2	−5	2	5	2	−4	8
3	2	−4	2	1	3	−5
4	2	1	−6	−2	−3	−1

TABLE 7: Optimization results of word sentiment values.

Individual number	Make complaints	Speechless	Happiness	Awesome	Fond dream	Life
1	−9	−3	9	7	5	2
2	−5	−2	5	2	−4	−1
3	−2	−7	6	1	2	0
4	−5	−8	5	5	3	−1

6.2.1. Mediating Effect Test.

The mediation effect was examined using the nonparametric percentile Bootstrap method with the PROCESS 2.16 plug-in installed in SPSS 24.0 software. The results showed that desire mediated the relationship between perceived behavioral control, sense of relatedness, and willingness to act; desire partially mediated the relationship between perceived behavioral control and willingness to act; and desire fully mediated the relationship between sense of relatedness and willingness to act (see Table 3).

In the genetic algorithm-based method for calculating affective tendencies in ELT, the objective function is used to select the individual with the least variance in the optimized population as the current topic word affective value, as shown in Table 8.

After the classification of ELT topic contexts, there may be the same words in different topics, and this classification method is consistent with the reality that there are differences in affective tendencies of the same word in different topic contexts, and the affective values of the same word in different topics are shown in Table 9.

After the classification of topics, the emotional tendency of ELT is judged by calculating the sum of the emotional values of ELT subwords under the topic, and when the sum of the emotional values of ELT subwords is greater than or equal to 0, the emotional tendency of ELT is positive, and when it is less than 0, the emotional tendency of ELT is negative.

6.3. Comparison of Methods.

The present method (LDA-GA) was compared with LDA, plain Bayesian classification (NB), random forest (RF), and decision tree (DT) for the

TABLE 8: Example of optimal word sentiment values.

Make complaints	Speechless	Happiness	Awesome	Fond dream	Life
−2	−7	6	1	2	0

TABLE 9: Examples of sentiment values for the same word in different themes.

Terms	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5
Good	8	5	2	6	2
Large	4	1	−1	3	−7
Yes	2	0	1	3	−2
Sad	−9	−8	−2	−5	−9

calculation of affective disposition in ELT. Precision (P), recall (R) and F_1 values were used as evaluation indicators, and the comparison results are shown in Table 10.

The experimental results show that the F_1 values of the LDA method are higher than those of the DT, NB, and RF algorithms. The reason is that the LDA method for ELT sentiment analysis is based on semantic "text-word" topic classification, while the DT, NB, and RF algorithms convert words into word vectors without considering the semantic information of words, which leads to the less-than-optimal results of microbial sentiment calculation. The accuracy of the LDA-GA method in this paper is not satisfactory. The accuracy, recall, and F_1 value of the LDA-GA method in this paper are higher than those of the LDA method. The reason is that the LDA method only uses feature sentiment words to calculate ELT sentiment polarity, while the method in this paper uses a genetic algorithm to calculate all word

Table 10: Comparison experimental results of emotional disposition calculation.

Method	Accuracy (%)	Recall (%)	F_1 (%)
LDA	60.63	78.02	68.23
DT	52.44	73.06	61.07
NB	54.31	85.00	66.23
RF	58.67	74.63	65.69
LDA-GA	63.75	87.32	73.69

sentiment values, unites feature sentiment words and nonfeature sentiment words to calculate ELT sentiment polarity, and calculates word sentiment values according to different topic contexts to distinguish the sentiment polarity of the same word in different topic contexts.

7. Conclusions

In this paper, we propose an ELT topic analysis method based on contextual classification and genetic algorithms. The method first constructs ELT topic sets and ELT topic word sets using the LDA model, then applies a genetic algorithm to each ELT topic word set one by one to automatically iterate and calculate the sentiment value of words in the word sets, and finally calculates the sentiment polarity of ELT text by the sentiment value of words in the topic word sets. The experimental results show that the accuracy, recall, and F_1 of this method are improved compared with those of LDA, plain Bayesian classification, random forest, and decision tree-based ELT sentiment analysis methods. The method in this paper requires repeated iterative computation in the genetic algorithm, which is time-consuming, and the next research work is to consider the problem of genetic algorithm acceleration.

In the future, we will optimize the algorithm to make it more robust and stable; we will optimize its scalability so that it can be applied in educational scenarios in different fields.

Data Availability

The data used in this paper are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest regarding this work.

References

[1] M. H. Santosa, "Learning approaches of Indonesian EFL gen Z students in a flipped learning context," *Journal of English and Foreign Languages*, vol. 7, no. 2, pp. 183–208, 2017.

[2] Z. Wang, "Study on teaching of foreign languages reading: under thematic progression model," *Open Journal of Modern Linguistics*, vol. 05, no. 1, pp. 73–78, 2015.

[3] M. Abrar, A. Mukminin, A. Habibi, and A. Fadhil, "If our foreign languages isn't a language, what is it? Indonesian EFL Student Teachers' Challenges Speaking foreign languages," *Qualitative Report*, vol. 23, no. 1, pp. 129–145, 2018.

[4] A. Aunurrahman, A. Hikmayanti, and Y. Yuliana, "Teaching English using a genre pedagogy to Islamic junior high school students," *Journal of English and Foreign Languages*, vol. 10, no. 1, pp. 1–24, 2020.

[5] S. Kahn and A. R. Lewis, "Survey on teaching science to K-12 students with disabilities: teacher preparedness and attitudes," *Journal of Science Teacher Education*, vol. 25, no. 8, pp. 885–910, 2014.

[6] O. Jonathan, S. Shirley, C. Andri, C. H. Richardson, and K. Richardson, "Learning to argue: a study of four schools and their attempt to develop the use of argumentation as a common instructional practice and its impact on students," *Journal of Research in Science Teaching*, vol. 50, no. 3, pp. 315–347, 2013.

[7] T. Xie, C. Zhang, Z. Zhang, and Y. Kai, "Utilizing active sensor nodes in smart environments for optimal communication coverage," *IEEE Access*, vol. 7, pp. 11338–11348, 2018.

[8] S. Ahn and H. Han, "US student teachers' cross-cultural teaching abroad experience at a Korean high school," *Journal of Education & Culture*, vol. 24, no. 5, pp. 733–758, 2018.

[9] A. Andriani, V. D. Yuniar, and F. Abdullah, "Teaching English grammar in an Indonesian junior high school," *AL-ISHLAH: Jurnal Pendidikan*, vol. 13, no. 2, pp. 1046–1056, 2021.

[10] M. Y. Abdullah, H. M. H. A. Ghafri, and K. S. H. A. Yahyai, "A qualitative study on the best motivational teaching strategies in the context of Oman: perspectives of EFL teachers," *English Language Teaching*, vol. 12, no. 3, pp. 57–64, 2019.

[11] A. Leis, "Study abroad and willingness to communicate: a case study at junior high school," *The Language Teacher*, vol. 39, no. 3, pp. 3–9, 2015.

[12] M. Ramos, "Teaching foreign languages for medical translation: a corpus- based approach," *Language Teaching Research Quarterly*, vol. 8, no. 2, pp. 25–40, 2020.

[13] T. D. Holmlund, K. Lesseig, and D. Slavit, "Making sense of "STEM education" in K-12 contexts," *International journal of STEM education*, vol. 5, no. 1, pp. 1–18, 2018.

[14] C. Zhang, T. Xie, K. Yang et al., "Positioning optimisation based on particle quality prediction in wireless sensor networks," *IET Networks*, vol. 8, no. 2, pp. 107–113, 2019.

[15] M. M. Rahman, M. K. M. Singh, M. K. M. Singh, and A. Pandian, "Exploring ESL teacher beliefs and classroom practices of CLT: a case study," *International Journal of Instruction*, vol. 11, no. 1, pp. 295–310, 2018.

[16] M. Ram, S. Mahendra, R. Campus, and N. Kathmandu, "College teachers' experiences in classroom ecology of language teaching: a study from the phenomenological perspective," *Journal of foreign languages Education and Teaching*, vol. 04, no. 4, pp. 2685–2743, 2020.

[17] R. S. Fernanda and J. G. C. Claudinei, "The use of active methodology in nursing care and teaching in national productions: an integrative review," *Revista da Escola de Enfermagem da USP*, vol. 46, no. 1, pp. 208–218, 2012.

[18] D. M. M. Zin, S. Mohamed, K. A. Bakar, and N. K. Ismail, "Further study on implementing thematic teaching in preschool: a needs analysis research," *Creative Education*, vol. 10, no. 12, pp. 2887–2898, 2019.

[19] L. Wang, C. Zhang, Q. Chen et al., "A communication strategy of proactive nodes based on loop theorem in wireless sensor networks," in *Proceedings of the 2018 Ninth International Conference on Intelligent Control and Information Processing (ICICIP)*, pp. 160–167, IEEE, Wanzhou, China, November 2018.

[20] K. Maree, C. Aldous, and A. Hattingh, "Predictors of learner performance in mathematics and science according to a large-

Retraction

Retracted: Analysis of the 3D Application Evaluation System of Landscape Based on Hybrid Cooperation of VR and AI

Scientific Programming

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope

(2) Discrepancies in the description of the research reported

(3) Discrepancies between the availability of data and the research described

(4) Inappropriate citations

(5) Incoherent, meaningless and/or irrelevant content included in the article

(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] L. Niu and F. Han, "Analysis of the 3D Application Evaluation System of Landscape Based on Hybrid Cooperation of VR and AI," *Scientific Programming*, vol. 2021, Article ID 4595808, 11 pages, 2021.

Research Article

Analysis of the 3D Application Evaluation System of Landscape Based on Hybrid Cooperation of VR and AI

Lina Niu[1] **and Feng Han** [ID] [2]

[1]*Jinzhong College of Information, Jinzhong 030800, Shanxi, China*
[2]*Taiyuan University of Technology, Taiyuan 030024, Shanxi, China*

Correspondence should be addressed to Feng Han; babynn1014@163.com

The traditional urban landscape planning and design method cannot integrate the design into the whole planning scene and cannot express the planning effect of multiple schemes and large scenes in an efficient, intuitive, and coordinated way. With the development of computer science and technology, virtual reality technology based on AI is playing an important role in many aspects of urban landscape planning and design. In this paper, on the basis of 3D visualization landscape planning and design, 3D visualization urban landscape modeling is carried out using VR combined with the constructed deep neural network (DNN) to establish a 3D database and virtual scene model of the landscape, and 3D representation of the virtual scene landscape is realized. The designed solution allows the user to grab the object and release it into the workspace. If any placement is needed, the user can pick up the layer again and move it to a new location. Finally, the simulation experiment and its result analysis were conducted to effectively analyze the site of the real model and its matching rate, to realize the integration of the 3D image, 3D stereo, and system integration technology, and to finally complete the landscape design task.

1. Introduction

Virtual reality technology emerged at the end of the 20th century and is a comprehensive information technology integrating a series of technologies such as computer technology, multimedia technology, network technology, parallel processing technology, and sensor technology [1, 2]. People create and experience a variety of virtual worlds through virtual reality technology. With the growing maturity of virtual reality technology, this technology has slowly started to be used in various design fields, providing great convenience for designers to realize their own design ideas, improving design quality and design efficiency [3]. On the other hand, today's constant innovations in science and technology, especially in computers, hold promise for the use of computers for large image data processing. Image digital recognition technology is an innovative technology that uses digital image processing technology as its own basis for the recognition of experimental content using a computer [4]. As technology in the field of AI, the main process is divided into three steps, image acquisition, image processing, and image recognition [5]. The features of the landscape in landscape images can be collected using special tools, such as cameras, and then the information collected can be analyzed using computer technology to complete the identification or design of the landscape images [6].

In the actual landscape design, all aspects should be fully considered, not only to realize the effect of the design structure but also to coordinate and match the surrounding environment as well as the landscape design, etc. [7, 8]. Once the final design plan is determined, then extensive modifications will increase the cost. If virtual reality technology is used, it allows users to experience the program more realistically, and once it is determined, extensive modifications can be avoided and costs reduced [9]. Therefore, before the realization of landscape design, the design plan should be improved, and virtual reality technology should be used to promote the scientificity and feasibility of landscape design.

Only by integrating landscape design and virtual reality technology, 3D characteristics can be highlighted;, in addition, the designers concerned to clarify whether the design scheme can use virtual reality technology to achieve the final effect, to ensure the operability of the design content. The use of virtual reality technology to design the landscape can effectively save a lot of labor and greatly reduce costs.

Convolutional neural networks (CNNs) are one of the most popular and innovative techniques in the context of big data, and they contain unique processing methods that have been utilized in many fields and scenarios [10], especially in the area of image processing discussed in this paper. Landscaping technology is the use of computers to collect image features contained in images such as landscape images or photographs and use these features for landscape recognition. Landscaping is categorized under the category of image recognition, where the image to be experimented with is viewed as a high-dimensional random vector, and the resulting data are mapped to a low-dimensional feature space [11].

Landscape architects concerned can expand the scheme according to its own specificity or change the details of the design scheme to make it more reasonable and then improve and update it to form the final design scheme [12]. In addition, although virtual reality technology cannot be fully realized but can be more comprehensive, a more perfect display of the content of the drawings guides users to experience the landscape design content, experience the process of guiding users according to their actual needs and the landscape design content of the unreasonable, and puts forward views and opinions, so that landscape designers can make users and designs achieve good interaction, so that designers can reshape the design works [12].

The use of virtual reality technology combined with deep learning technology to design the landscape can effectively save a lot of manpower, greatly saving costs, while the designer can also adjust or modify the plan according to their own needs, and thus closer to the actual needs of the user for the landscape effect, to achieve good interaction between the user and the design [13]. Such an interactive approach will greatly reduce landscape design time, make the design more responsive to the actual needs of users, improve the design more comprehensively, and make it easier for designers to reshape the design.

In summary, this paper uses virtual reality technology combined with the constructed deep neural network (DNN) to establish a 3D database and virtual scene model of the landscape in the 3D visualized landscape planning and design and carries out 3D visualized urban landscape modeling to achieve a 3D representation of the virtual scene landscape [14, 15]. Finally, the simulation experiment and its result analysis were conducted to effectively analyze the site of the real model and its matching rate, to realize the integration of 3D image, 3D stereo, and system integration technology, and to finally complete the landscape design task.

2. Related Work

Currently, the model design of landscape scenes, through VR or AI, contributes to advance innovation. In a large number of problems, the configuration of landscape scenes is upgraded and solved using computer innovation [5, 16]. Landscape scene measurement recreates, utilizes the view, and visualizes the scene of a rising degree [8]. Rafiq et al. [17] uses intelligent landscaping rather than intuitive perceptual structures to think about design ideas; when faced with similar static picture designs, many believe this would be a range of alternatives. A survey of intuitive 3D landscape scene models has likewise shown to be useful for media representation of different landscape types [18]. Portman et al. [19] examined GIS-based layout dialogues for 3D landscape models and feasible improvement markers using smart VR.

We use deep learning models to drive landscaping research, creating large-scale realistic video explorations and models while evaluating the soundness of a landscaping design [20]. For example, hand-drawn viewpoints are drawn using CNN for watercolor painting, advanced photo acquisition, and discrepancy in confirmation using semantic contrast to cross 3D models [21, 22]. We introduce intelligent visual rendering, based on data-driven deep learning models suggesting an influence on the final choice of creation. Meanwhile, in the field of computer-generated reality, many researchers have conducted related studies [23, 24]. Stimulation with Unity 3D allows us to access the sky, territory, water, clarification of development strategies for trees and flowers, and professional help in the use of VR innovations in nurseries [25]. Building models for landscape design based on GAN-generated natural scenes, vivid visual dramatization, and task scenes provide an assessment of human biological systems [26].

3. Materials and Methods

Virtual reality technology creates a virtual reality environment for the learner. The corresponding learning tasks are designed based on the existing experience and knowledge of the environmental landscape in the VR environment [6]. To accept a task, we select a character to enter the learning environment and use a VR device to complete the task; the proposed AI combined with the VR scheme in this paper is shown in Figure 1. In this environment, each learner is an individual and their organs are completely immersed in the virtual reality environment, isolated from the real world. It allows us, for example, to realize the true meaning of the wind and the natural world. Immersion in virtual reality technology is an experience in a situation that the learner cannot experience in the real world.

3.1. Environmental Image Preprocessing. The virtual reality environment is a minimal distraction and easy [27]. The user controls the gripping, pitching, and positioning of the virtual reality object. VR user interface with various tools was used for in-depth learning of the model around the building. It grabs the layer from the left side of the tool holder and places it in the workspace shown in Figure 2. Users can run it standing or sitting, using applications and operations. It has a calm environment that allows the user to fix points of the
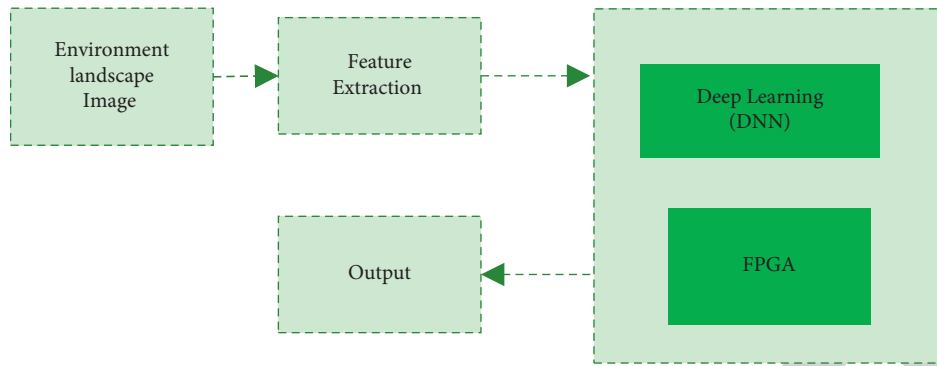
Figure 1: Block diagram of the proposal.
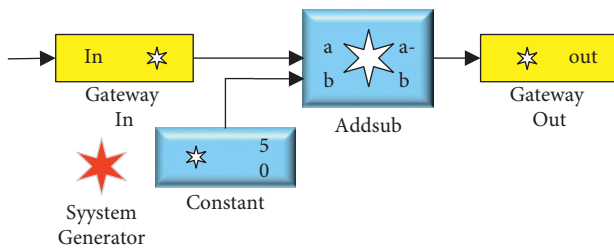
Figure 2: Virtual image processing.

baseline and turn on the virtual reality. VR was used with various tools for user interface for in-depth learning around the building model.

3.2. Feature Extraction. Structural elements are also a prerequisite for building deep models based on virtual reality technology and are central to the integration of virtual reality technology and deep learning concepts. These components include participants, virtual reality interaction platforms, interactive tools, resource sharing, and common purpose [28]. Specifically, participants refer to learners, instructors, and industry experts. Learners complete the core and subject matter of the learning task through learning. It is collected by teachers and industry professionals to assess student learning, provide guidance and support to learners, and actively explore and guide students to deepen their knowledge. Virtual reality interactive platform is a learning environment that uses information technology and virtual reality devices to provide learners with online communication. It is also a profession for sharing learning resources. The interactive tool is a medium of communication between participants. It can provide functions such as online search, transfer, and download resources to build knowledge and help participants. Resource sharing includes resource building and resource sharing, as shown in Figure 3.

The resource structure collects learning videos, documents, images, and other resources through the learning platform to form systematic learning materials for learners. Resource sharing is a learning discussion forum where learners share their knowledge and experience [16, 25].

From deep learning groups for daily tasks and common goals by learning the same topics and objectives, learners learn knowledge based on a specific learning path. Team members fully demonstrate their subjective efforts to achieve the ultimate learning goal.

3.3. Deep Learning Classification. The core aspect of this paper is DNN built as shown in Figure 4. This feature allows the user to grasp objects and release them into the workspace. If any placement is needed, the user can pick up the layer again and move it to a new location. In this feature, it is simple to ensure that the user can interact with the application in a familiar way. The element of familiarity is especially important for new virtual reality users. All new users are unfamiliar with the environment as it is designed to interact to minimize the learning curve. The layer is discarded. The user picks it up and then lifts it in any direction. It applies to the physical object at that point, causing it to drop to the ground, touch the ground, and disappear. In this way, the user can quickly discard irrelevant layers and intuitively change the structure of the model. DNN is shown in Figure 4.

For the characteristics and features of landscape design images, a training sample, validation sample, model layers, and fusion method are proposed to build a complete CNN model. The model is able to extract potential information about the grayscale spatial distribution of target pixels in landscape design images and achieve accurate recognition of targets. The target image data are composed of pixel gray values [29], and the target features are mainly reflected by the spatial distribution information of different pixel gray values, such as the distribution relationship of target pixels and the regional distribution relationship of target and clutter. This feature extraction method acquires the landscape design image containing the complete target and then uses it as the input layer feature image. Therefore, the designed model, which is based on the improvement of VGG16 [30], and different data features are selected separately during the training process, and then the blocks in the model are fused, which improves the accuracy of target recognition. And based on this, the prediction module is added to improve the recognition ability of the model. In these models, the convolutional layers all use the same convolutional kernel parameters; for example, the convolutional kernel size (kernel size) used in the convolutional
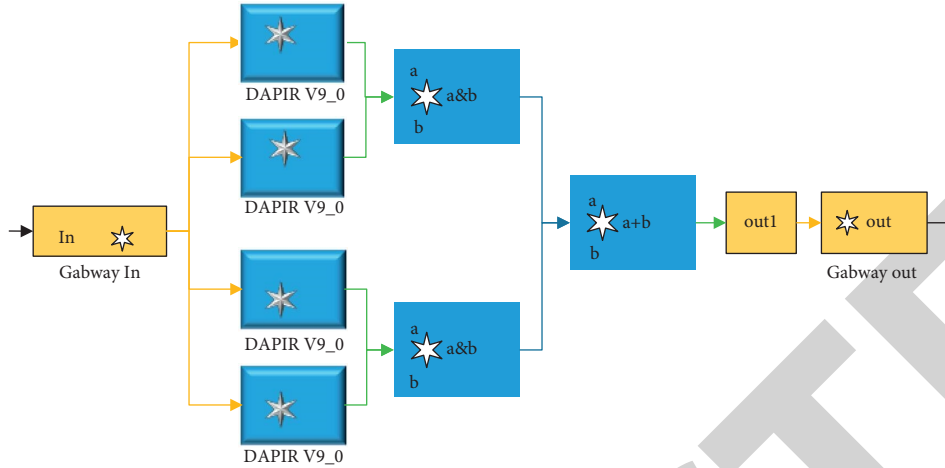
Figure 3: FPGA-based feature extraction.
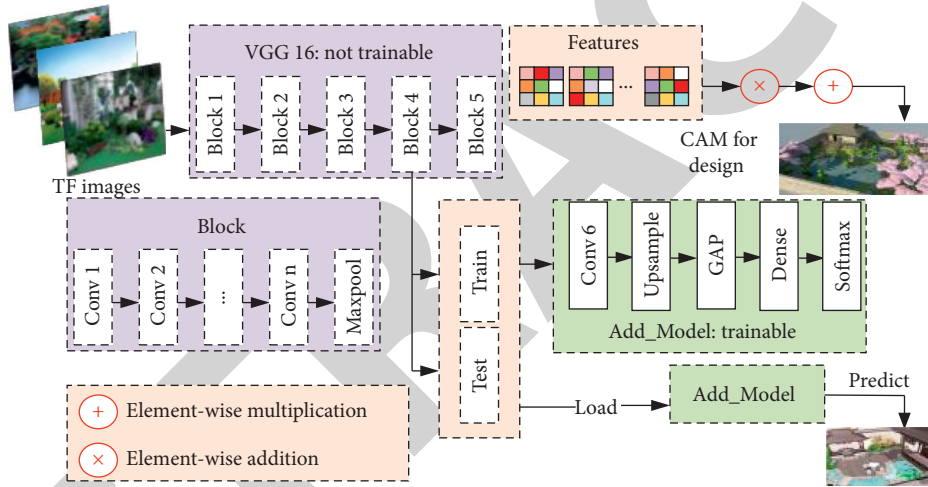
Figure 4: Flowchart of CNN-based landscape design images (TF: time-frequency; CAM: class activation map).

layers is 3; i.e., both width (width) and height (height) are 3, and 3 ∗ 3 is the small convolutional kernel size, which is used together with other parameters (padding = same), so that each convolutional layer (tensor) keeps the same width and height as the previous layer (tensor) with the same width and height as the previous layer (tensor). Using the same pooling kernel parameters, the pooling kernel parameters of the CNN model are chosen to perform target detection based on the distribution of target pixel grayscale values [31]. Pixel points of the image should focus on and highlight the transport target; in other cases, the mean value of the target can be extracted more accurately, so a mean pool is used to train the CNN target detection model. The model consists of multiple convolutional and pooling layers (stack), which makes it easier to form a deep network structure. The advantages of the overall model can be summarized as simple filtering, easy to analyze, and easy to use.

The steps of landscape design algorithm is given as follows:

Step 1: initialize the input image.

Step 2: apply the preprocessed images for analysis, cluster the images to evaluate the values, and partition the images.

$$V_i = \frac{\sum_{K=1}^{N} (U_{ik})^m X_k}{\sum_{K=1}^{N} (U_{ik})^m}, \tag{1}$$

V_i is the input cluster center. U_{ik} is the grayscale value.

Step 3: find the image values and neighborhood values and identify the output data. Train and check the values using deep neural networks.

$$u_{ik} \begin{cases} 1 \\ 0 \end{cases} d_{ik}^{(t)} = \min_{1 \le j \le nc} \left\{ d_{ik}^{(t)} \right\}, \tag{2}$$

u_{ik} is the data partition degree value.

Step 4: check the condition of the convergence value.

$$\| Ut - 1 - Ut \| < \varepsilon. \tag{3}$$

Step 5: analyze the image and classify the environmental landscape values.

Step 6: end.

3.4. FPGA. To provide more accurate results and real-time object recognition, for example, in applications such as robotics and self-driving cars, the size of CNN needs to be increased by adding more neural network layers [13, 31]. However, the changing new NN layers lead to more complex DNN structures and deep DNN models. Thus, millions or billions of operations and parameters as well as large computational resources need to be trained, and the resulting large DNNs need to be evaluated. Such a request represents a General Purpose Processor (GPP) computational task. Therefore, hardware has been used to improve the throughput of gas pedal DNNs, such as graphics processing specialized integrated circuits (ASICs), field-programmable gate arrays (FPGAs), and units (GPUs) [20]. DNNs are trained offline using a back-propagation process. The offline trained cellular neural network is then used for feedforward processing to complete the cognitive task. Therefore, the speed of feedforward processing is crucial. A convenient feature of CNN is that each feature-mapped neuron shares its weight with all other neurons. Performance comparison of the model is shown in Table 1.

The authors' argument is that access to mobile data is not a result of computing the highest energy consumption of off-chip DRAM memory [32]. In other words, the energy cost of memory access and data movement (which adds up to most DNN operations) typically exceeds the energy cost of computation.

Therefore, FPGA circuits for DNN gas pedals require careful consideration of this to achieve efficient architectures in terms of time and power. In this work, the current state of the art in implementing deep learning networks using FPGAs as gas pedals is reviewed. It highlights implementation challenges and design directions used to address these challenges. It also provides future recommendations to maximize the performance of FPGAs as deep learning network gas pedals.

4. Site Analysis Based on the 3D Realistic Model

In this paper, we use SuperMap software [33] to browse and analyze the 3D real-world model.

4.1. 3D Measurement. The model can be measured by using the measurement tools, such as horizontal distance, ground-dependent distance, spatial distance, height, spatial area, and ground-dependent area (Figure 5). The elevation of a point can be measured, and the relative height difference of a complex terrain can be measured directly. By clicking 2 or more points, you can measure the distance between 2 points, and you can also measure the distance or area between multiple points.

4.2. Daylight Analysis. Sunlight is an important influencing factor in landscape garden design, which has an impact on planning the spatial layout of the site, road landscape,

TABLE 1: Performance comparison of the model.

Technique	PCA	SVM	Our model
Sensitivity (%)	71.9	76.6	95.3
Specificity (%)	70.4	77.1	91.3
Accuracy (%)	85.8	91.5	96.7

planting design, and selection of garden materials [34]. Through sunshine analysis, the distribution pattern of the sunshine intensity of the site can be obtained, providing a basis for making full use of the favorable conditions of sunshine and avoiding unfavorable factors. Setting the time in the software, such as the winter solstice or the summer solstice, the sunshine situation at different times of day can be automatically simulated in the software. Figure 6 shows the results of regional local insolation analysis.

4.3. Line of Sight, Field of View Analysis. The main purpose of the visibility analysis is to ensure that there is a clear line of sight between important viewpoints and the landscape and that the appropriate spatial scale is maintained between buildings [35]. There are 2 main ways of analysis: one is to discuss the line of sight between 2 points, whether the target point can be seen from the viewpoint location; the other is the view area analysis, which analyzes all the visible range of the viewpoint. The green part in Figure 7 is the area that can be seen, and the red part is the area that is blocked from view; the result shows that most of the site can see the top of the tower, but only the local site can see the complete tower, and these sites will be important overlook points and viewpoints within the design scope.

4.4. Profile Analysis. Using the profile analysis tool, you can quickly get a profile of the site. By drawing a straight line under the 3D scene, a profile line that is consistent with the surface change of the real model data can be output. If the profile line encounters plants or buildings, the contour line of plants and the contour line of buildings will also be formed in the final profile (Figure 8). This is the most direct way to study geomorphology [36], feature contours, and topographic changes of the site and is more convenient and efficient than the traditional method of using contour lines to draw profiles.

4.5. Program Showcase. Whether the landscape planning scheme can harmoniously coexist with the surrounding environment and whether the vertical design is closely integrated with the current topography of the site are the key concerns of landscape planning and design. The traditional expression is mainly based on effect drawings, which place the planning and design scheme into a fictitious environment, and the effect is often gorgeous, but it differs greatly from the field situation of the site and loses the meaning of practical application [19, 33]. And in a real three-dimensional scene, the designer can better grasp the design direction, so that the design plan is appropriate to the site,

Figure 5: Distance measurement.

Figure 6: Daylight analysis.

Figure 7: View field analyses based on the top of the tower.
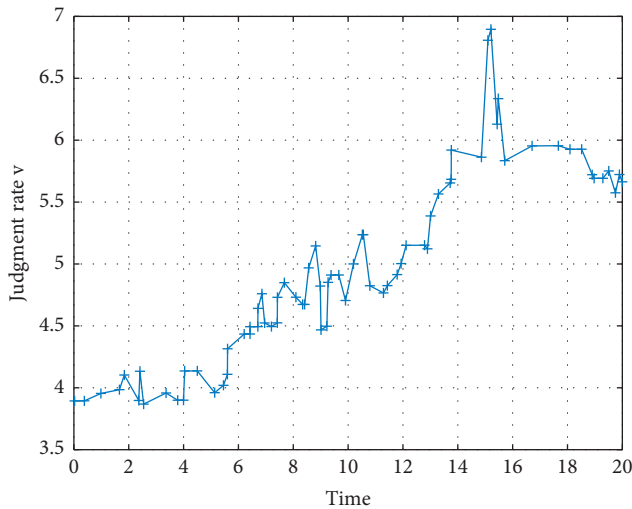
FIGURE 8: Planing surface analysis.

FIGURE 9: Landscape planning scheme overlays.

according to the characteristics of the site design, and in the style is also easier to maintain consistency and thus early detection of defects in the planning scheme. Traditional design using 3dMax model or SketchUp model [37] cannot be directly imported into GlS software [18]; through the AI software corresponding plug-in, the coordinates will be unified, both to achieve the overlay display of the planning scheme and the status qua scheme and to use the roll-up function to compare multiple schemes. Figure 9 shows the effect of overlaying the landscape design scheme with the current situation model.

The Chinese garden designed by the model in this paper is shown in Figure 10. It can be clearly seen that the design focuses on the Chinese "courtyard three entry" layout, creating a triple entry ritual, one garden gate, two courtyard gates, and three house gates. The public and private spaces are reinforced by the gate and court to define the group and shape the sense of homecoming ceremony. The landscape is symmetrically laid out with the landscape nucleus as the axis, "one center, one axis, and three zones" to create a sense of ritual and ceremonial landscape. The design of this project follows the Chinese garden's emphasis on "the origin of the place and the origin of the water" and the rule of "although made by man, it is like the opening of heaven": the water system, regardless of its size, must be twisted and turned. The "gathering is broad, scattering is lingering" of the garden in Jiangnan is interpreted to the fullest. In the design of this case, the traditional and modern elements are combined in the design of the Chinese garden, and this concept is carried through the whole design, from the front garden to the backyard, and the rich cultural temperament of the Chinese garden is exquisite and meticulous. Through the waterscape and then into the courtyard, the layout of the inner garden of "Qingquan Shi Liu" is unpredictable, with rocky hills, pavilions, and willows and pines. It is like a landscape painting made by waving ink. Different from the painting, "garden" is a three-dimensional space and contains flowers, trees, ponds, fish, houses, stacked stones, and other combinations formed by each other. Regarding the density of objects in

different landscape designs, it is necessary to grasp the rationality of the design. For example, a compact place does not appear crowded, and there are some objects in an open place. Such a design will bring a relaxing mood and vision, visual enjoyment.

5. Experiment

In order to prove the validity of the proposed method for judging the rationality of landscape design and 3D image simulation, we need to design experiments for verification. The 3D image modeling analysis platform for the rationality of landscape distribution was constructed in Windows 8 environments. Since the Point Grey Flea 2 camera [20] has a resolution of 640 pi, 60 frames of images provided by the Landscape Institute were selected for the experiment.

Through 3D image for reasonable acquisition identification analysis and calculating different matrix objective function values, if the function value is larger, then the garden landscape judgment analysis is more accurate; if the function value is smaller, it means that the garden landscape design reasonableness judgment is not accurate. Based on this, the design judgment process is shown in Figure 11.

5.1. Image Feature Points' Matching Number and Matching Rate Results and Analysis. The reasonableness of landscape design is compared using the 3D image simulation judgment method and the traditional deep estimation judgment method, respectively. Whether the number of image feature points extracted can judge the accuracy of 3D image simulation of landscape, the key factor is how to accurately match the corresponding attribute feature points of a landscape garden image [18]. Also, these two judgment methods are used to compare the number of feature point matching and matching rate of landscape garden, and the results are shown in Table 2.

As can be seen from Table 2, the number of feature point matches and the matching rate of the landscape obtained using the traditional deep estimation judgment method are lower than those of the 3D image simulation method used in this paper, and the number of image feature points is simpler in the steps performed in the matching process. The 3D

Figure 10: Chinese courtyard landscape design.
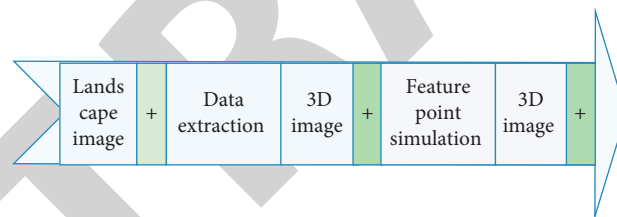
Figure 11: Judgment task flow.

Table 2: Matching number of 3D image feature points and matching rates.

Group	Traditional methods		Our method	
	Number of matches (pc)	Matching rate (%)	Number of matches (pc)	Matching rate (%)
A	540	30.25	540	15.28
B	350	25.61	350	13.61
C	360	24.32	360	20.12
D	320	27.02	320	14.32
E	280	21.34	280	20.15
F	256	30.16	256	9.38

image simulation method has a high feature point matching rate, which fully justifies the use of this judgment method.

Using the comparison results in Table 2 as the basis, the two judgment methods were used to compare the effect of the 3D image simulation of the garden landscape, and the results are shown in Figure 12.

From Figure 12, it can be seen that the distribution of the landscape obtained by using the 3D image simulation judgment method is more reasonable, the main reason is that, in the judgment process, the feature points are matched

with high precision, and the feature points are assigned to clusters, and the simulation points are analyzed according to the assigned image features, and the judgment of the rationality of the landscape design is completed based on the analysis results.

5.2. Judgment Rate and Accuracy Comparison Results and Analysis. The data were collected in the landscape environment, and the rate analyses of the conventional method

(a)

(b)

FIGURE 12: Comparison of the effect of two judgment methods on the simulation of the garden landscape. (a) 3D image simulation judgment method. (b) Traditional judgment method.
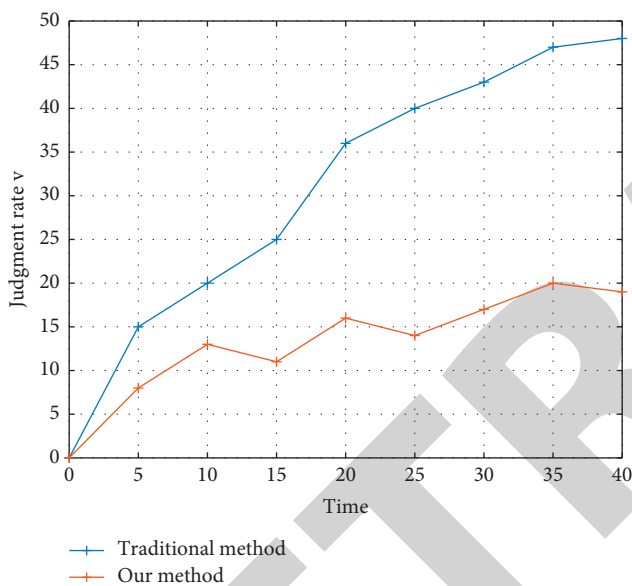
FIGURE 13: Two methods to determine the rate comparison results.

FIGURE 14: Results of accuracy.

and the method of this paper were judged, as shown in Figure 13.

As can be seen from Figure 13, when the time is in the range of 0–10, the judgment rate of the traditional method gradually increases, but it is always smaller than that of the 3D simulation method; when the time is in the range of 10–15, the judgment rate of the traditional method starts to decrease. In the subsequent time, the traditional method has been showing an up-and-down trend, and the fastest rate is only about 1/2 of the fastest rate of the 3D image simulation. From this, it can be seen that the speed is slow and low when using the traditional method to judge the reasonableness of the landscape, while the rate is faster with the 3D image simulation judgment method. The comparison results of judgment accuracy are shown in Figure 13. As can be seen from Figure 13, when the judgment range is within 20%, the accuracy of the traditional method and the method of this paper is not much different, and the accuracy of the traditional method is higher than the method of this paper; when the judgment range is 20% to 45%, the accuracy of the traditional method is significantly higher than the method of this paper. However, with the gradual expansion of the range, the accuracy of the three-dimensional image simulation judgment method used in this paper gradually improves, while the accuracy of the traditional method is always maintained at about 40%. Thus, it can be seen that the 3D image simulation judgment method is more accurate than the traditional method of judging the reasonableness of landscape design.

5.3. Experimental Conclusions. As observed from Figure 14, the number of matching feature points and the matching rate of the landscape obtained by the traditional deep estimation judgment method are lower than the 3D image simulation method used in this paper, the speed is slower, and the rate is lower when judging the reasonableness of the landscape, while using the 3D image simulation method has a higher matching rate of feature points and a faster judgment rate, which fully proves the reasonableness of using this judgment method.

6. Conclusions

With the development of computer science and technology, VR based on AI is playing an important role in many aspects of urban landscape planning and design. In this paper, in 3D visualization of landscape planning and design, 3D visualization of urban landscape modeling is carried out using virtual reality technology combined with the constructed DNN to establish a 3D database and virtual scene model of the landscape and achieve a 3D representation of the virtual scene landscape. The simulation experiments show that the effective site analysis of the real scene model and its matching rate is improved to realize the integration of 3D image, as well as system integration technologies.

Data Availability

The dataset used in this paper are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest regarding this work.

Acknowledgments

References

[1] W. Chen, L. Zhao, Q. Kang, and F. Di, "Systematizing heterogeneous expert knowledge, scenarios and goals via a goal-reasoning artificial intelligence agent for democratic urban land use planning," *Cities*, vol. 101, Article ID 102703, 2020.

[2] Z. Zhang and B. Bowes, "The future of artificial intelligence (AI) and machine learning (ml) in landscape design: a case study in coastal Virginia, USA," *Journal of Digital Landscape Architecture*, vol. 4, pp. 2–9, 2019.

[3] P. Shan and W. Sun, "Research on 3D urban landscape design and evaluation based on geographic information system," *Environmental Earth Sciences*, vol. 80, no. 17, pp. 1–15, 2021.

[4] B. Cantrell, Z. Zhang, and X. Liu, "Artificial intelligence and machine learning in landscape architecture," *The Routledge Companion to Artificial Intelligence in Architecture*, Routledge, London, UK, 2021.

[5] R. Li, Q. Liu, J. Gui, D. Gu, and H. Hu, "Indoor relocalization in challenging environments with dual-stream convolutional neural networks," *IEEE Transactions on Automation Science and Engineering*, vol. 15, pp. 651–662, 2017.

[6] Z. Zhao, D. A. Klindt, A. Maia Chagas et al., "The temporal structure of the inner retina at a single glance," *Scientific Reports*, vol. 10, no. 1, p. 4399, 2020.

[7] C. Shousong, G. Xiaomin, W. Xiaoguang, and C. Ying, "Notice of violation of IEEE publication principles: research on urban land price assessment based on artificial neuralnetwork model," *IEEE Access*, vol. 7, pp. 180738–180748, 2019.

[8] K. Raaphorst, I. Duchhart, W. Van Der Knaap, G. Roeleveld, and A. Van Den Brink, "The semiotics of landscape design communication: towards a critical visual research approach in landscape architecture," *Landscape Research*, vol. 42, no. 1, pp. 120–133, 2017.

[9] T. Yigitcanlar, K. C. Desouza, L. Butler, and F. Roozkhosh, "Contributions and risks of artificial intelligence (AI) in building smarter cities: insights from a systematic review of the literature," *Energies*, vol. 13, no. 6, p. 1473, 2020.

[10] R. Righi, S. Samoili, M. López Cobo, M. Vázquez-Prada Baillet, M. Cardona, and G. De Prato, "The AI techno-economic complex system: w," *Telecommunications Policy*, vol. 44, no. 6, Article ID 101943, 2020.

[11] D. Li, H. Li, W. Li, J. G. Jun, and E. Z. Li, "Application of flipped classroom based on the rain classroom in the teaching of computer-aided landscape design," *Computer Applications in Engineering Education*, vol. 28, no. 2, pp. 357–366, 2020.

[12] F. Xu and Y. Wang, "Color effect of low-cost plant landscape design under computer-aided collaborative design system," *Computer-Aided Design and Applications*, vol. 19, no. S3, pp. 23–32, 2021.

[13] M. Zhang and X. Deng, "Color effect of landscape architecture design under computer aided collaborative design system," *Computer-Aided Design and Applications*, vol. 19, no. S3, pp. 13–22, 2021.

[14] X. Ma, P. Guo, J. Zhu, and J. Zhao, "An optimization method for urban landscape design and artistic value analysis by using vr technology," *Boletin Tecnico/Technical Bulletin*, vol. 55, no. 11, pp. 533–539, 2017.

[15] X.-B. Fu, S.-L. Yue, and D.-Y. Pan, "Camera-based basketball scoring detection using convolutional neural network," *International Journal of Automation and Computing*, vol. 18, no. 2, pp. 266–276, 2021.

[16] R. A. Minhas, A. Javed, A. Irtaza, M. T. Mahmood, and Y. B. Joo, "Shot classification of field sports videos using alexnet convolutional neural network," *Applied Sciences*, vol. 9, no. 3, p. 483, 2019.

[17] M. Rafiq, G. Rafiq, R. Agyeman, G. S. Choi, and S.-I. Jin, "Scene classification for sports video summarization using transfer learning," *Sensors*, vol. 20, no. 6, p. 1702, 2020.

[18] Z. Lei, S. Shimizu, N. Ota, Y. Ito, and Y. Zhang, "Construction of urban design support system using cloud computing type virtual reality and case study," *International Review for Spatial Planning and Sustainable Development*, vol. 5, no. 1, pp. 15–28, 2017.

[19] M. E. Portman, A. Natapov, and D. Fisher-Gewirtzman, "To go where no man has gone before: virtual reality in architecture, landscape architecture and environmental planning," *Computers, Environment and Urban Systems*, vol. 54, pp. 376–384, 2015.

[20] A. Haldorai, A. Ramu, and S. Murugan, "artificial intelligence and machine learning for future urban development," *Computing And Communication Systems in Urban Development*, Springer, Berlin, Germany, 2019.

[21] E. G. Xydas and L. S. Louca, "Identification of kinematic traits in 3d reaching tasks with the use of a haptic nine-hole pegboard test: comparison between healthy people and people with multiple sclerosis," in *Proceedings of the IEEE Ras & Embs International Conference on Biomedical Robotics & Biomechatronics*, IEEE, Scottsdale, AZ, USA, October 2008.

[22] W. Liu, C. C. Yan, J. Liu, and H. Ma, "Deep learning based basketball video analysis for intelligent arena application," *Multimedia Tools and Applications*, vol. 76, no. 23, pp. 24983–25001, 2017.

Research Article

Design of the Poster Image System Based on Human Vision

Xiaolifei Sun [ORCID]

Academy of Arts and Crafts, Inner Mongolia Normal University, Hohhot, Inner Mongolia Autonomous Region 01000, China

Correspondence should be addressed to Xiaolifei Sun; 20150072@stu.nun.edu.cn

At present, the human visual perception system is the most effective, accurate, and fast image processing system in the world. This is because human eyes have some special visual features, but the features closely related to image enhancement include color constancy and brightness constancy. This paper presents a new image enhancement framework and computational model which can better simulate human visual features. It is based on the analysis of color constancy and luminance constancy and Retinex theory. And, this is a new image enhancement method in the compressed domain based on Retinex theory. In Retinex theory, DCT coefficients consist of incident components (DC coefficients) and reflection components (AC coefficients). By adjusting the dynamic range of DC coefficients, carefully adjusting AC coefficients, and using the threshold method for block suppression, the compressed domain image can be enhanced. On the basis of Retinex theory, the incident light and reflected light components are considered synthetically, the dynamic range (DC coefficient) of the incident light component and the details of the reflected light component (AC coefficient) are adjusted, and then the incident light component is reexamined. Moreover, it achieves a better image enhancement effect and avoids the blocking effect.

1. Introduction

With the development of modern society, posters have gradually become an important medium of cultural communication. Poster design conveys its message through its unique expression. Posters have a natural and moving charm. The development of poster design is not stagnant but keeps pace with the times, keeping a high degree of concern and forward-looking thinking on society and things. Nowadays, the poster design industry has embarked on a broader road of exploration under the guidance of new art aesthetics and has experimental characteristics. In the future, the design forms of image media in poster design will be richer and more diverse, and the development trend will be better and better. This may be a design experience with audience participation, or it may be "out of nothing" that advocates simplicity first. With the progress of the times, the carrier of posters is constantly changing, and the new media is constantly developing. Poster design has also moved from plane to space full of greater development possibilities, which can better communicate and interact with the audience.

The study in [1] defined a human visual system model which describes the monochrome features of each spatial location image as encoding the luminance and chrominance of the corresponding three colors of each spatial location image. There are many requirements for human perception [2], but the functions of fast analysis and selective coding of information in scenes in image compression are very suitable for human perception requirements. The basic information types of image processing are human brain perception, processing, and interpretation [3]. The study in [4] improved by building edge-enhanced images and original sketches. Spatial filters used in the human visual system seem to be a low flux of color stimulation and band pass of photometric stimulation [5]. Image processing technology focuses on two main tasks [6], such as improving image information for human interpretation and processing of image data and for storage, transmission, and representation of automatic machine perception. Disjointed information [7] is used as an indicator of image comparison and applied to perceptual image quality assessment, image registration, and video tracking. Recent developments in digital technologies utilizing large-scale image and video data such as

the Internet, multimedia systems, digital television, digital cinemas, and mobile/wireless devices have increased the demand for efficient processing [8], storage, and transmission of visual information. Image processing apparatus [9] can quickly detect objects. In the image processing apparatus of [10], an input unit is used for inputting image information. The purpose of the image relocation algorithm of [11] is to adjust the image to a display screen that is different from the aspect ratio or size of the image by uneven and perceptible content image size adjustment. Algorithms are capable of computing and clustering types of photos and videos for precise and nearly repetitive media items [12]. The study in [13] relied on more and more high-resolution multispectral digital art images (mainly painting and painting) and complex computer vision methods. This paper proposes a new research method [14], which takes the physical illumination model as the core, that is, dynamically calculating HDR illumination in the virtual environment. Our method allows the reuse of existing virtual environments as input and computes HDR images in photometric units. The study in [15] had some common difficulties in many visual learning tasks. One of them is the lack of supervision information because labels can be boring, expensive, or even impossible. Another difficulty is the high dimension of visual data. However, these difficulties can be alleviated by mixing labeled and unlabeled training data.

Based on the analysis of color constancy, brightness constancy, and Retinex theory, an enhancement framework and computational model that can better simulate human visual characteristics are proposed in this paper, and it is applied to poster processing to verify its effectiveness. The threshold method is used to suppress the block effect, thus realizing image enhancement in the compressed domain. Simultaneous processing of incident and reflection components can make full use of the low-frequency and high-frequency information of the original image to enhance the details, preserve the color, and suppress the block effect.

2. Theoretical Framework

2.1. The Primary Function of Posters.

In the basic theory of social symbol theory, symbol theory holds that all symbol systems are social, which is reflected in the social purpose of symbol systems. Therefore, posters also have their social purposes. Information dissemination is one of the most obvious social functions of posters, especially in the era of underdeveloped information technology. However, great changes have taken place after the industrial revolution. As we all know, the industrial revolution has greatly improved the level of productivity development, and for the first time in human history, products exceeded demand. In this context, various disciplines are discussing the methods of influencing consumers' decision-making, which directly lead to the development of new disciplines including marketing and advertising. Therefore, the social function of posters has also changed greatly, from the initial information dissemination to the persuasion of the audience. It can be said that persuading the audience to accept the information conveyed by posters is the primary role of posters in today's

society. Based on the above understanding of the main social functions of posters, it is the key to explain how posters convey information and persuade audiences through the image system.

2.2. Human Visual Properties.

The human visual system is a complex concept of biology and related disciplines. People initially studied and utilized the visual characteristics of HVS to manufacture products or put forward standards. For example, in the process of transmitting TV signals, the color signal is compressed to a smaller bandwidth, which reduces the overall signal transmission bandwidth, so that color TV can be popularized under the condition of limited bandwidth at that time. Another example of applying human visual features is image compression, such as the JPEG compression standard, which takes advantage of the insensitivity of human eyes to the details of high-frequency signals. In the compression process, most high-frequency coefficients are quantized to zero, so reducing the required storage space will only bring imperceptible image quality loss. The human visual system is a comprehensive three-dimensional world perception system. The human eye receives reflected light from surrounding objects through retinal cells. HVS processing starts with the eyes and is mostly processed by the brain. At present, how the brain processes this information is still unknown. The research on HVS mainly focuses on the visual characteristics of HVS, which are called human visual characteristics. Through computer science, they try to simulate these characteristics and functions of human eyes and apply human visual characteristics to information processing fields such as image enhancement.

3. Algorithm

3.1. Human Eye Structure and Visual Processing Process.

The eye is the starting position of HVS information processing. Light passes through the lens and falls on the retina behind the eyeball. The eye structure diagram is shown in Figure 1. Six muscles located on the outside of the eye are connected with the eye so that the eye can focus on any target and adjust the shape of the lens. The shape of the lens determines the focal length. By adjusting the shape of the lens, the scene captured by the eye can be focused behind the retina. The discoid vascular membrane composed of two annular muscles is called an iris, and the gap in the center of the iris is called a pupil, which is located in front of lens. The diameter of the pupil determines the amount of light entering the eye. One of the two annular muscles is used to enlarge the diameter of the pupil, and the other is used to reduce the diameter of the pupil.

The retina is essentially composed of a group of receptors that can sense the light entering the eye. The structure is divided into three main layers: the receptors are located at the bottom of the retina, the central cells of the retina are located at the top of the retina, and the polar cells are located in the middle of these two layers. The light entering the eye passes through the upper two layers before reaching the
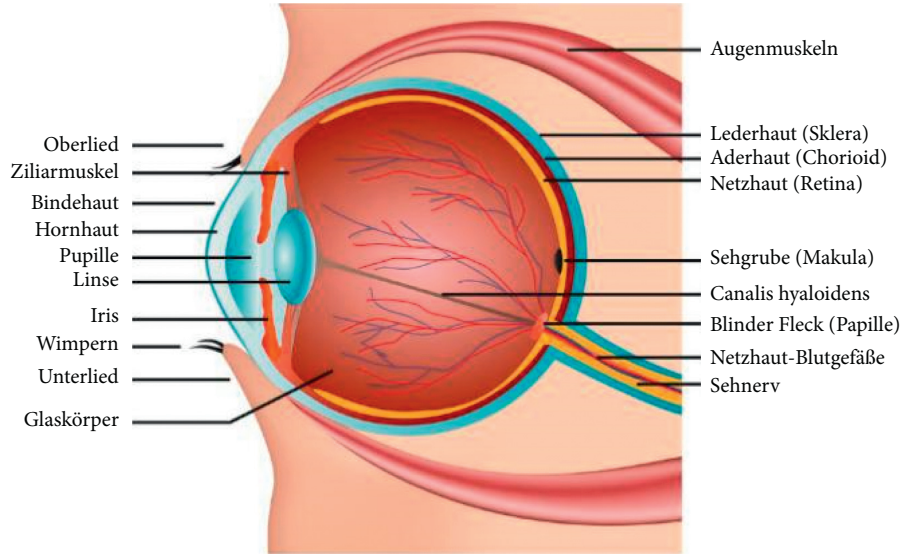
FIGURE 1: Eyeball structure diagram.

receptor located behind the retina. The light-absorbing material is attached to the lower part of the receptor layer, which can effectively prevent the scattering of light received by the retina due to astigmatism. Rod cells and cone cells are two very important photoreceptor cells in the eyes. Cone cells are used to distinguish colors under strong light conditions. They are the organs of bright vision. Rod cells are dark vision organs, which work under the condition of less light.

Rod cells are generally about 25 times more sensitive than cone cells because of their larger diameter and longer length. Generally, cones can be divided into three categories. The first category responds to the red part of the visible spectrum, the second category mainly responds to the green part, and the third category mainly responds to the blue part. This constitutes the physiological basis of the tricolor theory of color vision. The tricolor theory explains why the three primary colors must be red, green, and blue, while other colors can be composed of these three colors.

The receptor is the main organ for measuring the amount of light entering the retina. The amount of light measured by a single receptor (expressed in energy) can be calculated by integrating all wavelengths λ. It is known that three types of cones are sensitive to the red, green, and blue parts of the spectrum. Record $S_i(\lambda)$ as the response curve of these three receptors, where $i \in \{r, g, b\}$. $E(\lambda)$ is used to denote the light that falls on a receptor and has a given wavelength λ, so that the light energy Q_i measured by a single receptor can be obtained by the following formula:

$$Q_i = \int S_i(\lambda) E(\lambda) d\lambda. \tag{1}$$

When observing an object with an unsmooth surface, the light ray E perceived by the receiver can be considered as the common product of the reflected light R of the object (light reflected from the object surface into the eye) and the incident light L irradiated on the object surface; that is to say,

the light ray E perceived by the human eye is composed of the reflected light E and the incident light L, as shown by the following formula:

$$E(\lambda) = R(\lambda) L(\lambda). \tag{2}$$

Then, the complete calculation formula of energy Q_i is as follows:

$$Q_i = \int S_i(\lambda) R(\lambda) L(\lambda) d\lambda. \tag{3}$$

The receptors transmit the measured energy information to the polar cells located in the second layer, and then the polar cells use the central retinal cells to form some chromosome binding connections to stimulate nerve endings to transmit the information to the cerebral cortex. Of the 126 million receptors in the retina, rod cells account for the vast majority of about 120 million, and the remaining 6 million are cone cells. Most pyramidal cells are located in a very small area called the fovea. This area is located in the center of the retina. The resolution of the center of the retina is much higher than that of its edge. Although there are about 126 million receptors in the retina, only about 1 million central retinal cells carry this information to the cortex through nerve endings, so this information needs to be compressed in some way to pass through the axons of the central cells. Processing begins with the retina. A human visual system called the luminance visual model has been developed to describe the process of light entering the eye and reaching the cerebral cortex through the retina, as shown in Figure 2.

In order to establish the model quickly, this paper simplifies the visual model of human eyes into a linear system. The light reflection in Figure 2 refers to the light response system of human eyes. Lamant, Rohler, and others measured the spatial frequency characteristics of the human eye system, which are also called low-pass characteristics. Spectral response refers to the spectral response system of
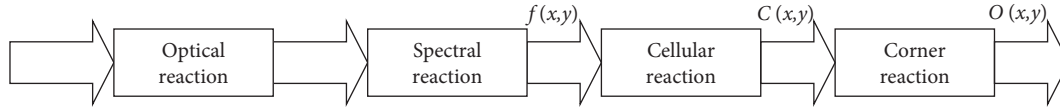
FIGURE 2: Human visual model diagram.

human eyes. Internal imaging $f(x, y)$ is formed by light passing through two reaction systems: optical reaction and spectral reaction. Imaging signal $f(x, y)$ acts on receptors in the retina, including rod cells and cone cells. Corner compression refers to the lateral inhibition effect of the visual neural network, because the number of receptors is much larger than the number of retinal central cells, so this information must be compressed before it can continue to be transmitted. The lateral inhibition effect shows that the visual signal perceived by human eyes is formed by the weighted sum of the outputs of other adjacent receptors, not just a single receptor. The internal imaging signal $f(x, y)$ produces signal $c(x, y)$ through cell reaction, and then the signal $o(x, y)$ produced by corner compression is the final visual subjective luminance signal, which is processed by the cerebral cortex and finally causes visual perception.

3.2. Human Visual Characteristics and Poster Image Enhancement

3.2.1. Dynamic Range of Vision.
The dynamic range of vision is the range of light intensity that human eyes can perceive. The whole dynamic range of human eyes ranges from 10 to $2\,\mathrm{cm/m^2}$ to $106\,\mathrm{cd/m^2}$, which can be said to be relatively wide. Such a wide dynamic range is beyond the capabilities of machinery, displays, and even photosensors. Take the gray image that can be displayed by the display as an example. Generally, the dynamic range of an 8-bit gray image is only 256 gray levels from 0 to 255. The smaller the gap in each level, the smoother, more delicate, and natural the brightness change, which indicates the wider the dynamic range.

3.2.2. Spatial Resolution.
Spatial resolution is the ability of human eyes to observe and distinguish details in scenes. The inherent properties of human eyes determine that the resolution of human eyes depends on the physiological structure of eyes.

3.2.3. Image Contrast.
Contrast refers to the measurement of different brightness levels between the brightest white and the darkest black in an image, which is also called the gray contrast of an image. The larger the difference range of image gray contrast, the greater the contrast. The smaller the difference range of image gray contrast, the smaller the contrast. When the contrast reaches 120 : 1, it is good, and vivid and rich colors can be easily displayed at this time. When the contrast reaches 300 : 1, it can support all levels of colors.

3.2.4. Relative Visual Acuity Curve.
The human eye has different sensitivity to different wavelengths of light. Light with different wavelengths but the same radiation power can be seen in different brightness and colors in human eyes. Therefore, the general object's perception of the brightness of the spatial and wavelength distribution $I(x, y, \lambda)$ of optical radiation is

$$f(x, y) = \int_0^{\infty} I(x, y, \lambda) V(\lambda) d\lambda \,, \tag{4}$$

where $V(\lambda)$ is called the relative luminous efficiency function of the visual system. For human eyes, $V(\lambda)$ is a bell-shaped curve, and the relative visual acuity curves of rod cells and cone cells are different.

3.2.5. Contrast Sensitivity.
Under the uniform background with brightness I, there is an area with brightness ΔI higher than the background brightness; that is, the brightness of this area is $I + \Delta I$, and the brightness difference I that human eyes can distinguish is a function of I. With the increase of background brightness I, the two-degree difference ΔI must also increase, so that human eyes can distinguish the area with brightness $I + \Delta I$. Experiments show that brightness I can change in a considerable range, $\Delta I/I$ can still keep a fixed value, and the ratio is constant, which is called the Weber ratio. Weberby also believed that the visual characteristics of human eyes have a certain degree of logarithmic characteristics because if the brightness is logarithmic and then differentiated, the following formula can be obtained:

$$d[\ln(I)] = \frac{dI}{I}. \tag{5}$$

3.3. Retinex Theory and Algorithm.
Land believed that color constancy is the result of the joint action of retina and cerebral cortex. So, he put forward the Retinex theory, that is, retina + cortex. Land proposed the application of the Retinex theory in the field of digital poster images, where long-wave, medium-wave, and short-wave wavelengths correspond to three color channels of digital poster RGB images, respectively. The image to be enhanced in R, G, and B channels is considered to be composed of two parts: the reflected light component and the incident light component. Estimating incident light components by comparing pixel values between pixels is called illuminance estimation. The illuminance component is removed from the image, and the reflected light component is obtained to restore the original appearance of the object, thus realizing the enhancement of the original image. The specific process is as follows (the three channels of color image are the same, taking single-channel processing as an example).

Think of the image as the product of the incident light component and the reflected light component:

$$I(x, y) = L(x, y) \cdot R(x, y), \quad (6)$$

where $I(x, y)$ is the input image, $L(x, y)$ is the reflected light component, and $R(x, y)$ is the reflected light component. Then, the incident light component of the scene is estimated by various methods to obtain the illuminance estimation $\widetilde{L}(x, y)$, and the reflection estimation value $\widetilde{R}(x, y)$ is obtained by the following formula:

$$\widetilde{R}(x, y) = \frac{I(x, y) + \theta}{\widetilde{L}(x, y) + \theta}, \quad (7)$$

where θ is a positive number infinitely close to 0 to ensure that the divisor is not 0.

How to estimate the incident light component of the scene to get the illuminance estimation $\widetilde{L}(x, y)$ is the core step of the Retinex algorithm. In the process of comparing the pixel values between pixels, path selection, that is, how to select the compared pixels, is the most important link, and it is also one of the common bases for distinguishing different Retinex algorithms. Path selection determines which pixels in the image to compare and how to compare, which directly affects the accuracy of estimating the incident light component of the scene and further determines the enhancement effect.

Provenzi's RSR method is an improvement of the Random Path Retinex method proposed by Land, which changes the path selection from a one-dimensional path to a two-dimensional region and can better reflect the influence of surrounding pixels on target pixels. MSR and MSRCR algorithms are proposed by Jobson. They were first applied to space pictures taken by NASA and achieved a good enhancement effect, which attracted wide attention. The implementation process of MSR is as follows (take single-channel processing as an example).

Step 1. Divide the image into two light quantities: incident light component and reflected light component, as shown in formula (6).

Step 2. Take logarithms on both sides of formula (6) and separate the incident and reflected components as follows:

$$log I(x, y) = log L(x, y) + log R(x, y). \quad (8)$$

Step 3. Filter the original image using Gaussian smoothing filter as follows:

$$\widetilde{L}(x, y) = I(x, y) * F(x, y), \quad (9)$$

where the symbol $*$ is a convolution operation, a low-pass filtered image $\widetilde{L}(x, y)$ obtained after smoothing processing;

that is, an illuminance estimate of an incident component. $F(x, y)$ is a Gaussian filter, and the expression is as follows:

$$F(x, y) = \frac{1}{\sqrt{2\pi}\sigma} exp\left(-\frac{(x^2 + y^2)}{2\sigma^2}\right). \quad (10)$$

Step 4. Subtract the original image from $\widetilde{L}(x, y)$ in the logarithmic field as follows:

$$R'(x, y) = \log I(x, y) - \log\widetilde{L}(x, y), \quad (11)$$

where $R'(x, y)$ is the logarithmic domain reflected light component.

Step 5. MSR algorithm selects several standard deviations of Gaussian functions for linear weighted average:

$$R'(i, j) = \sum_{i=1}^{k} w_i\{\log I(i, j) - \log[I(i, j) * F_i(i, j)]\}, \quad (12)$$

where k is the number of different standard deviations σ and generally three are taken; that is, $w_i, i = 1, 2, 3$ are weights, σ_i is three different standard deviations, and $F_i(i, j)$ is a Gaussian function with parameter σ_i.

Step 6. Take an opposition number to $R'(x, y)$ to obtain an enhanced image, namely,

$$\widetilde{R}(x, y) = exp(R'(x, y)). \quad (13)$$

In equation (10), the standard deviation σ parameter of Gaussian function is called scale parameter, and the enhancement effect σ is determined by its size. The stronger the detail enhancement ability, the smaller the value of σ, and the worse the color fidelity ability and vice versa. MSR algorithm selects three scales from small to large for weighted fusion to achieve a better balance effect. However, the MSR algorithm does not consider the local information of the image and uses three fixed scale parameters to fuse the whole image, so it cannot enhance the local information of the image pertinently.

3.4. Image Compression in DCT Domain. Image compression in the DCT domain is one of the main contents of the JPEG image coding standard. The basic idea of JPEG image coding can be easily extended to other image or video compression schemes based on DCT (such as MPEG and H.26X). Firstly, the image is divided into nonoverlapping subblocks of 8×8 pixels, and then each subblock is transformed in spatial frequency domain by DCT transform. The DCT transform of the 8×8 subblock $I(i, j)$ is defined as

$$d(u, v) = \frac{1}{4}c(u)c(v)\sum_{i=0}^{7}\sum_{j=0}^{7} I(i, j) \cos\frac{(2i + 1)u\pi}{16}\cos\frac{(2j + 1)v\pi}{16} i, j = 0, 1, \cdots\cdots, 7. \quad (14)$$

Among them,

$$c(\eta) = \begin{cases} \dfrac{1}{\sqrt{2}}, \eta = 0 \\ \\ 1, others \end{cases} \tag{15}$$

After quantizing the DCT coefficients, each coefficient is divided by the corresponding quantization parameter and the result is rounded to the nearest integer. The initial value of the quantization parameter depends on the compression standard. The quantized DCT coefficients are stored in a row or column matrix in a zigzag order for average information encoding. Coefficients will be compressed to zero after quantization, leaving only a few nonzero coefficients in subblocks. When stored in zigzag order, these nonzero coefficients will be in front of the queue. To reconstruct the original image from the decompression process, the compressed image can be decoded first and multiplied by the corresponding quantization coefficients; then the quantification is inversed, and finally the image is transformed by DCT. The inverse DCT transform IDCT is defined as

$$I(i,j) = \frac{1}{4} c(u)c(v) \sum_{i=0}^{7} \sum_{j=0}^{7} d(u,v) \cos\frac{(2i+1)u\pi}{16} \cos\frac{(2j+1)v\pi}{16} u, v = 0, 1, \cdots\cdots, 7. \tag{16}$$

3.5. YC_bC_r Color Space.

YC_bC_r color space is widely used in digital poster images in which luminance information is represented by a single component Y and color information is stored by two chromatic aberration components Cb and Cr. The difference between the blue component and the reference value is component Cb and component Cr is the difference between the red component and a reference value. YCbCr, HSI (hue, saturation, brightness) color space is more in line with the human description and interpretation of color. At the same time, to simplify the processing and reduce the amount of computation, this algorithm transforms the image to YC_bC_r color space for processing. Assuming that each color component of YC_bC_r space and RGB space images is stored in 8 bits, the relationship between the two spaces is as follows:

$$\begin{cases} Y = 0.503G + 0.097B + 0.254R \\ Cb = -0.291G + 0.437B - 0.14R + 128 \\ Cr = -0.367G - 0.072G + 0.441R + 128 \end{cases} \tag{17}$$

Although the relationship between YC_bC_r space and RGB space is not linear, from the point of view of algorithm implementation, we can define a YC_bC_r space, where $Cb' = Cb - 128, Cr' = Cr - 128$ has a linear relationship with RGB space. After conversion to the Y, C_b, C_r space, the contrast of the image can be enhanced only by processing the luminance component Y of the image and the C_b and C_r component operations are also required to preserve the color information.

It is assumed that, in the YC_bC_r color space, the DCT coefficient of the luminance component Y of an image I of length and width $N \times N$ is $Y = \{Y(i,j) | 0 \le i, j \le N - 1\}$, k_{dc} is an adjustment scaling factor for quantizing DC coefficients, and k_{ac} is an adjustment scaling factor for quantizing AC coefficients. The DCT coefficients of the adjusted component $\widetilde{Y}$ are given by

$$\widetilde{Y}(i,j) = \begin{cases} k_{dc} \cdot Y(i,j), i, j = 0 \\ k_{ac} \cdot Y(i,j), \text{others} \end{cases} \tag{18}$$

The contrast of the enhanced image is k_{ac}/k_{dc} times that of the original image. However, if only the parameters of the Y component are carried out, the color of the original image cannot be preserved. In RGB color space, the three components R, G, B of each pixel in the original image have the same effect on the color, and image enhancement algorithms based on RGB color space often process these three channels in the same way. This algorithm adopts a similar strategy: while adjusting the parameters of the Y component, the parameters of the Cb' and Cr' components are simpler than those of the Y component, and DC coefficients and AC coefficients are treated separately in the process. Assuming that the DCT coefficients of the Cb' and Cr' components of the image are D, respectively, and the adjusted scaling factors of the Y component are k, the adjusted DCT coefficients $\widetilde{U}$ and $\widetilde{V}$ of the Cb' and Cr' components are

$$\widetilde{U}(i,j) = \begin{cases} N\left(k\left(\dfrac{U(i,j)}{N} - 128\right) + 128\right), i, j = 0 \\ \\ k \cdot U(i,j), others \end{cases},$$

$$\tag{19}$$

$$\widetilde{V}(i,j) = \begin{cases} N\left(k\left(\dfrac{V(i,j)}{N} - 128\right) + 128\right), i, j = 0 \\ \\ k \cdot V(i,j), others \end{cases}$$

The function of U and V is to calculate DCT coefficients, realize compression function in calculating image processing, reduce data amount, and improve transmission efficiency.

4. Experiment

4.1. Experimental Flow.

The Retinex theory defines image $I(x,y)$ as the product of incident light component $L(x,y)$ and reflected light component $R(x,y)$ of the scene: namely,

$$I(x,y) = L(x,y) \cdot R(x,y). \tag{20}$$

Then, various methods are used to estimate the incident light component of the scene to obtain the illumination estimation $\widetilde{L}(x,y)$, the illumination estimation value is removed by the following formula, and then the reflection estimation $\bar{R}(x,y)$ is obtained as the final output result:

$$\widetilde{R}(x, y) = \frac{I(x, y)}{\widetilde{L}(x, y)}. \tag{21}$$

To improve the overall contrast of the poster image, the dynamic range adjustment of the incident light component must be carried out to obtain $L_{out}(x, y)$:

$$L_{out}(x, y) = G(\widetilde{L}(x, y)). \tag{22}$$

To increase the clarity of the poster image, the reflected light component must be enhanced in detail, and the following results are obtained $R_{out}(x, y)$:

$$R_{out}(x, y) = F(\widetilde{R}(x, y)). \tag{23}$$

Finally, the output image $O(x, y)$ is obtained by multiplying $L_{out}(x, y)$ and $R_{out}(x, y)$, and the poster image is enhanced:

$$O(x, y) = L_{out}(x, y) \cdot R_{out}(x, y). \tag{24}$$

4.2. Experimental Results

4.2.1. Comparison with Traditional Image Enhancement Algorithms in Compressed Domain.
Comparing the algorithm with the traditional compression domain image enhancement algorithm, Jinshan Tang proposed a compression domain image enhancement algorithm with DCT domain frequency band adjustment as the core. The algorithm transforms the image into the compressed domain and uses the same coefficients for each frequency band to adjust. The processing results are compared as shown in Figures 3–5.

The quadratic function $\tau(x)$ has no parameters and is easy to realize. Experiments on different images show that $\tau(x)$ has a good enhancement effect on dark areas with low contrast. The parametric function $\eta(x)$ has an adjustable parameter γ near 1, which has a good dynamic range compression effect for both dark and overbright areas.

The mean value, standard deviation, and entropy of each processing result in Figures 3–5 are given below, and the R channel of each image is selected for comparative explanation in Tables 1–3.

All the algorithms involved in the comparison are image enhancement algorithms in the compressed domain. It is more targeted to choose the JPEG image quality standard proposed by Wang as the evaluation standard. As shown in Table 4, the higher the value, the better the image quality.

As can be seen from Figure 6, compared with the traditional compressed domain image enhancement algorithm, the image quality enhanced by this algorithm has been significantly improved. The results show that the image quality of this algorithm is better than that of the traditional algorithm, which is more suitable for the visual characteristics of human eyes and accords with the subjective evaluation of human eyes.

4.2.2. Comparison with the Traditional Retinex Algorithm.
The processing effect of this algorithm is compared with the processing effect of MSR, as shown in Figures 7–9.

In the above experimental results, (a) of each image is the original image, (b) is the processing result of MSR, (c) is the processing result of mapping function $\tau(x)$ in the dynamic range adjustment process of illumination component, and (d) is the processing result of mapping function $\eta(x)$ in this algorithm. As can be seen from Figures 7–9, the color of MSR processing results is darker, and the brightness and color of the processing results are considered in Y, C_b, C_r color space. The processing effect is better than MSR, and the color is more natural. The details are also richer.

Objective evaluation of image quality, such as average value, standard deviation, and entropy of processing results, is shown in Tables 5–7. The mean value reflects the average brightness of the image, the standard deviation reflects the contrast of the image, and the entropy reflects the amount of information contained in the image.

It can be seen from Tables 5–7 that the brightness, contrast, and information content of the images processed by this algorithm are relatively good. The following image quality evaluation method based on statistics proposed by Jobson is used to further compare the processing results. The statistical results show that when the average gray level is between 90 and 190 and the standard variance is between 30 and 85, the image quality is better, as shown in Figure 10.

The specific method of statistics is to divide the image into nonoverlapping subblocks with the same size (generally 50×50 or 60×60); then calculate the standard deviation of each subblock, respectively, and then average the standard deviation of the subblocks; and multiply the average value of standard deviation $\overline{\sigma}$ with the gray value $\overline{I}$ of the image to get the evaluation result. The larger the result value, the better the image quality. The statistical results of the images in this chapter are shown in Table 8 and Figure 11.

As can be seen from Table 8, the quality evaluation results of the images processed by this algorithm are generally higher than those of the traditional MSR algorithm. The original image $\overline{I} * \overline{\sigma}$ value is relatively small, and the brightness and contrast of the image are not good. The quality evaluation data processed by this algorithm is improved compared with that processed by the traditional MSR algorithm.

4.2.3. Comparison with HE and MSR Algorithms.
The poster image enhanced by this algorithm is compared with HE and MSR algorithms, and the results are as follows:

Compared with the existing HE and MSR methods, the proposed algorithm has clearer processing results. The calculation of the incident component is more accurate than the traditional MSR algorithm, and the constancy of brightness and color is considered. The combination of the two methods further shows that the adaptive filter selection strategy and weighted fusion computing framework adopted in this algorithm are superior to the cumulative distribution equalization of the existing HE algorithm and the calibration filter of the MSR algorithm.

At the same time, it can be seen from Figures 12–14 that the processing results of this algorithm and the existing HE and MSR algorithms have reached the function of stretching
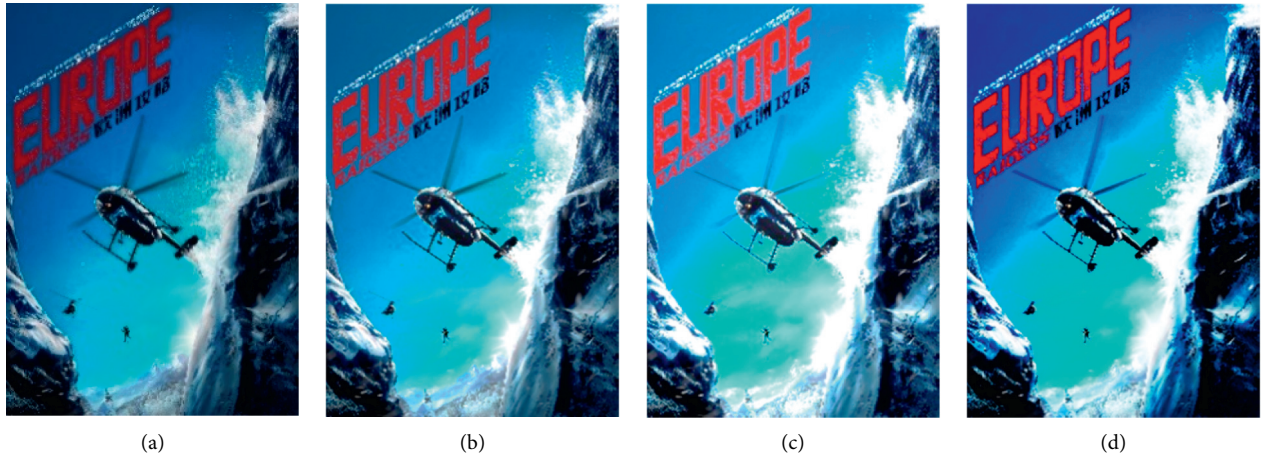
FIGURE 3: Comparison of poster 1 processing results. (a) Original drawing. (b) Result of TANG processing. (c) $\tau(x)$ outcome of treatment (d) $\eta(x)$ outcome of treatment.

FIGURE 4: Comparison of poster 2 processing results. (a) Original drawing. (b) Result of TANG processing. (c) $\tau(x)$ outcome of treatment. (d) $\eta(x)$ outcome of treatment.
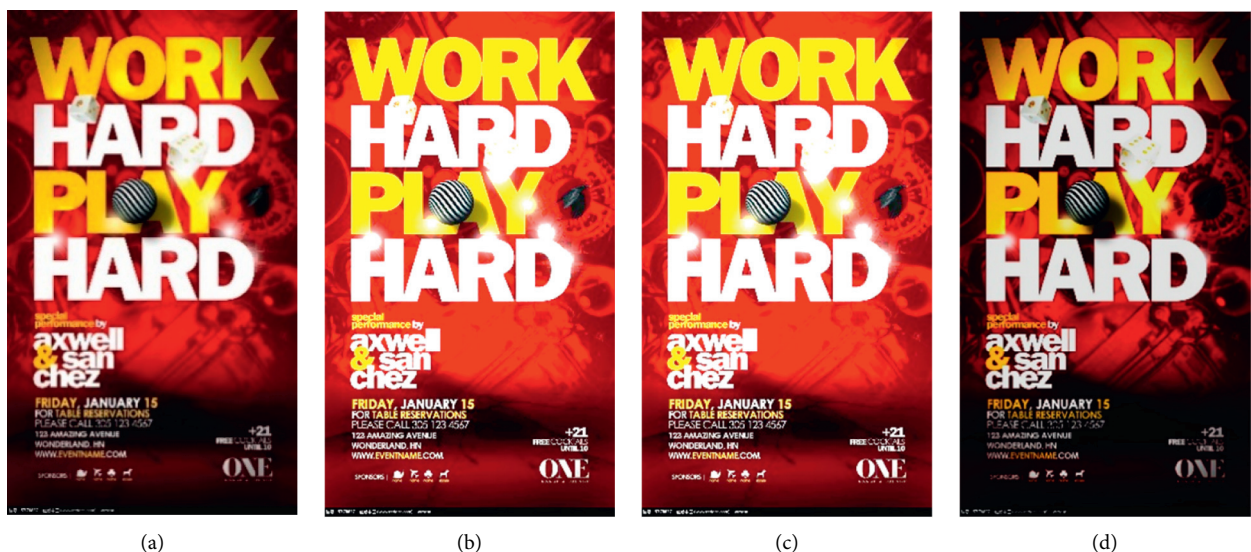
FIGURE 5: Comparison of processing results of poster 3. (a) Original drawing. (b) Result of TANG processing. (c) $\tau(x)$ outcome of treatment, (d) $\eta(x)$ outcome of treatment.

TABLE 1: Information content of poster 1.

Poster 1	Mean value	Standard deviation	Entropy
Original picture	35.7022	31.4674	6.4167
DCT enhancement	94.7360	70.1275	7.5659
$\tau(x)$	144.8949	51.6553	7.4089
$\eta(x)$	112.5389	29.0730	6.6685

TABLE 2: Information content of poster 2.

Poster 2	Mean value	Standard deviation	Entropy
Original picture	34.4844	39.3881	6.0319
DCT enhancement	78.7088	34.3615	6.8045
$\tau(x)$	135.0494	65.7795	6.7101
$\eta(x)$	127.8555	35.9261	6.3682

TABLE 3: Information content of poster 3.

Poster 2	Mean value	Standard deviation	Entropy
Original picture	91.3097	74.9361	7.3624
DCT enhancement	103.3698	76.9084	7.5544
$\tau(x)$	189.1378	83.7927	4.7532
$\eta(x)$	175.9137	52.7271	6.9573

TABLE 4: Image quality evaluation.

Poster	DCT enhancement	$\tau(x)$	$\eta(x)$
Poster 1	8.3945	7.7954	9.6317
Poster 2	8.7314	8.4625	9.7674
Poster 3	7.1804	6.6191	8.7442
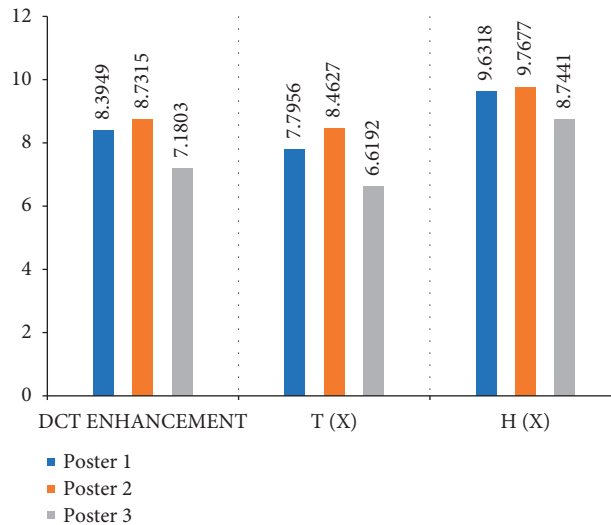
FIGURE 6: Image quality evaluation.

the dynamic range of the gray scale, but the gray scale distribution of the results processed by this algorithm is concentrated between 120 and 200, which is more in line with the average range of gray scale required by the best human vision. The gray mean and variance of the statistical image are used to verify the gray distribution, as shown in Tables 9 and 10.

As shown in Table 9 and Table 10, the average gray level of the algorithm processing results in this chapter is generally higher than that of the original image, HE, and MSR processing results, and most of them are concentrated between 140 and 180 gray levels, which is more in line with the observation range of human eyes. However, the standard deviation of the gray level of the algorithm processing results
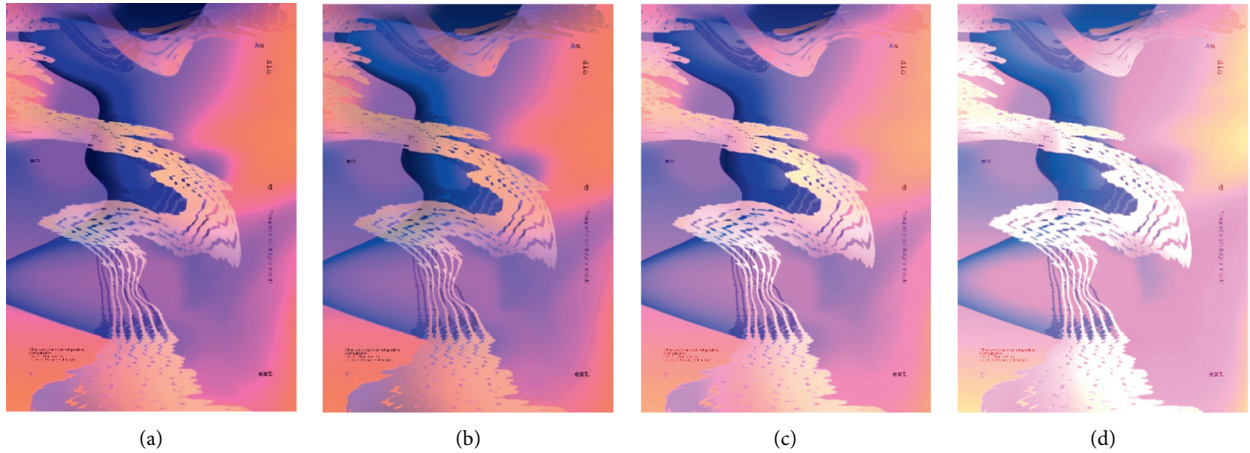
Figure 7: Poster 4 processing results. (a) Original drawing. (b) Result of MSR processing. (c) $\tau(x)$ outcome of treatment. (d) $\eta(x)$ outcome of treatment.
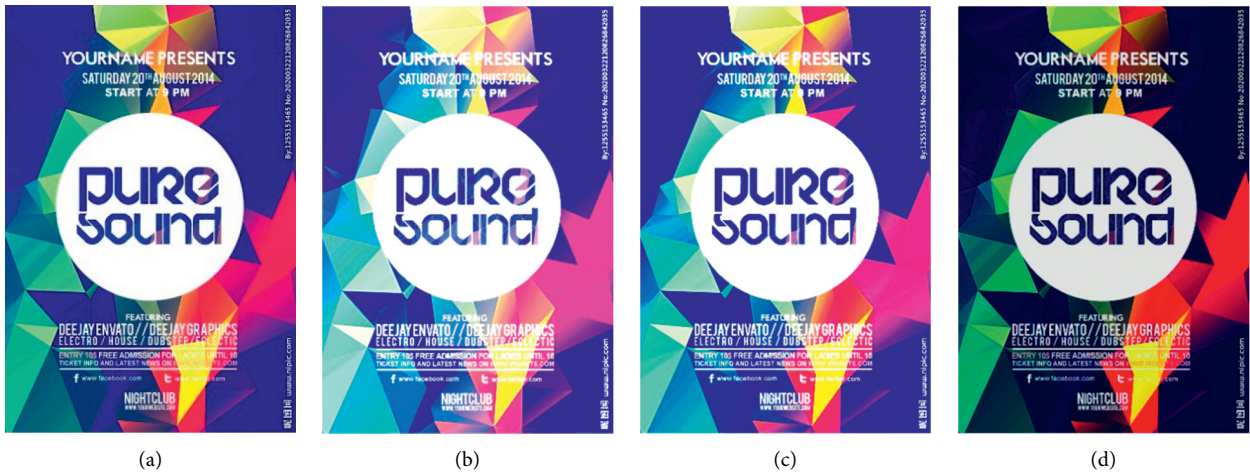
Figure 8: Poster 5 processing results. (a) Original drawing. (b) Result of MSR processing. (c) $\tau(x)$ outcome of treatment. (d) $\eta(x)$ outcome of treatment.

Figure 9: Poster 6 processing results. (a) Original drawing. (b) Result of MSR processing. (c) $\tau(x)$ outcome of treatment. (d) $\eta(x)$ outcome of treatment.

TABLE 5: Information content of poster 4.

Poster 4	Mean value	Standard deviation	Entropy
Original picture	37.1123	44.8613	5.8187
MSR	109.7563	44.3257	6.1724
$\tau(x)$	153.3104	56.3232	6.2354
$\eta(x)$	129.6176	30.2824	6.1662

TABLE 6: Information content of poster 5.

Poster 5	Mean value	Standard deviation	Entropy
Original picture	34.4853	39.3891	6.0335
MSR	78.7095	34.3628	6.8066
$\tau(x)$	135.0505	65.781	6.7121
$\eta(x)$	127.8564	35.9277	6.3697

TABLE 7: Information content of poster 6.

Poster 5	Mean value	Standard deviation	Entropy
Original picture	91.3105	74.9373	7.3646
MSR	103.3711	76.9092	7.5559
$\tau(x)$	189.1386	83.7939	4.7548
$\eta(x)$	175.9148	52.7286	6.9587
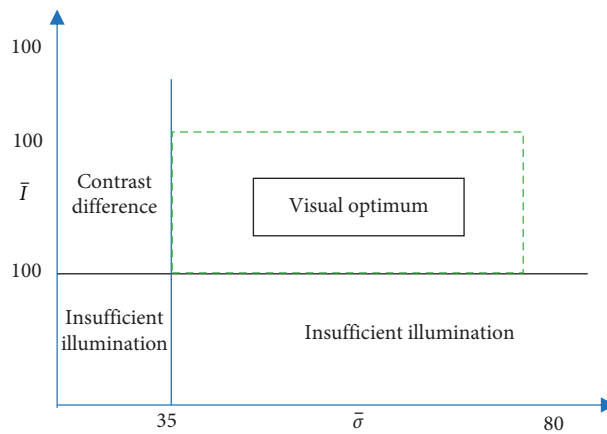
FIGURE 10: Visual effect distribution.

TABLE 8: Image quality evaluation $\overline{I}*\overline{\sigma}$.

Image quality evaluation	Original picture	$\tau(x)$ MSR	$\tau(x)$	$\eta(x)$
Poster 4 $\overline{I}*\overline{\sigma}$	3.5243e+004	2.8317e+004	1.2893e+005	7.3034e+004
Poster 5 $\overline{I}*\overline{\sigma}$	9.1194e+004	1.6066e+004	5.2845e+005	2.8071e+005
Poster 6 $\overline{I}*\overline{\sigma}$	3.6497e+004	4.7245e+004	6.7868e+004	5.7020e+004

in this chapter is reduced, which shows that the brightness difference between pixels is reduced, and the gray level is more concentrated near the median value, which is more suitable for human vision to observe. Experimental results show that the contrast and texture of low illumination images with different degradation degrees can be effectively enhanced, and better visual effects can be achieved in brightness, color, and detail. Finally, the output image is obtained by multiplying the incident component and the reflection component, which enhances the contrast while keeping the color information of the original image, and makes the enhanced image more natural in color. In the future, the quality evaluation of the processed image is often a relatively subjective view in enhancement processing. According to different application requirements, the evaluation criteria may vary greatly. Most of the existing image evaluation criteria are calculated according to the gray level of pixels, which to some extent ignores that pixels with the same gray level may have different meanings in the image. In addition, color information often reflects the quality of an
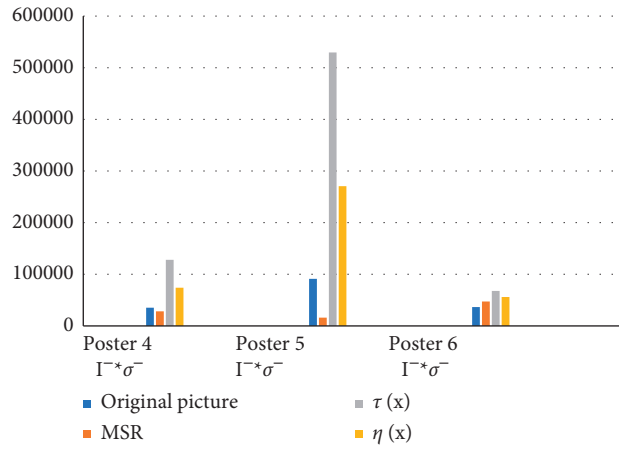
FIGURE 11: Histogram of evaluation criteria.

(a)   (b)   (c)   (d)

FIGURE 12: Poster 7 processing results. (a) Original drawings. (b) Results of HE treatment. (c) Results of MSR treatment. (d) The processing result of the algorithm.

(a)   (b)   (c)   (d)

FIGURE 13: Poster 8 processing results. (a) Original drawings. (b) Results of HE treatment. (c) Results of MSR treatment. (d) The processing result of the algorithm.
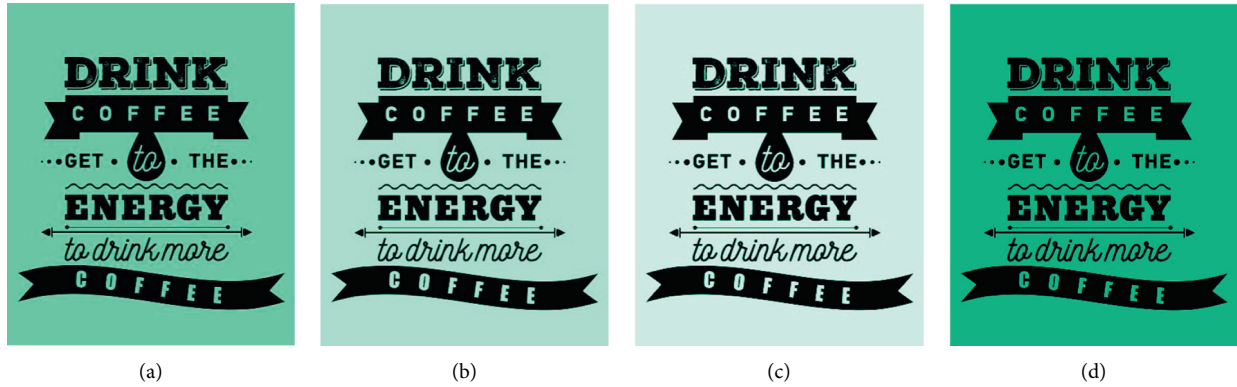
FIGURE 14: Poster 9 processing results. (a) Original drawings. (b) Results of HE treatment. (c) Results of MSR treatment. (d) The processing result of the algorithm.

TABLE 9: Comparison of image gray mean value.

Image	Original picture	HE treatment results	MSR processing result	The processing result of this algorithm
The first group of mean values	132.5610	127.5293	164.0854	150.0141
The second group of mean values	162.8083	127.4974	151.4334	154.8484
The third group of mean values	128.2582	127.4641	188.0133	149.6046

TABLE 10: Comparison of image gray variance.

Image	Original picture	HE treatment results	MSR processing result	The processing result of this algorithm
The first group of variances	29.1894	74.7411	62.1202	18.9998
The second group of variances	39.8166	74.6716	58.1282	30.566
The third group of variances	23.7563	74.8365	45.2645	18.5144

image better than brightness information. Therefore, how to put forward a quality evaluation standard with more visual significance and taking into account color information will become an important part of the follow-up research.

5. Conclusion

Aiming at the problem that the existing image enhancement algorithms are not ideal in the spatial domain, the image contrast is enhanced in the compressed domain, but the details and color information of the image cannot be preserved. An image enhancement algorithm in DCT compressed domain based on retina theory is proposed. Different from the traditional Retinex algorithm, which discards the incident light component and only enhances the reflection component, this algorithm processes the incident and reflection components at the same time, fully considers the image details, and defines a new spectral content ratio. By manipulating the spectral content ratio to process the reflection component, the details of the image can be effectively enhanced. Finally, the output image is obtained by multiplying the incident component and the reflection component, which enhances the contrast while keeping the color

information of the original image, and makes the enhanced image more natural in color. In this algorithm, the threshold method is used to suppress the block effect, and the information of adjacent subblocks can be used in the subblock processing in the DCT domain, so that halo artifacts can no longer be produced around the boundary with obvious light and shade jump. Experimental results show that the algorithm can effectively enhance the contrast and texture of low illumination images with different degradation degrees and achieve good visual effects in brightness, color, and detail. With compressed images, especially JPEG images, becoming the popular data transmission standard in the network, a new image enhancement algorithm based on retina theory is proposed in the compressed domain to make full use of the incident light component information of images. Based on the Retinex theory, DCT coefficients are divided into an incident component (DC coefficient) and a reflection component (AC coefficient). By adjusting the dynamic range of DC coefficients, carefully adjusting AC coefficients, and using the threshold method for block suppression, the compressed domain image can be enhanced. To enhance the compressed image better, we must make full use of the information contained in the DC coefficient.

Data Availability

The experimental data used to support the findings of this study are available from the author upon request.

Conflicts of Interest

The author declares that no conflicts of interest regarding this work.

References

[1] U. Anitha, R. Narmadha, D. R. Sumanth, and D. N. Kumar, "Robust human action recognition system via image processing," *Procedia Computer Science*, vol. 167, pp. 870–877, 2020.

[2] P. J. Burt, "Multiresolution techniques for image representation, analysis, and 'smart' transmission," *Visual Communications and Image Processing IV*, SPIE, Bellingham, Washington, US, pp. 2–15, 1989.

[3] R. A. Shaikh, L. I. Jian-Ping, A. Khan, and I. Memon, "Biomedical image processing and analysis using Markov Random Fields," in *Proceedings of the The International Computer Conference on Wavelet Active Media Technology and Information Processing 2015 Conference Dates: December 18-20*, pp. 179–183, IEEE, Chengdu, China, 2015.

[4] J. Yang, Q. Meng, M. Murroni, S. Wang, and F. Shao, "IEEE access special section editorial: biologically inspired image processing challenges and future directions," *IEEE Access*, vol. 8, pp. 147459–147462, 2020.

[5] Euroconfidentiel, "The directory of EU information sources," *Journal of the Optical Society of America A Optics Image Science & Vision*, Euroconfidentiel, vol. 15, no. 3, pp. 563–569, 1997.

[6] M. Selvapriya and D. J. Komalalakshmi, "Face recognition using image processing techniques: a survey," *ISSN*, vol. 3, no. 12, pp. 9704–9711, 2014.

[7] H. Ruan, J. Xu, and C. Yang, "Optical information transmission through complex scattering media with optical-channel-based intensity streaming," *Nature Communications*, vol. 12, no. 1, 2021.

[8] A. Beghdadi, A. Bouzerdoum, K. M. Iftekharuddin, and M.-C. Larabi, "Biologically inspired approaches for visual information processing and analysis," *Signal Processing: Image Communication*, vol. 28, no. 8, pp. 809-810, 2013.

[9] S. Yashiro, "Image processing apparatus for detecting object from image and method thereof," *United States Patent*, vol. 15, no. 4, pp. 55–57, 2010.

[10] T. Yamazaki, M. Sato, and H. Kajiwara, "Image processing apparatus and method, and storage medium used therewith," *U.S.*vol. 10, no. 7, pp. 7601–7604, 2007.

[11] S. Qi, "University of Florida," *ProQuest Dissertations Publishing*, vol. 35, no. 8, pp. 63–65, 2013.

[12] T. Steiner, R. Verborgh, J. Gabarro, E. Mannens, and R. Van de Walle, "Clustering media items stemming from multiple social networks," *The Computer Journal*, vol. 58, no. 9, pp. 1861–1875, 2015.

[13] M. Tong, K. Shao, X. Luo, and H. Duan, "Application of a fractional grey prediction model based on a filtering algorithm in image processing," *Mathematical Problems in Engineering*, vol. 2020, no. 11, pp. 1–18, 2020.

[14] J. Petit and R. Brémond, "A high dynamic range rendering pipeline for interactive applications," *The Visual Computer*, vol. 26, no. 6, pp. 533–542, 2010.

[15] Y. Tang, M. Zhao, and L. Li, "Secure and efficient image compression-encryption scheme using new chaotic structure and compressive sensing," *Security and Communication Networks*, vol. 2020, no. 2, pp. 1–15, 2020.

Research Article

Domain Adaptation-Based Automatic Modulation Recognition

Tong Li[1] and Yingzhe Xiao [2]

[1]*Department of Computer Science and Information, Shanxi Engineering Vocational College, Taiyuan 030032, China*
[2]*College of Information and Computer, Taiyuan University of Technology, Taiyuan 030600, China*

Correspondence should be addressed to Yingzhe Xiao; xiaoyingzhe@tyut.edu.cn

Deep learning-based Automatic Modulation Recognition (AMR) can improve the recognition rate compared with traditional AMR methods. However, in practical applications, as training samples and real scenario signal samples have different distributions in practical applications, the recognition rate for target domain samples can deteriorate significantly. This paper proposed an unsupervised domain adaptation based AMR method, which can enhance the recognition performance by adopting labeled samples from the source domain and unlabeled samples from the target domain. The proposed method is validated through signal samples generated from the open-sourced Software Defined Radio (SDR) GNU Radio. The training dataset is composed of labeled samples in the source domain and unlabeled samples in the target domain. In the testing dataset, the samples are from the target domain to simulate the real scenario. Through the experiment, the proposed method has a recognition rate increase of about 88% under the CNN network structure and 91% under the ResNet network structure.

1. Introduction

Automatic Modulation Recognition (AMR) plays an important role in the wireless communication field. For example, in cognitive radio tasks [1, 2], the transmission source can adaptively change modulation scheme according to the current channel state to improve communication efficiency. Therefore, the receiver needs to do AMR as a basic procedure prior to demodulation. In spectrum sensing tasks [3, 4], modulation information based on AMR is considered as basic information for describing the spectrum conditions, which can provide references for further spectrum management.

Traditional AMR methods have adopted carefully handcrafted features for classification [5, 6]. Normally, there are two procedures. Firstly, features with explicit meanings are extracted to form a feature vector or feature space. And then, according to the features extracted from the baseband sample signals, the modulation scheme is recognized through some machine learning based classification methods. The features from the first phase include the following: (1) spectral based feature, which can exploit spectral properties from different signal components; (2) wavelet based features from wavelet transformation; (3) high-order statistics, which are effective for classification of digital modulations; (4) cyclic features from cyclostationary analysis. In the second phase, many existing classifiers are adopted, including Support Vector Machine (SVM), Decision Tree (DT), and k Nearest Neighbor (KNN) [7–9]. These traditional methods are well described and proved to be effective in AMR tasks. However, many recent researches have shown that Deep Learning (DL) based methods can also be adopted to the AMR tasks [10, 11]. Even with a simple few-layer Convolutional Neural Network (CNN), the classification performance is better than the traditional handcrafted-feature based methods.

Deep learning based methods adopt training samples to enable automatic learning of effective features for modulation classification. These methods can be categorized according to different aspects including samples and network models. From the aspect of samples adopted, some methods adopt In-phase and Quadrature (IQ) samples in the time domain, while others adopt spectrograms. For the first type, as the modulation

information including phase and amplitude information in the IQ samples remains intact, it is in principle capable of classifying any modulation types. Also, this type of methods has the advantages of less input data volume, smaller set of network parameters, and so on, while, for the methods adopting spectrograms as the input, limitations lie in the classification of Phase Shift Keying (PSK) signals, due to the loss of phase information from the spectrograms. However, it is much easier to transfer the training models or parameters in image recognition tasks to AMR tasks, because spectrograms can be regarded as special image samples. The authors in [12] proposed a method that has a two-classification-stage structure. In the first stage, the spectrograms are adopted to discriminate PSK signals and other modulation signals. Then, in the second stage, IQ baseband samples are adopted for modulation classification within the PSK signals. From the aspect of network models, many try to enhance performance by adopting different network models. The authors in [13] have proposed a simple CNN network based framework, which is proved to have better performance than traditional handcrafted feature based methods. The authors in [14] adopted a Recurrent Neural Network (RNN) model for AMR tasks. Paper [15] gave comprehensive performance comparisons between many popular network models in AMR. The models include Residual Network (ResNet), inception network, and Convolutional Long-short Deep Neural Network (CLDNN). Among these networks, the authors conclude that the CLDNN model has the best recognition rate. [16, 17] proposed an optimized ternarized CNN, which is implemented in FPGA based hardware design. The method can bring a high data throughput with low classification latency at the cost of a few recognition rate reduction. The authors in [18] proposed a complex-valued network, which redefine the convolution layer, pooling layer, and so on from Euclidean spaces to non-Euclidean signal spaces utilizing weighted Fréchet mean. The complex-valued model can achieve state-of-the-art performance with much less parameter size (less than 1% of CNN). The network can significantly accelerate training process and can have less computational burden. The authors in [19] proposed a radio transformer network, which added a transformer structure to a common CNN network to cope with the problem of time, frequency, and phase offsets in the IQ baseband samples. The method basically introduced the notion of synchronization to design the transformer, which in principle can be regarded as a way to adopt a priori knowledge for classification. A fine-tuning based transfer learning method is proposed in [20], which has adopted partially labeled target domain data samples for tuning the parameters of the trained model with source domain samples. A complex-valued network is proposed in [21] for modulation recognition. Compared with its real-valued counterparts, which process the data in the Euclidean space, the complex-valued network proposes to process the data on the Riemannian manifold, which can better preserve the geometric structure of the complex-

valued signals. The experimental results show that it can enhance performance compared to real-valued networks. However, it is much more difficult to train than real-valued networks.

The mentioned methods, although effective, do not take into consideration that the signal samples can have significantly different distributions among training and actual classification. The reasons are twofold: (1) the channel varies between the training samples and samples in real implementation; (2) the estimated parameters differ between training samples and samples in real implementation; e.g., the estimated bandwidth and center frequency can affect the process of sampling. The different distributions between the training samples and the samples in real implementation can significantly deteriorate the performance. Therefore, strategies should be added to alleviate the problem. There are two basic strategies: (1) trying to enhance the training dataset by introducing signal samples with new channels and estimation parameters; (2) adopting transfer learning to transfer existing network to adapt new data with different distributions. For the first strategy, as the properties in new channels can be unknown, it is very hard to ensure that the enhanced dataset has the same distribution with the samples in real implementation. Also, many redundancies can be introduced to the dataset, which can make it significantly harder for the training process to converge. As for the second strategy, transfer learning does not necessarily need a large number of labeled samples with new distribution. Therefore, transfer learning is much appropriate for solving the mentioned problem.

This paper proposed an unsupervised domain adaptation based AMR method, which can enhance the recognition performance by adopting labeled samples from the source domain and unlabeled samples from the target domain. The proposed method can cope with the problem that, in real modulation recognition tasks, the channels and parameters (including the band width and center frequency) are varying. The proposed method has the following advantages: (1) the network structure has only added a small-scale subnetwork (the domain discriminator); thus, the changes along with the complexity increase compared with the original neural network are minor; (2) the method is unsupervised, meaning that, for the samples in the target domain, the label information is exemplified. This is especially suitable for real implementations, where samples without labels can be acquired much easier than those with labels; (3) the proposed method is compatible with existing network and, thus, can inherit the favorable features of existing network structure. The proposed method is validated adopting signal samples generated from the open-sourced Software Defined Radio (SDR) framework GNU Radio. In the training dataset, there are labeled samples in the source domain and unlabeled samples in the target domain. In the testing dataset, the samples are from the target domain to simulate the real scenario. Through the experiment, the proposed method has a recognition rate increase of about 88% under the CNN network structure and 91% under the ResNet network structure.

2. Related Works

In this part, firstly, we give a simplified introduction about the signal models and different factors, which can affect the performance for AMR. Then, some classical AMC methods including feature based and deep learning based methods are introduced.

2.1. Simplified Signal Model. A simplified received signal model can be described as follows:

$$r(n) = e^{j2\pi f_{\text{bias}}nT} * n_{crw}(t) \sum_{i=1}^{N-1} s(i)h(nT - iT - \epsilon T) + w(n),$$

$$(1)$$

where f_{bias} denotes the frequency offset when down converting to baseband, T denotes the symbol interval, $n_{crw}(t)$ denotes the time varying residual carrier random walk, $s(i)$ denotes the sending symbol sequence, $h(t)$ denotes the channel impulse response, ϵ denotes the symbol rate deviation, and $w(n)$ denotes white Gaussian noise. Noting that the mentioned model is only a simplified one. In real scenario applications, the signal propagation model can be very complex and can be time-frequency varying. Therefore, it is analytically hard to have a closed form of signal model. Therefore, when evaluating an AMR method, it is necessary to take into consideration whether changing factors will affect the robustness of the method. For traditional handcrafted feature based AMR methods, as the features are inferred from a enforced simple signal model assumption, the performance can deteriorate significantly when the assumption is not fulfilled or encountered changing factors.

In this paper, we have considered the four factors that can affect the IQ signal samples. The four factors are listed in detail in [22]. Here, only a simplified description is given. The four factors are as follows: (1) sample rate offset, which denotes the sample clock offset in the receiver side; (2) center frequency offset, which normally resulted from the errors in carrier frequency estimation; (3) selective fading model, which can be regarded as effects of multipath propagation; (4) white noise, which affects the received signal noise ratio.

2.2. Classical Deep Learning-Based Modulation Recognition. As deep learning can enable feature learning from training data, it can also be adopted in learning signal sample features for AMR applications. Compared with handcrafted feature based methods, the features leaned from training samples are more adopted for classification and may not clearly related to expert features. Here, some classical deep learning based methods are listed, which has shown better classification performance than traditional handcrafted feature based methods.

A classical CNN-based method is proposed in [13], which apply a CNN network for classification. Its network structure is shown in Figure 1, which has two convolution layers and two dense layers. The convolutional layers in the method can be regarded as "matched filter" in the receiver. If the training samples consisted of samples with different
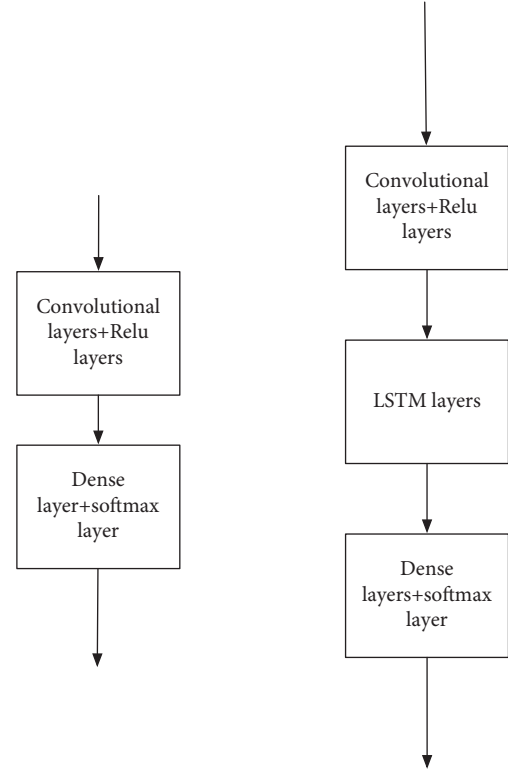
FIGURE 1: The CNN structure (left) and the CLDNN structure (rights) for AMR.

pulse shaping filter parameters, such as different roll-off factors and different types of shaping filter, the convolutional layers can cope with changes and recover constellation information. The convolutional layer can also recover the signals samples from different channel effects (act as compensation filter) if there are enough training samples from different channels.

Another classical deep architecture for modulation recognition is the Convolutional Long-short term Deep Neural Networks (CLDNN) network [15]. As shown in Figure 1, the network consisted of two convolutional layers and tow recurrent layers (adopting Long Shot Term Memory LSTM cells). The advantage for this structure is that it has considered not only the feature extraction in different scales, but also features extraction along time. This structure can correspond to procedures for common demodulation. The convolutional layers can serve as the matched filter, and the LSTM cells can serve as the synchronizer and sampler in time based processing. The new structure with layers corresponding to standard demodulators is proved to have better recognition performance than the CNN structure mentioned previously. However, the cost is a significant increase in training time.

In paper [19], a radio transformer based network is proposed. The radio transformer, inspired by spatial transformer network adopted for image recognition tasks, has added a parameter regression substructure to a simple CNN network. This substructure, noted as transformer, is designed to estimate the received radio signal related parameters, such as time offset, sampling offset, phase offset,

and frequency offset. The transformer, then, is adopted to compensate these offsets to enable better classification. As the radio transformer network has adopted known models, including radio signal synchronization and normalization, the recognition performance is enhanced.

3. Materials and Methods

As mentioned previously, in actual modulation recognition tasks, the received signal samples may encounter channel effects, which are different from those of the training sample. In other words, the distribution of the training samples is different from that of the testing samples. Here, we call the labeled signal samples the source domain, while the unlabeled signal samples encountered in real implementation will be called the target domain. The source domain differs from the target domain due to different radio propagation channels. In this paper, we propose a domain adaptation based network, which can adopt labeled source domain samples and unlabeled target domain samples for modulation recognition. In the context of this paper, the relationship of source domain and target domain is shown in Figure 2. In this section, the proposed method is introduced from the following three parts: the model details, the training and optimization process, and the dataset generation details.

3.1. Model Description. The proposed network structure has a model shown in Figure 3. As can be seen, the network is composed of three substructures: the feature extraction subnetwork, the modulation prediction subnetwork, and the domain classifier subnetwork. The feature extraction along with the modulation prediction subnetwork can be regarded as a traditional neural network, where the feature extraction subnetwork can be composed of several convolutional layers, and the modulation prediction subnetwork can be composed of a dense layer with softmax layer. Except for the layers mentioned above, the feature extraction together with the modulation prediction subnetwork can make up any classical neural network, such as a ResNet and CLDNN network. Inspired by the method proposed in [23], the domain classifier subnetwork is added to the existing neural network, which can be adopted for domain adaptation. Note that the goal of domain adaptation is that the features extracted are less domain sensitive. Equivalently, the extracted features are shared within both the source domain and the target domain. We will discuss the three subnetworks in the following descriptions.

(1) The feature extraction network. The subnetwork can be denoted as $H_f(.; \theta_f)$, where the subindex f denotes that it is related to feature extraction, and θ_f denotes the network parameters in it. In our implementation, the structure of the feature extractor can be similar to any existing networks. For example, it can make up of a few convolutional layers (similar to CNN network), a few residual diagram structures (similar to ResNet), and a few convolutional layers along with a few LSTM cells (similar to CLDNN).
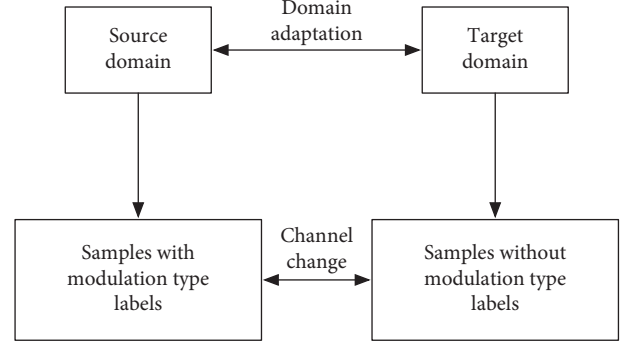
FIGURE 2: Source and target domain in the context of modulation recognition.
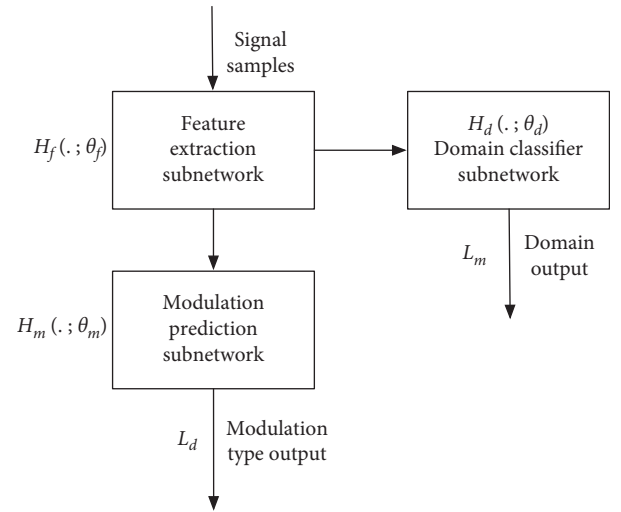
FIGURE 3: The network structure of the proposed model.

(2) The modulation prediction network. This can be denoted as $H_m(.; \theta_m)$, where the subindex m denotes that it is related to modulation prediction, and θ_m denotes the network parameters for prediction. The modulation prediction structure is similar to many neural works, with some linear dense layers and softmax layer for prediction. Here, we have adopted the classical cross-entropy loss L_m as the loss function from this structure, which can be denoted as

$$L_m = -\sum c_i \log p_i, \tag{2}$$

where c_i denotes the one-hot encoded results adopting known labels, while p_i denotes the output of the softmax layer. As the cross-entropy loss is commonly adopted in many deep learning based methods, only a simple introduction is given here.

(3) The domain classifier substructure. It can be denoted as $H_d(.; \theta_d)$, where θ_d denotes the network parameters. This substructure can be added to many existing neural networks for domain adaption. In this structure, we want the network to tell whether the input data are from the source or target domain,

so that the network can be diverged when training. However, we also want the domain classifier to feedback to the feature extraction subnetwork, so that the features extracted are shared in both the source and target domains. To make the mentioned two goals happen at the same time, a gradient reverse layer is added to the domain classifier. Mathematically, this means finding the optimal θ_d, which makes the domain classifier loss L_d minimum, while finding the optimal θ_f, which makes the samples domain invariant. L_d denotes the domain classifier loss function. It is also a cross-entropy loss similar to L_m. The difference is that, for Lm, the classification types equal the number of modulation types, while, for L_d, the classification type number is 2, denoting the source domain type or the target domain type. The detailed information for the reverse gradient layer will be introduced in the next section.

As mentioned previously, the model has a three-subnetwork structure, including the feature extraction, modulation prediction, and domain classifier subnetwork. The forward propagation function can be denoted as $H_f(.;\theta_f)$, $H_m(.;\theta_m)$ and $H_d(.;\theta_d)$, respectively. The cost function is listed as follows:

$$F_{\text{loss}}\big(\theta_f, \theta_m, \theta_d\big) = L_m\big(\theta_f, \theta_m\big) - \eta L_d\big(\theta_f, \theta_d\big), \qquad (3)$$

which are composed of two terms. The term $L_m(\theta_f, \theta_m)$ denotes the cross-entropy cost function from modulation recognition, and the term $L_d(\theta_f, \theta_d)$ denotes the domain classification error. η is a parameter denoting the respective weight for the two terms responsible for the overall cost function. Worth noting is that there is a minus sign in front of the domain classifier related cost function, which means that we want the domain classification error to grow larger during training. Again, this can make the feature extraction subnetwork prone to favor features shared by both source and target domain. The two terms of the cost function can be expanded as

$$L_m\big(\theta_f, \theta_m\big) = \sum_i L_{m,i}\big(H_m\big(H_f\big(\mathbf{s}_i; \theta_f\big); \theta_m\big), m_i\big),$$
$$L_d\big(\theta_f, \theta_d\big) = \sum_i L_{d,i}\big(H_d\big(H_f\big(\mathbf{s}_i; \theta_f\big); \theta_d\big), d_i\big), \qquad (4)$$

where the subindex i denotes the corresponding training signal sample index, $\mathbf{s}_i$ denotes the training sample, m_i denotes the modulation label the sample, and d_i denotes the domain label of the corresponding sample. As can be seen from the overall cost function, signals samples of different modulation types and domains are all needed for training.

3.2. Training and Optimization. The overall structure of the proposed method is shown in Figure 3. The key for the proposed method is the domain adaptation subnetwork (also denoted as the domain classification subnetwork). As mentioned, to make the network converge, the domain classification loss should be minimized during optimization. However, this will courage the feature extraction

subnetwork to learn features effective to discriminate the data domain, which is not our intention. To make the feature extraction subnetwork learn shared features in source and target domain, while keeping the network converged, a gradient reversal layer is added between the feature extraction subnetwork and the domain classification subnetwork. This layer is a special one with no parameters to train and optimize, while it only has a hyperparameter η. The η parameter will be multiplied with a minus during optimization shown as follows:

$$\theta_f = \theta_f - \mu\left(\frac{\partial L_m}{\partial \theta_f} - \eta\frac{\partial L_d}{\partial \theta_f}\right), \qquad (5)$$

$$\theta_m = \theta_m - \mu\frac{\partial L_m}{\partial \theta_m}, \qquad (6)$$

$$\theta_d = \theta_d - \mu\frac{\partial L_d}{\partial \theta_d}. \qquad (7)$$

Equations (5)–(7) have shown the updates of the parameters from the subnetworks during training, where μ denotes the update rate, θ_f, θ_m and θ_d denote the feature extraction parameters, the modulation classification parameters, and the domain classification parameters, and the partial signs denote the calculated partial gradient of different losses. From the mentioned parameter update equations, we can see that the mathematical expression of the gradient reversal layer can be written as

$$\mathbf{R}(\mathbf{x}) = \mathbf{x},$$
$$\frac{\partial \mathbf{R}}{\partial \mathbf{x}} = -\eta\mathbf{I}. \qquad (8)$$

Note that, here, $\mathbf{I}$ denotes the identity matrix. The backward expression has a partial gradient of $-\eta\mathbf{I}$. The reason is that the parameter η is a hyperparameter and is not related to the losses during optimization.

4. Results and Discussion

4.1. Dataset Generation. The dataset adopted here is generated from the open-sourced Software Defined Radio GNU radio. Here, we are inspired by the open-sourced dataset from the authors in [22]. Based on that dataset, we enhanced it with different frequency offset in the carrier, along with different channel responses. By generating data with different carrier frequency offsets and channel responses, the target domain data are produced, which can be adopted to validate the proposed method. The modulation dataset is composed of signal samples with 5 different Signal Noise Ratios (SNR), from 0 to 20 dB. The dataset includes 9 commonly seen modulation types: AM, FM, GFSK, BPSK, QPSK, 8PSK, OQPSK, 8QAM, and 16QAM. Each signal sample is made of 128 baseband IQ samples. As mentioned, the differences between the source and target domain data are twofold as shown in Table 1. The differences between the source and target domain for the dataset mainly lie in three aspects: (1) for the source domain, no carrier frequency

TABLE 1: Source domain and target domain differences.

	Source domain	Target domain
Frequency offset	No	Yes
Channel condition	Gaussian channel	Rayleigh channel
With labels	Yes	No

offset is added, while it is added in the target domain. Specifically, the maximum carrier frequency offset added to the target domain is 0.05 ∗ pi in the digital frequency. The mentioned situation is commonly seen when performing signal detection in a wide band and down convert it to baseband; (2) the source domain is with a typical Gaussian channel, while the target domain is with the Rayleigh channel; (3) during training, the source domain data are with labels, while the target domain data are without labels. The differences between the source and target domain data are listed in Table 1.

Figure 4 has shown some signal samples from the dataset. As can be seen, the signal samples are baseband IQ data. During the preprocessing procedure, all the signal samples are normalized according to the total energy.

4.2. Hyper Parameter Settings. As mentioned previously, both source and target domain samples are adopted in the training processes. The difference is that, for the target domain training samples, the samples are without modulation type labels.

The equation from 1 has a hyperparameter η_p, which denotes the ratio of the two losses: the cost function modulation recognition and the cost function of domain classifier. In this paper, the parameter η_p is set as

$$\eta_p = \frac{2}{1 + \exp(-\varepsilon p)} - 1, \tag{9}$$

where η_p denotes the hyperparameter (it has a value between 0 and 1), and p denotes the real time recognition rate. To set the value of η_p like this makes the network much easier to train. The reason is as follows: in the beginning, when the recognition rate p is near 0, the training process only aims at having higher modulation recognition rate. When p is higher, the network cost function also starts to consider the effects of domain classification. Parameter ε denotes how fast the training goal changes from only higher recognition rate to a balance between higher recognition rate and domain relevance. To choose the best hyperparameter, we have compared the recognition rate and the training time of different values of ε. The results are shown in Figure 5. We can see that the training time reaches the minimum when $\varepsilon = 15$, while the recognition rate reaches its maximum when $\varepsilon = 20$. As a matter of fact, we set the value of ε as 20, to make the network have a maximum recognition rate.

5. Methods Comparisons

In order to fully illustrate the effectiveness of the proposed method, in the experiment, the traditional CNN network and ResNet network were changed. The domain classification subnetwork is added according to the method described above. During the training of the network, source domain samples with labels and unlabeled target domain samples are used. On the basis of the reconstructed network, the method of non-transfer learning, the supervised transfer learning method, and the fine-tuning based transfer learning method proposed in [20] are compared. The test datasets adopted in the experiment for method comparison are the same, including samples of both the source and target domains. The training sets adopted in the four methods compared are different, and their characteristics are shown in Table 2. Among them, the method proposed in this paper has adopted source domain samples with modulation type labels and target domain samples without modulation type labels for training. The non-transfer learning method has only adopted source domain samples with modulation type labels for training. For the supervised transfer learning method, both labeled source domain samples and labeled target domain samples are adopted. The fine-tuning based transfer learning method firstly adopts labeled source domain data for training and then a small part of labeled source domain data for parameter fine-tuning. Table 2 gives recognition rate comparisons between the mentioned four methods using CNN and ResNet as the basic neural network.

As can be seen from Table 3, ResNet, as a whole, has a higher recognition rate than the CNN network, which is consistent with previous publications. Among the CNN-based methods, the method proposed in this paper has increased the recognition rate by about 48% and 6% compared with the method without transfer learning and the method of fine-tuning based transfer learning method. This has demonstrated that the proposed unsupervised transfer learning method (here, the unsupervised notion refers to the source domain data that are unlabeled) can make full use of the distribution of the unlabeled target domain data to adjust the neural network, so that the feature extraction subnetwork can adapt to the input of both source and target domain samples, which are distributed differently. However, the fine-tuning based method only relies on a few labeled target domain data and cannot make full use of all target domain data. Therefore, its recognition rate is lower than that of the method proposed in this paper. Compared with the supervised method, the recognition rate of the method proposed in this paper is only 3% lower, indicating that the method proposed in this paper has almost reached the upper limit of the recognition rate (i.e., both labeled source and target domains are used for training). In practical applications, usually only a small amount of labeled source domain data and a large amount of unlabeled target domain data can be obtained. Therefore, the proposed method is more practical than the supervised method. In the comparison of methods based on ResNet, the method in this paper has improved the recognition rate by about 46% and 8% compared with the method without transfer learning and the method of fine-tuned transfer learning, respectively. Compared with the recognition rate of supervised methods, ours is only 3% lower, which denotes a similar conclusion.
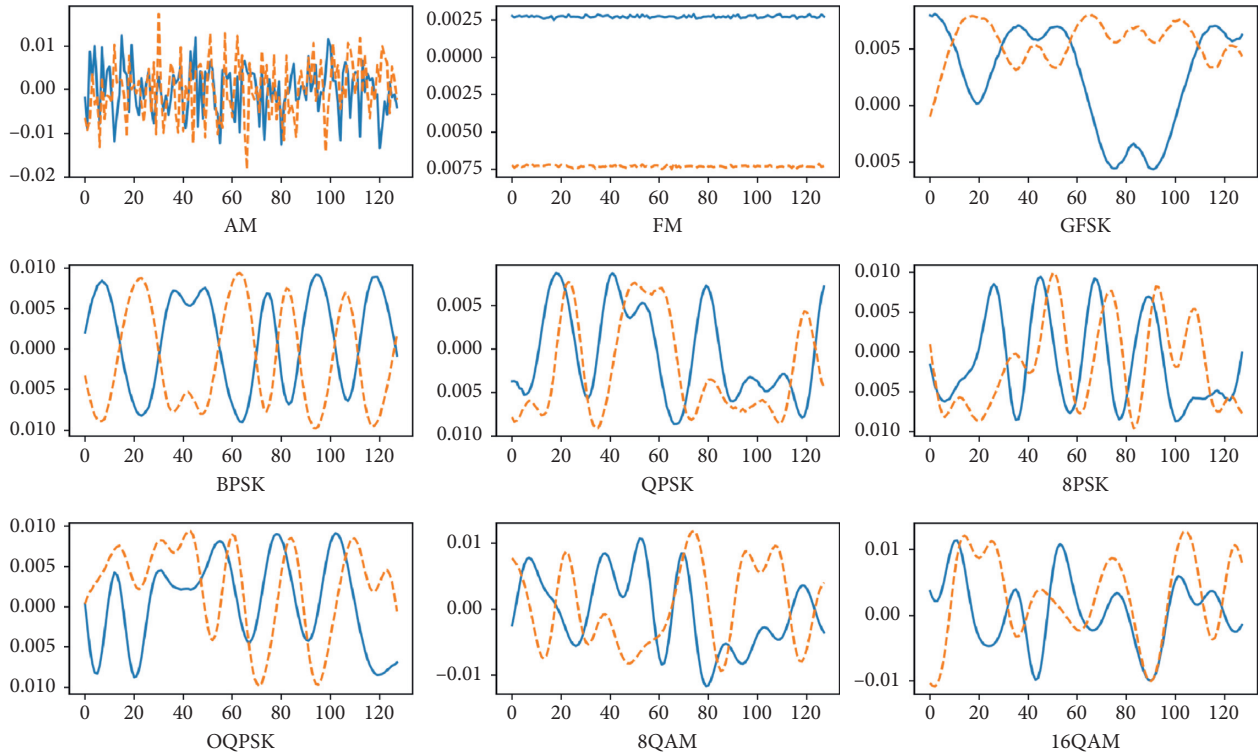
FIGURE 4: Several signal samples of different modulation types in the dataset.
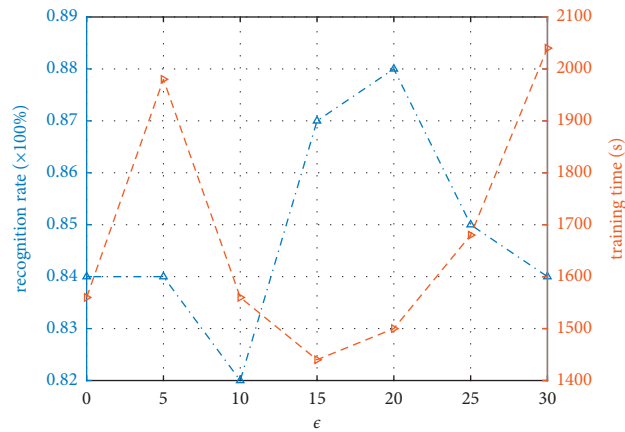
FIGURE 5: The effects of different ε over training time and recognition rate.

TABLE 2: The differences between the training dataset including source and target domain in the four methods.

Method	Source domain	Target domain
Proposed method	With label	Without label
Without transfer learning	With label	No data needed
Fine-tuning based method	With label	Partially with labels
Supervised method	With label	With label

Table 3: The recognition rate comparisons between the four methods.

Method	CNN based (%)	ResNet based (%)
Proposed method	88	91
Without transfer learning	40	45
Fine-tuning based method	82	83
Supervised method	91	94

6. Conclusions

In the application of deep learning based modulation recognition, the purpose of improving the recognition rate is often achieved by changing the deep neural network structure. However, in practical applications, training samples and real scenario signal samples in practical applications have different distributions due to different channels and different frequency offsets. This can make the neural network achieve a high recognition rate for source domain samples, while having a poor recognition rate for target domain samples. This paper proposed an adaptive unsupervised learning modulation recognition method, which can use labeled data in the source domain and unlabeled target domain data for training and can be directly implemented adopting existing deep learning network with minor structural changes. Through the simulation data set generated by the open source GNU Radio software, it proves the proposed method in this paper, compared with the method without transfer learning and the method with fine-tuning, and the recognition rate has increased by 48% and 6%, respectively, under the condition of CNN as the basic network. Under the condition of ResNet as the basic network, the recognition rate has increased by 46% and 8%, respectively. Moreover, the method proposed in this paper is close to the upper limit of the modulation recognition rate after training adopting both labeled target and labeled source domain data. Under the condition of CNN and ResNet as the basic network, it only drops by 3% and 3%, respectively.

Data Availability

The RadioML dataset is publicly available.

Conflicts of Interest

There are no conflicts of interest among the authors.

Acknowledgments

References

[1] H. Anandakumar and K. Umamaheswari, "Supervised machine learning techniques in cognitive radio networks during cooperative spectrum handovers," *Cluster Computing*, vol. 20, no. 2, pp. 1505–1515, 2017.

[2] F. A. Awin, Y. M. Alginahi, E. Abdel-Raheem, and K. Tepe, "Technical issues on cognitive radio-based Internet of Things systems: a survey," *IEEE access*, vol. 7, pp. 97887–97908, 2019.

[3] W. Lee, M. Kim, and D.-H. Cho, "Deep cooperative sensing: cooperative spectrum sensing based on convolutional neural networks," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 3, pp. 3005–3009, 2019.

[4] M. S. Gupta, K. Kumar, and K. Kumar, "Progression on spectrum sensing for cognitive radio networks: a survey, classification, challenges and future research issues," *Journal of Network and Computer Applications*, vol. 143, pp. 47–76, 2019.

[5] J. Huang, S. Huang, Y. Zeng, H. Chen, S. Chang, and Y. Zhang, "Hierarchical digital modulation classification using cascaded convolutional neural network," *Journal of Communications and Information Networks*, vol. 6, no. 1, pp. 72–81, 2021.

[6] D. H. Al-Nuaimi, I. A. Hashim, I. S. Zainal Abidin, L. B. Salman, and N. A. Mat Isa, "Performance of feature-based techniques for automatic digital modulation recognition and classification-A review," *Electronics*, vol. 8, no. 12, p. 1407, 2019.

[7] Z. Zhu and A. K. Nandi, *Automatic Modulation Classification: Principles, Algorithms and Applications*, John Wiley & Sons, Hoboken, NJ, USA, 2015.

[8] E. Azzouz and A. K. Nandi, *Automatic Modulation Recognition of Communication Signals*, Springer, New York, NY, USA, 2013.

[9] F. Wang and X. Wang, "Fast and robust modulation classification via Kolmogorov-Smirnov test," *IEEE Transactions on Communications*, vol. 58, no. 8, pp. 2324–2332, 2010.

[10] Y. Wang, M. Liu, J. Yang, and G. Gui, "Data-driven deep learning for automatic modulation recognition in cognitive radios," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 4, pp. 4074–4077, 2019.

[11] Y. Tu and Y. Lin, "Deep neural network compression technique towards efficient digital signal modulation recognition in edge device," *IEEE Access*, vol. 7, pp. 58113–58119, 2019.

[12] X. Zha, X. Qin, Y. Zhou, and H. Peng, "Power of deep learning for amplitude-phase signal modulation recognition," in *Proceedings of the 2019 IEEE 8th Joint International Information Technology and Artificial Intelligence Conference (ITAIC)*, May 2019.

[13] T. J. O'Shea, J. Corgan, and T. Charles Clancy, "Convolutional radio modulation recognition networks," in *Proceedings of the International Conference on Engineering Applications of Neural Networks*, Springer, Aberdeen, UK, September 2016.

[14] C. Yang, Z. He, Y. Peng, Y. Wang, and J. Yang, "Deep learning aided method for automatic modulation recognition," *IEEE Access*, vol. 7, pp. 109063–109068, 2019.

[15] N. E. West and T. O'Shea., "Deep architectures for modulation recognition," in *Proceedings of the 2017 IEEE International Symposium on Dynamic Spectrum Access Networks (DySPAN)*, March 2017.

[16] S. Tridgell, D. Boland, P. H.W. Leong, and S. Siddhartha, "Real-time automatic modulation classification," in *Proceedings of the 2019 International Conference on Field-Programmable Technology (ICFPT)*, December 2019.

[17] S. Tridgell, D. Boland, P. H.W. Leong, R. Kastner, A. Khodamoradi, and S. Siddhartha, "Real-time automatic modulation classification using RFSoC," in *Proceedings of the 2020 IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW)*, May 2020.

[18] R. Chakraborty, Y. Xing, and S. Yu, "SurReal: complex-valued learning as principled transformations on a scaling and rotation manifold," 2019, https://arxiv.org/abs/1910.11334.

[19] T. J. O'Shea, L. Pemula, D. Batra, and T. C. Clancy, "Radio transformer networks: attention models for learning to synchronize in wireless systems," in *Proceedings of the 2016 50th Asilomar Conference on Signals, Systems and Computers*, November 2016.

[20] Z. Lei, "Research on deep learning based modulation recognition technologies,"Master thesis, University of Electronic Science and Technology of Chinalocation, Chengdu, China, 2019.

[21] R. Chakraborty, J. Wang, and S. X. Yu, "Sur-real: frechet mean and distance transform for complex-valued deep learning," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, Long Beach, CA, USA, June 2019.

[22] O'shea, J. Timothy, and W.. Nathan, "Radio machine learning dataset generation with gnu radio," *Proceedings of the GNU Radio Conference*, vol. 1, no. 1, 2016.

[23] Y. Ganin and V. Lempitsky, "Unsupervised domain adaptation by backpropagation," in *Proceedings of the International Conference on Machine Learning*, PMLR, Lille, France, July 2015.

Hindawi

Retraction

Retracted: Willingness and Evaluation Model of College Students' Online Learning Behavior Based on Distributed Cognition

Scientific Programming

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope

(2) Discrepancies in the description of the research reported

(3) Discrepancies between the availability of data and the research described

(4) Inappropriate citations

(5) Incoherent, meaningless and/or irrelevant content included in the article

(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] L. Chen and S. Huang, "Willingness and Evaluation Model of College Students' Online Learning Behavior Based on Distributed Cognition," *Scientific Programming*, vol. 2021, Article ID 6386455, 12 pages, 2021.

Research Article

Willingness and Evaluation Model of College Students' Online Learning Behavior Based on Distributed Cognition

Lingjing Chen [1] and **Shuying Huang**[2]

[1]*Students' Affairs Division, Yiwu Industrial & Commercial College, Yiwu 322000, Zhejiang, China*
[2]*Mental Health Center, Yiwu Industrial & Commercial College, Yiwu 322000, Zhejiang, China*

Correspondence should be addressed to Lingjing Chen; chen@ywicc.edu.cn

Guided by distributed cognition theory, we analyze the influential elements of content, tools, and contextual interactions in the online learning process through research and case studies to explore the characteristics and evaluation of college students' willingness to engage in online learning behavior under distributed cognition and provide guidance for the experience design of online education platforms. Based on distributed cognition, this paper designs a convolutional neural network model based on InceptionNet, which uses a global average pooling layer instead of a fully connected layer to reduce the number of parameters, and InceptionNet increases the depth and width of the network by branching to improve the performance of the network and avoid overfitting. Distributed cognitive theory emphasizes the distributed nature of cognition, and the intrinsic variables that influence the willingness to participate in online learning communities from a systemic viewpoint are mainly attitudes, subjective norms, expected emotions, competence, sense of relatedness, desire, and perceived behavioral control. In addition, perceived behavioral control has a direct positive effect on the willingness to participate in online learning communities.

1. Introduction

Online education platforms have become an important tool to assist learning in recent years, and the number of online users and the variety and number of online education platforms continue to grow. The online education business is gradually developing in the direction of multidimensionality, depth, personalization, and segmentation, but the learning experience still suffers from the problems of designing a large number of cognitive activities around the individual cognitive ability of learners and resulting in low learning autonomy, lack of classroom participation, and little collaboration and communication among learners [1, 2].

The rapid development and widespread use of Internet technology have led to dramatic changes in the way people live, work, learn, and communicate. More and more employees of enterprises, government officials, school teachers, and social workers are adopting online learning to realize lifelong learning

[3]. The informal online learning communities built by the Internet are favored by adult learners and play an indispensable role in their learning process because they are not constrained by time and place [4]. For example, it can solve the problems of traditional learning communities, such as time and space constraints, difficulties in accessing resources, and lack of long-distance interaction; it can increase information exchange, mutual support, team commitment, cooperation, and satisfaction with the learning experience; it enables members to find like-minded and progressive learning partners on the Internet, and a common learning vision can stimulate their learning initiative and creativity and enhance their sense of belonging. It is conducive to constructing knowledge and exchanging emotions and paves the way for members' common growth and the reduction of the knowledge gap [5]. With the deep integration of "Internet + education," the significance and value of adult online learning communities have been highlighted, and their practical applications in different professional fields are becoming more and more widespread [6, 7].

However, in practice, problems such as insufficient willingness of adults to participate in online learning communities, insufficient acceptance and use of mobile learning, and insufficient interaction among members still abound. What factors affect the willingness to participate in online learning communities? What is the relationship between them? The results of Bette's study of Alberta Community Adult Learning Council coordinators' participation in online learning communities show that their motivations for participation include learning new skills and gaining access to work practices, strengthening ties with colleagues and careers, and reducing isolation due to job functions and geographic location [8, 9]. Results of Karunanayaka's case study of teachers suggest that interactive and collaborative course design, the role of online tutors, and the various strategies used by online learners have an impact on the learning outcomes of online learning communities. The authors of [10, 11] argued that learners' conceptualization of prior knowledge and experience, focus on computer-mediated communication, and participation in courses and projects can enhance their perception of and social engagement in online learning communities. The study in [12] showed that whether adults choose to join online communities of practice is related to intrinsic factors such as stress relief, increased social interaction, community service, and improved work performance. The study in [13] argued that the role of the facilitator is important in virtual learning community learning activities, both to help form effective communication among participants and to play an active role in effectively managing the process of the activity. The study in [14] noted that, in informal online learning communities, social presence influences participants' satisfaction and sustained attention.

Although there is no shortage of studies that explore the behavioral decision-making process of adult participation in online learning communities from motivational factors and identify the key factors that influence adult learning behaviors, the psychological mechanisms of adult participation in online learning communities from the perspective of motivation are rarely examined. Given that adults' behavioral decisions to participate in online learning communities are mostly accompanied by clear goals and that adult learning is characterized by distinctive self-direction, this study will introduce goal-oriented behavior theory and self-determination theory to clarify the drivers of adult participation in online learning communities and their underlying mechanisms from a microscopic perspective and provide theoretical support for enhancing the effectiveness of participation.

The contributions of this paper are as follows:

(1) This paper designs a convolution neural network model based on InceptionNet, which uses the global average pooling layer instead of the full connection layer to reduce the number of parameters. At the same time, InceptionNet increases the depth and width of the network by branching, which can improve the performance of the network and avoid overfitting.

(2) Based on distributed cognitive theory, this paper analyzes the internal variables affecting the willingness to participate in online learning community from the perspective of system view. It also explores the relationship between these internal variables and students' behavior intention.

(3) This paper makes a large number of experimental arguments and finds that the sense of relevance and perceived behavior control have a significant positive impact on the desire to participate in the online learning community, and desire further positively affects the generation of willingness to participate in behavior.

2. Overview of Distributed Cognitive Theory

The American psychologist Hutchens first introduced the concept of distributed cognition; that is, cognition is distributed in nature and cognitive activity depends on the joint functioning of all elements in the whole context [15]. Scholar Salomon proposed that individual cognition is the core of distributed cognition and that the interaction of elements in distributed cognitive activities resulted in the spiral development of individual cognition [16]. The theory gives designers a systematic perspective on research and aims to shift the researcher's unit of analysis from the individual involved in cognition in the mind to consider the relationship between the individuals involved in cognitive activities in the external environment and the factors that together participate in cognitive activities [17].

Some scholars have long applied distributed cognition theory to educational practice; the study in [18] introduced distributed cognition theory to the analysis and design of offline classroom learning activities, arguing that it is important to focus not only on the design of learning content and teaching tools but also on the collaborative communication activities among learners in the learning environment, which provides a new research perspective on the design of course learning. The study in [19] applied distributed cognition theory to the practice of online learning, using the individual, media, environmental, cultural, and social elements of distributed cognition to establish mapping relationships between learners, instructional organizers, learning resources, collaborative tools, and learning contexts in distributed learning environments as the entry point, as shown in Figure 1. The activities of the e-learning platform are classified into content interaction, tool interaction, and contextual interaction in three dimensions: interactive content, interactive behavior, and interactive environment, and strategies to promote collaborative learning in distributed e-learning environments are proposed [20]. However, none of the above studies have developed a detailed description of how to optimize the experience of e-learning practices through distributed cognitive theory from a design perspective.

3. Case Study of Online Education Application Design

In order to ensure that the influential elements of online learning interactions are evidenced in real design practice, we use the Coursera platform [4, 21], a typical overseas
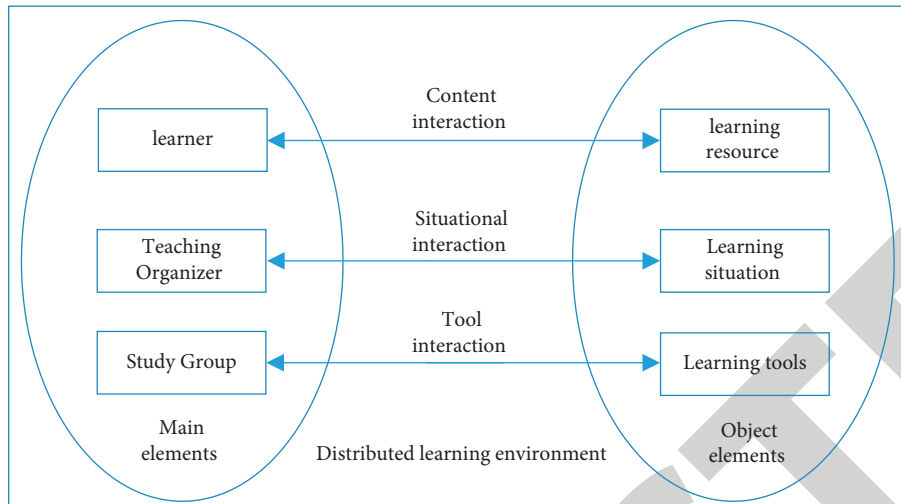
FIGURE 1: Model of collaborative learning interaction process in a distributed learning environment.

online education platform, as a case study to explore the application of distributed cognitive theory in the design of online education platforms. Coursera is one of the largest online education platforms in the world, which supports students to obtain the certificate of completion of corresponding subjects through independent study. Currently, the platform has more than 30 million registered users and offers more than 2000 courses.

3.1. Distributed Online Education. Under the distributed cognition perspective, learner-learning content interaction is influenced by the content quality of learning resources, social interaction of online videos, knowledge structure relationship of content, and interface design of learning resources. Individual cognition is the core of distributed cognition, so learning resources designed based on learners' purposes and needs are the first step in the design of online education platforms [22]. The Coursera platform provides a large number of excellent course resources in popular fields such as computer science, data science, information technology, mathematics, language learning, art and the humanities, and business. Learning materials mainly include videos, reading materials, graded tests, practice tests, and other forms, as shown in Figure 2. The instructor will provide learners with expandable learning materials including people, books, articles, and website links, based on their true level, to promote the construction of personal knowledge. Course materials set up systematic knowledge into staged course, and learners can discover the relevance of knowledge and form a systematic knowledge framework through different forms of learning materials according to the course they want to study. In terms of interface design, Coursera's visual elements are consistent, with titles and navigation, content information, and so forth clearly visible and overall easy for learners to use, without excessive visual interference, in line with learners' cognitive needs.

The Coursera platform is based on a humanistic and emotional need for instructor-led course content design, where the online videos show the instructor interacting with

the online learners, rather than just a one-way transfer of knowledge. The lecturer's teaching style, language, and teaching behavior will enhance learners' interest in learning. Learning materials of online classes in many domestic online education platforms are purely recorded screens, where only knowledge points appear without a lecturer, lacking classroom interactivity and making it difficult to establish learners' trust and thus reducing learning initiative [8]. Some videos are recorded offline lectures where the lecturer appears on the screen. However, traditional offline lectures face offline students, while online lectures face learners beyond the screen, so they are not designed for the characteristics and needs of the online learning population. However, courses on Coursera platform are clearly optimized for online teaching, such as well-designed slides, short videos for mobile learning, live demonstrations of course recordings, and quizzes that pause in the middle of video playback.

3.2. Experience Design Strategies for Online Education Platforms from a Distributed Cognitive Perspective. Through the above analysis of content interaction, tool interaction, and contextual interaction, and combined with the case study of Coursera platform, the experience design strategy of online education platform from the perspective of distributed cognition was finally refined by substituting into specific contexts, as shown in Figure 3.

4. Online Learning Algorithms

4.1. Preference Function and Objective Function. For a pair of candidate jobs X_i and X_j, the preference function between them is represented by (1) and results in a real number between [0, 1]; that is,

$$F: X_i^d \times X_j^d \longrightarrow [0, 1]. \tag{1}$$

A return value close to 1 indicates that the first candidate job has a higher preference for the second, and when k different candidate jobs are given, the pairwise preference

function can be applied to each possible combination of candidate job pairs to determine the best job. Using (1), the sum of the differences of the preferences of candidate jobs X_i in the set of candidate jobs under state θ is found, $v(X_i, J_\theta)$. The job with the largest value X_i is the best job in the current candidate job set J_θ, so for a given state θ the assignment policy π in the candidate job set J_θ is shown in

$$v(X_i|J_\theta) = \sum_{X_i, X_j \in J_\theta} \left(F(X_i, X_j) - F(X_j, X_i) \right), \qquad (2)$$

$$\pi(J_\theta) = \mathrm{argmax}_{X_i \in J_\theta} v(X_i|J_\theta).$$

The two main objectives to be achieved in the experiment are minimizing the average job time of QC and the idle distance of learning; we should measure the merit of the intentional decision in this regard and choose T_n and D_n minimum as the evaluation index to establish the objective function $f(n)$. Assuming that n learner jobs have been processed using this intentional strategy, the objective function (10) is used to measure the performance of this intentional decision.

$$f(n) = \omega_T \times T_n + \omega_D \times D_n, \qquad (3)$$

where T_n denotes the average job time of QC processing n learners, D_n denotes the average idle distance of learning to transport n learners, and ω_T and ω_D are the respective weights. Here T_n and D_n are calculated as follows:

$$T_n = \frac{Q}{n} \times (t_n - s),$$

$$D_n = \frac{1}{n} \sum_{q \in Q} \sum_{J \in F_{q,n}} e_J, \qquad (4)$$

where Q is the set of QCs, t_n is the job time of the QC when the nth job is completed, s is the start time F_q of all jobs, n is the set of jobs completed by the QC up to time t_n, and e_J is the idle distance of learning for job J. ω_T and ω_D can change depending on the relative importance of the two objectives.

4.2. Training and Evaluation Function of the Sample. After each assignment of learner assignments to learn using the above preference function-based intentional policy, the preference function is updated for the next round of learner assignments. That is, each time an intention decision is completed, a short over-the-top simulation is performed to evaluate each candidate job, assuming that the assignment is made only among k candidate jobs, treating each job in k as the best job once, and then performing a simulation in which the assignment rule ω remains unchanged, and after over-the-top simulation of k candidate jobs, the following objective function is used to evaluate the candidate jobs and obtain an evaluation matrix to find the initial job with the smallest evaluation function in the evaluation matrix [23]. In the first learning assignment, since there is no sample for learning, we use a random strategy for training sample acquisition.

The evaluation function in the over-the-top simulation process is as follows:

$$f(n + k) = \omega_T \times T_{n+k} + \omega_D \times D_{n+k}, \qquad (5)$$

where T_{n+k} and D_{n+k} are calculated as follows:

$$T_{n+k} = \frac{t_{n+k} - s}{\min_{q \in Q} |F_{q,n+k}|},$$

$$D_{n+k} = \frac{1}{n+k} \sum_{q \in Q} \sum_{J \in F_{q,n+k}} e_J, \qquad (6)$$

where T_{n+k} is calculated in the same way as D_n, but T_{n+k} is obtained by calculating the average assignment time of the QC that handles the least number of learner assignments, because if equation (5) is used to calculate it in a short-term over-the-top simulation, it will make the QC go first to those learners with shorter assignments to get a smaller average QC assignment time, but this bias towards learners with shorter assignments bias would cause learner assignments with longer processing times to not be selected until the end, which would eventually lead to delays in assigning QC for these assignments. So, using equation (6) for evaluation would drive a relatively average job schedule for all QCs.

The initial job $J*$ with the smallest evaluation function in the evaluation matrix is found, and then the best job $J*$ is paired two-by-two with other candidate jobs in k to generate $(k-1)$ training samples. Since the total number of $(k-1)$ training samples generated is usually not enough for the model to reliably learn the entire new function, the online preference learning algorithm maintains a set of recent sample pools T, and these new training samples have been accumulated in the sample pool T [24]. All samples in this pool T then learn new preferences and thus update the preference function for the next job assignment.

5. Online Learning Based on Improved Convolutional Neural Networks

The preference selection between candidate jobs can be attributed to a multiclassification problem, and the preference function is learned by an improved convolutional neural network using an incremental learning strategy [25]. The improved convolutional neural network model is shown in Figure 4, which combines InceptionNet and traditional convolutional neural network to construct a new classification model in terms of network structure and parameter optimization. First, each job is compared two by two into a 15×7 matrix input to obtain the first layer, and three 15×16 convolutions are obtained by feature transformation through three different branches of 1×3 convolution, 1×5 convolution, and 1×7 convolution, followed by an aggregation operation to merge the results of the three branches to obtain the final filter as the input of the next convolutional layer [26]. According to the same logic as the previous layer, three 1×3 convolutions, 1×5 convolutions, and 1×7 convolutions are transformed, respectively, and three 8×32 convolutions are obtained through the global average pooling layer, and, finally, the results of the three branches are combined on the output channel through dimensional aggregation to output a vector of length 6, because there are
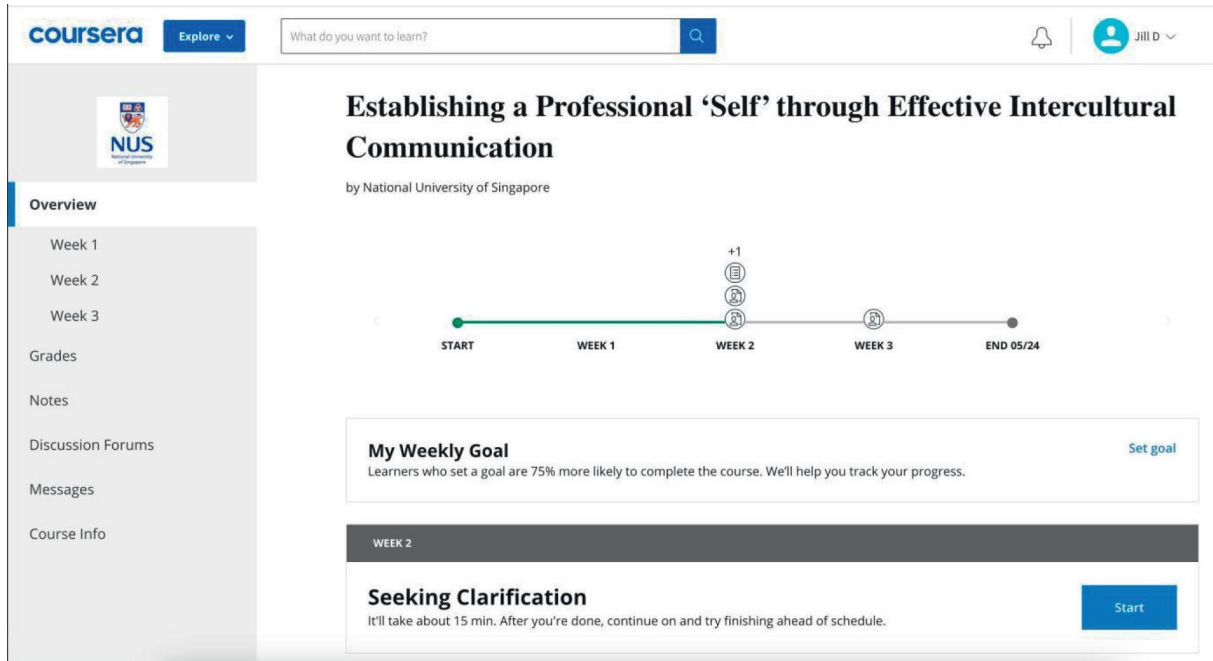
Figure 2: Coursera platform home page and course selection page.
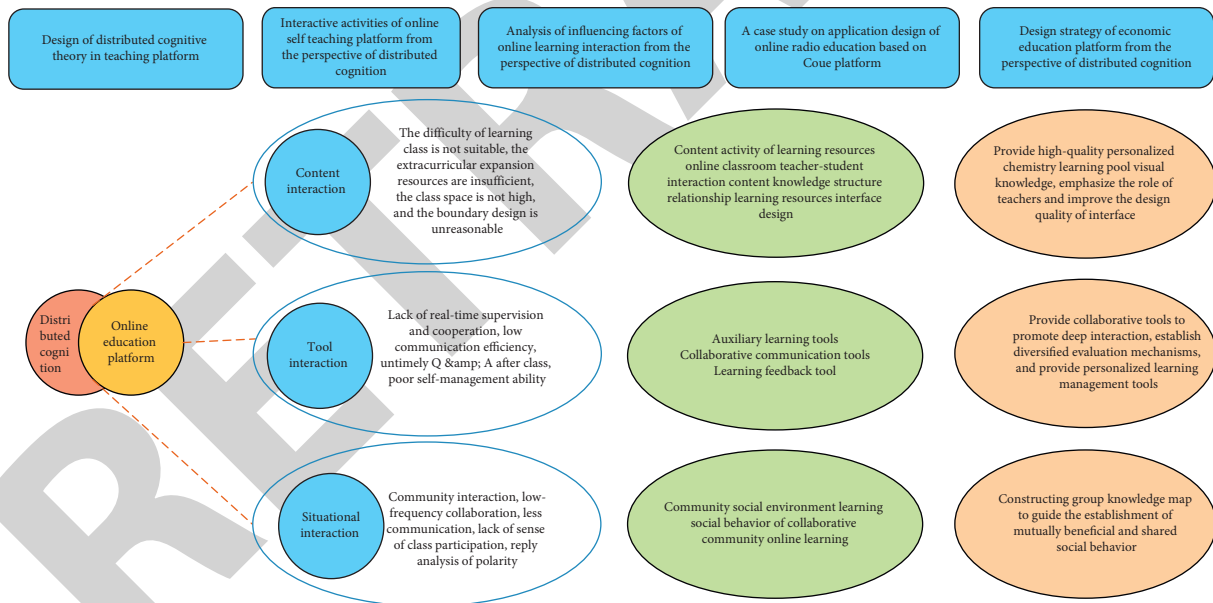
Figure 3: Experience design models of online education platform from the perspective of distributed cognition.

six operations for classification at the same time, and the activation function is adopted as the Softmax function is used. Compared with the traditional convolutional neural network, the modified convolutional neural network based on InceptionNet [27] increases the number of network layers by adding three 1×3 convolutions, 1×5 convolutions, and 1×7 convolutions in front of each convolutional layer but removes the final fully connected layer and uses the global average pooling layer instead of the fully connected layer, which in turn reduces the number of parameters, increases the operation speed, and reduces overfitting. The size of the convolution kernels in the network varies, which increases the adaptability of the neural network to different scales.

6. Research Hypothesis and Model Construction

According to the goal-directed behavior model, four variables, namely, attitude, subjective norm, expected emotion, and perceived behavioral control, were identified as the main

factors influencing desire as an intermediate variable; based on self-determination theory, feelings of competence and relatedness were identified as the other two main factors influencing desire [15, 25, 28]. The following research hypotheses were proposed in this study:

H1: attitudes have a positive impact on the desire to participate in adult online learning communities

H2: subjective norms have a positive impact on the desire to participate in adult online learning communities

H3: positive anticipatory emotions have a positive impact on the desire to participate in adult online learning communities

H4: competence has a positive impact on the desire to participate in adult online learning communities

H5: a sense of connectedness has a positive impact on the desire to participate in adult online learning communities

H6: perceived behavioral control has a positive effect on the desire to participate in adult online learning communities

H7: perceived behavioral control has a positive effect on adults' willingness to participate in online learning communities

H8: desire has a positive impact on willingness to participate in adult online learning communities

Based on the above hypothetical relationships, I constructed a theoretical model of the mechanisms underlying adults' willingness to participate in online learning communities as shown in Figure 5.

7. Experimental Scenarios and Analysis of Results

7.1. Experimental Scenes. The experiment is implemented in Python language, and the computer parameters for the experiment are Intel(R) Core(TM) i7-7700HQ CPU @ 2.80 GHz, 16 G RAM, Windows 10.

Suppose that there are 6 QCs, the number of learner yards is 14, there are 900 learner jobs in total, learning times simulated experimentally are 12, 18, and 24, the speed $v = 4$ m/s, ω_T: $\omega_D = 1:1$, and the number of yards for the purpose of learner jobs handled by QCs obeys a normal distribution $N(\mu_i, \sigma_i^2)$. If $\mu_4 = 9$, it means that the learners handled by QC4 are stored near yard B9, and as σ_4 increases or decreases, the learners are stored in a larger or smaller range of B9, so different experimental scenarios can be simulated by changing the values of μ_i and σ_i [29].

Two experimental scenarios were designed, each containing 900 learner assignments. Scenario 1 was set up as the learners handled by each QC were mostly distributed around the nearest yard, and the learning load distances were relatively uniform and short, at which time the learner distribution parameters were set to $(\mu_1, \mu_2, \mu_3, \mu_4, \mu_5, \mu_6) = (2, 4, 6, 9, 11, 13)$, $\sigma_i = 1$. Scenario 2 was set up as the learners were widely distributed in each yard, and the learning load

distances were uneven, at which time the learner distribution parameters were set to $(\mu_1, \mu_2, \mu_3, \mu_4, \mu_5, \mu_6) = (4, 5, 6, 7, 8, 9)$, $\sigma_i = 5$; scenario 2 is shown in Figure 6.

8. Analysis of Experimental Results

8.1. Analysis of the Results of Experiment 1

8.1.1. Model Goodness-of-Fit Test. After completing validation of each variable, AMOS 24.0 was used to test the fit of the hypothesized model to the measured data. The fit indices of this model were as follows: CMIN/DF = 4.040, GFI = 0.727, AGFI = 0.671, RMSEA = 0.100, TLI = 0.839, and CFI = 0.847 (see Table 1). Among them, CFI and TLI are very close to the critical value of 0.9, and RMSEA is greater than the critical value of 0.08, indicating that there is some room for improvement of the coefficients of the fitness indicators of this model. In this study, the model fit indices were corrected by the Bollen-Stine Bootstrap formula, and the corrected fit indices were as follows: CMIN/DF = 1.318, GFI = 0.937, AGFI = 0.921, RMSEA = 0.032, TLI = 0.987, and CFI = 0.984 [30].

8.1.2. Hypothesis Testing. The hypotheses were tested by measuring the standardized path coefficients among the variables. Table 2 shows that the *P* values of H1, H2, H3, and H4 are not less than 0.001, indicating that the hypotheses do not pass the test; the *P* values of H5, H6, H7, and H8 are less than 0.001, indicating that the hypotheses pass the test. In addition, according to the magnitude of β values of the path coefficients of each hypothesis of the structural equation model, it can be seen that the effects of each variable on desire are, in descending order, relatedness, perceived behavioral control, competence, attitude, expected emotion, and subjective norm, while the effects on behavioral intention are, in descending order, perceived behavioral control and desire.

8.1.3. Mediation Effect Test. The mediation effect was examined using the nonparametric percentile Bootstrap method with the PROCESS 2.16 plug-in installed in SPSS 24.0 software. The results showed that desire mediated the relationship between perceived behavioral control, sense of relatedness, and willingness to act; desire partially mediated the relationship between perceived behavioral control and willingness to act and fully mediated the relationship between sense of relatedness and willingness to act (see Table 3).

8.2. Analysis of the Results of Experiment 2. The intention results of learning under scenario 1 and scenario 2 for 12 to 24 units of different sizes are compared.

The experimental results are shown in Table 4, where the rate of decrease of *Tn* and *Dn* gradually slows down when the learning times gradually increase from 12 to 24, because the same may be accompanied by conflicts between learning as the learning times increase.
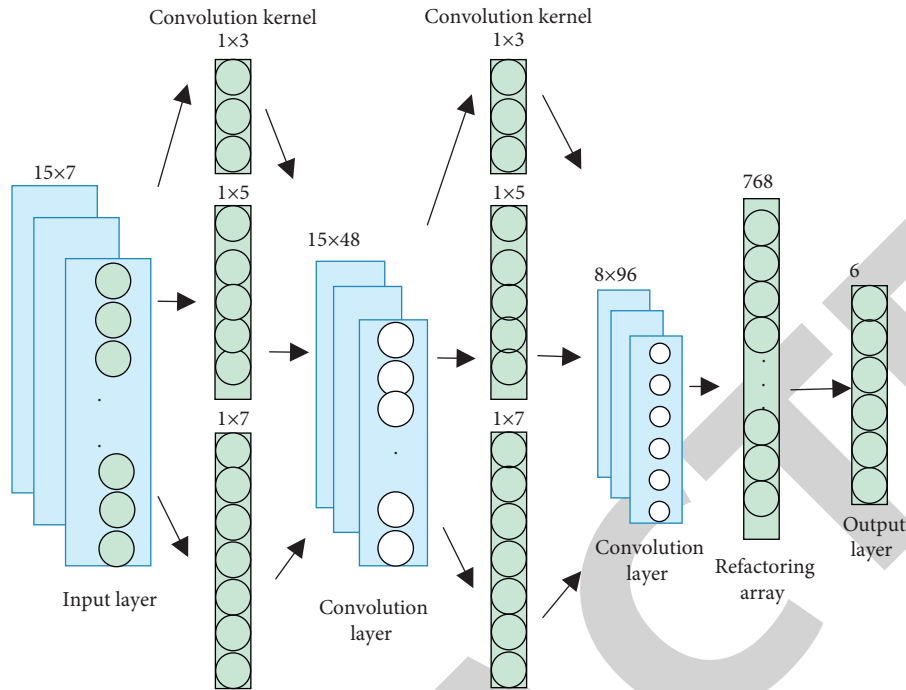
FIGURE 4: Network structure of the improved convolutional neural network.
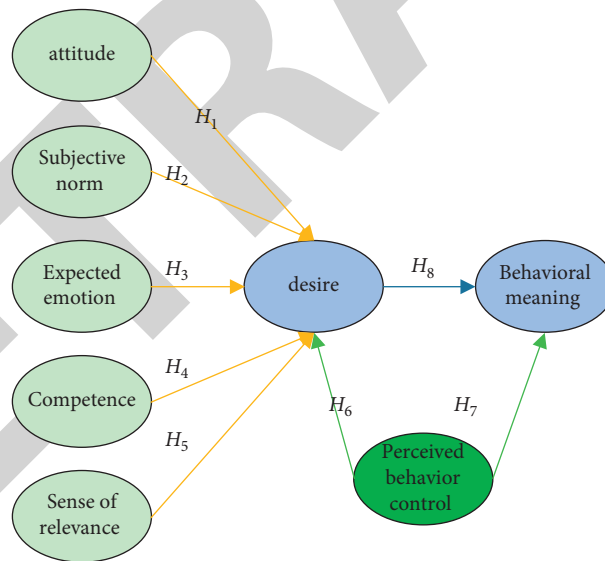
FIGURE 5: Research conceptual model.

(1) The sums under scenario 2 are higher than those under scenario 1, which is because the learners in scenario 2 are scattered among the heap fields. The dynamic degree is higher, and the learning load distance is uneven.

(2) When the numbers of learner assignments are the same, the one under scenario 1 with 18 learning runs is close to the one under scenario 2 with 24 learning runs. Therefore, the assignment strategy can be adjusted according to the information collection of QC and learning and the preference of learner assignments, thus reducing the number of learning inputs.

As can be seen from Figure 7, as the number of iterations of this paper's model increases, the students' learning intentions increase in different distributed scenarios, and the highest learning intention satisfaction can reach 138 GI. In different student interest context, this paper's model can make students interested in online learning, which is our improvement of the model, digging deeper into the data
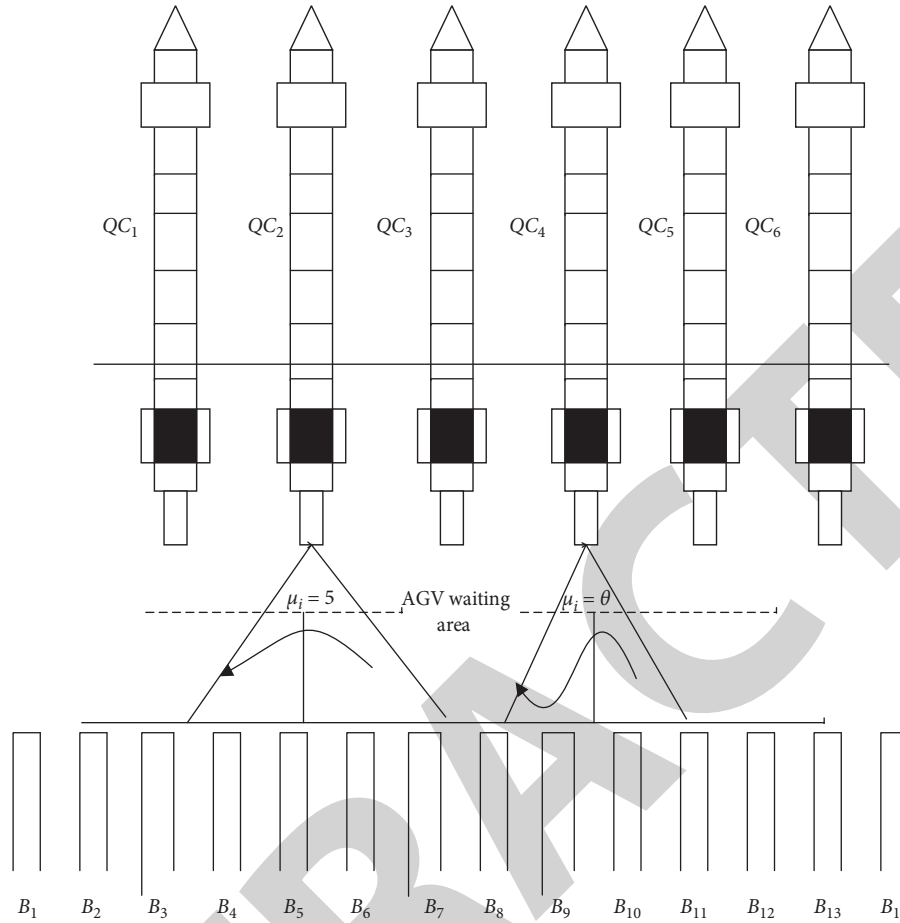
Figure 6: Automated learner dock layout.

Table 1: Model fit test indicators.

Allocation index	CMIN/DF	GFI	AGFI	RMSEA	TLI	CFI
Standard value	<5	>0.9	>0.8	<0.08	>0.9	>0.9
Index coefficient of this paper	4.040	0.727	0.671	0.100	0.839	0.847
Corrected index coefficient	0.318	0.937	0.921	0.032	0.987	0.984

features. It is also known that, according to each hypothesis of the knot model, it can be seen that the effect of each variable on desire from high to low is the sense of association, perceived behavioral control, sense of competence, attitude, expected emotion, and subjective norm, while the effect of the effect of willingness to act from high to low is the perceived behavioral control and desire.

8.3. Analysis of the Results of Experiment 3. The intentional results of using a multifeature-based online learning assignment strategy with a shortest drive assignment strategy by varying the number of learner assignments are compared.

As can be seen in Figure 8, the total completion time with the multifeature online learning assignment strategy is approximately 13% greater than that with the shortest drive assignment strategy D_n, but the difference between the multifeature online learning assignment strategy and the shortest drive assignment strategy remains approximately the same as the number of learners handled by the QC increases, D_n. The total completion time increases with the number of learners for both assignment strategies, and the total completion time with the multifeature online learning assignment strategy is about 12% shorter than the total completion time with the shortest drive assignment strategy on average, and the total completion time with the multifeature online learning assignment strategy saves about 21% when each QC handles about 225 TEUs on average.

As shown in Table 5, the *Tn* differences of QC under the two assignment strategies are 3.9 s, 26.2 s, 33.8 s, 56.4 s, and so forth, and the *Tn* differences between the online learning assignment strategy and the shortest drive assignment strategy keep increasing as the number of learners increases, indicating that the online learning assignment strategy based on multiple features can effectively reduce the average assignment time and minimize the completion time of QC.

TABLE 2: Structural equation model hypothesis testing results.

Hypothesis	Path relationship	P	Path coefficient β	Conclusion
H1	Attitude ⟶ desire	0.024	0.124	No support
H2	Subjective norm ⟶ desire	0.371	-0.048	No support
H3	Expected emotion ⟶ desire	0.266	0.062	No support
H4	Competence ⟶ desire	0.015	0.131	No support
H5	Relevance ⟶ desire	***	0.544	Support
H6	Perceived behavior control ⟶ desire	***	0.417	Support
H7	Perceived behavioral control ⟶ behavioral intention	***	0.492	Support
H8	Desire ⟶ willingness to act	***	0.425	Support

TABLE 3: Results of the mediation effect test.

		Effect coefficient	Standardization coefficient	T	P	Intermediate effect confidence interval (estimated upper limit)	Intermediate effect confidence interval (estimated lower bound)
Perceived behavior control	Total effect	0.663	0.058	11.367	***	0.548	0.777
	Direct effect	0.513	0.059	8.678	***	0.397	0.629
	Indirect effect	0.150	0.039	3.846	***	0.083	0.234
Sense of relevance	Total effect	0.197	0.051	3.861	***	0.097	0.298
	Direct effect	0.042	0.053	0.787	0.432	-0.063	0.147
	Indirect effect	0.155	0.034	4.559	***	0.099	0.233

TABLE 4: Intention to learn in different scenarios.

	Number of AGVS		
Scenario 1	12	166.74	310.63
	48	161.95	295.32
	24	159.13	279.48
Scenario 1	12	217.35	452.48
	18	184.57	429.56
	24	164.03	421.37

From Figure 9, we can know that the method in this paper helps students' intentional interest in online learning the most, and the blue intentional interest points have been distributed above the other two schemes from the beginning of the iteration. It shows that our scheme is effective. Traditional CNN is simple, but it can also dig deeper into the data, so it also has some effect.

Using structural equation modeling, this study examined the pathway relationships among adults' attitudes, subjective norms, expected emotions, competence, relatedness, perceived behavioral control, desires, and behavioral intentions to participate in online learning communities and analyzed the mechanisms that generate behavioral intentions to participate in online learning communities, focusing on the mechanisms that influence behavioral intentions. Four main conclusions were drawn from this study.

First, attitudes, subjective norms, expected emotions, and competence did not play a significant antecedent role in the process of adults' desire to participate in the online learning community. In particular, subjective norms had little effect on desire ($\beta = -0.048$), suggesting that the factors influencing adults' participation in online learning communities were more focused on the community organizational environment and self-control.

Second, a sense of connectedness has a positive impact on the desire to participate in an adult online learning community. A sense of connectedness can be viewed as an internal psychological experience for individuals, the key to which is to feel cared for and supported by other members of the organization. The acquisition of a sense of connectedness not only enhances identity and mutual understanding between individuals but also provides a sense of pleasure and satisfaction. Conversely, it can lead to feelings of isolation.

Third, perceived behavioral control affects adults' behavioral intentions not only directly but also indirectly through desires. The total effect β of perceived behavioral
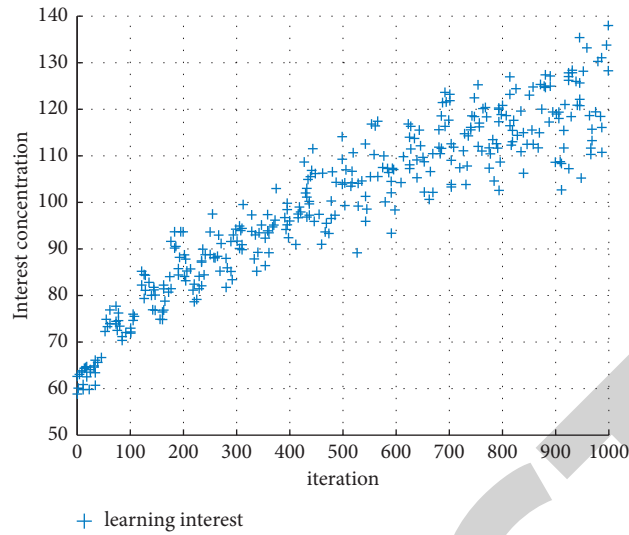
Figure 7: Student learning intention.
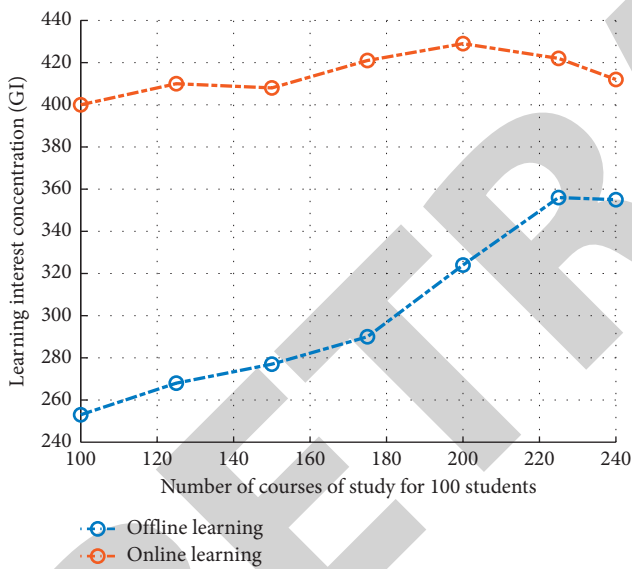
(a)
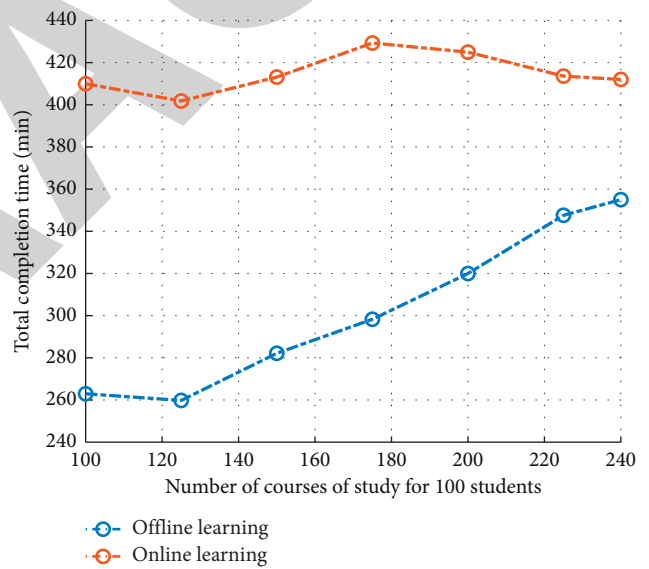
(b)

Figure 8: Intent results under different assignment rules.

Table 5: Average QC completion time under different assignment rules.

Number of containers handled by each QC	100 (s)	125 (s)	150 (s)	175 (s)	200 (s)	225 (s)
Online learning assignment	179.8	174.2	177.8	172.2	170.9	169.3
Shortest distance assignment	183.7	200.4	211.6	228.6	248.4	273.2
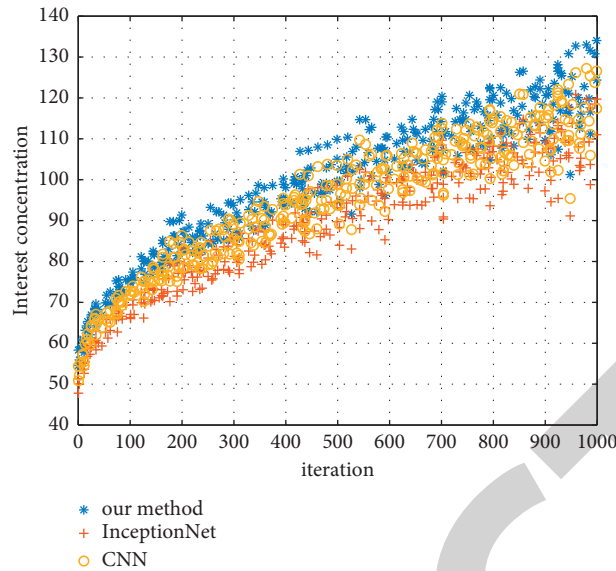
Figure 9: Comparison of learning intentions of different programs.

control on behavioral intentions reached 0.663, indicating that perceived behavioral control is an important variable influencing the generation of behavioral intentions in adults.

Fourth, adults' desire to learn has a significantly positive effect on behavioral intentions.

9. Conclusions

Given that behavioral decisions to participate in learning are mostly accompanied by clear goals and students' learning has distinctive self-directed characteristics, it is a new perspective and an important way to advance research in this field to investigate the intrinsic influencing mechanisms of adults' behavioral willingness to participate in online learning communities with the guidance of goal-oriented behavior theory and self-determination theory. Based on distributed cognition, this paper designs a convolutional neural network model based on InceptionNet to analyze the intrinsic variables influencing willingness to participate in online learning communities from a systems view perspective, mainly attitudes, subjective norms, expected emotions, competence, sense of relatedness, desire, and perceived behavioral control. The relationship between these intrinsic variables and behavioral intentions was further explored through model hypotheses, scale development, reliability analysis, and testing of research hypotheses using structural equation modeling [31].

Data Availability

The datasets used in this paper are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest regarding this work.

Acknowledgments

References

[1] C.-H. Wang, D. M. Shannon, and M. E. Ross, "Students' characteristics, self-regulated learning, technology self-efficacy, and course outcomes in online learning," *Distance Education*, vol. 34, no. 3, pp. 302–323, 2013.

[2] F. M. Sáez, C. E. Bustos, M. V. Pérez, J. A. Mella, K. A. Lobos, and A. E. Díaz, "Willingness to study, self-efficacy and causal attributions in Chilean university students," *Journal of Educational Psychology-Propósitos y Representaciones*, vol. 6, no. 1, pp. 223–245, 2018.

[3] M. Aparicio, F. Bacao, and T. Oliveira, "An e-learning theoretical framework," *An e-learning theoretical framework*, vol. 19, no. 1, pp. 292–307, 2016.

[4] M. Erez, A. Lisak, R. Harush, E. Glikson, R. Nouri, and E. Shokef, "Going global: developing management students' cultural intelligence and global identity in culturally diverse virtual teams," *The Academy of Management Learning and Education*, vol. 12, no. 3, pp. 330–355, 2013.

[5] M. Kurucay and F. A. Inan, "Examining the effects of learner-learner interactions on satisfaction and learning in an online undergraduate course," *Computers & Education*, vol. 115, pp. 20–37, 2017.

[6] S. W. van Rooij and K. Zirkle, "Balancing pedagogy, student readiness and accessibility: a case study in collaborative online course development," *The Internet and Higher Education*, vol. 28, pp. 1–7, 2016.

[7] M. A. Almaiah, M. A. Jalil, and M. Man, "Extending the TAM to examine the effects of quality features on mobile learning acceptance," *Journal of Computers in Education*, vol. 3, no. 4, pp. 453–485, 2016.

Retraction

Retracted: Application of Computer Simulation Analysis in Green City Garden Plant Landscape Design

Scientific Programming

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope

(2) Discrepancies in the description of the research reported

(3) Discrepancies between the availability of data and the research described

(4) Inappropriate citations

(5) Incoherent, meaningless and/or irrelevant content included in the article

(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] J. Zhang, "Application of Computer Simulation Analysis in Green City Garden Plant Landscape Design," *Scientific Programming*, vol. 2021, Article ID 9422417, 9 pages, 2021.

Research Article

Application of Computer Simulation Analysis in Green City Garden Plant Landscape Design

Jing Zhang 🅸🅳

Inner Mongolia Technical College of Construction, Hohhot 010070, Inner Mongolia, China

Correspondence should be addressed to Jing Zhang; 20070100233@imaa.edu.cn

As an important component of urban landscape elements, garden plants have the role of improving urban ecological environment and forming landscape effects. Digital landscape is a high-simulation model of the system combined with computer hardware and software systems, which allow analysis of the application of computer simulation in the analysis of the value of urban landscape design. In this paper, we propose a "Task-Crowd-Process-Evaluation" computer simulation analysis framework to support management decision, which can form 2D or 3D spatial data, realize real-time and statistical analysis, make use of the multidimensional, efficient, and humanized human living environment, and make the environment landscape planning more reasonable and practical. The simulation shows that the effective use of urban vegetation can not only beautify the urban landscape, but also play the function of rainwater utilization.

1. Introduction

Modern landscape design, instead of being "scenic" as in the case of human vision, should be creatively transformed with the spirit of place to create a subversive, avant-garde, and modern landscape [1]. While urban landscapes are often based on the modeling of mountains and water to create natural and pristine landscapes, the integration of landscape plants helps to create a "new landscape" that belongs to the urban space [2]. In this picture, based on the artificial landscape in the site, in order to achieve the integration of landscape, life, and landscape, different colorful garden plants are planted in the upper space of the rockery and trees are used to make the rockery more vibrant. Around the landscape, colorful flowering plants are arranged to create a relaxed and natural courtyard atmosphere, making the landscape have a rich and natural form [3, 4].

In the design of urban plant scene, some local designs are very important, such as plant size, structure, surface, and color effect, for example, a mixture of craftsmanship, plant network guidelines, and area conditions [5]. Therefore, vegetation design is the key technology in the use area

of urban leisure center. The analysis of urban life in urban climate data, fast-paced panic sense and thought weight of artworks, and urban display climate can reduce people's physical and psychological weight, relaxation, and urban scenes and can lead individuals to engage in healthy exercise, thus enhancing the consumption of urban scenes [6]. On this premise, the design of urban vegetation landscape and urban social population structure are further analyzed, which can improve the quality of life of urban residents.

After completing the project research and forecast, computer simulation analysis can help planners in the process of completing the landscape master planning and design, with the help of certain models, quickly analyze the actual situation of the base, grasp the entire project in all aspects, and accurately control each process [7, 8]. Construction layout and implementation clearly understand the details of the process. The concept of landscape planning and design is a starting point compared with the conceptual design and landscape planning and design drawings [9, 10]. Designers use immersion computer simulation analysis to draw concept drawings for landscape planning and design [11]. In the interaction between virtual and reality, the

landscape designer puts himself in the virtual environment of landscape design and is able to find information about the design plan quickly and accurately, so that he can establish an analysis and evaluation model to master the design more accurately and improve the success rate of the design [12, 13].

Based on this, this paper proposes a "Task-Crowd-Process-Evaluation" computer simulation [14] and analysis framework to support management decision making, which can form spatial data, realize real-time and statistical analysis, and utilize the multidimensional, efficient and humanized human living environment. It makes the environmental landscape planning more reasonable and practical. Effective use of urban vegetation for the construction process of the urban landscape as well as landscaping and visual beautification optimizes the ecosystem; these functions in the landscape design, rain gardens, can achieve roof gardens and urban greening. Designated style evaluation of urban park plant landscape and how to carry out urban park plant display have hypothetical and realistic guidance.

The contributions of this article are as follows:

(1) A "Task-Crowd-Process-Evaluation" computer simulation analysis framework is proposed, which can form two-dimensional or three-dimensional spatial data, realize real-time and statistical analysis, and make use of the multidimensional, efficient, and humanized human living environment to make environmental landscape planning more reasonable and practical.

(2) In the process of individual innovation amplitude, the element dimension introduces the variable of individual innovation amplitude, and the number of decision-making dimensions for each individual adjustment problem solution is represented by IA. Individuals searching for optimal problem solutions will randomly adjust IA dimensions of their own problem solutions.

(3) We have done a lot of experiments to verify the effectiveness of the scheme in this paper, and the results can produce 3D simulation diagrams. And the solution in this article can be the guide of the designer.

1.1. Related Work. To improve the complex social living environment, urban landscape is closely related to human survival. Many have explored explicitly in the design of courtyard plants in a hypothetical premise meadow that exploits the possibilities of environmental areas, which are important to guide and regulate the use of decree aspirations [6, 15]. There is knowledge of urban botanical landscapes and recreation centers. The most effective method of making urban park botanical landscapes is resourceful initiative and specific assumptions [16]. Throughout the design, 94 examples were coordinated by determining the structure of the tree arrangement little by little [17]. A number of photographs were taken from each point at each checkpoint, and arbitrarily selected photographs were used as evaluation tests [18]. There is a tendency to show their work in various ways, namely, landscape effects. Panoramic renderings of gardens were drawn by hand or by computer [19]. It also provides a strong verbal communication bridge for the viewer in a visual way.

2. Simulation Model Construction

2.1. Basic Model. In the study of how groundsheet supports management decisions, a "Task-Crowd-Process-Evaluation" (TCPE) framework [14] for groundsheet operations was proposed. Process and evaluation are shown in Figure 1. This framework incorporates three subjects: the groundsheet issuer, the platform, and the recipient, and can fully depict the essential characteristics of the groundsheet operation process, which mainly consists of four stages:

(1) *Crowdsourcing Task Generation.* The issuer chooses a problem that the enterprise needs to solve and entrusts it to a groundsheet platform, which organizes and analyzes it to form a groundsheet task [20].

(2) *Crowdsourcing Participant Identification.* The groundsheet platform publishes the task on its website or pushes it directly to the market [21].

(3) *Crowdsourcing Participants Determination.* The groundsheet platform publishes the task on its website or pushes it directly to the problem solvers on the platform (the receiver), the groundsheet participants choose whether to accept the task, and all participants who accept the groundsheet task form a network to solve the problem [22, 23].

(4) *Participants Solve the Problem.* All participants who accept the task start to solve the problem and in the process of solving the problem, they can communicate and share knowledge with other problem solvers through the platform function [24, 25].

(5) *Crowdsourcing Performance Evaluation.* Within the specified time of completion of the groundsheet task, participants submit the final problem solution to the groundsheet platform, and the platform evaluates it according to certain criteria and requirements and sends one or several best solutions back to the issuer [26].

This study adopts a model-oriented modeling approach, combining the NK model [27] and the TCPE framework, and constructs a simulation model to model the process of groundsheet problem solving based on the four basic elements of the framework: task, quality, process, and evaluation:

(1) *Modeling of Task Elements.* If a new product consists of N parts, it is essentially an N-dimensional array of decisions. Since each decision of a task has several different choices, in order to simplify the problem space structure, two choices are usually set [11, 28], denoted by 0 and 1, so the groundsheet task has 2 N potential problem solutions and forms the problem space of the groundsheet task. The complexity of the
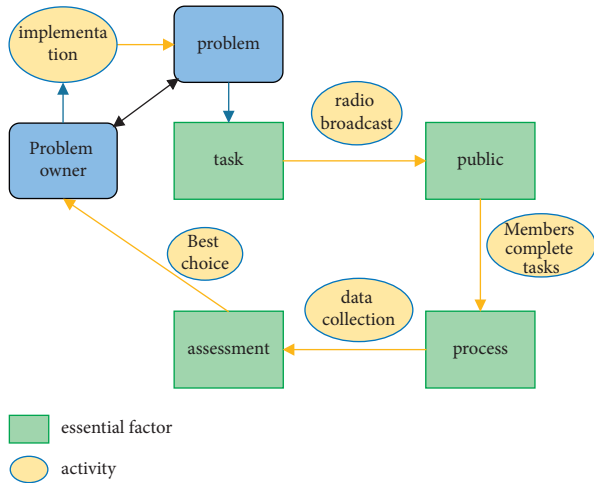
FIGURE 1: Framework of TCPE groundsheet operation process.

essential factor

activity

task is determined by the value of K. The larger the value of K, the higher the complexity of the task.

(2) *Mass Element Modeling.* When a groundsheet task is released, it is assumed that there are M individuals participating in this groundsheet activity, and each individual has to submit a problem solution. Since individual problem solutions correspond to groundsheet tasks, an N-dimensional array $O = \{O_1 O_2 \cdots O \cdots O_N\}$ is used to represent them, where $O_1 \in \{0, 1\}$.

(3) *Modeling of Process Elements.* The process is a series of activities that participants perform after receiving the task. In competition-based groundsheet, participants randomly adjust one or several dimensions of the existing problem solution through several iterations of innovative search based on their initial problem solution, and if a high-quality problem solution is formed, it replaces the old one; otherwise, it remains unchanged [29].

(4) *Modeling of Assessment Elements.* The assessment includes both individual problem solution quality assessment and groundsheet performance assessment. In general, the quality of an individual problem solution can be represented by the performance of the solution of the problem space. Assume that a decision i in a problem solution has a value of a_i. Since the contribution of each decision to the performance of the problem solution c_i is influenced by the values of other K decisions in addition to its own value [18, 30] and the values of K decisions are represented by vector a_k, the contribution of decision i to the performance of the problem solution is calculated as follows:

$$P_i = c_i (a_i, a_k). \tag{1}$$

The overall performance of a problem solution is represented by the average contribution size of N decisions [26], which is calculated as follows:

$$P = \frac{\sum_{i=1}^{N} (a_i, a_k)}{N}. \tag{2}$$

Regarding groundsheet performance assessment, this paper draws on Javadi Khasraghi's (2014) study [31], which uses the average quality of participants' solution submissions to reflect groundsheet performance. Reference [32] argues that the sender does not always select the best solution, so the average performance of participants' submissions is a robust measure to circumvent the failure to select the best solution. The calculation formula is as follows:

$$CP = \frac{\sum_{i=1}^{N} P}{N}. \tag{3}$$

2.2. New Features

(1) *Individual Innovation Amplitude.* In the process factor dimension, the individual innovation magnitude variable is introduced, which is defined as the number of problem solution decision dimensions that an individual adjusts each time and is expressed as IA. The solution of the optimal problem can be searched through individual random adjustment, and the performance of the adjusted solution can be observed for improvement. The adjusted problem solution is adopted; otherwise, it remains unchanged.

(2) *Individual Limited Rationality Level.* In the mass factor dimension, the individual limited rationality level variable is introduced. According to existing research, finite rationality is defined as an individual's partial attention to the problem space [24]. Individuals have a parsimonious space whose dimension is smaller than the problem space, and they construct the optimal parsimonious space based on their past knowledge and experience, i.e., individual cognitive representations, which reflect the source of their competitive advantage. The individual cognitive representation dimension is assumed to be N1, and the ratio of cognitive representation dimensions (N1/N) is used to represent the individual's limited rationality level (BRL).

Based on [32] modeling of the finite rationality level of individuals, each decision configuration in the cognitive representation space has one decision configuration in the groundsheet task problem space corresponding to it, and it is assumed that the fitness value of each point in the cognitive representation space is equal to the average fitness value of the points in the groundsheet task problem space corresponding to that point, and individuals choose the one with the largest fitness value in the cognitive representation space as their cognitive representation. The individual chooses one of the decision configurations with the highest fitness value as his or her cognitive representation.

(3) *Individual Limited Rationality Bias*. The variable was introduced in the mass factor dimension. Due to different regions, personal upbringing, and professional backgrounds, individuals have different knowledge and experience, i.e., different individual cognitive representations. According to the existing literature, individual limited rationality bias is defined as the degree of accuracy of cognitive representations [18, 33], which is expressed by BRB.

Generally, N1 consists of some key decision dimensions and other random decision dimensions, where the number of key decision dimensions is determined by BRB; BRB = $N1^*$(1-BRB). Individuals select $N1^*$(1-BRB) key decision dimensions from N dimensions and $N1^*$BRB random decision dimensions from the remaining (N–$N1^*$(1–BRB)-dimensions to construct cognitive representations. Among them, the key decision dimensions are identified based on the influence of each decision item in the groundsheet task, and the influence is determined by the number of other decisions influenced by each decision. If several decisions have the same influence, one decision is randomly selected as the key decision dimension to build the cognitive representation.

3. Green Urban Landscaping Design Process

3.1. Landscape Design Analysis. Through the classification of urban land use, the land types around the city can be divided into forest land, cultivated land, water area and wetland, other urban construction land, road traffic land, and other land use types. The land change can be seen from Figure 2. We can clearly know that urban land can be planned systematically. As can be seen from the data change in Figure 3, from 1995 to 2019, with the increase in urban construction area, the road network density gradually increases and the forest land area decreases. Among them, the water area decreased rapidly, by 31.93%, respectively. The construction land of other cities in Yuhang district increased by 25.96%, and the land type in Yuhang District changed little. In terms of municipal landscaping, the construction and development of urban landscaping have directly driven the rapid growth of the landscaping industry. From 2003 to 2010, the average annual growth rate was 38.34%, growing very rapidly; from 2011 to 2014, the investment in fixed assets of landscaping in Hangzhou was generally maintained at a high level.

Cities are in a period of rapid industrialization and urbanization. On the one hand, there is a huge demand for high energy consumption products such as steel and cement; on the other hand, the improvement of residents' quality of life is accompanied by the increase in living energy consumption, which makes China's carbon emission reduction face arduous challenges. Therefore, how to effectively develop low-carbon economy and balance the relationship between China's urbanization process and carbon emission will become China's sustainable development in the future. This paper takes China's urbanization and carbon dioxide emission as an example [23]. With the increasingly prominent contradiction between environmental problems and economic development, "low-carbon economy" has become a topic of increasing concern by governments and academic circles. The general situation of urbanization and carbon dioxide emission is analyzed through statistical data. As shown in Figure 4, the stability test of urbanization rate and carbon dioxide emission during 1995–2019 is carried out. The empirical test results show that there is a long-term equilibrium relationship between the two. Urbanization plays a positive role in the fluctuation of carbon dioxide. For different regions, due to the different economic structure and development status of different provinces and cities, the force of urbanization on carbon dioxide emission is also different.

From Figure 3, it can be seen that before 2000, high carbon emission areas were mainly concentrated in the central area of the main urban area. After 2000, patches gradually began to migrate to the southeast and northeast. Until 2019, the high carbon spots no longer appear in the central area of the main urban area. The study also found that high carbon showed spatial aggregation before 2000 and gradually dispersed in space after 2000.

As can be seen from Figure 5, before 2010, high carbon emission areas were mainly concentrated in the main urban areas. After 2010, patches began to migrate. In 2019, high carbon points will no longer appear in the main urban area. Such data provide a reference for our urban greening landscape design and then design the greening density of the city according to different carbon emissions [34].

As can be seen from Figure 6, carbon emissions decreased gradually every year from 1995 to 2000 and basically disappeared in the main urban area after 2000. High carbon areas began to develop gradually in the northwest and southwest. The carbon emissions of urban residents' living consumption and residents' breathing can be distributed according to the urban vegetation density. According to the carbon emission of urban land, the road network and road vegetation density can be designed. Through the progress of urban greening, even if the carbon emission of construction land in various regions increases, the vegetation density increases more, which can make the city full of plants.

4. Simulation Results

Plants have cultural quality and symbolic meanings, and landscape architecture belongs to urban infrastructure. Combining the humanistic attributes of landscape architecture in urban landscape design and reasonably applying different garden plants can improve the overall landscape design layout, enhance the cultural value of the landscape, and create a landscape that is suitable for the humanistic urban environment. In Figure 7, combining the mood and characteristics of the buildings in the urban landscape, in order to better match the architectural landscape to achieve the effect of cultural nurturing, attention is paid to the biodiversity of the environment when incorporating garden plants, the essence of mountain and forest scenery is inherited and developed, and flowers or shrubs with brighter colors are combined with some taller and more luxuriant native plants, forming a beautiful landscape picture with

(a)

(b)

(c)

(d)

(e)

N

0   8 km

■ Unused land

■ Residential land

■ Forest land

■ Water area

■ Cultivated land

Figure 2: Urban land use change of (a) 1995; (b) 2000; (c) 2005; (d) 2010; (e) 2015.

Figure 3: Land use change by urban wards from 1995 to 2019.

Figure 4: Spatial variation of urban carbon emission intensity, 1995–2019.

Carbon dioxide emission density/ (kh/m²)

▬ <0.78	▬ 1.70–3.68
▬ 0.78–1.19	▬ >3.68
▬ 1.19–1.70	

(a)

Carbon dioxide emission density/ (kh/m²)

▬ <0.78	▬ 1.70–3.68
▬ 0.78–1.19	▬ >3.68
▬ 1.19–1.70	

(b)

Figure 5: Continued.

(c)

Figure 5: Spatial variation of urban carbon sink density (1995, 2010, and 2019).

Figure 6: Carbon emissions changes of 1995–2019.

landscape and reflect the vitality that changes in the four seasons. In this urban landscape design, with the help of tree species composition and topographic changes, regular planting techniques are used, and evergreen herbaceous plants are selected for landscape ground cover, which can further improve the ornamental effect while enhancing the urban green space coverage [25]. On the basis of color block shaping space, we follow the principle of locality and choose seasonal plants to match in the rear view treatment, which helps to form the landscape phenomenon of changing seasons and show the landscape characteristics of regular order, making the overall space more dynamic and interesting.

As shown in Figure 8, the scene simulated in this paper has a shaded grass and trees of different colors, where different trees and plants are set against each other, allowing people to enjoy the visual changes while enjoying the elegant scenery. Specifically, different planting techniques are used, and the main form of planning is the art of arrangement so that the plants become lines of points and form several corridors of sight lines with large areas of open flat land. In order to strengthen the sense of ritual and mystery of the space and to give full play to the effect of height and undulation, plants are planted on the slope with a certain horizontal height, and the difference of height and perspective links the space into a unified whole, creating a richly layered landscape space. The use of garden plants to create ornamental attractions is the value of garden plants in urban landscape design highlights.

Plant elements can be combined with other spatial elements to create interesting spatial relationships, forming a

poetic sentiment [15]. It allows visitors to experience the cultural heritage of the city while traveling and sightseeing.

Seasonality is an important characteristic of garden plants. Different plants will show different states in different seasons, and the same plants will show different forms in different seasons. Therefore, in urban landscape design, designers select different kinds of plants for landscape creation to increase the ornamental nature of the urban

FIGURE 7: Urban vegetation landscapes.

FIGURE 8: Vegetation generation pattern.

visually impeccable landscape. In the picture, the central fountain is the key landscape in the urban landscape. In order to alleviate the visual monotony and tedium, the designer has constructed a geometrically interesting plant landscape directly in front of it to create a fresh environment. In addition, a large number of courtyard plants with different colors are also incorporated around it, and the main planting form is column planting, creating the visual effect of a romantic garden together with the central fountain landscape, forming a three-dimensional spatial hierarchy of high and low staggering.

5. Conclusions

Digital information is particularly important for landscape design, allowing for human-computer interaction. In this paper, we propose a "Task-Crowd-Process-Evaluation" computer simulation and analysis framework to study how groundsheet can support management decisions, which can form two-dimensional or three-dimensional spatial data, realize real-time and statistical analysis, and make use of the multidimensional, efficient, and humanized human living environment to make the environmental landscape planning more reasonable and practical. Reasonable application of different garden plants can enhance the overall landscape design layout, enhance the cultural value of the landscape, and create a suitable landscape with the human urban environment.

Data Availability

The raw data supporting the conclusions of this article are made available by the author, without undue reservation.

Conflicts of Interest

The author declares that there are no conflicts of interest regarding this work.

References

[1] H. U. Qingqing, "Application of Shanshui culture in campus landscape design," *Journal of Landscape Research*, vol. 7, no. 6, p. 72, 2015.

[2] A. C. Carmona-Ribeiro and B. Nascimento Carboni, "Mina Klabin and modern landscape design in Brazil," *Studies in the History of Gardens & Designed Landscapes*, vol. 39, no. 2, pp. 154–174, 2019.

[3] Z. Wang and Y. Gao, "Application of artificial intelligence and Zen space in modern landscape design and topology optimization," in *Proceedings of the 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC)*, pp. 173–176, IEEE, Coimbatore, India, July 2020.

[4] J. Jiao, H. Liu, and N. Zhang, "Research on the urban landscape design based on digital multimedia technology and virtual simulation," *International Journal of Smart Home*, vol. 10, no. 9, pp. 133–144, 2016.

Research Article

Point Set Registration Based on Improved KL Divergence

Guangfu Qu and Won Hyung Lee

Computer Game/Culture Technology Lab, Graduate School of Advanced Imaging Science, Multimedia & Film,
Chung-Ang University, Seoul 06974, Republic of Korea

Correspondence should be addressed to Guangfu Qu; quguangfu@shu.edu.cn

A point set registration algorithm based on improved Kullback–Leibler (KL) divergence is proposed. Each point in the point set is represented as a Gaussian distribution. The Gaussian distribution contains the position information of the candidate point and surrounding ones. In this way, the entire point set can be modeled as a Gaussian mixture model (GMM). The registration problem of two point sets is further converted as a minimization problem of the improved KL divergence between two GMMs, and the genetic algorithm is used to optimize the solution. Experimental results show that the proposed algorithm has strong robustness to noise, outliers, and missing points, which achieves better registration accuracy than some state-of-the-art methods.

1. Introduction

Point set registration is an important content of computer vision and pattern recognition, which is widely used in image registration [1], target recognition [2], image processing [3], etc. The task of point set registration is to find the corresponding relationship between the point patterns from two different point sets and solve the transformation parameters from one point set to another. However, in actual cases, the noises, outliers, and missing points exist, resulting in a sharp drop in the accuracy of point set registration. These problems have become a hot but difficult problem in current computer vision and pattern recognition field. Therefore, it is urgent to study and develop robust and high-precision point set registration algorithms.

Point set registration algorithms can be basically divided into two categories. One is to find the correspondence relationship by estimating the space transformation parameters between two point sets. This type of algorithm mainly includes iterative closest point (ICP) [4], the thin-plate spline [5], etc. Another type is based on point features. The transformation relationship between point sets is found through the point feature recognition. This type of algorithm mainly includes the methods based on invariant features [6], the methods based on shape context [7], etc., which are validated to be robust to noise, outliers, and missing points. The ICP algorithm is one of the most basic and most commonly used point set registration algorithms. The basic principle of the ICP algorithm is to find the correspondence between the points in the two point sets according to the nearest neighbor criterion and calculate the average point pair between the two point sets. The transformed point set is used as the input. The square error is calculated between two point sets. The ICP algorithm is a simple and intuitive point set registration method. When the initial positions of the two point sets are not much different and there is a clear "one-to-one" correspondence between the two point sets, the registration accuracy of the ICP algorithm is higher. However, in the presence of outliers, missing points, and noise, the accuracy of the ICP algorithm drops sharply. Some researchers have proposed some improved algorithms for the problems of the ICP algorithm [8–10], but they basically require the initial positions between the two point sets to be relatively close.

Different from the requirement for clear "one-to-one" correspondence, a registration algorithm with a "one-to-many" correspondence was proposed in [11]. The corresponding relationship between the two point sets is constructed into a matrix, and the solution of the optimization

problem is sought by the method of simulated annealing. Andriy and Xubo improved the algorithm in [11], using the expectation-maximization (EM) algorithm to solve the correspondence between transformation parameters and points. At the same time, the Gaussian radial basis function (GRBF) was employed instead of the thin-plate spline model to represent the nonrigid change of the point set. Experiments show that this method achieved better robustness in nonrigid registration. In recent years, under the framework of EM algorithm optimization, some robust point set registration algorithms have been proposed [10, 12, 13]. This type of algorithm models the two point sets as Gaussian mixture models (GMMs). Afterwards, the problem of finding the corresponding relationship between the two point sets can be regarded as the problem of maximizing the posterior probability. This kind of algorithm has good robustness to noise, missing points, and outliers with high registration accuracy. However, in this type of algorithm, the accuracy of the transformation parameters between the point sets is affected by the point set modeling. In response to these shortcomings, some registration algorithms that do not need to solve the correspondence between points were proposed [12, 14–17]. These methods also model the two point sets as GMM, respectively, and then the point set registration problem is transformed into the problem of finding the difference between two GMMs. The corresponding parameter when the difference between the two GMMs is the smallest is adopted as the transformation parameter of the point sets. Experiments have proved the robustness of this type of algorithm to noise, outliers, and missing points, but the accuracy of the algorithm will decrease when there are more noise and outliers. In addition, in synthetic aperture radar (SAR) images, non-uniform GMM was used to model the edge point set of water bodies, and high-precision registration of water bodies was achieved [18–20].

Based on the idea of robust point set registration algorithm, combined with the statistics of point set, this paper proposes a point set registration algorithm based on improved Kullback–Leibler (KL) divergence. The proposed method models the two point sets to be registered as GMMs, respectively, where each point in the point set is represented as a Gaussian distribution. The mean value is the position coordinate of the point, and the size of the root variance represents the influence of the surrounding points on this point size, using equal weight coefficients to mix Gaussian distribution. Based on the above modeling, the point set registration problem is transformed into a problem that minimizes the improved KL divergence between two GMMs. The transformation parameters of the registration are solved by genetic algorithm. Experiments have proved that the proposed method has good robustness to noise, outliers, and missing points in comparison with some existing point registration algorithms.

2. GMM of Point Set

In the point set registration process, it is necessary to fix a scene point set to be used as a registration template and another point set as a model set to perform registration according to changing transformation parameters. For the model set m and the scene set s, the numbers of points in the two point sets are usually different due to the presence of noise, outliers, and missing points. Assuming that the numbers of points in the point sets m and s are M and S, respectively, the GMMs are established for the two point sets as follows:

$$
\begin{aligned}
p(x|\theta) &= \sum_{i=1}^{M} \alpha_i \phi(x|\mu_i, \Sigma_i), \\
p(y|\Theta) &= \sum_{j=1}^{S} \beta_j \phi(y|\nu_j, \Gamma_j),
\end{aligned}
\tag{1}
$$

where $p(\mathbf{x}|\theta)$ and $p(\mathbf{y}|\Theta)$ represent the GMMs with the parameters θ and Θ, respectively, in which θ includes α_i, μ_i, and Σ_i and Θ includes β_j, ν_j, and Γ_j; $\phi(x|\mu_i, \Sigma_i)$ and $\phi(y|\nu_j, \Gamma_j)$ represent the ith and jth Gaussian distributions in the two GMMs, respectively, in which μ_i and ν_j are the corresponding mean values and Σ_i and Γ_j are the corresponding covariance matrices; and α_i and β_j are the mixing coefficients in GMMs.

In order to express the relationship between the points, μ_i and ν_j are denoted as the mean values and Σ_i and Γ_j are denoted as the covariance matrix $\Gamma_j = \Sigma_i = \sigma^2 I$. The Gaussian distribution can be modeled as follows [15]:

$$
\begin{aligned}
\phi(x|\mu_i, \Sigma_i) &= \frac{1}{2\pi\sigma^2} \exp\left\{ -\frac{\|x - \mu_i\|^2}{2\sigma^2} \right\}, \\
\psi(x|\nu_j, \Gamma_j) &= \frac{1}{2\pi\sigma^2} \exp\left\{ -\frac{\|x - \nu_j\|^2}{2\sigma^2} \right\},
\end{aligned}
\tag{2}
$$

where σ^2 is the variance of the Gaussian distribution and $\|\cdot\|$ calculates the norm of the vector.

Each point in the two point set is modeled as a Gaussian distribution as equation (2). M and S represent the numbers of points in the two point sets, which are also the numbers of GMM components. The above formula is formed for two-dimensional point sets, but it can be directly extended to three-dimensional in mathematics. The variance σ^2 in each Gaussian distribution indicates the influence of points in the neighborhood. The larger the variance, the greater the influence from the surrounding points. On the contrary, a smaller variance indicates lower influences. The mixing coefficients α_i and β_j are determined in the equal forms, i.e., $\alpha_i = 1/M$, $\beta_j = 1/S$.

3. Point Set Registration Based on Improved KL Divergence

3.1. Improved KL Divergence between Two GMMs. After getting the GMMs of two point sets, the registration problem is transformed into a minimization problem of optimally solving the difference between GMMs. In [21], the difference between GMMs is evaluated using a combination of KL divergence and distance to move. In [22], the KL divergence is used to calculate the difference between two GMMs. Since KL divergence does not satisfy symmetry and triangle inequality and is not a true distance measure, this paper employs an improved KL divergence, i.e., symmetry KL (SKL), to calculate the difference between two GMMs.

For two GMMs $p(x|\theta)$ and $q(x|\Theta)$, which are denoted as $p(x)$ and $q(x)$ in the following, the SKL divergence between them is calculated as follows:

$$
\mathrm{SKL}(p(x), q(x)) = \frac{1}{2}\{\mathrm{KL}(p(x), q(x)) + \mathrm{KL}(q(x), p(x))\},
\tag{3}
$$

where $\mathrm{KL}(p(x), q(x))$ represents the SKL divergence between the GMM models $p(x)$ and $q(x)$. The core of equation (3) is to calculate the SKL divergence between two GMM models $p(x)$ and $q(x)$. Because there is no analytical expression for the SKL divergence, a numerical approximation method is required for calculation, and a matching approximation method is used in this paper.

For two GMM models $p(x)$ and $q(x)$, they contain M and S Gaussian distributions as follows:

$$
\begin{aligned}
p(x) &= \sum_{i=1}^{M} \alpha_i f_i(x), \\
q(x) &= \sum_{j=1}^{S} \beta_j g_j(x).
\end{aligned}
\tag{4}
$$

The SKL divergence is given by the following equation:

$$
\mathrm{SKL}(p(x), q(x)) = \int p(x)\ln\frac{p(x)}{q(x)}\mathrm{d}x + \int q(x)\ln\frac{q(x)}{p(x)}\mathrm{d}x.
\tag{5}
$$

According to the properties of the convex function and substituting equation (4), the following inequality [23] can be obtained:

$$
\mathrm{SKL}\left(\sum_{i=1}^{M} \alpha_i f_i \,\middle\|\, \sum_{j=1}^{S} \beta_j g_j\right) \le \sum_{i,j} \alpha_i \beta_j \mathrm{SKL}\left(f_i \,\|\, g_j\right).
\tag{6}
$$

According to the principle of matching approximation, it can be further obtained as follows:

$$
\begin{aligned}
\mathrm{SKL}(p\|q) &\approx \sum_{i=1}^{n} \alpha_i \log \alpha_i f_i + \sum_{j=1}^{n} \beta_j \log \beta_j g_j \\
&\quad - \sum_{i=1}^{n} \alpha_i \max_{j} \int f_i \log \beta_j g_j \\
&\quad - \sum_{i=1}^{n} \beta_i \max_{j} \int g_i \log \alpha_j f_j \\
&= \sum_{i=1}^{n} \alpha_i \min_{j}\left(\mathrm{SKL}\left(f_i \,\|\, g_j\right) + \log\frac{\alpha_i}{\beta_j}\right).
\end{aligned}
\tag{7}
$$

According to the approximation of the equation (7), it can be seen that $p(\mathbf{x})$ and $q(\mathbf{x})$ have a mapping relationship $\pi: \{1, 2, \cdots, n\} \longrightarrow \{1, 2, \cdots, m\}$. In this paper, the corresponding relationship is described as follows:

$$
\pi(i) = \arg\min_{j}\left(SKL\left(f_i\|g_j\right) - \log \beta_j\right).
\tag{8}
$$

The corresponding relationship is substituted, and the simplification can be obtained as follows:

$$
\mathrm{SKL}_{\mathrm{match}}(f\|g) = \sum_{i=1}^{n} \alpha_i\left(\mathrm{SKL}\left(f_i\|g_{\pi(i)}\right) + \log\frac{\alpha_i}{\beta_{\pi(i)}}\right).
\tag{9}
$$

According to the chain rule of divergence [24, 25], the above formula can be further simplified as follows:

$$
\mathrm{SKL}_{\mathrm{match}}(p, q) = \sum_{i=1}^{M} \mathrm{SKL}\left(f_i, g_i\right) - \log\left(C_\pi\right),
\tag{10}
$$

where $C_\pi = \sum_{i=1}^{M} \beta_{\pi(i)}$. The above equation can be seen as an approximate calculation for the matching of SKL divergence between two GMMs.

3.2. Point Set Registration Algorithm. Based on the calculation of the SKL divergence between two GMMs, a transformation $T = \{R, s, t\}$ is given to the model set s. The transformed point set is expressed as $T(s)$. Then, the point set registration problem is to find the minimum value of the SKL divergence between the point set m and the point set s, namely, $\arg\min \mathrm{SKL}(m, T(s))$. In this paper, the genetic algorithm is used to solve the above optimization problem. The process of point set registration algorithm can be divided into the following steps:

Step 1: give an initial transformation to the point set s $R = I, s = 1, t = 0$.

Step 2: calculate the SKL divergence between the two GMMs and compare it with the preset threshold. If it is

(a)

(b)

(c)

(d)
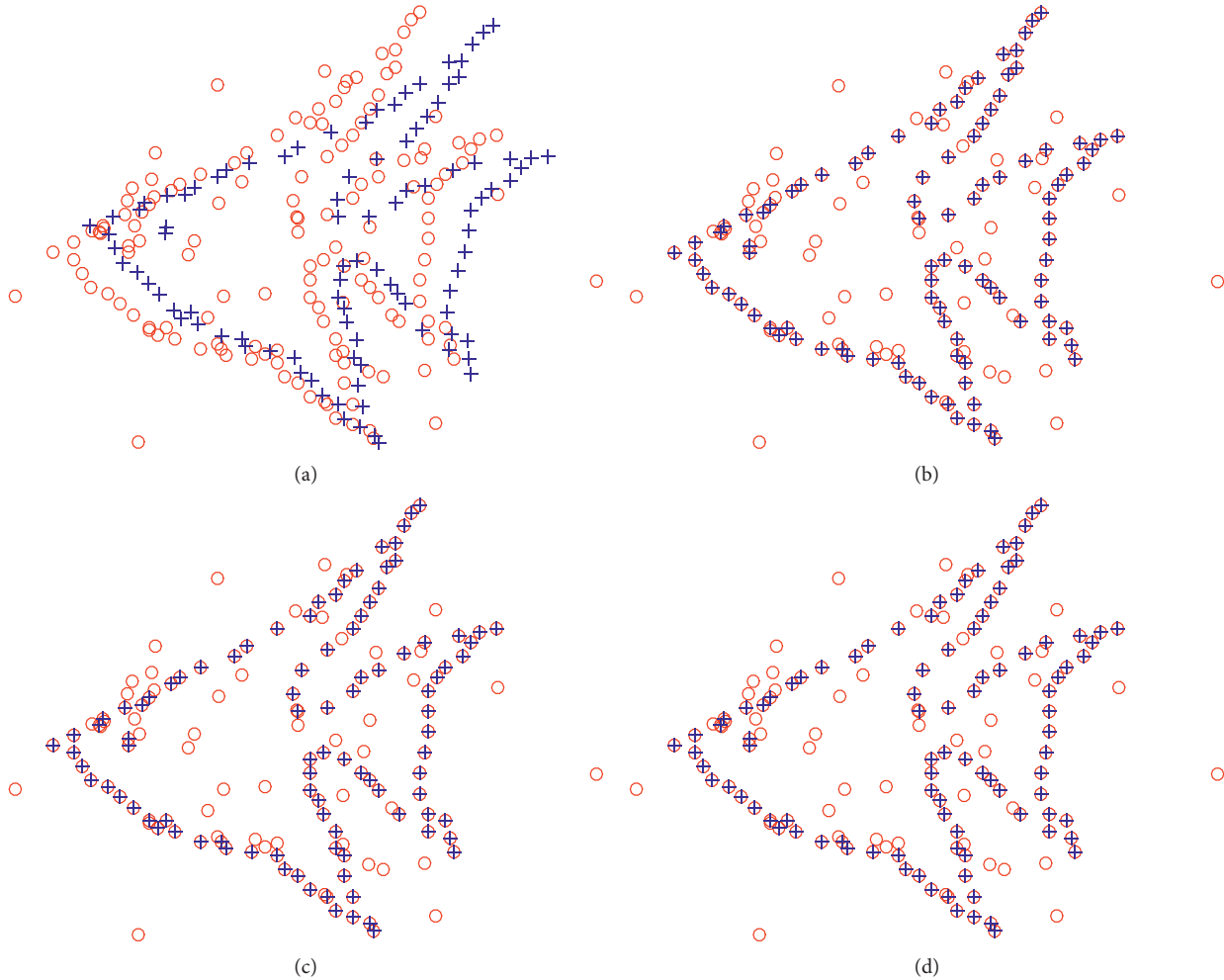
Figure 1: Registration results of different algorithms at 20% outliers. (a) Original point set. (b) Results of KC. (c) Results of CPD. (d) Results of the proposed method.

greater than the threshold, go to step 3. If it is less than the threshold, go to step 4.

Step 3: the genetic algorithm is used to solve the problem $\arg \min \mathrm{SKL}\,(m, T\,(s))$ with the optimization parameters.

Step 4: the point set registration result is obtained according to the optimized parameters.

4. Experiment and Analysis

In order to analyze and compare the robustness and effectiveness of the point set registration algorithm proposed in this paper, experiments are carried out using the point set data in the Chui dataset. At the same time, the results of rigid body point set registration based on SKL divergence are compared with ICP algorithm [4], kernel correlation (KC) algorithm [15], and coherent point drift (CPD) [11], respectively. The registration success rate (RSR) and MSE are used as indexes to measure registration accuracy.

The "fish" point set data in the Chui dataset are used as the template point set, and random outliers and noise with a ratio of 0.2 and 0.8 are added to the sample point set,

respectively. SKL algorithm and ICP are used for the template point set and the sample point set. The algorithm, KC algorithm and CPD algorithm are used for registration experiments. Because the registration results of the ICP algorithm are too poor when there are many outliers, the experimental results are not listed. The comparison results of the other algorithms are shown in Figures 1–4. It can be seen from the results that when the outliers and noise are not serious, the equation registration accuracy of CPD, KC, and SKL algorithms is relatively higher. However, when the outliers and noise are severe, the registration accuracy of CPD is poor. Both KC and SKL have better registration accuracy, and the registration result of SKL in this paper is better than that of KC.

In order to quantitatively compare the registration results of SKL, KC, CPD, and ICP, the experiments are conducted under "no outlier or noise" and different levels of interferences of 20%, 40%, 60%, 80%, and 100%. The registration success rate and MSE value of different registration algorithms are calculated. The results are shown in Tables 1 and 2. From the experimental results, it can be seen that when the ratio of outliers and noise is less than 60%, the

(a)

(b)

(c)

(d)
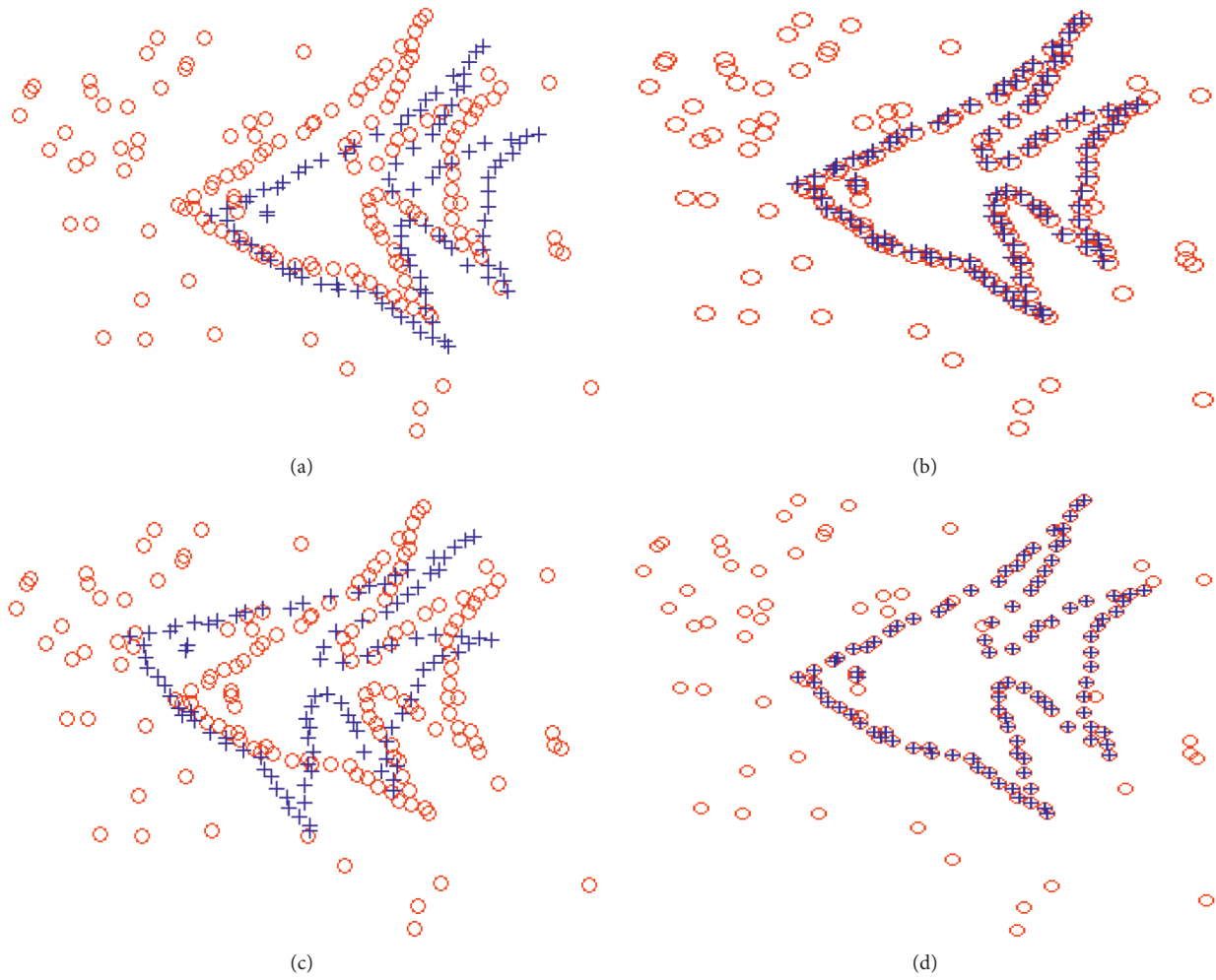
FIGURE 2: Registration results of different algorithms at 80% outliers. (a) Original point set. (b) Results of KC. (c) Results of CPD. (d) Results of the proposed method.

(a)

(b)
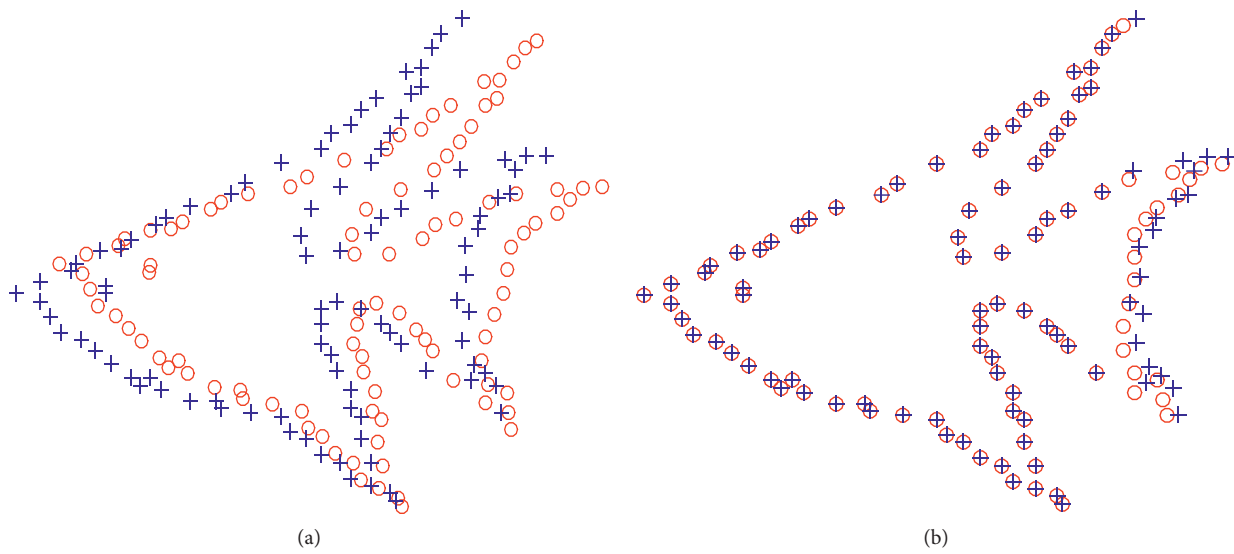
FIGURE 3: Continued.

(c)

(d)
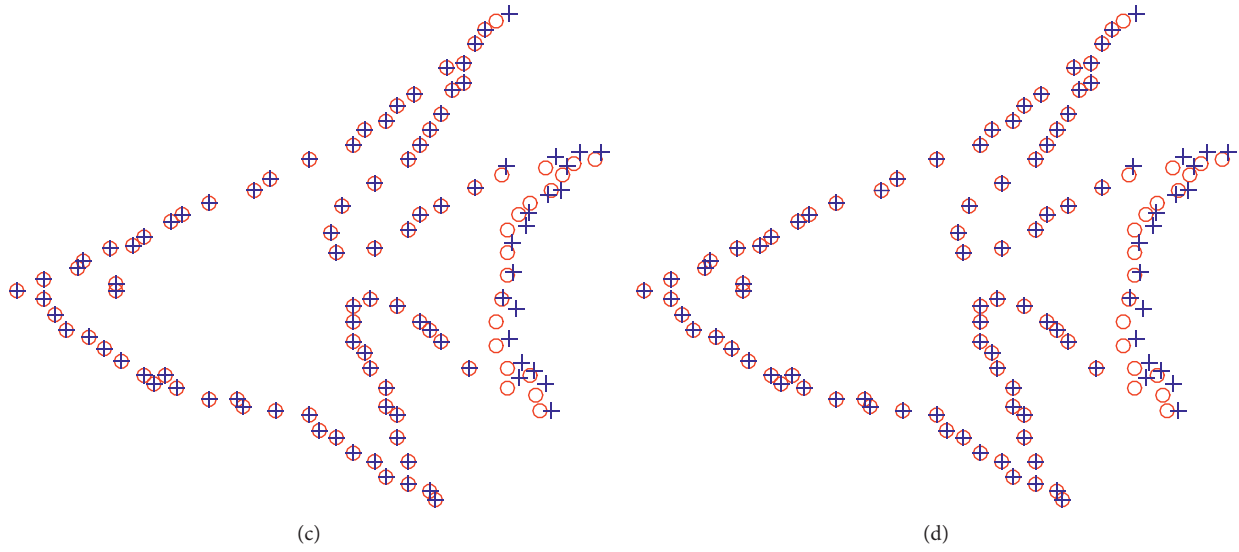
Figure 3: Registration results of different algorithms at 20% noise. (a) Original point set. (b) Results of KC. (c) Results of CPD. (d) Results of the proposed method.

(a)

(b)

(c)

(d)
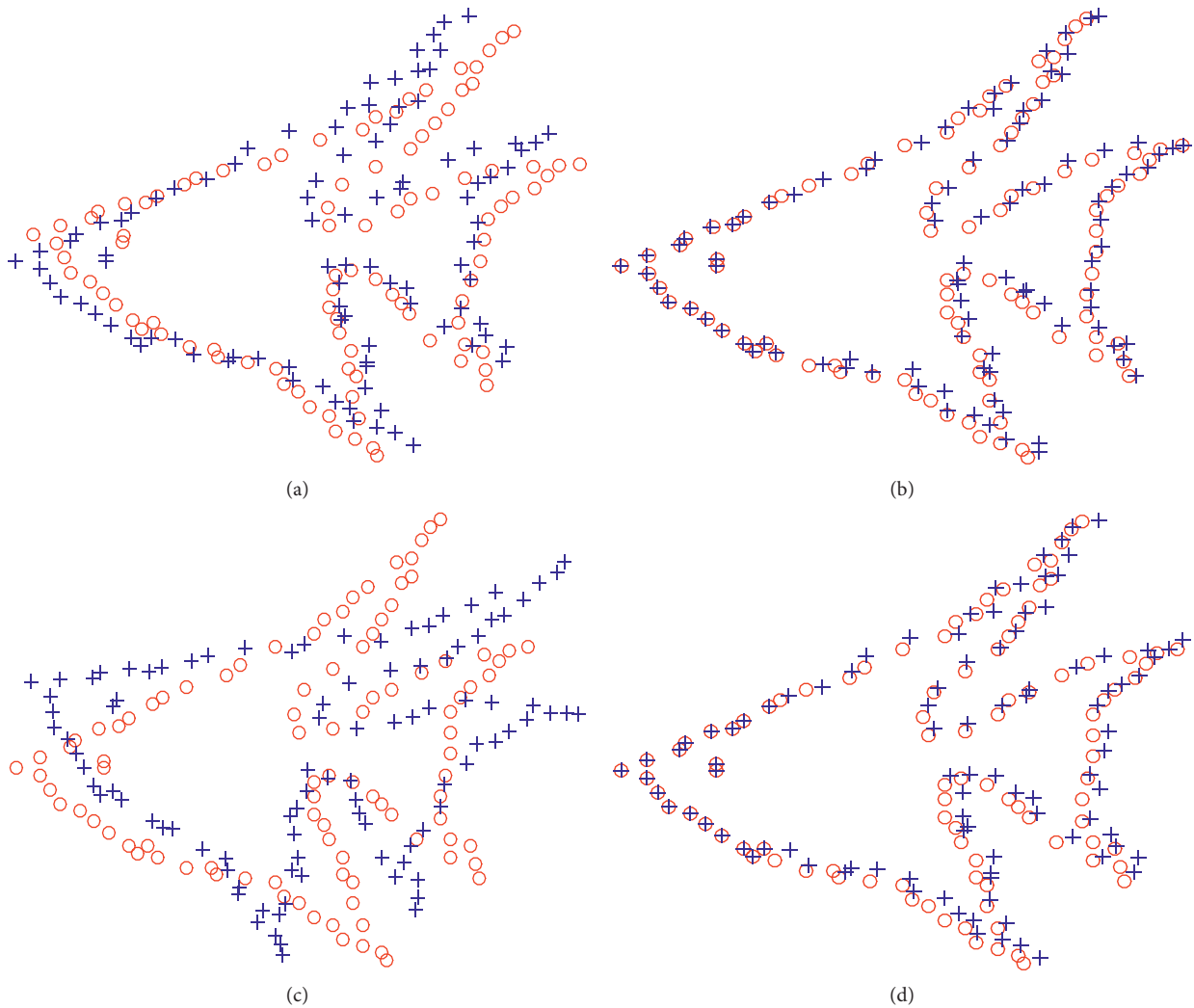
Figure 4: Registration results of different algorithms at 80% noise. (a) Original point set. (b) Results of KC. (c) Results of CPD. (d) Results of the proposed method.

TABLE 1: Registration accuracy at different outlier proportions.

| Method | Proportion of outliers | | | | | | | | | | | |
| | 0 | | 20% | | 40% | | 60% | | 80% | | 100% | |
	RSR	MSE	RSR	MSE	RSR	MSE	RSR	MSE	RSR	MSE	RSR	MSE
Proposed	0.98	0.02	0.93	0.08	0.87	0.15	0.79	0.20	0.74	0.26	0.59	0.35
ICP	0.96	0.03	0.92	0.10	0.85	0.17	0.76	0.23	0.70	0.28	0.57	0.39
KC	0.97	0.03	0.92	0.12	0.84	0.16	0.75	0.25	0.72	0.30	0.52	0.42
CPD	0.97	0.03	0.91	0.13	0.83	0.19	0.76	0.24	0.69	0.32	0.53	0.43

TABLE 2: Registration accuracy at different noise proportions.

| Method | Proportion of noises | | | | | | | | | | | |
| | 0 | | 20% | | 40% | | 60% | | 80% | | 100% | |
	RSR	MSE	RSR	MSE	RSR	MSE	RSR	MSE	RSR	MSE	RSR	MSE
Proposed	0.97	0.02	0.92	0.07	0.85	0.16	0.76	0.21	0.76	0.22	0.58	0.34
ICP	0.95	0.04	0.91	0.11	0.83	0.18	0.75	0.25	0.73	0.25	0.55	0.42
KC	0.94	0.03	0.90	0.11	0.85	0.17	0.74	0.23	0.73	0.31	0.55	0.46
CPD	0.94	0.02	0.91	0.12	0.82	0.21	0.74	0.22	0.74	0.30	0.54	0.45

accuracy RSR of SKL and KC algorithms is equivalent to the value of MSE. Also, both are better than the registration results of CPD. ICP has the lowest registration accuracy. When the ratio of outliers to noise is greater than 60%, the RSR of SKL is the highest, and the error of MSE is the smallest. The above experimental results fully verify the robustness and effectiveness of the registration algorithm proposed in this paper.

5. Conclusion

This paper proposes a point set registration algorithm based on improved KL divergence. By separately modeling two point sets as GMMs, the point set registration problem is transformed into a problem of finding the minimum KL divergence between two GMMs. The genetic algorithm is used to solve the transformation parameters of the point set registration, and the SKL divergence between the two Gaussian mixture models is calculated by the matching approximation method. The registration experiment is carried out under different outliers and noise ratios, and the registration achievement rate and MSE are used to measure the registration accuracy of the algorithm. The experimental results verify the effectiveness of the algorithm in this paper.

Data Availability

The dataset used in this study is publicly accessible.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] Y. N. Dwith Chenna, P. Ghassemi, T. J. Pfefer, J. Casamento, and Q. Wang, "Free-form deformation approach for registration of visible and infrared facial images in fever screening," *Sensors*, vol. 18, no. 1, pp. 125–138, 2018.

[2] U. Martinez Hernandez, T. J. Dodd, and T. J. Prescott, "Feeling the shape: active exploration behaviors for object recognition with a robotic hand," *IEEE Transactions on Systems Man and Cybernetics Systems*, vol. 23, no. 9, pp. 1–10, 2017.

[3] S. Papazoglou, J. Wrfel, F. Paul, A. Brandt, and M. Scheel, "Single-subject independent component analysis-based intensity normalization in non-quantitative multi-modal structural MRI," *Human Brain Mapping*, vol. 12, no. 5, pp. 15–24, 2017.

[4] P. J. Besl and N. D. Mckay, "A method for registration of 3-D shapes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 2, pp. 239–256, 1992.

[5] H. Chui and A. Rangarajan, "A new point matching algorithm for non-rigid registration," *Computer Vision and Image Understanding*, vol. 89, no. 2, pp. 114–141, 2003.

[6] C. Gope and N. Kehtarnavaz, "Affine invariant comparison of point-sets using convex hulls and hausdorff distances," *Pattern Recognition*, vol. 40, no. 1, pp. 309–320, 2007.

[7] M. Yang, K. Kidiyo, and R. Joseph, "Shape matching and object recognition using chord context," in *Proceedings of the Visualisation, International Conference*, vol. 1132, pp. 63–69, London, UK, July 2008.

[8] D. W. Shin and Y. S. Ho, "3D scene reconstruction using colorimetric and geometric constraints on iterative closest point method," *Multimedia Tools and Applications*, vol. 1, no. 6, pp. 1–26, 2017.

[9] L. Liang, M. Wei, A. Szymczak et al., "Nonrigid iterative closest points for registration of 3D biomedical surfaces," *Optics and Lasers in Engineering*, vol. 100, no. 1, pp. 41–54, 2018.

[10] H. Sobreira, C. M. Costa, I. Sousa, and L. Rocha, "Map-matching algorithms for robot self-localization: a comparison between perfect match, iterative closest point and normal distributions transform," *Journal of Intelligent and Robotic Systems*, vol. 2, no. 5, pp. 1–14, 2018.

[11] M. Andriy and S. Xubo, "Point set registration: coherent point drift," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 12, pp. 2262–2275, 2010.

[12] L. Liang, Y. Ming, C. Wang, and B. Wang, "Robust point set registration using signature quadratic form distance," *IEEE Transactions on Cybernetics*, vol. 99, no. 1, pp. 1–13, 2018.

[13] Y. Liu, D. Kong, D. Zhao, X. Gong, and G. Han, "A point cloud registration algorithm based on feature extraction and matching," *Mathematical Problems in Engineering*, vol. 2018, Article ID 7352691, 9 pages, 2018.

[14] M. Afgani, S. Sinanović, and H. Haas, "The information theoretic approach to signal anomaly detection for cognitive radio," *International Journal of Data Mining and Bioinformatics*, vol. 2010, Article ID 740594, 18 pages, 2010.

[15] N. A. Alqahtani and Z. I. Kalantan, "Gaussian mixture models based on principal components and applications," *Mathematical Problems in Engineering*, vol. 2020, Article ID 1202307, 13 pages, 2020.

[16] J. Qi, F. Xin, L. Yu, L. Haojie, L. Zhongxuan, and G. He, "Hierarchical projective invariant contexts for shape recognition," *Pattern Recognition*, vol. 52, no. 358, pp. 74–84, 2016.

[17] F. Chen, S. Wu, F. Liu, J. Ji, and Q. Lin, "A novel angular-guided particle swarm optimizer for many-objective optimization problems," *Complexity*, vol. 2020, Article ID 6238206, 18 pages, 2020.

[18] B. Zhang, H. Yang, and Z. Yin, "A spatial-constraint-based feature point matching algorithm for the positioning of multiple IC instances," *IEEE Transactions on Semiconductor Manufacturing*, vol. 29, no. 2, pp. 137–144, 2016.

[19] W. Kun Sun and K. Sun, "Robust point sets matching by fusing feature and spatial information using nonuniform Gaussian mixture models," *IEEE Transactions on Image Processing*, vol. 24, no. 11, pp. 3754–3767, 2015.

[20] G. Wang and Y. Chen, "SCM: spatially coherent matching with Gaussian field learning for nonrigid point set registration," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 1, pp. 203–213, 2021.

[21] B. Jiang, J. Tang, B. Luo, and L. Lin, "Robust feature point matching with sparse model," *IEEE Transactions on Image Processing*, vol. 23, no. 12, pp. 5175–5186, 2014.

[22] J. Konecny, M. Prauzek, P. Kromer, and P. Musilek, "Novel point-to-point scan matching algorithm based on cross-correlation," *Mobile Information Systems*, vol. 2016, Article ID 6463945, 11 pages, 2016.

[23] J. Goldberger, S. Gordon, and H. Greenspan, "An efficient image similarity measure based on approximations of KL-divergence between two Gaussian mixtures," in *Proceedings of the IEEE International Conference on Computer Vision*, vol. 1, no. 3, pp. 487–493, Nice, France, France, October 2003.

[24] F. Hu, Y. Li, and M. Feng, "Continuous point cloud stitch based on image feature matching constraint and score," *IEEE Transactions on Intelligent Vehicles*, vol. 4, no. 3, pp. 363–374, 2019.

[25] L. Chen, G. Xin, Y. Liu, and J. Huang, "Driver fatigue detection based on facial key points and LSTM," *Security and Communication Networks*, vol. 2021, Article ID 5383573, 9 pages, 2021.

Research Article

Research on Night Tourism Recommendation Based on Intelligent Image Processing Technology

Meng Li[1] and Ning Fan [ID][2]

[1]*Baotou Teachers' College, Inner Mongolia, Baotou 014030, China*
[2]*Capital Normal University, Beijing 100089, China*

Correspondence should be addressed to Ning Fan; 2190101023@cnu.edu.cn

The rapid development of the tourism industry and the Internet era has led to an increasingly severe problem of travel information overload, and travel recommendation methods are essential to solving the information overload problem. Traditional recommendation algorithms only target common travel scenarios during the daytime, combining the ratings and necessary attributes between users and items to calculate similarity for a recommendation. Still, the research on night travel recommendations is one of the few scenarios that needs to be explored urgently. This paper, based on histogram equalization, achieves better experimental results on image enhancement, combines convolutional neural network technology with night image and text comment feature extraction technology, and evaluates the resulting error with mean absolute error (MAE). This paper presents the first night travel recommendation system. It compares it with the traditional collaborative filtering method, and the model proposed in this paper can effectively reduce the prediction error.

1. Introduction

Online travel has become a daily travel application, primarily bringing convenience for people to travel and tour. However, with the increasing number of travelers and the explosive growth of information on the internet, it is difficult for travel users to quickly find the travel data they need in massive data. The emergence of recommendation systems effectively solved this problem of information overload. A recommendation system [1] refers to a website platform, which digs the explicit and implicit information of users to determine their preferences, and selects the N products that best match them, thus helping them to make efficient choices. In order to further study the recommendation system, people have developed various recommendation algorithms.

However, the research on the travel recommendation systems for night scenes has always been a problematic research field because the biggest difference between night and day is the significant difference in lighting conditions. It

is not perfect to use the data of daytime travel directly for a recommendation. Therefore, this paper presents a feasible solution to solve one of the challenges by using intelligent image processing technology. Image enhancement processing usually does not consider the reason of image quality degradation. However, it only uses specific techniques to selectively highlight the interesting features in the image according to the image features and processing purposes and may suppress another part of the information, so as to improve the visual effect of the image or facilitate the analysis and processing of the picture by humans or computers. Its main purpose is to improve the comprehensibility of the image.

Traditional image enhancement methods can be divided into spatial domain and frequency domain methods according to different scopes. The spatial domain method refers to the arithmetic processing of pixel gray values directly in the spatial domain. The commonly used spatial domain methods include grayscale transformation, histogram modification, image space smoothing and sharpening,

and pseudo-color processing. The frequency domain method enhances the transform value of the image in a certain transform domain and then obtains the enhanced image through inverse transformation.

Commonly used recommendation algorithms include content-based recommendation algorithm, collaborative filtering recommendation algorithm, CF recommendation algorithm, and mixed recommendation algorithm [2]. Content-based recommendation algorithm analyzes the users' historical content data, builds an interest model, and recommends similar interesting items to users. However, the biggest problem with this algorithm is that the recommendation domain is narrow. The collaborative filtering algorithm only relies on the historical data of users to obtain a group of users with similar interests and then recommends this group of users' preference items to the target users [3]. The hybrid recommendation algorithm is the most effective algorithm to solve the shortcomings of various algorithms in theory, but the design of this algorithm is relatively complex.

In recent years, artificial intelligence is in full swing. As an important research direction in artificial intelligence, deep learning has penetrated into various industries. Deep learning can deeply mine data features in image, text, and natural language processing and has made some progress, especially for hidden information which is easy to be ignored. In traditional tourism recommendation models, predictions are only based on the scores between tourists and scenic spots, which lack the information about the characteristics and attributes of tourists and scenic spots and tourists' evaluation of tourism experience. A multimodal recommendation algorithm based on deep multimodal learning is proposed, which establishes the feature representation between multimodal features in videos and effectively obtains more information. Gao and Xu [4] used the recursive neural network to process the time series data of users and movies, so as to recommend movies more accurately. They used long-term and short-term memory units to accumulate historical states at long distances, which ultimately improve the accuracy of recommendation.

However, most of the current standard recommendation systems are based on textual review features and recommendations made by daytime well-lit tourist attraction picture features. Despite these, travel suggestions under low light conditions at night are still tasted because of low visibility, or travel suggestions during the day are directly adopted. This paper mainly studies the complex problem of night travel recommendations under the general lighting conditions. In this paper, the night travel recommendation model is designed by using a histogram equalization-based night image enhancement algorithm and collaborative filtering combined with the convolutional neural network.

2. Materials and Methods

2.1. Common Methods and Quality Evaluation of Image Enhancement for Night Tourist Attraction. Usually, the acquired images cannot be processed directly by using the computer and must be converted into digital images before processing can be performed. The process of digitization consists of 3 steps: scanning, sampling, and quantization. From the computer's point of view, a digital image can be seen as an image obtained by sampling and quantizing a two-dimensional function $f(x, y)$, i.e., after discrete processing. Therefore, an image is usually represented by a two-dimensional matrix. When an image is generated from a physical process, its value is usually proportional to the radiated energy of the physical source (e.g., electromagnetic waves). Therefore, $f(x, y)$ must be nonzero and finite.

Digitizing an image will result in a two-dimensional matrix of integers, as in equation (1), which defines a digital image of size M^*N:

$$f(x, y) = \begin{bmatrix} f(0,0) & f(0,1) & \ldots & f(0, N-1) \\ f(1,0) & f(1,1) & \ldots & f(1, N-1) \\ \ldots & \ldots & \ldots & \ldots \\ f(M-1,0) & f(M-1,1) & \ldots & f(M-1, N-1) \end{bmatrix},$$

(1)

where each element of the matrix is called an image unit, image element, or pixel.

2.1.1. Night Image Enhancement Basic Method. Image enhancement techniques usually have two types of methods: spatial domain methods and frequency domain methods. Spatial domain enhancement methods target direct processing of pixel grayscale values. Common ones are grayscale transformations, image smoothing and sharpening in the null domain, histogram correction, pseudo-color processing, etc. Frequency domain enhancement methods are based on the Fourier transform of the image to enhance or suppress the desired spectrum, thus achieving an improvement of the image spectrum.

(1) Grayscale Transformation. Grayscale transformation is one of the simplest and most important means of image enhancement techniques. Grayscale transform extends the image contrast to obtain a clearer image. It corrects the grayscale of the input pixels, and the grayscale transform can be expressed as

$$g(x, y) = T(f(x, y)),$$

(2)

where T denotes the function relationship between the input gray value and the output gray value. According to different application requirements and the nature of the function, the following grayscale transformation methods can be used.

For a given grayscale space, let the grayscale transformation range of the original image $f(x, y)$ be $[a, b]$ and the grayscale extension of the transformed image $g(x, y)$ be $[c, d]$; then, a proportional linear grayscale change can be used to achieve.

$$g(x, y) = \frac{d - c}{a - b}(f(x, y) - a) + c.$$

(3)

The proportional linear grayscale transform operation is relatively simple, and it can effectively change the image visual effect.

And the segmented linear transform has a better enhancement effect in stretching the gray details of feature objects and suppressing the uninteresting gray levels with the following mathematical expressions:

$$g(x, y) = \begin{cases} \dfrac{c}{a} f(x, y) & 0 \leq f(x, y) \leq a, \\[2mm] \dfrac{d-c}{b-a}(f(x, y) - a) + c & a \leq f(x, y) \leq b, \\[2mm] \dfrac{f-d}{e-b}(f(x, y) - b) + d & b \leq f(x, y) \leq e. \end{cases} \quad (4)$$

(2) Histogram Correction. The histogram of an image represents the statistical relationship between each gray level and its frequency. Usually, the horizontal coordinate represents the gray level, and the vertical coordinate represents the frequency. According to the definition of the histogram, it can be expressed as

$$p(r_k) = \frac{n_k}{N} \quad (k = 0, 1, \ldots, L-1), \quad (5)$$

where N is the total number of pixels in an image, n_k denotes the number of pixels at the kth gray level, r_k is the kth gray level, L is the number of gray levels, and $p(r_k)$ denotes the relative frequency of the occurrence of that gray level.

Histogram is the main means to reflect the general characteristics of a digital image. It gives information about the grayscale range, the grayscale distribution, the average brightness, and the contrast between light and dark of a digital image, which is an important basis for further processing of digital images.

The histogram equalization method is a common grayscale enhancement algorithm that transforms a known grayscale probability density distribution into a new image with a uniform grayscale probability density distribution to enhance the overall contrast of the image.

Assuming that the number of pixels in an image is n, there are l gray levels, and n_k represents the number of pixels with a gray level of gas, the probability of the occurrence of the kth gray level can be expressed as

$$p(r_k) = \frac{n_k}{n}, \quad (6)$$

where $0 \leq r_k \leq 1$ and $k = 0, 1, \ldots, L-1$. The transformation function $T(r)$ is

$$s_k = T(r_k) = \sum_{j=0}^{k} P_r(r_j) = \sum_{j=0}^{k} \frac{n_j}{n}. \quad (7)$$

From the above calculation process, it can be seen that the statistical value of the histogram of the original input image can calculate the grayscale value of each pixel after equalization. The histogram equalization results in the stretching of the dense grayscale distribution and the compression of the sparse grayscale distribution, thus enhancing the overall contrast of an image.

However, the traditional histogram equalization algorithm is processed for the whole image, so there are some insurmountable defects: (1) the brightness information of the original image is not retained, and the grayscale is approximately uniformly distributed in the processed image, so the average brightness is always around the median of the grayscale range; (2) the phenomenon of parsimony occurs in the process, i.e., some low-frequency grayscale may be combined, and the total grayscale level will be lost, thus causing the loss of image details. For these drawbacks, some improvement methods are listed in the next section of this paper.

2.1.2. Image Quality Evaluation Criteria. Image quality evaluation [5] is one of the basic techniques of image information engineering, and there are two kinds of subjective and objective evaluations. Subjective evaluation is to let observers make quality judgments on test images according to visual effects based on some predefined evaluation scales or their own experiences and give quality scores, and finally, the scores given by all observers are weighted and averaged, and the result is the subjective quality evaluation of the image. The subjective evaluation is determined by some subjective factors such as individual knowledge background, observation purpose, and prevailing environment because different individuals evaluate the same image. The two commonly used subjective evaluation methods are the absolute scale grading method and the relative scale grading method.

(1) Mean Value. The mean value indicates the average value of image grayscale, which reflects the image illumination situation.

(2) Mean Square Error. The mean square error, also called the mean squared error of the image gray value, reflects the dynamic range of the image gray value. Let $f(i, j)$ denote the original image, $\hat{f}(i, j)$ denote the distorted image of $f(i, j)$, and the mean square error between $f(i, j)$ and $\hat{f}(i, j)$ for any i and j is

$$\text{MSE} = \frac{\sum_{i=0}^{M-1} \sum_{j=0}^{N-1} (f(i, j) - \hat{f}(i, j))^2}{MN}, \quad (8)$$

where M and N denote the height and width of the image, respectively. Obviously, the smaller the MSE value is, the more similar the two images are before and after information hiding.

(3) Information Entropy. For an image with n levels of gray, if the probability of the occurrence of its ith level of gray is p_i, then the entropy of that level of gray is

$$e(i) = -p_i \log p_i. \quad (9)$$

Then, the entropy of the whole image is

$$E = \sum_{i=0}^{n-1} e(i) = -\sum_{i=0}^{n-1} p_i \log p_i. \quad (10)$$

(4) Peak Signal-to-Noise Ratio. The peak signal-to-noise ratio is defined as follows:

$$PSNR = 10\lg\left(\frac{\sum_{i,j=1}^{H,W}255^2}{\sum_{i,j=1}^{H,W}(\hat{f}(i,j)-f(i,j))}\right), \quad (11)$$

where H and W indicate the height and width of the image, respectively. Usually, a larger PSNR value indicates a better image recovery, but it does not mean that the subjective quality is higher.

2.2. Night Image Improvement Algorithm Based on Histogram Equalization.

Histogram equalization is one of the very important algorithms in image enhancement. It uses basic knowledge of probability theory to perform gray point operations for the purpose of histogram transformation [6–8]. The essence of the image histogram equalization algorithm is to selectively enhance the low-frequency information and suppress the high-frequency information of the image. After histogram equalization of an image, the grayscale that originally possesses fewer pixels is most likely to be merged. In contrast, the boundaries between image regions that normally occupy fewer pixels contain important structural information, which may cause the loss of image details. In addition, various types of image noise are inevitably present in the image, and then the image noise will be amplified accordingly [7, 9]. Therefore, the traditional histogram equalization algorithm has the disadvantages of image detail information loss and noise amplification.

2.2.1. Improved Algorithm 1

(1) For each pixel in the original image, it is recorded into the corresponding pixel number accumulator r_i according to its grayscale value.

(2) Select the appropriate mapping relationship for the number of pixels. Here, $r_i = \log(r_{i+1})$ is used, and r_i and $Q = \sum_{i=0}^{m-1}r_i$ are calculated, respectively.

Calculate the converted grayscale value of grayscale i in the original image.

$$f_i = (L-1)\frac{\sum_{k=0}^{i-1}r_k}{Q-r_i}. \quad (12)$$

The grayscale conversion method of the above equation can extend the gray space of the original image so that the converted grayscale values are distributed over the entire gray space.

2.2.2. Improved Algorithm 2.

For the grayscale $f_i(x, y)$, at level i in the original image grayscale histogram, the position of its corrected grayscale $gi(x, y)$ is determined based on the ratio of $\sum_{k=0}^{i-1}P_r(r_k)$ and $\sum_{k=i+1}^{i-1}P_r(r_k)$ on its left and right sides using the following equation j:

$$j: (l-1-j) = \sum_{k=0}^{i-1}P_r(r_k): \sum_{k=i+1}^{i-1}P_r(r_k). \quad (13)$$

Collate to obtain

$$j = (l-1)\frac{\sum_{k=0}^{i-1}P_r(r_k)}{\sum_{k=0}^{i-1}P_r(r_k)+\sum_{k=i+1}^{i-1}P_r(r_k)}. \quad (14)$$

Furthermore, it is reduced to

$$j = (l-1)\frac{\sum_{k=0}^{i-1}P_r(r_k)}{1-P_r(r_i)} = (l-1)\frac{\sum_{k=0}^{i-1}n_k}{n-n_i}. \quad (15)$$

2.2.3. Improved Algorithm 3.

After histogram equalization of the image, the average brightness of the output image usually lies in the middle of the gray level of the image. By using histogram recursive decomposition and segmented equalization, we can achieve multilevel brightness control of the output image, thus avoiding the unnatural enhancement effect caused by too much brightness after enhancement.

The basic idea of histogram recursive decomposition and segmented equalization is to use the average luminance as the closed value, decompose the image recursively according to the specified depth r, so as to obtain 2^r subimages of different grayscale ranges, then make histogram equalization for each subimage, and finally merge the equalized subimages to obtain the equalization result of the night image. Obviously, when the recurrence depth r is 0, the input image is not decomposed in any way, i.e., in the conventional way. The case when r is 1 is illustrated below, and so on for other cases.

First, the original image X is divided into two subgraphs X_1 and X_2, with the grayscale mean value G_m as the threshold, i.e.,

$$\begin{aligned}X_1 &= \{x(i,j) \,|\, G(X) \leq G_m\},\\ X_2 &= \{x(i,j) \,|\, G(x) > G_m\}.\end{aligned} \quad (16)$$

Next, the histograms are counted for each of the two subplots and equalized in the grayscale range on each side of the respective mean.

$$\begin{cases} p_1(G_k) = \dfrac{n_1(k)}{n_1}, \quad k = 0, 1, 2, ..., m, \\[2ex] C_1(k) = \displaystyle\sum_{k=0}^{m} p_1(k), \\[2ex] Y_1 = f_1(X_1), \text{while}, G(y) = G_m C[G(x)]. \end{cases} \quad (17)$$

Then, the processed subgraphs are merged:

$$Y = Y_1 \cup Y_2. \quad (18)$$

Since the segmentation, processing, and merging are performed in grayscale dependence, the spatial location of the enhanced pixels does not change. Similarly, according to the recurrence depth, the image is decomposed into 2^r

subimages according to the average luminance, then each of them is histogram equalized, and finally, the equalized subimages are merged. Then, the average luminance of the output image is

$$E(Y) = \left(1 - \frac{1}{2^r}\right)G_m + \frac{1}{2^r}G_c. \tag{19}$$

From the above equation, we can see that the greater the depth of decomposing the image is, the closer the average brightness of the final image obtained will be to the average brightness of the input image. Therefore, we can use different r values to achieve luminance control. The above method can automatically extend the range of image gray levels and, at the same time, can control the average luminance of the output image with better enhancement effect. However, there is still the disadvantage of gray level merging, and the image gray level average value is still offset.

2.3. Night Tour Recommendation Model Design. This paper improves the feature recognition of night attraction images by combining the techniques of night image enhancement in the previous section. Combining convolutional neural networks and collaborative filtering methods, we extract and classify the factor features in night attraction images and visitor text reviews by convolutional neural networks and predict the user item ratings by using collaborative filtering methods. Among them, the convolutional neural network for mining features in reviews can effectively alleviate the data sparsity problem in the collaborative filtering method [10], and the method can retain the advantages of shared experience and personalization in the collaborative filtering method.

2.3.1. Collaborative Filtering Recommendation Algorithm. As the most widely used recommendation algorithm, the core idea of collaborative filtering is to calculate the similarity between users or items to achieve recommendations. User-based collaborative filtering recommendation algorithm calculates the similarity between users, establishes the nearest-neighbor user group, and then recommends the products of interest in the nearest-neighbor group to the target users; item-based collaborative filtering recommendation algorithm calculates the similarity between items; model-based collaborative filtering recommendation algorithm combines some intelligent models to train and test the data in order to obtain the user's preference information more effectively to achieve recommendations.

In order to find users or items in close proximity, similarity calculation is the key. The most commonly used methods to calculate similarity are Pearson's method and cosine method [11]. Among them, the Pearson similarity method is more applicable to the calculation of document data and is more comprehensive for different users, which is expressed by the following formula:

$$\text{sim}(a, b) = \frac{\sum_{i \in I_{a,b}} \left(R_{a,i} - \overline{R_a}\right)\left(R_{b,i} - \overline{R_b}\right)}{\sqrt{\sum_{i \in I_{a,b}} \left(R_{a,i} - \overline{R_a}\right)^2 \sum_{i \in I_{a,b}} \left(R_{b,i} - \overline{R_b}\right)^2}}. \tag{20}$$

Chen et al. [12] used a new dynamic evolutionary clustering algorithm to divide the set of users and items into K clusters and clustered the individuals with high similarity. Collaborative filtering technique is applied to predict among each of the K clusters, and recommendation predictions are made for target users based on the prediction levels obtained from clustering. Bo and Fei [2] used a combined collaborative filtering recommendation algorithm to improve the accuracy utilization of group buying websites using an offline testing method [13–16].

2.3.2. Convolutional Neural Network. Convolutional neural networks (CNNs) are one of the main focuses of research in deep learning [17, 18], and their applications in image processing are quite mature, and they are also widely used for processing text comments. As a feedforward neural network, the basic structure of the CNN is similar to that of other neural networks in that it consists of an input layer, an implicit layer [19–25], and an output layer for data feature extraction, and the implicit layer of the CNN consists of three layers, including convolution, pooling, and full connectivity [3]. The CNN structure is shown in Figure 1, including the convolution layer, pooling layer, and fully connected layer, which are the important neural network structures in this paper.

The purpose of this paper is to design a personalized night tour recommendation model, and to address the problem of data sparsity where it is difficult to mine deep information based on user collaborative filtering recommendation algorithms, convolutional neural networks are selected to classify the night attraction images and review information of tour data for deep extraction. The text comments of tourists on tourism contain emotional factors about the tourism process, which are divided into two categories: positive and negative. The basic information, behavioral information, and contextual information of tourist users are combined to form tourist user data to solve the problem of single data.

In this paper, in addition to the night landscape features after data enhancement, text features are also required in the tourism recommendation model, which mainly considers three major elements of user data: basic information, behavioral information, and contextual information. The vectorized tourism user review data matrix is sent to the convolutional neural network for training to get the positive and negative emotions of tourists towards the project, and then the processed tourism data are used for similarity calculation. The structure of the travel recommendation model is shown in Figure 2.

The basic information of the user is expressed as

$$P = \alpha_1 \text{age} + \alpha_2 \text{sex} + \alpha_3 \text{place}. \tag{21}$$

The behavioral information of travel users is expressed as
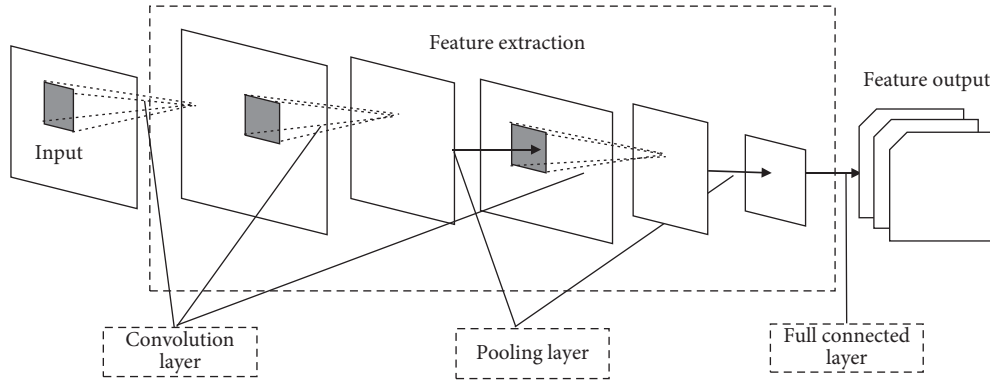
FIGURE 1: Structure of the convolutional neural network.
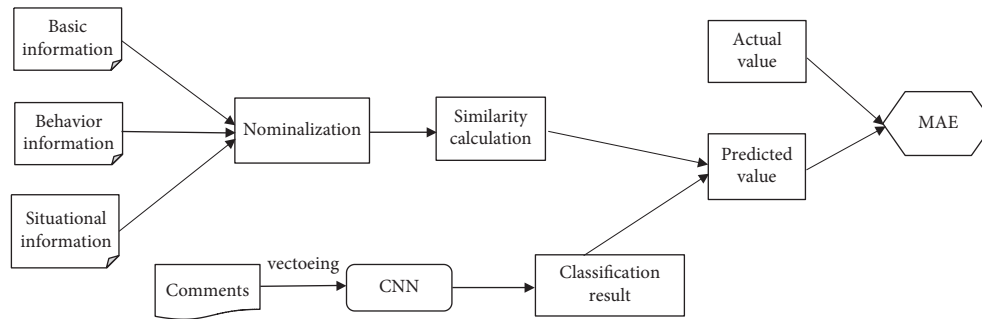
FIGURE 2: Structure of the recommended model.

$$S = \beta_1 \text{demand} + \beta_2 \text{prefer} + \beta_3 \text{consume}. \qquad (22)$$

The emotional information of travel users is expressed as

$$Q = \gamma_1 \text{time} + \gamma_2 \text{location} + \gamma_3 \text{intact}. \qquad (23)$$

The textual feature information of the night tour recommendation information is expressed as

$$TR = \alpha P + \beta S + \gamma Q. \qquad (24)$$

3. Results and Discussion

3.1. Experimental Results of a Night Image Improvement Algorithm Based on Histogram Equalization. In the experiment, the images of night tourist attractions are selected, and the traditional histogram equalization and the three improved histogram equalization methods proposed in this paper are carried out, respectively, and the corresponding histograms are given. The comparison graph of the image enhancement results of the three improved algorithms is shown in Figure 3.

In addition to the visual comparison of the results, this paper provides a table of numerical comparisons of different algorithms, as shown in Table 1. According to the results of the above experimental images, the improved histogram equalization method has a more obvious contrast of the enhanced images. From the experimental data, the enhanced histogram equalization method enlarges the grayscale range of the image. In addition, the average value of the image is closer to the actual value.

3.2. Analysis of Travel Recommendation Results after Night Image Enhancement. In this paper, we crawled the nighttime attraction pictures of China 5A and 4A, a total of 32 attractions on the Internet and the rating data set on tourism websites. Then, the text features are extracted: first, the duplicate data, blank data, and invalid data of irrelevant scenic spots are removed. Then, the noisy data are filtered; that is, the nontext symbols are removed. Then, we use Python's jieba to split the word for annotation data and filter stop words. Convolutional neural networks have certain requirements on the length of the input vectors, so the length of segmented sentences is standardized. Most of the processed word segmentation data are about 50 in length, so sentences with a length of about 50 are chosen. The data corresponding to each visitor and scenic spot are merged. Then, word2vec vectorization is performed on the merged word segmentation data. The vectorized multidimensional data are sent to the convolutional neural network to extract features.

In order to reflect the effectiveness of the convolutional neural network and collaborative filtering method more intuitively, this experiment compares this method with the traditional collaborative filtering method which only considers the scores between users and items. The experimental results are shown in Figure 4. It can be seen that the MAE value of this paper is smaller than that of the collaborative filtering method in the same scene, considering the behavior information and context data of tourists. As the data volume in the test set increases, the MAE value increases, but the overall trend decreases, which indicates that when the data volume is more significant, the quality of the model can be fully verified.

(a)

(b)

(c)
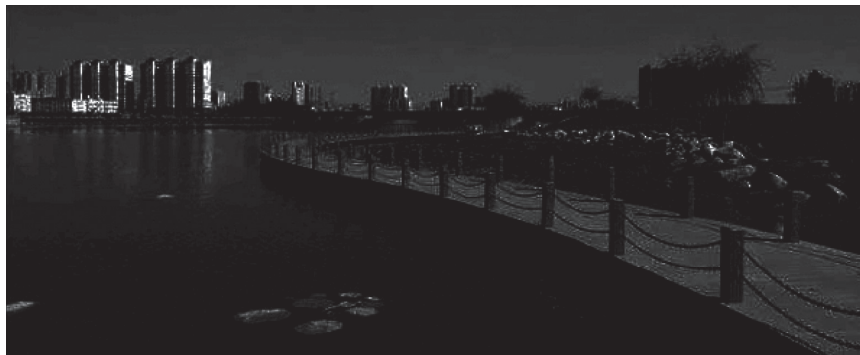
(d)

Figure 3: Comparison of image enhancement results of three improved algorithms. (a) Traditional algorithms. (b) Improved algorithm 1. (c) Improved algorithm 2. (d) Improved algorithm 3.

TABLE 1: Comparison of experimental data.

Parameters	Original image	Traditional algorithms	Improved algorithm 1	Improved algorithm 2	Improved algorithm 3
Number of shades of gray used	118	53	110	55	64
Maximum grayscale value	133	255	255	254	255
Minimum grayscale value	0	24	0	0	0
Average grayscale value	11	128	42	125	45
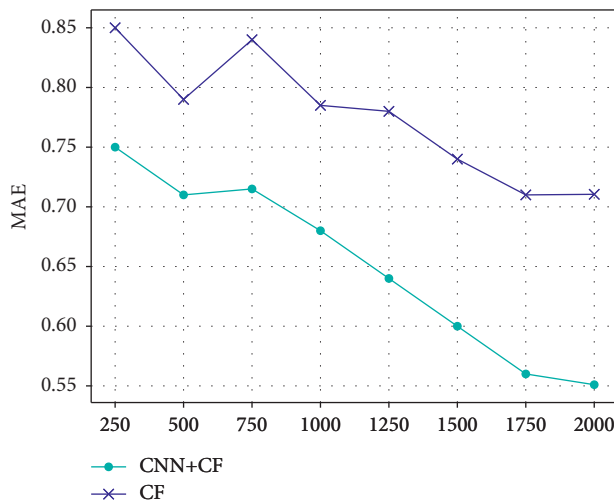
FIGURE 4: Comparison of experimental results.

4. Conclusions

Night travel recommendation, as an essential part of travel recommendation, is often ignored due to low light. However, with the optimization of intelligent image processing technology, it has been paid much attention. In digital image processing, image enhancement technology plays an important role in improving image quality. It is the pre-processing stage of image processing, which plays the role of carrying forward and preparing for the subsequent processing stages. Firstly, the basic theories and algorithms of image enhancement are introduced. Secondly, this paper lists the related improvements made in view of the shortcomings of traditional histogram equalization methods and analyzes the advantages of each improvement. Finally, a better night travel recommendation system is proposed by combining collaborative filtering and convolutional neural network.

Data Availability

The experimental data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest regarding this work.

References

[1] W. Lei, G. Chen, Z. Bei, and L. Yong, "Deep learning based social-aware location recommendation," *Journal of Terahertz Science and Electronic Information Technology*, vol. 17, no. 3, pp. 502–508, 2019.

[2] Y. Bo and Z. P. Fei, "Review of the art of recommendation algorithms," *Journal of Shanxi University (Natural Science Edition)*, vol. 34, no. 3, pp. 337–350, 2011.

[3] P. Sun, Z. Y. Li, Z. Y. Han, and F. Y. Wang, "An overview of collaborative filtering recommendation algorithm," *Advanced Materials Research*, vol. 756-759, pp. 3899–3903, 2013.

[4] M. T. Gao and B. Y. Xu, "Recommendation algorithm based on circular neural network," *Computer Engineering*, vol. 45, no. 8, pp. 198–202, 2019.

[5] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *Proceedings of the sixth international conference on computer vision*, pp. 839–846, IEEE, Bombay, India, January 1998.

[6] L. H. Xu and X. X. Lu, "Collaborative filtering recommendation system based on interest drifting and similar cloud," *Journal of Liaoning Technical University*, vol. 36, no. 7, pp. 761–766, 2017.

[7] N. Zhou, J. Du, Z. Xue, M. Liang, X. Yao, and W. Cui, "Security topics related microblogs search based on deep convolutional neural networks," *Neurocomputing*, vol. 395, pp. 199–211, 2020.

[8] A. Abdul, J. Chen, H.-Y. Liao, and S.-H. Chang, "An emotion-aware personalized music recommendation system using a convolutional neural networks approach," *Applied Sciences*, vol. 8, no. 7, pp. 1–16, 2018.

[9] X. Liu, J. Zhang, S. Jiang et al., "Accurate localization of tagged objects using mobile RFID-augmented robots," *IEEE Transactions on Mobile Computing*, vol. 20, no. 4, pp. 1273–1284, 2021.

[10] J. Su, R. Xu, S. Yu, B. Wang, and J. Wang, "Idle slots skipped mechanism based tag identification algorithm with enhanced collision detection," *KSII Transactions on Internet and Information Systems*, vol. 14, no. 5, pp. 2294–2309, 2020.

[11] H. G. Rong, S. X. Huo, C. H. Hu, and J. X. Mo, "User similarity-based collaborative filtering recommendation algorithm," *Journal on Communications*, vol. 35, no. 2, pp. 16–24, 2014.

[12] J. Chen, L. Wei, and L. Uliji, "Dynamic evolutionary clustering approach based on time weight and latent attributes for collaborative filtering recommendation," *Chaos, Solitons & Fractals*, vol. 114, no. 114, pp. 8–18, 2018.

[13] J. Su, Y. Chen, Z. Sheng, Z. Huang, and A. X. Liu, "From M-ary Query to Bit Query: a new strategy for efficient large-scale RFID identification," *IEEE Transactions on Communications*, vol. 68, no. 4, pp. 2381–2393, 2020.

[14] J. Su, Z. Sheng, A. X. Liu, Y. Han, and Y. Chen, "Capture-aware identification of mobile RFID tags with unreliable channels," *IEEE Transactions on Mobile Computing*, p. 1, 2020.

[15] Z. Kuang, L. Li, J. Gao, L. Zhao, and A. Liu, "Partial offloading scheduling and power allocation for mobile edge computing systems," *IEEE Internet of Things Journal*, vol. 6, no. 4, pp. 6774–6785, 2019.

[16] J. Su, Z. Sheng, A. X. Liu, Z. Fu, and C. Huang, "An efficient missing tag identification approach in RFID collisions," *IEEE Transactions on Mobile Computing*, p. 1, 2021.

[17] R. Zhang, W. P. Li, and T. Mo, "Review of deep learning," *Information and Control*, vol. 47, no. 4, pp. 385–397, 2018.

[18] F. Yang, H. Xie, and H. Li, "RETRACTED: v," *Applied Soft Computing*, vol. 82, no. 82, pp. 105597–105601, 2019.

[19] J. Su, Z. Sheng, V. C. M. Leung, and Y. Chen, "Energy efficient tag identification algorithms for RFID: survey, motivation and new design," *IEEE Wireless Communications*, vol. 26, no. 3, pp. 118–124, 2019.

[20] J. Su, X. Zhao, D. Hong, Z. Luo, and H. Chen, "Q-value fine-grained adjustment based RFID anti-collision algorithm," *IEICE - Transactions on Communications*, vol. E99-B, no. 7, pp. 1593–1598, 2016.

[21] K. Keping Yu, M. Arifuzzaman, Z. Zheng Wen, D. Di Zhang, and T. Sato, "A key management scheme for secure communications of information centric advanced metering infrastructure in smart grid," *IEEE Transactions on Instrumentation and Measurement*, vol. 64, no. 8, pp. 2072–2085, 2015.

[22] W. Wang, N. Kumar, J. Chen et al., "Realizing the potential of the Internet of things for smart tourism with 5G and AI," *IEEE Network*, vol. 34, no. 6, pp. 295–301, 2020.

[23] P. Wang, B. Xu, J. Xu, G. Tian, C.-L. Liu, and H. Hao, "Semantic expansion using word embedding clustering and convolutional neural network for improving short text classification," *Neurocomputing*, vol. 174, no. 174, pp. 806–814, 2016.

[24] L. V. Haiyan, "Design of individualized recommended algorithm in college online platform," *Cyber Security Intelligence and Analytics*, Springer, New York, NY, USA, 2021.

[25] M. Deepa and K. D. Sanil, "An efficient IP core of consultative committee for space data systems (CCSDS) recommended authenticated cryptographic algorithm," in *Proceedings of the 2020 8th International Symposium on Digital Forensics and Security (ISDFS)*, pp. 1–6, Beirut, Lebanon, June 2020.

Research Article

Design of Painting Art Style Rendering System Based on Convolutional Neural Network

Xingyu Xie and Bin Lv (iD)

School of Art, East China University of Technology, Nanchang, Jiangxi 330000, China

Correspondence should be addressed to Bin Lv; 201260007@ecut.edu.cn

Convolutional Neural Network- (CNN-) based GAN models mainly suffer from problems such as data set limitation and rendering efficiency in the segmentation and rendering of painting art. In order to solve these problems, this paper uses the improved cycle generative adversarial network (CycleGAN) to render the current image style. This method replaces the deep residual network (ResNet) of the original network generator with a dense connected convolutional network (DenseNet) and uses the perceptual loss function for adversarial training. The painting art style rendering system built in this paper is based on perceptual adversarial network (PAN) for the improved CycleGAN that suppresses the limitation of the network model on paired samples. The proposed method also improves the quality of the image generated by the artistic style of painting and further improves the stability and speeds up the network convergence speed. Experiments were conducted on the painting art style rendering system based on the proposed model. Experimental results have shown that the image style rendering method based on the perceptual adversarial error to improve the CycleGAN + PAN model can achieve better results. The PSNR value of the generated image is increased by 6.27% on average, and the SSIM values are all increased by about 10%. Therefore, the improved CycleGAN + PAN image painting art style rendering method produces better painting art style images, which has strong application value.

1. Introduction

In recent years, deep learning has been widely used in many fields such as medical imaging [1], remote sensing [2], and three-dimensional modeling [3] and has played an important role in promoting the application of artificial intelligence in multiple industries. In order to discover useful macro information in the data, the purpose of deep learning is to combine low-level features to form more abstract features with strong representation ability. There are many types of image styles, such as oil painting, ink painting, sketch painting, etc., and applying these styles to a given image increases the diversified presentation of the image. The use of deep learning methods for style rendering is one of the hotspots in the field of image research. Image style rendering is an image conversion method based on deep learning, which can be widely used in image processing,

computer picture synthesis, and computer vision. The original image style rendering is based on the optimization method proposed in [4], which uses the backpropagation of the convolutional neural network (CNN) and uses pixel-by-pixel comparison to obtain the optimal image conversion model.

Style rendering is a processing method for rendering the semantic content of an image in different styles. The style is extracted from the specified style image template x, and the extracted style features are mapped to y through mapping T without destroying the content of the image y. y performs style conversion; that is, it realizes the image style rendering of $T(x) \approx y$. The study in [5] combined deep learning and texture generation methods, using CNN to represent style images as content feature and style feature. The high-level convolutional layer of CNN model extracts global features of image content, and the low-level convolutional layer

describes the image style. This method can generate a more objective artistic style effect. However, this method has slow convergence speed, long rendering time, and sometimes poor style expression. Image stylization methods based on model iteration [6] include two types of methods based on generative models and image reconstruction decoders. The study in [7] used the perceptual loss function to train the feedforward network and used the VGGNet trained on ImageNet to simplify the loss function calculation process and directly generate stylized images. The efficiency was improved by three orders of magnitude. At the same time, the pixel-by-pixel difference loss function is improved to a perceptual loss function, and super-resolution reconstruction is efficient, but this method requires artificially constructing a complex loss function. The study in [8] proposed an image stylization method based on a conditional adversarial network, which does not require artificial construction of loss functions and mapping functions, which simplifies the image stylization process.

The network structure of the generative adversarial network (GAN) does not need to construct a complex loss function and achieves the global optimum through the minimax game process of the generator G and the discriminator D [9]. The conditional generative adversarial network (CGAN) uses different learning models. The GAN model is a mapping $G: z \longrightarrow y$ from the random noise vector z to the output image y, while the conditional GAN model is to learn from the random noise vector z and observe the image. The generative model of the mapping is from x to the output image y, namely, $G: (x, z) \longrightarrow y$. The GAN style rendering effect is better than the CGAN model, but there is a problem that the coloring effect is not ideal.

Academia researches have proposed many improved image style conversion methods based on the GAN model. The study in [10] proposed to add feature space loss (L_{fea}) and image space loss (L_2) when training GAN. The study in [11] took the feature map difference of the middle layer of the network as the perceptual loss function and used the style transfer and superdivision images obtained by GAN to achieve real-time stylization and four times of super-resolution. The study in [12] obtained the super-resolution GAN model from low-resolution images to super-resolution images based on GAN. The study in [13] proposed a perceptual adversarial network (PAN) model that combines perceptual loss and GAN model and realized a variety of image style conversion applications. In the research of image style rendering, the method in [14] has been proved to be able to obtain good results, but its perceptual loss network is a pretrained VGG-16 network. The loss in the network is not easy to optimize, and the network is mainly used for classification. Although the ability to recognize the subject of the image is stronger, the ability to retain the background is weak. In view of this, this paper proposes a GAN model training method based on the PAN model. This method combines adversarial loss, content loss, and style loss into a new perceptual loss function, which can alternately update the loss network and the image style conversion network, thereby replacing the fixed loss network [15].

In view of the abovementioned image stylization problems, this paper proposes to use a new type of cycle GAN model (CycleGAN) to achieve image style rendering and adopts a PAN model-based confrontation training method for CycleGAN model. The proposed method combines perceptual loss, content loss, and style loss into a new perceptual loss function, and the loss network and the image style conversion network can be alternately updated, thereby replacing the fixed loss network [16] and at the same time improving the original generator network structure. The experimental results prove that the proposed method can enhance the background definition of the image, make it closer to the original image in content and style, and at the same time increase the convergence speed, and the generated style rendering effect is more realistic.

The rest of the paper is organized as follows: Section 2 gives the related works about painting style rendering system based on CNN models, including conventional generative adversarial network, the perceptual adversarial network, and the image style rendering process. Section 3 illustrates the proposed improved CycleGAN + PAN model, image conversion network, and discriminator network. Section 4 gives the novel proposed combination of loss functions, such as content loss function, style loss function, and perceptual adversarial loss function. Moreover, the painting art segmentation rendering system is also given. Section 5 gives the painting art style rendering experiments and results on four different painting art forms. Section 6 concludes the model and results of the paper.

2. Related Works

2.1. Generative Adversarial Network.
In the process of generating painting style images in the GAN model, the generator G constantly learns and improves the ability to produce image data, and the discriminator D gradually improves the ability to discriminate data, and G and D reach a balance in the process of confronting the game. The adversarial relationship between G and D_Y in GAN is shown in the following equation:

$$\min_G \max_{D_Y} V(D_Y, G) = E_{y \Leftrightarrow P_{\text{data}}(y)}[\log D_Y(y)] + E_{z \Leftrightarrow P_z}[\log(1 - D(G(z)))], \tag{1}$$

where y represents the image data, z is the noise input to the G network, and $G(z)$ is the generated image. $P_{\text{data}}(y)$ is the distribution of the real image data, $P_z(y)$ represents the noise distribution of the input network G, and $\Leftrightarrow$ represents the data obeys the distribution relationship. E is the mathematical expectation function. The deep convolutional GAN model (DCGAN) [17] is a combination of CNN and GAN, which generates images randomly, and cannot meet the requirements of style rendering between images. Therefore, this paper selects the CycleGAN model that is based on the CNN model to construct the painting art style rendering system.

2.2. Perceptual Adversarial Network.
Inspired by the GAN model, on the basis of the existing perceptual loss [18] research, reference [19] proposed the perceptual adversarial

network (PAN) model. PAN model combines perceptual loss and generative adversarial loss and conducts adversarial training between the image style conversion network and the discriminant network. Such model can continuously and automatically find the difference between the output and the real image that has not been reduced. Therefore, when the difference measured in the high-dimensional space is small, the hidden layer of the discriminant network model will be updated to a higher level, and the difference between the new output image and the real image will continue to be searched for until converging to the optimal solution.

The innovation of PAN is that there is no longer a need for a complex loss function constructed based on human experience like traditional image models. This method automatically learns the mapping from input to output pictures through the adversarial network and applies it to the image conversion problem to achieve a generalization model. The PAN model is based on GAN model, which is combined with perceptual loss for adversarial training, and enhances the naturalness and realism of the image. The PAN model can realize a variety of image conversion tasks, such as image super-resolution, denoising, semantic segmentation, automatic completion, etc. Therefore, in this paper, we used the PAN model to improve the performance and efficiency of the CycleGAN model for rendering painting art style images.

2.3. Image Style Rendering Process. The image style rendering process is divided into two stages: training phase and performing phase [20]. In the training phase, the system selects a style map Y_s and trains a transformation network model based on each style map. The samples of the content map X are continuously obtained from the training set through iterations. In each iteration, the conversion style network converts the content image X to Y and randomly transports X and Y to the discriminant network D. The network D judges between X and Y (content) and Y_s and Y (style) through the adversarial loss function, and the difference between the two is fed back to the network T. T adjusts weights and parameters and enters the next iteration. At the same time, the network D is continuously optimized to find more differences. The final purpose is to generate an image conversion network model with Y style. In the execution stage, the system inputs any content map into the well-trained Y style conversion model and converts the content map into a Y style effect image in real time, and the original content and structure remain unchanged. Therefore, this paper improves the image painting art style rendering network, as shown in Figure 1.

3. Improved Image Style Rendering Network Structure

Although the VGG-16 loss network can be well trained with the ImageNet data set, the loss in the VGG-16 network is not easy to optimize. If a supervision item can be added to the hidden layer of D to measure the effect of generation, the loss

can be changed at any time when the network T is updated. Based on this setting, it is easier to get a better T.

3.1. Cycle Generative Adversarial Network. The CycleGAN model contains two generators and discriminators to realize the mutual mapping between images x and y [21]. Suppose the mapping of $X \longrightarrow Y$ is the generator F, and the image sample $x \in X$ generates an image similar to the sample $y \in Y$ through $F(x)$. CycleGAN uses the discriminator D_Y to determine whether the generated image is a real image or not. The loss function L between the generator F and the discriminator D_Y is defined as

$$
\begin{aligned}
L_{\text{GAN}}\left(F_{X \longrightarrow Y}, D_Y\right) = {} & E_{y \Leftrightarrow P_{\text{data}}(y)}\left[\log D_Y(y)\right] \\
& + E_{x \Leftrightarrow P_{\text{data}}(x)}\left[\log\left(1 - D_Y\left(F_{X \longrightarrow Y}(x)\right)\right)\right],
\end{aligned}
\tag{2}
$$

where $\Leftrightarrow$, P_{data}, and E are defined with equation (1). The single adversarial loss function L cannot guarantee that the learning function maps the input x to the expected output y; that is, it is not possible to train GAN only with equation (2). CycleGAN introduces the mapping from Y to X as $G_{Y \longrightarrow X}$, and the discriminator D_X judges whether the image similar to x generated by $y \in Y$ through $G(y)$ is true or not. We can deduce the mapping loss function L between $G_{Y \longrightarrow X}$ and the discriminator D_X similar to equation (2), as shown in the following equation:

$$
\begin{aligned}
L_{\text{GAN}}\left(F_{X \longrightarrow Y}, D_X\right) = {} & E_{x \Leftrightarrow P_{\text{data}}(x)}\left[\log D_X(x)\right] \\
& + E_{y \Leftrightarrow P_{\text{data}}(y)}\left[\log\left(1 - D_X\left(G_{Y \longrightarrow X}(y)\right)\right)\right].
\end{aligned}
\tag{3}
$$

CycleGAN introduces the loss function of cycle consistency loss (CCL) and learns the two mappings $F_{X \longrightarrow Y}$ and $G_{Y \longrightarrow X}$ at the same time. After x is converted to y, from y to $\hat{x}$, the loss between x and $\hat{x}$ is calculated to avoid, so that it is possible that all images in X are mapped to the same image in Y. Figure 2 shows the training process of the cycle consistency loss function.

Figure 2 shows the forward consistency and backward consistency of CycleGAN model. The forward consistency and backward consistency constitute the overall consistency of CycleGAN model. The overall consistency is the key of the CycleGAN model. In this paper, we use $F(G(y)) \approx y$ and $G(F(x)) \approx x$ to define the cycle consistency loss L_{CCL} as

$$
\begin{aligned}
L_{\text{CCL}}\left(F_{X \longrightarrow Y}, G_{Y \longrightarrow X}\right) = {} & E_{x \Leftrightarrow P_{\text{data}}(x)}\left(\left\|G_{Y \longrightarrow X}\left(F_{X \longrightarrow Y}(x)\right) - x\right\|_1\right) \\
& + E_{y \Leftrightarrow P_{\text{data}}(y)}\left(\left\|F_{X \longrightarrow Y}\left(G_{Y \longrightarrow X}(y)\right) - y\right\|_1\right).
\end{aligned}
\tag{4}
$$

From equations (2)–(4), the objective function L' of CycleGAN shown in equation (5) is obtained:

$$
\begin{aligned}
L'\left(F_{X \longrightarrow Y}, G_{Y \longrightarrow X}\right) = {} & L_{\text{GAN}}\left(F_{X \longrightarrow Y}, D_Y\right) + L_{\text{GAN}}\left(G_{Y \longrightarrow X}, D_X\right) \\
& + \lambda_{\text{CCL}} L_{\text{CCL}}\left(F_{X \longrightarrow Y}, G_{Y \longrightarrow X}\right),
\end{aligned}
\tag{5}
$$
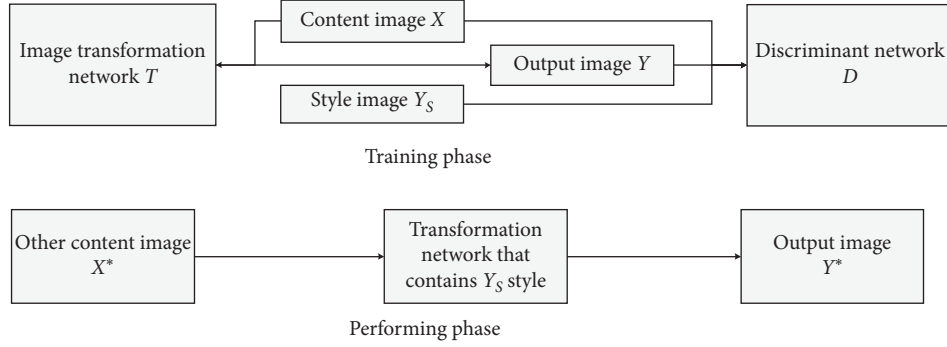
Figure 1: Image painting art style rendering process.
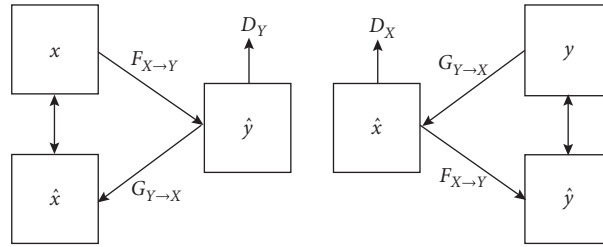
Figure 2: CycleGAN backward consistency and backward consistency.

where λ_{CCL} is the weight of adjusting the loss L_{CCL} in the objective function.

3.2. Image Conversion Network.
The image style conversion network T improves on the network proposed in [22], and the image style conversion network structure is shown in Figure 3. The overall structure follows its deep residual network (ResNet), which has three convolutional layers and five residual blocks, as shown in Figure 3(a). Except that the output layer uses a scaled Tan*h* function to ensure that the output pixels are in range of [0, 255], all other nonresidual convolutional layers use ReLU activation functions. Since no pooling layer is used and stride convolution or microstride convolution is used for upsampling and downsampling, this operation not only reduces the parameters, but also maintains a large field of view and avoids excessive deformation of image objects [23].

Considering that each nonresidual convolutional layer in the network is followed by a spatial batch normalization to achieve accelerated convergence and the study in [24] showing that instance regularization is better than batch normalization, which can significantly improve performance, this paper uses instance normalization instead of batch normalization, as shown in Figure 3(b). It can be seen that the instance normalization is actually to modify the batches to 1, and the normalization is used on a single image instead of a batch of images. Therefore, the network improved by the instance normalization performs better in the test phase of painting art style rendering.

3.3. Discriminator Network.
The discriminant network D is a multilayer convolutional neural network, as shown in Figure 4. Behind each hidden layer are added the batch

normalization and LeakyReLU linear activation functions. The first, fourth, sixth, and eighth layers are used to measure the perceptual adversarial loss between the generated image and the style image. The judgment network outputs a probability of the images, that is, whether the image is from the real data set (True) or generated by the style conversion network (Fake) [10].

4. Painting Art Style Rendering Based on the CycleGAN Model and Perceptual Confrontation Loss

4.1. Perceptual Adversarial Loss.
Although the PAN model [25] has been proven to measure the difference between images from a high-dimensional visual perception level, how to extract effective feature differences through the hidden layer is still an open question. The key to the problem is how to minimize the difference between the generated image and the real image in each high-dimensional level [26]. For this reason, this paper combines the PAN model and the perceptual loss proposed in [27] and defines the perceptual adversarial loss to consist of content feature loss, style loss, and adversarial loss. In the N-layer discriminant network, the image features are regarded as feature maps of N dimensions. If the size of each layer of feature maps is $H_i \times W_i$, the size of the feature map is $C_i \times H_i \times W_i$, where C represents the number of feature maps. Since each grid position of the image can be regarded as an independent sample, it can capture the key features. Perceptual adversarial loss is the weighted sum of content loss and style loss. It can punish the difference between the generated image and the style image when the first, fourth, sixth, and eighth hidden layers of the network are dynamically updated, so that the
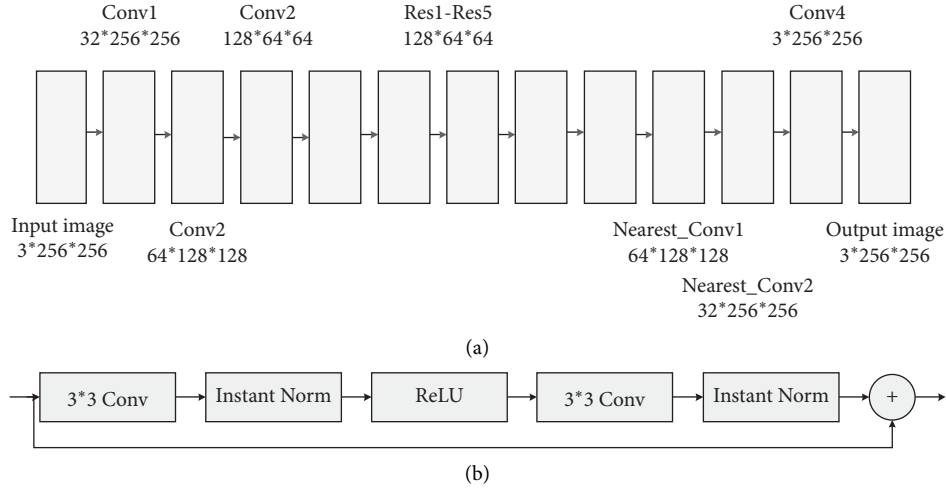
FIGURE 3: Image conversion network structure. (a) Artistic painting style rendering system model. (b) The residual structure used in the model.
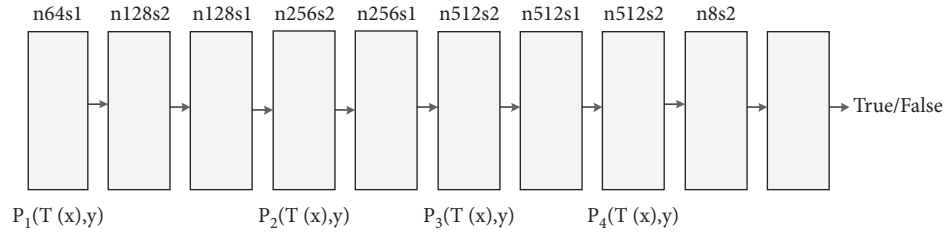
FIGURE 4: Discriminant network structure of the painting art style rendering system.

generated image has the best excellent content and style synthesis effect.

4.1.1. Content Loss Function.

The content loss function P_i uses Manhattan distance to calculate the image space loss (L_2) of the hidden generated image Y and the real painting art style image X as follows:

$$P_i = \frac{1}{N} \|H_i(Y) - H_i(X)\|, \tag{6}$$

where H_i represents the L_2 value of the i-th hidden layer of the discrimination network.

Based on the above definition, multiple levels of content loss can be expressed as

$$L_{\text{content}}(X, Y) = \sum_{i=1}^{N} \lambda_i P_i(X, Y), \tag{7}$$

where λ_i represents the balance factor of discriminating the N hidden layers i of the network. By minimizing the perceptual loss function L_{content}, the generated image and the content image have a similar content structure.

4.1.2. Style Loss Function.

Considering that the style loss function is used to penalize the deviation of the output image in style, including color and texture, this paper uses the style reconstruction method [28], which is obtained by the distance between the output image and the style image Gram matrix. Set $\varphi_i(x)$ as the feature map of the i-th hidden layer, so that the shape of $\varphi_i(x)$ is $C_i \times (H_i \times W_i)$, and the style loss value of the feature map of the i-th layer of the discriminant network can be expressed as

$$\text{Gram}_i(Y) = \frac{\varphi_i(Y)^* \varphi_i(Y)^T}{C_i \times H_i \times W_i},$$

$$G_i(Y_s, Y) = \|\text{Gram}_i(Y) - \text{Gram}_i(Y_s)\|. \tag{8}$$

In order to represent the style reconstruction from multiple levels, this paper defines $G_i(Y_s, Y)$ as a set of losses (sum of losses for each layer) as

$$L_{\text{style}}(Y_s, Y) = \sum_{i=1}^{n} G_i(Y_s, Y). \tag{9}$$

4.1.3. Perceptual Adversarial Loss Function.

The overall perceptual loss consists of a combination of content loss and style loss as a linear function, which can be defined as follows:

$$L_{\text{style}}(Y_s, Y) = \sum_{i=1}^{n} G_i(Y_s, Y), \tag{10}$$

where λ_c, λ_s are the weight parameters set based on human experience. The style conversion network T and the

discriminant network D are alternately optimized based on the overall perceptual loss value. The alternate optimization between such two network models is based on the method of PAN [29] model to realize the min-max confrontation. For the generated image Y, content image X, and style image Y_s, the loss function of network T and the loss function of network D are defined as follows:

$$L_T = \ln(1 - D(T(x))) + L_{\text{perc}},$$
$$L_D = -\ln(D(y)) - \ln(1 - D(T(x))) + \left[m - L_{\text{perc}}\right]^+. \tag{11}$$

We set a positive boundary value m in equation (11). Because the positive boundary value m can make the third term of L_D achieve gradient descent, we minimize L_T through the parameters of the network T, which can maximize the second and third terms of L_D at the same time. When L_T is less than m, the loss function L_D will update the discriminant network to a new high-dimensional level to calculate the remaining difference. Therefore, by perceiving against loss, the diversified differences between the generated image and the style image can be continuously perceived and explored.

4.2. Painting Art Segmentation Rendering System.

The generator of the traditional CycleGAN model adopts ResNet [21]. The generator is a fully convolutional connection type [30] and consists of an encoder, a converter, and a decoder. On the basis of the perceptual loss error, this paper improves the traditional CycleGAN network and uses CycleGAN with the DensNet [31] generator to achieve the style rendering of the image, as shown in Figure 5.

4.2.1. Encoder.

The CNN is used to extract features from the input image. The number of filters in the first convolutional layer (Conv_Layer_1) is set to 64. The output of the first convolutional layer is sent to the second convolutional layer (Conv_Layer_2) to continue to extract features, and the number of filters in the second convolutional layer is set to 128. The connection and data transmission method of the second convolutional layer and the third convolutional layer (Conv_Layer_3) is similar to that of the first and second convolutional layers. The number of filters in the third convolutional layer is set to 256. The image size input to the encoder is 256×256, and the encoder extracts 256 feature vectors with a size of 64×64.

4.2.2. Converter.

The purpose of converter is to combine the different extracted features and use these features to determine how to convert the feature vector of the image from the X domain (image style template) to the feature vector of the Y domain (generated image). The original CycleGAN network converter uses a 6-layer ResNet block to convert the feature vector, as shown in the following equation:

$$f_l = [H_l(f_{l-1})] + [f_{l-1}], \tag{12}$$

where H_l is a nonlinear transformation function. f_{l-1} and f_l are the input features and style conversion output features of the first layer as a ResNet block, including batch normalization layer [25], convolutional layer, and ReLU layer [32]. The converter designed in this paper uses DenseNet model to replace the traditional ResNet model. DenseNet model can reduce the disappearance of the gradient, enhance the feature transfer ability, and reduce the number of parameters to a certain extent. DenseNet model connects modules to each other, which improves the coupling capability of information flow between modules. The l-th module of the DenseNet model accepts the feature mapping from the previous module, as shown in the following equation:

$$f_l = H_l([f_0, f_1, \ldots, f_{l-1}]), \tag{13}$$

where $[f_0, f_1, \ldots, f_{l-1}]$ is formed by the series of generation features of $0, 1, \ldots, l-1$ layers. This paper integrates the DenseNet module in the converter to reduce model parameters while avoiding overfitting and reducing the amount of calculation.

4.2.3. Decoder.

The decoder process is opposite to that of the encoder, using a three-level deconvolution layer (Deconv_Layer) [33] to restore the low-level features of the image from the feature vector step by step, until the image is generated. The detailed internal composition of encoder, converter, and decoder is shown in Table 1.

4.2.4. Discriminator.

The discriminator of the improved network adopts the PatchGAN [34] classifier. In this model, we input the image to the discriminator, which discriminates whether the image is the original image or the image generated by the generator. The image input to the discriminator is divided into multiple 70×70 image blocks. In the discriminator's discrimination process, the network convolves the input image layer by layer, discriminates each image block through the one-dimensional output convolutional layer, and takes the average of the judgment results of all image blocks as the image judgment result.

5. Simulation Experiment and Result Analysis

5.1. Experiment Initialization.

The experimental environment used in this paper is as follows: CPU: Intel(R) Core(TM) i7-9700K@3.20 GHZ, 16 GB memory, operating system: 64 bit Windows 10, deep learning framework: Pytorch 1.10. This open source framework is mainly written in Python and can be used across platforms to implement various common CNN models and GAN models.

In order to verify the feasibility and effectiveness, the following model training and testing data sets are constructed for the perceptual adversarial loss error and the improved CycleGAN + PAN model. In the network training process, 983 oil painting images, 933 sketch painting images, and 1000 260×260 RGB ink wash paintings were selected as the image painting art style rendering data set. The training data used 1142 architectural images. There is no designated corresponding pairing relationship between different types
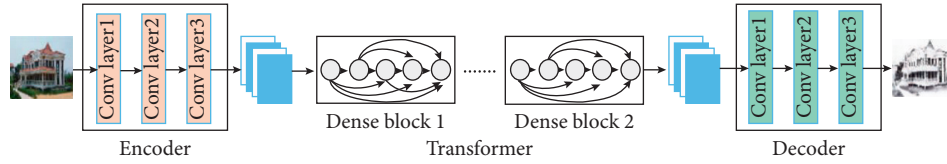
FIGURE 5: Generation network structure of the painting art style rendering system.

TABLE 1: Internal structure of the generation network of the painting art style rendering system.

Modules	Descriptions in detail			
Encoder	Convolutional kernel	Ways of network connecting	Number of filters	Stride of operations
1	7 * 7	Conv, BN, ReLU	64	1
2	5 * 5	Conv, BN, ReLU	128	1
3	3 * 3	Conv, BN, ReLU	256	2
Transformer	Modules			
1	Dense block 1			
. . .	. . . . . .			
5	Dense block 5			
Decoder	Convolutional kernel	Ways of network connecting	Number of filters	Stride of operations
1	3 * 3	De-conv, BN, ReLU	128	1/2
2	3 * 3	De-conv, BN, ReLU	64	1/2
3	7 * 7	De-conv, BN, ReLU	3	1

of data sets during the training process. As shown in Figure 6, they are CycleGAN, CycleGAN + PAN, and the image style rendering algorithm based on the improved CycleGAN + PAN that is proposed in this paper, Ground-truth ⟶ Oil painting, Ground-truth ⟶ Sketch painting, and Ground-truth ⟶ Ink wash painting for three styles of rendering model training.

5.2. Qualitative Results Analysis. In order to verify the feasibility and effectiveness of the rendering model proposed in this paper, the objective functions of CycleGAN, Cycle-GAN + PAN, and the improved CycleGAN + PAN are used in the experiment to perform image painting art style rendering experiments. The experimental results are shown in Figure 7. The first column in Figure 7 is the original image, the second column is the style image, and the third, fourth, and fifth columns are, respectively, experimental results of image style rendering for the CycleGAN model, Cycle-GAN + PAN model, and the improved CycleGAN + PAN model.

The first row, the second row, and the third row are experiments of three comparative models using oil painting, ink painting, and sketch painting as rendering styles. From the comparison of experimental results, it can be found that, under the same number of iterations, the proposed model can achieve painting art style rendering faster and achieve more realistic style rendering effects. Under the same experimental configuration and settings, this paper compares the generation results after 50,000 iterations. Table 2 shows the image style rendering results of CycleGAN, Cycle-GAN + PAN, and the proposed model using SSIM and PSNR image quality evaluation indicators for three different styles' rendering.

It can be seen from Table 2 that, in the results of the oil painting style rendering experiment, the PSNR of Cycle-GAN + PAN model is slightly higher than the proposed model. However, in the experimental results of the style rendering of ink and drawing styles, the SSIM and PSNR of the improved painting art style rendering of the proposed model are higher than other models.

5.3. Quantitative Results Analysis. CycleGAN introduces a cycle consistency loss function on the basis of traditional GAN unidirectional mapping and avoids model collapse to a certain extent through bidirectional mapping. The improved CycleGAN + PAN model proposed in this paper introduces the same mapping loss and perceptual loss function. Compared with CycleGAN + PAN model, the image style rendering process is more stable and the model is not easy to collapse. Figure 8 shows the comparative experimental results of some detailed painting art rendering between CycleGAN + PAN and the improved CycleGAN + PAN model proposed in this paper. Under the same number of iterations, the results of image style rendering are compared for two hard situations, which are marked as red box in Figure 8.

It can be seen from Figure 8 that CycleGAN + PAN uses two-way mapping to reduce the probability of model collapse, but instability sometimes occurs during training. As shown in Figure 8, there are black blocks in the image style rendering with the CycleGAN + PAN training model, and the style is lost to a certain extent. The third column shows the style rendering image generated by the improved CycleGAN + PAN. The result is that the image rendering style is well preserved, and there is no model crash. Therefore, the improved CycleGAN + PAN model proposed
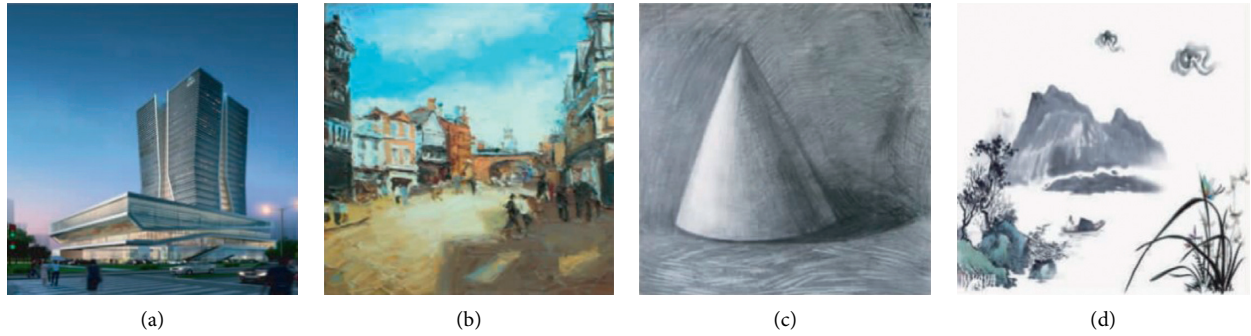
FIGURE 6: An example display of the experimental data set of painting art style rendering. (a) Ground-truth. (b) Oil painting. (c) Sketch painting. (d) Ink wash painting.

FIGURE 7: Experimental results of painting art style rendering. (a) Oil painting-based painting art style rendering. (b) Sketch painting-based painting art style rendering. (c) Ink wash painting-based painting art style rendering.

TABLE 2: Quality evaluation of the painting art style rendering experiment results.

Different styles	Indexes	CycleGAN	CycleGAN + PAN	Proposed method
Oil painting	SSIM	0.5451	0.7216	0.7456
	PSNR	18.6212	20.4515	20.4425
Sketch painting	SSIM	0.6315	0.6418	0.6987
	PSNR	18.5187	18.9251	20.2147
Ink wash painting	SSIM	0.6542	0.6781	0.7121
	PSNR	17.8154	18.1214	19.0241

| CycleGAN+PAN | Proposed | CycleGAN+PAN | Proposed | CycleGAN+PAN | Proposed |
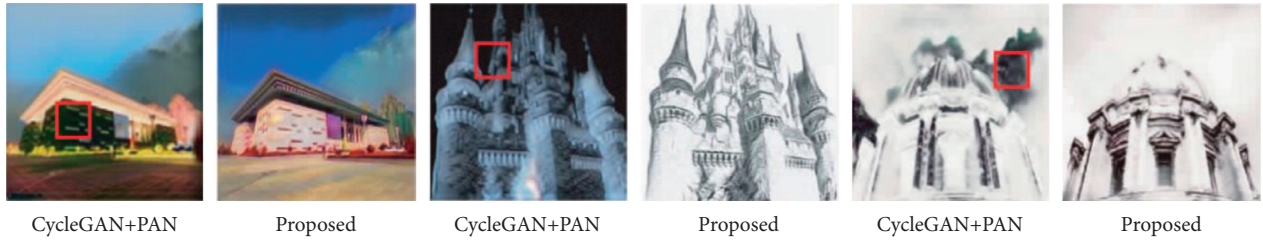
FIGURE 8: Effect of the abnormal situation of painting art style rendering between the original CycleGAN + PAN model and the improved CycleGAN + PAN model proposed in this paper.

(a)

(b)

(c)
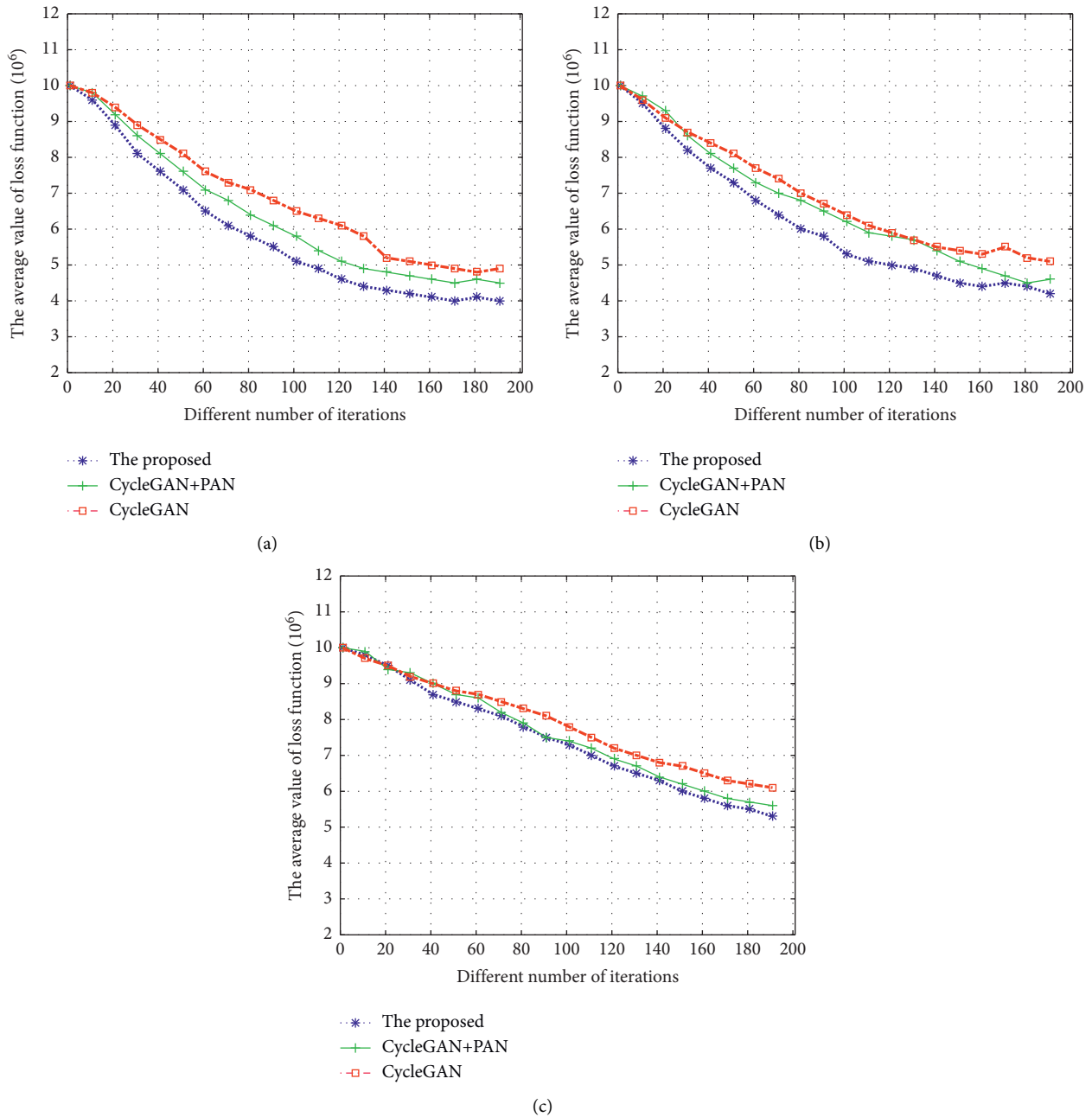
FIGURE 9: The comparison results of the loss function of the improved model in this article, the classic CycleGAN + PAN model, and the CycleGAN model. (a) 256 ∗ 256. (b) 515 ∗ 512. (c) 1024 ∗ 1024.
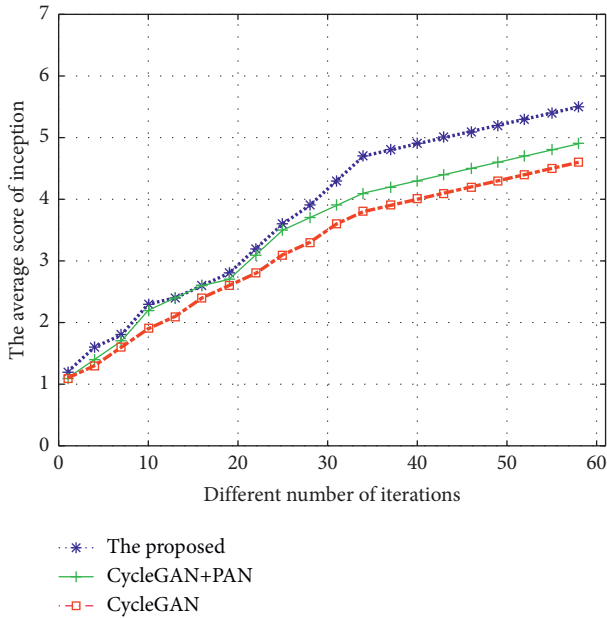
FIGURE 10: Comparison of the efficiency of the three comparison models in the rendering of artistic style of painting.

in the paper has good training conditions and is not easy to collapse under the same number of iterations.

The purpose of considering the painting art style rendering method is to achieve style loss and content loss, that is, to minimize equation (10). Therefore, the method proposed in this paper can be compared with the original CycleGAN + PAN [21] model in terms of measuring the extent of the successful reduction of the loss function. This paper uses the three compared methods to train 50, 100, and 200 images to record the value of equation (10) for each group of training. The loss function values of the three different sizes of images are shown in Figure 9. It can be seen from Figure 9 that, no matter for low-resolution or high-resolution images, the perceptual adversarial network can more effectively minimize the loss function.

The method proposed in this paper improves the convergence speed more than the original CycleGAN model. As shown in Figure 10, CycleGAN model, CycleGAN + PAN model, and improved CycleGAN + PAN model are trained on the same data set, and the inception score is used to measure the quality and variety of style rendering images. In the case of the same number of iterations, the higher the evaluation score is, the better the quality of the style rendering image is and the faster the convergence speed is obtained. In the experimental results of the three models, the improved CycleGAN + PAN model can converge faster and obtain higher evaluation scores. Through the convergence speed comparison experiment, we can see that the improved CycleGAN + PAN model proposed in this paper has a faster convergence speed and better stylized image quality.

6. Conclusion

In this paper, we select the characteristics of the PAN model and CycleGAN model and combine the advantages of such two models to construct the image painting art style rendering system. The discriminator updates parameters to explore the difference between generated images and style image, and the image conversion network uses the discriminant network to proceed before feed training until the difference is minimized to the optimal image conversion model. Such two models are alternately updated according to the perceptual loss. Experimental results have shown that, compared with the original CycleGAN network with the ResNet module, this paper adopts the improved CycleGAN + PAN image style rendering system based on perceptual confrontation loss. The style rendering effect is better, and the learned style can be rendered to new painting art style image to obtain a clearer performance effect. In future research, further optimization of the network structure and loss function will be considered in our researches. In terms of color and texture, the structural similarity and semantic consistency of content maps and style maps will be enhanced for novel combination of network structure and loss functions.

Data Availability

The experimental data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest regarding this work.

References

[1] P. Andreini, S. Bonechi, M. Bianchini, M. Alessandro, and S. Franco, "Image generation by GAN and style transfer for agar plate image segmentation," *Computer Methods and Programs in Biomedicine*, vol. 184, Article ID 105268, 2020.

[2] C. T. Lin, S. W. Huang, Y. Y. Wu, and H. L. Shang, "GAN-based day-to-night image style transfer for nighttime vehicle detection," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 2, pp. 951–963, 2020.

[3] S. Yang, Z. Wang, L. Jiaying, Z. Gao, N. Xu, and J. Liu, "Controllable artistic text style transfer via shape-matching gan," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 4442–4451, IEEE, Seoul, South Korea, October 2019.

[4] Y. Jing, Y. Yang, Z. Feng, Y. Jingwen, Y. Yizhou, and S. Mingli, "Neural style transfer: a review," *IEEE Transactions on Visualization and Computer Graphics*, vol. 26, no. 11, pp. 3365–3385, 2019.

[5] X. Chen, C. Xu, X. Yang, S. Li, and T. Dacheng, "Gated-gan: adversarial gated networks for multi-collection style transfer," *IEEE Transactions on Image Processing*, vol. 28, no. 2, pp. 546–560, 2018.

[6] L. Zhang, Y. Ji, and X. Lin, "Style transfer for anime sketches with enhanced residual u-net and auxiliary classifier gan," in *Proceedings of the 2017 4th IAPR Asian Conference on Pattern Recognition (ACPR)*, pp. 506–511, IEEE, Nanjing, China, December 2017.

[7] X. Deng, R. Yang, M. Xu, and L. D. Pier, "Wavelet domain style transfer for an effective perception-distortion tradeoff in

single image super-resolution," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 3076–3085, IEEE, Seoul, South Korea, October 2019.

[8] X. Deng, "Enhancing image quality via style transfer for single image super-resolution," *IEEE Signal Processing Letters*, vol. 25, no. 4, pp. 571–575, 2018.

[9] Y. Liu, "Improved generative adversarial network and its application in image oil painting style transfer," *Image and Vision Computing*, vol. 105, Article ID 104087, 2021.

[10] H. Fang, W. Deng, Y. Zhong, and H. Jiani, "Triple-GAN: progressive face aging with triple translation loss," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pp. 804-805, IEEE, Seattle, WA, USA, July 2020.

[11] J. Ye, Y. Ji, X. Wang, G. Xin, and M. Song, "Data-free knowledge amalgamation via group-stack dual-gan," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, August 2020, Article ID 12516.

[12] H. Emami, M. M. Aliabadi, M. Dong, and B. C. Ratna, "Spa-gan: spatial attention gan for image-to-image translation," *IEEE Transactions on Multimedia*, vol. 23, pp. 391–401, 2020.

[13] H. Liu, P. N. Michelini, D. Zhu, and G. Artsy, "A style transfer system with improved quality, diversity and performance," in *Proceedings of the 2018 24th International Conference on Pattern Recognition (ICPR)*, pp. 79–84, IEEE, Beijing, China, November 2018.

[14] A. A. Atapour and T. P. Breckon, "Real-time monocular depth estimation using synthetic data with domain adaptation via image style transfer," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2800–2810, IEEE, Salt Lake City, UT, USA, December 2018.

[15] R. Li, C. H. Wu, S. Liu, W. Jue, and W. Guangfu, "SDP-GAN: saliency detail preservation generative adversarial networks for high perceptual quality style transfer," *IEEE Transactions on Image Processing*, vol. 30, pp. 374–385, 2020.

[16] J. Svoboda, A. Anoosheh, C. Osendorfer, and M. Jonathan, "Two-stage peer-regularized feature recombination for arbitrary image style transfer," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, August 2020, Article ID 13816.

[17] Y. Zhang, Y. Zhang, and W. Cai, "Separating style and content for generalized style transfer," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 8447–8455, IEEE, Salt Lake City, UT, USA, June 2018.

[18] P. Li and M. Yang, "Semantic GAN: application for cross-domain image style transfer," in *Proceedings of the 2019 IEEE International Conference on Multimedia and Expo (ICME)*, pp. 910–915, IEEE, Shanghai, China, August 2019.

[19] Z. Ma, J. Li, N. Wang, and G. Xinbo, "Semantic-related image style transfer with dual-consistency loss," *Neurocomputing*, vol. 406, pp. 135–149, 2020.

[20] T. Wang, W. Q. Toh, H. Zhang, and S. Xiuchao, "RoboCoDraw: robotic avatar drawing with GAN-based style transfer and time-efficient path optimization," in *Proceedings of the AAAI Conference on Artificial Intelligence*, New York, NY, USA, February 2020.

[21] C. Liu, X. Chang, and Y. D. Shen, "Unity style transfer for person re-identification," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 6887–6896, IEEE, Seattle, WA, USA, August 2020.

[22] R. Kips, P. Gori, M. Perrot, and B. Isabelle, "Ca-gan: weakly supervised color aware gan for controllable makeup transfer," in *Proceedings of the European Conference on Computer Vision*, pp. 280–296, Springer, Glasgow, UK, August 2020.

[23] P. Wang, Y. Li, and N. Vasconcelos, "Rethinking and improving the robustness of image style transfer," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 124–133, New Orleans, LA, USA, November 2021.

[24] Y. Zhou, K. Yu, M. Wang et al., "Speckle noise reduction for oct images based on image style transfer and conditional GAN," *IEEE Journal of Biomedical and Health Informatics*, 2021.

[25] Y. Chong, C. Peng, J. Zhang, and P. Shaoming, "Style transfer for unsupervised domain-adaptive person re-identification," *Neurocomputing*, vol. 422, pp. 314–321, 2021.

[26] Z. Yang, Z. Hu, C. Dyer, and P. X. Eric, "Unsupervised text style transfer using language models as discriminators," in *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, pp. 7298–7309, Montréal, Canada, December 2018.

[27] S. Yang, J. Liu, W. Wang, and G. Zongming, "TET-GAN: text effects transfer via stylization and destylization," in *Proceedings of the AAAI Conference on Artificial Intelligence*, Honolulu, HI, USA, February 2019.

[28] Y. Men, Y. Mao, Y. Jiang, Y. M. Wei, and L. Zhouhui, "Controllable person image synthesis with attribute-decomposed gan," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5084–5093, IEEE, Seattle, WA, USA, August 2020.

[29] A. Zhu, X. Lu, X. Bai, U. Seiichi, K. I. Brian, and X. Shengwu, "Few-shot text style transfer via deep feature similarity," *IEEE Transactions on Image Processing*, vol. 29, pp. 6932–6946, 2020.

[30] Y. Xu, Y. Li, and B. S. Shin, "Medical image processing with contextual style transfer," *Human-centric Computing and Information Sciences*, vol. 10, no. 1, pp. 1–16, 2020.

[31] Z. Hu, J. Jia, B. Liu, F. Jianlong, and B. Yaohua, "Aesthetic-aware image style transfer," in *Proceedings of the 28th ACM International Conference on Multimedia*, pp. 3320–3329, Seattle, WA, USA, October 2020.

[32] M. Lu, H. Zhao, A. Yao, C. Yurong, and Z. Li, "Decoder network over lightweight reconstructed feature for fast semantic style transfer," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2469–2477, IEEE, Venice, Italy, December 2017.

[33] S. Zhou, M. Ke, and P. Luo, "Multi-camera transfer GAN for person re-identification," *Journal of Visual Communication and Image Representation*, vol. 59, pp. 393–400, 2019.

[34] Y. Huang, M. He, L. Jin, and W. Yongpan, "RD-GAN: few/zero-shot Chinese character style transfer via radical decomposition and rendering," in *Proceedings of the European Conference on Computer Vision*, pp. 156–172, Springer, Glasgow, UK, August 2020.

Research Article

Image Denoising Based on Improved Gaussian Mixture Model

Hui Wei [ID] **and Wei Zheng**

College of Electronic and Information Engineering, Henan Institute of Technology, Xinxiang 453000, China

Correspondence should be addressed to Hui Wei; dxweihui@hait.edu.cn

An image denoising method is proposed based on the improved Gaussian mixture model to reduce the noises and enhance the image quality. Unlike the traditional image denoising methods, the proposed method models the pixel information in the neighborhood around each pixel in the image. The Gaussian mixture model is employed to measure the similarity between pixels by calculating the L2 norm between the Gaussian mixture models corresponding to the two pixels. The Gaussian mixture model can model the statistical information such as the mean and variance of the pixel information in the image area. The L2 norm between the two Gaussian mixture models represents the difference in the local grayscale intensity and the richness of the details of the pixel information around the two pixels. In this sense, the L2 norm between Gaussian mixture models can more accurately measure the similarity between pixels. The experimental results show that the proposed method can improve the denoising performance of the images while retaining the detailed information of the image.

1. Introduction

Due to the effects of different illumination and changes in sensor temperature during image acquisition, transmission, and digitization, random noises will be introduced into the acquired images, and these noises often exhibit Gaussian characteristics [1–4]. The existence of noise will not only affect the visual effect of the image but also further affect the subsequent processing of the image, such as image feature extraction, image classification and recognition, and so on. Therefore, before image processing, it is necessary to perform denoising processing on the acquired image to improve the quality of the image and facilitate the postprocessing of the image. There are many methods of image denoising, which are often divided into frequency-domain filtering and spatial-domain filtering. Common frequency-domain filtering methods include wavelet denoising [5], high-pass filtering, Wiener filtering [6], etc. The spatial filtering methods are also rich, such as partial differential equations, variational methods [7], statistical methods [8], and so on, which are widely used in practical applications. The classic partial differential equation method includes mean filtering, which is an isotropic

filtering algorithm that can remove noise very well. However, because the image information is also averaged while averaging the noises, the effective information in the original images becomes vague at the same time. To handle this problem, many researchers have improved the isotropy. A typical example is the anisotropic filtering method proposed by Perona, which can change the weight coefficient when averaging the noise and image information. In this way, the smoothing effect of the noise is good while the effective information can be maintained. The anisotropic filtering can remove noises very well, but because the image information is less smooth, its denoising effect is poor in areas with rich details. In addition, the blocking effects are prone to appear in these methods. As a remedy, Le Montagner defined the similarity between pixels according to the difference of pixel gray levels, the so-called Yaroslavsky filter [9]. The Yaroslavsky filter can change the smoothing weight coefficient according to the pixel information of the image, and its denoising effect is better than the general anisotropic filter. Moreover, it can well retain the detailed information of the image. On the basis of Yaroslavsky filtering, Tomasi and Manduchi combined the grayscale difference and spatial distance between pixels to

define the similarity between pixels and obtained the bilateral filtering [10]. Bilateral filtering not only considers the impact of grayscale differences on the weights but also considers the impact of pixels at different distances. Compared with mean filtering and Gaussian filtering, the bilateral filtering can be achieved in both flat areas of the image and areas with rich details. At the same time, the detailed information of the image can be well retained. However, the denoising performance of the bilateral filtering is limited in areas with rich image texture. In recent years, some improved algorithms have also appeared for the bilateral filtering [11–15], mainly focusing on the selection of filtering parameters and the improvement of algorithm efficiency. Among them, Ghosh and Chaudhury [16] proposed the concept of distance kernel and applied it to the definition of pixel similarity weights, especially in areas with rich details. The denoising performance of the algorithm is greatly improved, and the calculation efficiency of the algorithm is also optimized. In recent years, new signal processing algorithms represented by compressed sensing and machine learning algorithms represented by deep learning have also been widely used in the field of image denoising. The works in [17–20] are based on the theory of compressed sensing and realize image reconstruction and denoising through the method of sparse representation. The works in [21–24] employ a variety of deep learning models for noise image processing and achieve a good denoising effect.

The above methods mainly use information such as pixel gray difference and spatial geometric distance to measure the similarity between pixels. This paper uses the Gaussian mixture model to model the pixel information in the neighborhood around the pixel. The L2 norm and the spatial distance between Gaussian mixture models are combined to define the similarity weight between pixels. The Gaussian mixture model of the pixel information in the image area represents the pixel grayscale and the richness of details in the local area of the image. Based on the spatial distance between pixels, the pixel gray intensity and the richness of details in the local area of the image can be more accurately measured. The similarity between the two improves the denoising performance of the algorithm and maintains the detailed information of the image. In the experiment, some classic image samples are used to test the proposed method, and the adaptability of the method to different noise samples is tested through the noise level. After comparative analysis with several existing methods, the experimental results verify the superior performance of the proposed method for image denoising.

2. Image Denoising Based on Gaussian Mixture Model

The basic idea of image denoising based on the Gaussian mixture model is as follows. First, the pixel information of the neighborhood around each pixel is used to estimate the parameters of the Gaussian mixture model. The pixel information is modeled as a Gaussian mixture model to obtain a prefiltered image. Afterwards, the L2 norm between the

Gaussian mixture models corresponding to the two pixels is calculated. The L2 norm and the spatial position distance between the two pixels are combined to define the similarity weight between the two pixels. Finally, the weighted smoothing filtering is performed on each pixel to obtain a denoised image.

Different from the traditional spatial denoising methods, the image denoising method based on the Gaussian mixture model defines the similarity weight between pixels according to the statistical difference of the information around the pixels. The detailed steps of the proposed method can be summarized as follows:

Step 1: for a certain pixel, the pixel information in its surrounding neighborhood is used to estimate the parameters of the Gaussian mixture model to obtain a Gaussian mixture model.

Step 2: a Gaussian mixture model is estimated for each pixel in the image, and all Gaussian mixture models constitute a prefiltered image.

Step 3: the L2 norm between the Gaussian mixture models corresponding to two pixels is calculated, which is combined with the spatial distance between the two pixels to define the similarity between the pixels.

Step 4: each pixel in the image is subjected to weighted smoothing filtering to obtain a filtered image.

It can be seen from the above steps that image denoising based on the Gaussian mixture model is based on two pixels x and y. The spatial distance between the two pixels and the L2 norm between the corresponding Gaussian mixture models define the similarity weight $w(x, y)$. Afterwards, the weighted smoothing and filtering are performed to obtain the filtered image as follows:

$$f(x) = \frac{1}{S(x)} \int_{\Psi(x)} w(x, y) I(y) \mathrm{d}y, \qquad (1)$$

where $S(x) = \int_{\Psi(x)} w(x, y) \mathrm{d}y$ is to normalize the integral value; $\Psi(x)$ is the neighborhood window with the center of the pixel x; and $I(y)$ represents the gray value of the pixels y. The similarity weight $w(x, y)$ in equation (1) is calculated as follows:

$$w(x, y) = \exp\left\{-\frac{d(x, y) L2(G(x), G(y))}{r^2}\right\}, \qquad (2)$$

where $G(x)$ and $G(y)$ represent the Gaussian mixture models of the pixels x and y, respectively; $d^2(x, y) = |x - y|^2$ calculates the spatial distance between x and y; and the parameter r is the filter control coefficient. The larger the value of r, the better the smoothing effect of the image and the higher the loss of the image's detail information. On the contrary, the smaller the value of r, the worse the smoothing effect of the image and the smaller the loss of image's detail information. Therefore, it is necessary to select appropriate filter control coefficients in the filtering process. The basic idea of the image denoising method based on the Gaussian mixture model is shown in Figure 1.
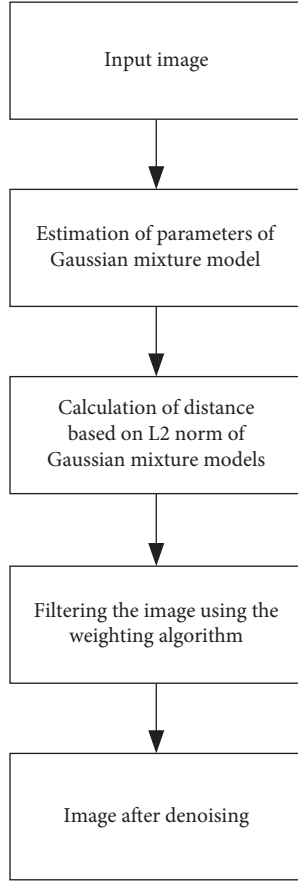
Figure 1: Main steps of image denoising based on the Gaussian mixture model.

3. L2 Norm between Gaussian Mixture Models

For each pixel, the pixel information in its surrounding neighborhood is used to estimate the parameters of the Gaussian mixture model to obtain a Gaussian mixture model. The Gaussian mixture model corresponding to the pixel indicates the local grayscale intensity and the richness of the details of the pixel information in the neighborhood around the pixel. The L2 norm between two Gaussian mixture models and the spatial distance between the pixels are used to define the similarity between pixels. According to the similarity weights, each pixel in the image is weighted and smoothed. The following mainly introduces the Gaussian mixture modeling of pixel information and the calculation of the L2 norm between two Gaussian mixture models.

3.1. Gaussian Mixture Model Estimation of Pixel Information.
For a certain pixel x, the neighboring window $\Psi(x)$ has the size of $M \times M$ including M^2 pixels. The gray value of the pixel at the location (i, j) is denoted as $I(i, j)$. The information of these pixel is used to estimate a Gaussian mixture model as follows:

$$G(x) = \sum_{m=1}^{K} \alpha_m N\left(\mu_m, \sigma_m^2\right)\pi, \tag{3}$$

where α_m is the mixed weight coefficient; $N(\mu_m, \sigma_m^2)$ represents a Gaussian distribution with the mean of μ_m and variance of σ_m^2; and K represents the number of mixture components of the Gaussian mixture model. In Gaussian mixture modeling, the number of mixture components is usually set in advance.

In the following, the expectation-maximization (EM) algorithm is employed to estimate the parameters of the Gaussian mixture model. The EM algorithm provides an iterative estimation way to maximize the posterior probability. Assuming that the parameters of the Gaussian mixture model are $\theta = \{\alpha_1, \ldots, \alpha_K, \mu_1, \ldots, \mu_K, \sigma_1^2, \ldots, \sigma_K^2\}$, there are n^2 samples to perform the estimation, where y_i, $i = 1, \ldots, n^2$ represents the ith sample value and γ_{im} denotes that sample y_i belongs to the mth class. Given y and θ, γ_{im} can be calculated according to the maximum posterior probability as follows:

$$\begin{aligned}
\widehat{\gamma}_{im} &= P(z = m \mid y_i, \theta) \\
&= \frac{P(z = m, y_i | \theta)}{P(y_i | \theta)} \\
&= \frac{P(z = m | \theta) P(y_i | z = m, \theta)}{\sum_{m=1}^{K} P(z = m | \theta) P(y_i | z = m, \theta)} \\
&= \frac{\alpha_m N(y | \theta_m)}{\sum_{m=1}^{K} \alpha_m N(y | \theta_m)},
\end{aligned} \tag{4}$$

where z is the class label and $P(\cdot)$ represents the probability. Equation (4) gives the way to obtain $\gamma_{im}, i = 1, \ldots, n^2, m = 1, \ldots, K$, and the remaining parameters can be calculated as follows:

$$\begin{aligned}
\alpha_m &= \frac{\sum_{i=1}^{n^2} \widehat{\gamma}_{im}}{n^2}, \\
\mu_m &= \frac{\sum_{i=1}^{n^2} y_i \widehat{\gamma}_{im}}{\sum_{i=1}^{n^2} \widehat{\gamma}_{im}}, \\
\sigma_m^2 &= \frac{\sum_{i=1}^{n^2} \widehat{\gamma}_{im} (y_i - \mu_m)^2}{\sum_{i=1}^{n^2} \widehat{\gamma}_{im}}.
\end{aligned} \tag{5}$$

According to equations (4) and (5), the EM algorithm can be used to iteratively estimate the parameters of the Gaussian mixture model. The detailed steps of parameter estimation are summarized as follows:

Step 1: initialize value of the parameter set θ.

Step 2: according to the maximum posterior probability, $\widehat{\gamma}_{im}$ can be calculated according to equation (4).

Step 3: based on the estimation of $\widehat{\gamma}_{im}$ from step 2, the remaining parameters $\{\alpha, \mu, \sigma^2\}$ of the Gaussian mixture model are calculated according to equation (5).

Repeat step 2 and step 3 until convergence.

Based on the above steps, the parameters of the Gaussian mixture model of each pixel can be estimated. The Gaussian

mixture model corresponding to all pixels constitutes the prefiltered image. In the prefiltered image, the L2 range between the Gaussian mixture models corresponding to the two pixels and the spatial distance are combined to calculate the similarity weight between two pixels.

3.2. Distance Measure for Gaussian Mixture Models.
Assuming that the pixel information around each pixel is modeled as K mixed components, the Gaussian mixture models for two pixels x and y are formulated as follows:

$$
\begin{aligned}
G(x; \theta_x) &= \sum_{i=1}^{K} \alpha_i N(t; \mu_i, \sigma_i^2), \\
G(y; \theta_y) &= \sum_{i=1}^{K} \beta_j N(t; \nu_j, \varepsilon_j^2).
\end{aligned}
\tag{6}
$$

Actually, there are many calculation methods for the distance measurement, such as the KL divergence, Monte Carlo method, and so on. However, the KL divergence does not have an analytical expression and can only be calculated in an approximate way. The Monte Carlo method requires too much calculation, which is not suitable in real-time processing. As a remedy, this paper uses the L2 norm to calculate the difference between two Gaussian mixture models. The L2 norm between the Gaussian mixture models has an analytical expression, which is convenient for the implementation of the algorithm.

For the above two Gaussian mixture models, the L2 norm can be calculated as follows:

$$
d\big(G(x; \theta_x), G(y; \theta_y)\big) = \int_0^{\infty} \big\{ G(x; \theta_x) - G(y; \theta_y) \big\}^2 \mathrm{d}t.
\tag{7}
$$

It can be expanded as follows:

$$
\begin{aligned}
d\big(G(x; \theta_x), G(y; \theta_y)\big) &= \int_0^{\infty} \big\{ G(x; \theta_x) - G(y; \theta_y) \big\}^2 \mathrm{d}t \\
&= \int_0^{\infty} \big\{ G^2(x; \theta_x) + G^2(y; \theta_y) - 2G(x; \theta_x) G(y; \theta_y) \big\} \mathrm{d}t \\
&= \int_0^{\infty} G^2(x; \theta_x) \mathrm{d}t + \int_0^{\infty} G^2(y; \theta_y) \mathrm{d}t - 2 \times \int_0^{\infty} G(x; \theta_x) G(y; \theta_y) \mathrm{d}t.
\end{aligned}
\tag{8}
$$

Substituting equation (6) into equation (8), the simplification can be obtained as follows:

$$
\begin{aligned}
d\big(G(x; \theta_x), G(y; \theta_y)\big) &= \sum_{i=1}^{K} \sum_{j=1}^{K} \alpha_i \alpha_j \int_0^{\infty} N(t; \mu_i, \sigma_i^2) N(t; \mu_j, \sigma_j^2) \mathrm{d}t + \sum_{i=1}^{K} \sum_{j=1}^{K} \beta_i \beta_j \int_0^{\infty} N(t; \nu_i, \varepsilon_i^2) N(t; \nu_j, \varepsilon_j^2) \mathrm{d}t \\
&\quad - 2 \times \sum_{i=1}^{K} \sum_{j=1}^{K} \alpha_i \beta_j \int_0^{\infty} N(t; \mu_i, \sigma_i^2) N(t; \nu_j, \varepsilon_j^2) \mathrm{d}t.
\end{aligned}
\tag{9}
$$

According to the integral nature of Gaussian distribution, the following can be obtained:

$$
\int_0^{\infty} N(t; \mu_i, \sigma_i^2) N(t; \nu_j, \varepsilon_j^2) \mathrm{d}t = N(0; \mu_i - \nu_j, \sigma_i^2 + \varepsilon_j^2).
\tag{10}
$$

Equation (10) can be further simplified as follows:

$$
\begin{aligned}
d\big(G(x; \theta_x), G(y; \theta_y)\big) &= \sum_{i=1}^{K} \sum_{j=1}^{K} \alpha_i \alpha_j N(0; \mu_i - \mu_j, \sigma_i^2 + \sigma_j^2) + \sum_{i=1}^{K} \sum_{j=1}^{K} \beta_i \beta_j N(0; \nu_i - \nu_j, \varepsilon_i^2 + \varepsilon_j^2) \\
&\quad - 2 \times \sum_{i=1}^{K} \sum_{j=1}^{K} \alpha_i \beta_j N(0; \mu_i - \nu_j, \sigma_i^2 + \varepsilon_j^2).
\end{aligned}
\tag{11}
$$

According to equation (11), the L2 norm distance between two Gaussian mixture models can be smoothly calculated.

4. Experiment and Discussion

4.1. Evaluation Indexes. The performance of image denoising algorithms can be generally evaluated by two aspects, i.e., how much noise is removed and how much image detail information is lost. The more the noise is removed, the stronger the denoising ability of the denoising algorithm is. On the contrary, the less the noise is removed, the worse the denoising ability of the algorithm is. At the same time, the more the image detail information is retained, the better the image quality is. The less the image detail information is retained, the worse the image quality is. While removing image noise, the detail information of the image should be preserved as much as possible. This paper adopts the peak signal-to-noise ratio (PSNR) to measure the denoising ability of the algorithm. At the same time, the structure similarity index measure (SSIM) is used to evaluate the quality of the image after denoising, that is, the ability to retain the detailed information of the image. With a large PSNR and a SSIM approaching 1, the denoising performance of the method is good.

Generally, the PSNR can be calculated as follows:

$$\text{PSNR} = 10 \ \log_{10}\left(\frac{v_{\max}^2}{\text{mean}\left\{ \left[v(i) - u(i) \right]^2 \right\}} \right), \tag{12}$$

where $v_{\max}$ indicates the maximum pixel value in the image (for an image with grayscale value from 0 to 255, $v_{\max} = 255$); $v(i)$ and $u(i)$ represent the gray values of the noisy image and the noise-free image, respectively; and $\text{mean}[v(i) - u(i)]^2$ refers to the average power of noise.

SSIM is an evaluation index for image denoising defined according to the human visual mechanism. From a visual point of view, the more detailed the information of the image, the clearer the image, and vice versa. SSIM is an evaluation index that combines the three characteristics of image contrast, edge structure, and image brightness and is obtained by weighted product. SSIM can be calculated by the following equation [25]:

$$\text{SSIM}(i, j) = \frac{\left(2m_i m_j + c_1 \right)\left(2s_{ij} + c_2 \right)}{\left(m_i^2 + m_j^2 + c_1 \right)\left(s_i^2 + s_j^2 + c_2 \right)}, \tag{13}$$

where m_i and m_j are the mean values corresponding to the pixel positions i and j, respectively; s_i^2 and s_j^2 are the variances; and s_{ij} represents the covariance.

In this paper, $c_1 = (k_1 L)^2$, $c_2 = (k_2 L)^2$, L indicates the range of pixel values, and k_1 and k_2 are the weight coefficients, which are generally set as $k_1 = 0.01$ and $k_2 = 0.03$. The index mSSIM is the average of the SSIMs from different windows in an image.

4.2. Analysis of Experimental Results. In order to verify the effectiveness of the proposed algorithm (denoted as GMMD), several reference methods are employed for comparison including the bilateral filtering (BF), mean filtering (MF), kernel bilateral filtering (KBF), sparse representation (SR), and deep learning (DL). The indexes PSNR and SSIM are used to evaluate the denoising performance of different methods. The following experiment uses 4 images: "Lena," "Academy," "Einstein," and "Mandrill," as shown in Figure 2, which are classical samples for the test and evaluation of image denoising algorithms.

Figure 3 shows the denoising results achieved by the proposed method for a local area from the "Lena" image. It is clearly shown that the proposed method can effectively remove the noises while maintaining the detail information in the image. Furthermore, different levels of noises (denoted by the noise variance) are added in the 4 samples and different denoising methods are examined. The performance of those methods is summarized in Tables 1 and 2, which use the PSNR and mSSIM, respectively.

From the results in Tables 1 and 2, it can be seen that the denoising and detail information retention ability of the GMMD algorithm and the KBF algorithm is better than that of the MF and BF algorithms. In the relatively flat image denoising results of "Lena" and "Einstein," GMMD's denoising ability and detail information retention ability are better than those of the KBF algorithm. However, in the images "Academy" and "Mandrill" with richer image detail information, the denoising performance improvement of the GMMD algorithm is not obvious. The results show that the combination of the local average gray intensity of image pixel information and the richness of details to define the similarity weight between pixels is more accurate compared to traditional methods. Compared with the two emerging methods of SR and DL, the proposed method is very close to their denoising performance but has a slight advantage, reflecting its effectiveness. It shows the advantages of the Gaussian mixture model in image representation and similarity measurement. In addition, this paper also calculates the average time required for these three types of methods to process a single noisy image on the same hardware platform, which is 1.8 ms, 2.3 ms, and 4.7 ms, respectively. In contrast, the proposed method has better computing efficiency and more practicality. The DL method generally requires an offline training process, which reduces its overall efficiency.

(a)

(b)

(c)

(d)
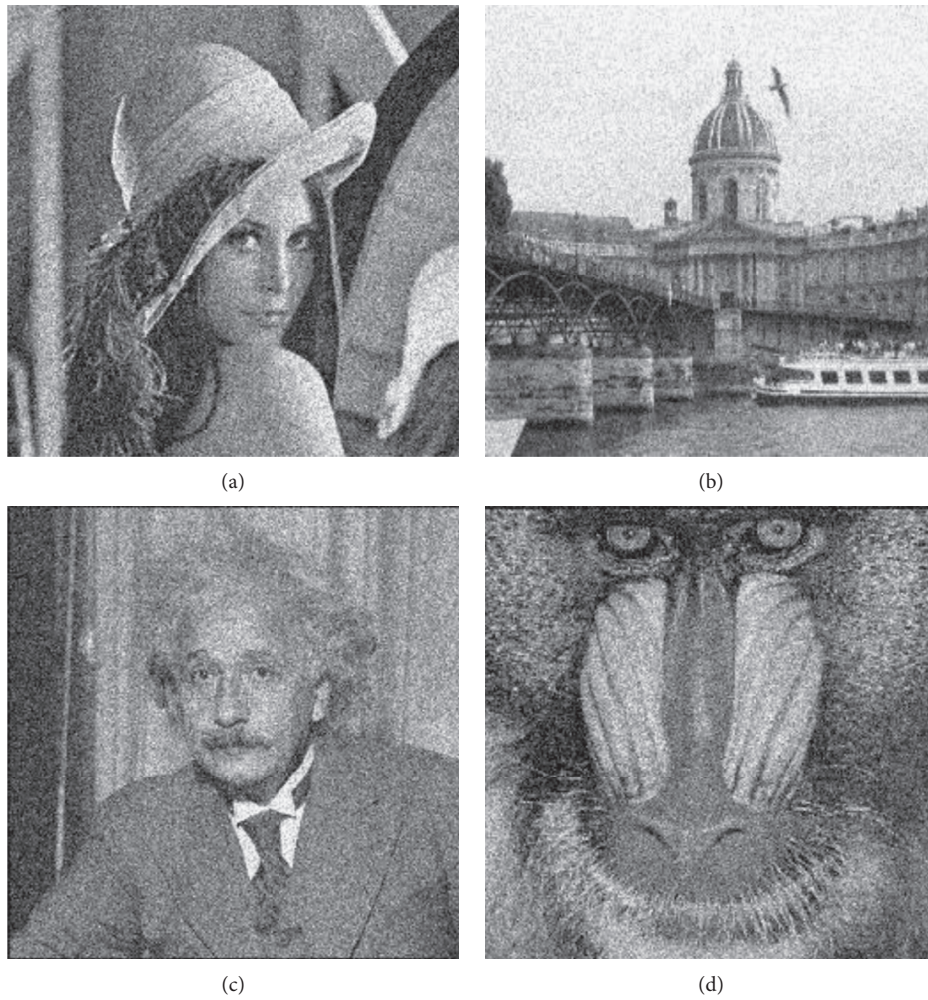
FIGURE 2: Illustration of the noisy images. (a) Lena. (b) Academy. (c) Einstein. (d) Mandrill.

(a)

(b)
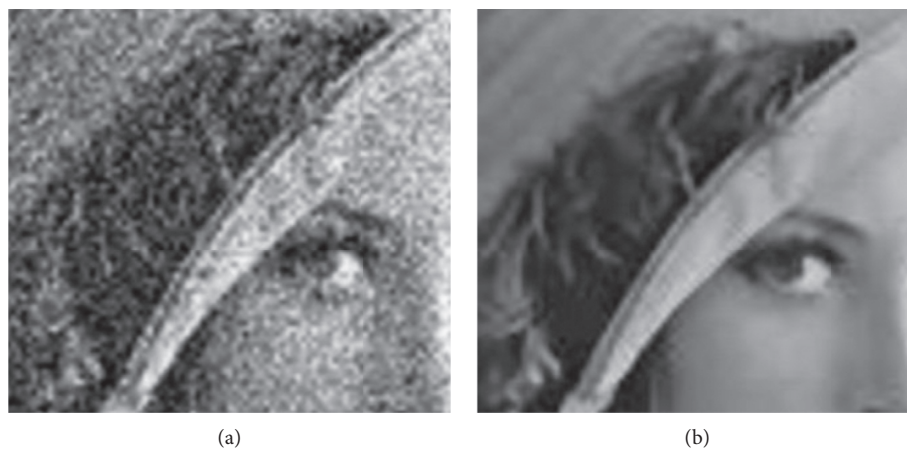
FIGURE 3: Denoising results of local area of "Lena" image using the proposed method. (a) Original noisy image. (b) After denoising.

TABLE 1: PSNR achieved by different denoising methods (dB).

| Image label | Noise level | Noisy image | Denoising image | | | | | |
			BF	MF	KBF	SR	DL	GMMD
Lena	0.2	11.017	11.773	11.721	11.652	11.823	11.903	12.008
	0.15	12.833	14.102	13.979	14.198	14.262	14.277	14.330
	0.1	15.933	17.211	16.727	17.018	17.232	17.265	17.347
	0.05	17.068	19.595	19.780	20.928	21.602	21.674	21.770
Academy	0.2	15.262	16.652	15.644	16.456	16.442	16.543	16.969
	0.15	16.381	18.520	17.007	18.290	18.764	18.788	18.959
	0.1	17.423	20.019	18.394	20.451	20.721	20.734	20.805
	0.05	18.222	19.437	19.445	22.663	22.887	22.924	23.043
Einstein	0.2	8.121	8.444	8.554	8.405	8.532	8.568	8.649
	0.15	11.025	11.828	11.834	11.719	12.014	12.087	12.168
	0.1	12.897	14.182	14.129	14.029	14.226	14.276	14.425
	0.05	14.978	17.266	17.001	17.076	17.231	17.239	17.328
Mandrill	0.2	14.119	15.518	14.953	15.421	15.672	15.693	15.867
	0.15	15.471	17.409	16.419	17.148	17.842	17.902	17.961
	0.1	11.065	11.774	11.496	11.604	11.789	11.802	11.843
	0.05	12.894	14.074	13.526	13.790	14.614	14.645	14.711

TABLE 2: mSSIM achieved by different denoising methods.

| Image label | Denoising image | | | | | |
	BF	MF	KBF	SR	DL	GMMD
Lena	0.782	0.777	0.714	0.812	0.817	0.821
Academy	0.862	0.742	0.835	0.876	0.882	0.885
Einstein	0.736	0.703	0.673	0.762	0.769	0.776
Mandrill	0.694	0.477	0.757	0.782	0.785	0.789

5. Conclusion

This paper proposes an image denoising method based on the Gaussian mixture model. The statistical characteristics of the local gray information of pixels around the pixel are modeled as a Gaussian mixture model. The L2 norm between Gaussian mixture models and the spatial distance of pixel positions are combined to define the similarity weight between pixels. The weighted smoothing filter is used to denoise the image. The experimental results show that the denoising performance of the proposed method is more obvious in the relatively flat area of the image. In addition, the denoising performance of the proposed method is improved to a certain extent in the area with rich detailed information. The comparison between the proposed method and several existing image denoising algorithms validates the superior effectiveness of the used method.

Data Availability

The images used in this study are publicly available.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

References

[1] Z. A. Mustafa and Y. M. Kadah, "Multi resolution bilateral filter for MR image denoising," in *Proceedings of the 2011 1st Middle East Conference on Biomedical Engineering*, pp. 180–184, Sharjah, UAE, February 2011.

[2] J. Li, J. Yu, L. Xu et al., "A cascaded algorithm for image quality assessment and image denoising based on cnn for image security and authorization," *Security and Communication Networks*, vol. 2018, Article ID 8176984, 13 pages, 2018.

[3] Y. Feng, Yu Shi, and D. Sun, "Blind poissonian image deblurring regularized by a denoiser constraint and deep image prior," *Mathematical Problems in Engineering*, vol. 2020, Article ID 9483521, 15 pages, 2020.

[4] J. V. Manjón, N. A. Thacker, J. J. Lull, and G. Garcia-Martí, "Multicomponent MR image denoising," *International Journal of Biomedical Imaging*, vol. 2009, Article ID 756897, 10 pages, 2009.

[5] Da Zhang and R. Liu, "Laser Doppler signal denoising based on wavelet packet thresholding method," *International Journal of Optics*, vol. 2019, Article ID 1097292, 11 pages, 2019.

[6] F. Wu, W. Yang, L. Xiao, and J. Zhu, "Adaptive wiener filter and natural noise to eliminate adversarial perturbation," *Electronics*, vol. 9, no. 10, p. 1634, 2020.

[7] H. Salehi, J. Vahidi, T. Abdeljawad, A. Khan, and S. Y. B. Rad, "A sar image despeckling method based on an extended adaptive wiener filter and extended guided filter," *Remote Sensing*, vol. 12, no. 15, p. 2371, 2020.

[8] L. Bedini and A. Tonazzini, "Image restoration preserving discontinuities: the Bayesian approach and neural networks," *Image and Vision Computing*, vol. 10, no. 2, pp. 108–118, 1992.

[9] Y. Le Montagner, E. D. Angelini, and J.-C. Olivo-Marin, "An unbiased risk estimator for image denoising in the presence of mixed Poisson-Gaussian noise," *IEEE Transactions on Image Processing*, vol. 23, no. 3, pp. 1255–1268, 2014.

[10] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *Proceedings of the 6th International Conference on Computer Vision (IEEE Cat. No.98CH36271)*, pp. 839–846, Bombay, India, January 1998.

[11] A. Me and B. Ks, "Optimal bilateral filter and convolutional neural network based denoising method of medical image measurements," *Measurement*, vol. 143, pp. 125–135, 2019.

[12] S. Ghosh, P. Nair, and K. N. Chaudhury, "Optimized fourier bilateral filtering," *IEEE Signal Processing Letters*, vol. 25, no. 10, pp. 1555–1559, 2018.

[13] S.-J. Jang and Y. Hwang, "Noise-aware and light-weight vlsi design of bilateral filter for robust and fast image denoising in mobile systems," *Sensors*, vol. 20, no. 17, p. 4722, 2020.

[14] S. Ghosh, R. G. Gavaskar, D. Panda, and K. N. Chaudhury, "Fast scale-adaptive bilateral texture smoothing," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 7, pp. 2015–2026, 2020.

[15] J. Seetha and S. S. Raja, "Denoising of MRI images using filtering methods," in *Proceedings of the 2016 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET)*, pp. 765–769, Chennai, India, March 2016.

[16] S. Ghosh and K. N. Chaudhury, "On fast bilateral filtering using fourier kernels," *IEEE Signal Processing Letters*, vol. 23, no. 5, pp. 570–573, 2016.

[17] M. Kang, M. Jung, and M. Kang, "Rician denoising and deblurring using sparse representation prior and nonconvex total variation," *Journal of Visual Communication and Image Representation*, vol. 54, pp. 80–99, 2018.

[18] S. Li, H. Yin, and L. Fang, "Group-sparse representation with dictionary learning for medical image denoising and fusion," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 12, pp. 3450–3459, 2012.

[19] G. Ma, T.-Z. Huang, J. Huang, and C.-C. Zheng, "Local low-rank and sparse representation for hyperspectral image denoising," *IEEE Access*, vol. 7, pp. 79850–79865, 2019.

[20] S. Q. Liu, S. H. Xiao, and Y. L. An, "Bayesian Shearlet shrinkage for SAR image de-noising via sparse representation," *Multidimensional Systems and Signal Processing*, vol. 25, no. 4, pp. 683–701, 2014.

[21] S. Chen, D. Shi, M. Sadiq, and X. Cheng, "Image denoising with generative adversarial networks and its application to cell image enhancement," *IEEE Access*, vol. 8, pp. 82819–82831, 2020.

[22] A. J. Reader, G. Corda, A. Mehranian, C. d. Costa-Luis, S. Ellis, and J. A. Schnabel, "Deep learning for PET image reconstruction," *IEEE Transactions on Radiation and Plasma Medical Sciences*, vol. 5, no. 1, pp. 1–25, 2021.

[23] C. Tang, Q. Zhu, W. Wu, W. Huang, C. Hong, and X. Niu, "PLANET: improved convolutional neural networks with image enhancement for image classification," *Mathematical Problems in Engineering*, vol. 2020, Article ID 1245924, 10 pages, 2020.

[24] J. Hou, Y. Si, and X. Yu, "A novel and effective image super-resolution reconstruction technique via fast global and local residual learning model," *Applied Sciences*, vol. 10, no. 5, p. 1856, 2020.

[25] L. Wang, W. Zheng, X. Ma, and S. Lin, "Denoising speech based on deep learning and wavelet decomposition," *Scientific Programming*, vol. 2021, Article ID 8677043, 10 pages, 2021.

Research Article

Prediction of Urban and Rural Tourism Economic Forecast Based on Machine Learning

Wusheng Zhou [ID]

Business School, Guangxi University, Nanning, Guangxi 530004, China

Correspondence should be addressed to Wusheng Zhou; robin@gxu.edu.cn

With the rapid development of tourism, tourism revenue, as one of the important indicators to measure the development of the tourism economy, has high research value. The quasi-prediction of tourism revenue can drive the development of a series of related industries and accelerate the development of the domestic economy. When forecasting tourism income, it is necessary to examine the causal relationship between tourism income and local economic development. The traditional cointegration analysis method is to extract the promotion characteristics of tourism income to the local economy and construct a tourism income prediction model, but it cannot accurately describe the causal relationship between tourism income and local economic development and cannot accurately predict tourism income. We propose an optimized forecasting method of tourism revenue based on time series. This method first conducts a cointegration test on the time series data of the relationship between tourism income and local economic development, constructs a two-variable autoregressive model of tourism income and local economy, and uses the swarm intelligence method to test the causal relationship and the relationship between tourism income and local economic development, calculate the proportion of tourism industry, define the calculation result as the direct influence factor of tourism industry on the local economy, calculate the relevant effect of local tourism development and economic income, and construct tourism income optimization forecast model. The simulation results show that the model used can accurately predict tourism revenue.

1. Introduction

Tourism economic forecasting [1–3] serves tourism economic decision-making and planning management. It is the premise of scientific decision-making and planning management and directly affects the accuracy and reliability of tourism economic decision-making [4, 5] and planning management [6, 7]. It is impossible to make an optimized tourism decision and planning without a tourism economic forecast that conforms to the objective reality. Tourism economic forecast participates in tourism economic decision-making and planning management and affects decision-making and planning. This important role is mainly reflected in the following aspects: first, through forecasting, reveal the changing trend of tourism economic development in the future [8], for the purpose of formulating tourism economic development. The strategy provides a reliable

basis. The formulation of a tourism economic development strategy is the most important tourism economic decision, and every link and every factor that constitutes a tourism economic development strategy, including development goals [9], implementation steps [10], and measures, cannot be separated from the prediction of future trends. To formulate a tourism economic development strategy, first of all, it is necessary to predict and make reasonable predictions about a series of unknowns, such as the overall development of the national economy [11, 12], changes in economic structure, changes in national policies, and changes in population quantity and quality, in order to grasp the possibility of the development of the tourism economy of the country and propose feasible development goals; secondly, it is necessary to predict the changes in the market within a certain period of time, the changes in the industrial structure of the tourism economy, the changes in the

product structure, and the changes in reception capacity. Only by making scientific predictions on the status and level of tourism economic development can we accurately divide the development stages and strategic steps [13] and determine the approximate execution sequence and time range for related work; again, the flow direction and flow of the future tourism market [14] and the tourism economy [15] must be determined. Only by predicting the changing trends of the quality and abundance of regional resources can we reasonably formulate the strategic layout of tourism economic development and complete the optimal spatial configuration of tourism productivity [16, 17]. Obviously, leaving the basis of tourism economic forecasting, the entire tourism economic development strategy has become a castle in the sky. Second, through forecasting, reveal the various situations that may occur in the development of the tourism economy: mainstream and tributary [18], favorable and unfavorable factors [19], opportunities and risks [20], and achievements and problems [21], so as to be confident and avoid blindness and one-sidedness in decision-making [22–24]. For example, in 1982, Hong Kong's tourism industry still maintained a momentum of development despite the global economic downturn, and all walks of life were deeply affected. One of the important reasons is that the Hong Kong tourism authority [25] predicted the trend of the world economic downturn and made corresponding preparations in advance. Third, predicting the economic benefits of a number of alternative tourism development programs can provide a practical basis for choosing the best program. In other words, the tourism economy forecast not only proposes a variety of ways and plans for the development of the tourism economy but also analyzes and analyzes the possible losses and benefits of each plan and the possible consequences and impacts of each proposed tourism policy. Demonstration is to make a decision on the premise of comprehensively weighing the pros and cons. Fourth, the role of forecasting not only is limited to speculating on the economic process specified by tourism economic decision-making and planning management but also includes foreseeing the changes and prospects of the external environment related to it. Planning management provides more ambitious information in order to make tourism economic decision-making and planning management more comprehensive. The external environment mentioned here mainly refers to various external noneconomic factors that may have an impact on the development of the tourism economy, such as the trend of global climate warming, the peaceful trend of the international political environment, and the negative impact of the SARS epidemic. Forecasting the changes and prospects of the external environment of tourism is particularly important for macro decision-making.

In order to effectively play the role of tourism economic forecasting, machine learning algorithms [26] have played a big role. The application of machine learning in the economic field mainly includes helping scholars obtain data that was difficult to obtain in the past, exploring the correlation between variables and making predictions, predicting counterfactuals, and then identifying cause and effect. From the perspective of predictive ability, machine learning is a predictive method with strong applicability, good accuracy, and high efficiency. First of all, machine learning [27] is not limited to "interpretability"; it can flexibly choose functional forms to fit data, study highly nonlinear, unexplainable models, and make out-of-sample predictions. Its predictive power surpasses traditional econometric methods [28]. Although mainstream empirical methods mostly use econometric models based on causality, these models have strict application conditions. Even with the support of the correct economic theory, they are often unusable in research, or even though they can be used, they end up in failure. Secondly, machine learning can make full use of the value of big data, directly mining the relationship between data and "discovering nontrivial knowledge that is of interest to specific users from the database." Finally, when the machine classifies, almost no human judgment is added, so the objectivity is high.

The contribution of this paper is to study the effectiveness of machine learning methods, promote the application of machine learning methods in financial forecasting, and provide ideas and references for the intelligent and digital transformation of tourism economic forecasting. The research results of this article prove that machine learning is an accurate, simple, and objective forecasting tool suitable for listed companies in my country, and different models have their own strengths in tourism economic forecasting. As a forecasting tool that "advances with the times," machine learning can be self-optimized with the continuous enrichment of future tourism economic data so that the forecasting method can be constructed "once and forever" and "excellent" in terms of results. Therefore, machine learning can help companies discover financial problems in time to take remedial measures; provide investors, corporate partners, and other stakeholders with more financial information to optimize investment decisions; and provide effective methods for regulators to reduce human and material costs and improve market supervision.

2. Related Works

Machine learning methods have improved the economics research paradigm [29], and the academic results of applying them to financial forecasting have become increasingly abundant. Scholars at home and abroad have done more research on financial distress forecasting, but there are few results involving performance explosions. On the whole, machine learning provides ideas and methods for the prevention and discovery of financial problems of listed companies [30], and the predictive model trained by it provides an effective practical reference for the stakeholders of listed companies [31]. Chen et al. [32] used data from listed companies in Taiwan and found that the closer to the time point of financial distress, the higher the prediction accuracy of the prediction data and the accuracy of the neural network model is higher than that of the machine learning (clustering) model. Sun et al. [33] used Chinese ST companies as samples and used regression trees (CART)

[34], support vector machines (SVM) [35], K-nearest neighbors (KNN) [36], multiple discriminant analysis (MDA) [37], logistic regression [38], and other methods to make predictions. The results showed the prediction effect of the CART model. Sun et al. [33] also used Chinese ST companies as cases of financial distress, single-factor testing (SAT) [39], and decision trees as weak learners, and used the AdaBoost method [40] to integrate weak learners to predict the company's financial distress situation; combined with separate decision tree models, SVM is compared; and it is found that the AdaBoost method, which uses SAT as a weak learner, has the highest prediction accuracy. Sun and Lie [33] took Chinese listed companies as a sample and defined financial distress as two consecutive years of loss or the loss of the most recent year exceeding the registered capital, constructed a dynamic financial distress prediction model, and used minority oversampling technology (SMOTE) to solve the problem of sample imbalance. Financial fraud forecasts: there are also many research results in this area. Given the limited resources, it is unlikely to find all financial frauds, and the possibility of exploring causality is limited. The use of machine learning to predict is practical. Nasir et al. [41] used the support vector machine model to detect the financial fraud of listed companies. After adding a specific "core" to the model, the model worked well. Al-Hashedi et al. [42] summarized and compared the technologies and methods of financial fraud detection and found that there are applicable technologies and methods for different types of fraud.

In terms of predictors, scientific and complete data are necessary conditions for the success of machine learning models. The data of listed companies is relatively rich, and there are many variables that can be collected. Therefore, the selection of predictors has become a research topic. Liengaard et al. [43] believed that there are generally two methods for selecting predictor variables: one is to select variables based on financial accounting theory and the other is to select variables based on machine learning. They studied the financial distress prediction problem [44, 45] of listed companies in mainland China (using ST as the standard) and found that the selection of predictive variables based on data mining models has the same effect as the selection based on expert financial accounting knowledge, and ROA is the best predictor variables.

3. Methodology

In the process of forecasting tourism income, first, calculate the regional tourism income and tourism income growth rate indicators, obtain the promotion characteristics of tourism income to the local economy, describe the law of change between tourism income and economic growth, and build a tourism income prediction model. The specific steps are described in detail as follows: assuming that N_p represents the tourism output value, C_{cs} represents the connection between the tourism industry and other industries, and DW represents the elastic value of tourism income, then use equations (1) and (2) to calculate regional tourism income and tourism income growth rate:

$$Q_{JP} = \frac{D_{DS} \times w}{C_{cs}} \times \text{TES} \times N_p \text{DW}, \tag{1}$$

$$W^* = \frac{D_{DS} \times w}{C_{cs}} \times \frac{\text{TES} \times N_p Q_{JP}}{\text{DW}}. \tag{2}$$

Here, N_p represents the added value of tourism, w'' represents the final demand for tourism in the place, and TES represents the income effect of the tourism economy.

Assuming that X represents the degree of dependence of tourism on other industries, formula (3) is used to obtain the promotion characteristics of tourism income to the local economy:

$$\phi''' = \left(Q_{JP} \times W^*\right) \times \frac{\varphi \times \kappa}{(H \cdot \chi)} \times \varsigma. \tag{3}$$

In the formula, φ represents the balanced relationship between tourism income and tourism growth, κ represents the income effect of tourism, H represents the factors influencing the development of tourism, χ represents the comprehensive employment coefficient of tourism to other industries, and W^* represents tourism income and the lag structure between the two variables of economic growth. The tourism income forecast model is shown in the following:

$$\text{NA} = \frac{\nu\gamma \times \phi'''}{W^*} \otimes Q_{JP} \cdot \text{DW}. \tag{4}$$

In the formula, VR represents the law of change between the two variables of tourism income and economic growth. However, traditional methods cannot accurately describe the causal relationship between tourism revenue and local economic development and cannot accurately predict tourism revenue. This paper proposes an optimized forecasting method of tourism income based on time series.

In the process of modeling and modeling of tourism revenue optimization forecast, obtain time-series data of tourism revenue and local economic growth, conduct cointegration test on tourism revenue and economic development time-series data, and build a vector autoregressive model of tourism revenue and economic development variables. To test the causal relationship between the models, the specific steps are described in detail as follows: Before the cointegration test on tourism income and economic development time-series data, the stability of tourism income and economic development time series data should be tested, respectively. Its function is to avoid false regressions with high $R2$ values between tourism income and economic development time series variables. Assuming that B represents the criterion for judging regional economic growth, use equation (5) to calculate the economic growth level of a tourist area:

$$p_\alpha(X, A) = \frac{B \cdot \Omega}{\mu_n \cdot R2} \times \frac{\varsigma^\gamma \cdot U_n}{M^{\not{A}}}. \tag{5}$$

In the formula, Ω represents the total tourism income, μn represents the dynamic impact of random disturbances on the variable system, ς^γ represents the impact of economic lagging variable in the economic region on the current

Table 1: Index system of influencing factors of tourism income.

Explained variable	First level indicator	Secondary indicators
Domestic tourism income	Socioeconomic factors	Added-value of tertiary industry
		Regional per capita production value
	Residents' living standards	Per capita disposable income of urban residents
		Per capita consumption expenditure of urban residents
		The total retail sales of social consumer goods
	Traffic convenience	Passenger turnover
		Passenger volume
		Kilometers, railway density
	Tourism resources and services	Number of star-rated hotels
		Number of travel agencies
		Number of A-level and above sceneries
	Environmental quality factors	Green area rate of built-up area
		Harmless treatment rate of domestic garbage
	Regional demographic factors	Permanent population at the end of the year
		Number of students in the university

variables, U_n represents the availability of variable data, and $M^{\mathbb{H}}$ represents the tourism order of the time series of income and economic growth.

Assuming that the tourism income time series represented by ξ is a first-order single integer and l represents the second-order difference sequence of the economic growth data series, then use equation (6) to obtain stable tourism income and economic development time series variable data:

$$p\left(\frac{\gamma}{\eta_1}\right) = \frac{\xi\left[\sum_{i=1}^{k} \eta_{1i}\right]}{\Gamma(\eta_{1i})} \gamma(\eta_{1i} - 1). \tag{6}$$

In the formula, n_{1i} represents the cointegration relationship between tourism income and regional GDP, Γ represents the random disturbance term, Y represents the n-dimensional endogenous variable, and k represents the lag period of economic growth:

$$P^L = \frac{\mu_{1n} - \mu_{jn}}{t_{ES} \times P(\gamma/\eta_1)} \cdot P_{\propto}(X, A). \tag{7}$$

The income effect of tourism is defined as the impact of tourism on domestic per capita income. Since the expenses spent by tourists on tourism in tourist destinations will directly become the income of local enterprises, tourism income will gradually be based on the correlation of its related industries. Infiltrate the local economic system, thereby driving the improvement of the overall local economy. The tourism effect of a place can be expressed as the direct and indirect impact of the tourism industry on the local economy. Obtain the causal relationship between local tourism revenue and economic growth as a basis, calculate the proportion of tourism industry, define the result of the calculation as the direct influence factor of tourism on the local economy, integrate the related effects of the local tourism industry, and establish the model of tourism revenue optimization forecast.

4. Experiment and Results

In order to prove the effectiveness of the proposed time-series-based tourism revenue optimization forecasting modeling method, an experiment is needed. The experiment takes Yan'n from 2009 to 2019 as an example to empirically demonstrate the relationship between tourism income and economic growth in Xi'an. The simulation tool for the experiment is python.

In view of the complexity of the tourism system, there are two internal and external systems for its impact factors. Taking into account the availability and quantification of data, an indicator system is constructed from the internal system of domestic tourism revenue impact factors. Table 1 shows the selected indicators.

According to the method proposed in the third part, Table 2 shows the prediction results of this method, including the original value, predicted value, error, relative error, and level error.

It can be seen from Figure 1 that the average relative error is 0.073923, the average grade ratio deviation is 0.1148, both are less than 0.2, and the posterior difference ratio C value is 0.021 less than 0.35, which means that the model accuracy meets the requirements. The combined effect is good, and the model can be used for prediction.

In order to reflect the superiority of our algorithm, we use the algorithm of this paper, SVM, and Naive Bayes algorithm to construct tourism revenue optimization forecasting models and compare different algorithm models to optimize tourism revenue forecast accuracy. The comparison results are shown in Figure 2.

It can be seen from Figure 2 that the accuracy of using the algorithm model in this paper for different algorithm models to optimize tourism revenue is better than the accuracy of the SVM algorithm model for time series testing. This is mainly because the algorithm is used to establish the model first. The time-series data of tourism income and local economy are tested for cointegration, which guarantees the accuracy of the algorithm in this paper to optimize the forecast of tourism income by using different algorithm models.

The algorithm of this paper, SVM algorithm, and Naive Bayes algorithm is used to construct tourism revenue optimization forecasting model, and the error rate (%), stability (%), and time efficiency (%) of tourism revenue optimization forecasting models of three different algorithms are

TABLE 2: The predicted value of our proposed method.

Year	Original value	Predictive value	Error	Relative error	Step ratio deviation
2009	53.1	53.1	0	0	53.1
2010	67.3	51.177	16.123	0.23957	67.3
2011	83.81	75.115	8.695	0.10375	83.81
2012	103.39	100.532	2.858	0.02764	103.39
2013	126.55	127.521	−0.971	−0.0077	126.55
2014	141.56	156.179	−14.619	−0.1033	141.56
2015	160.01	186.609	−26.599	−0.1662	160.01
2016	207.33	218.919	−11.589	−0.0559	207.33
2017	275.22	253.288	21.932	0.07969	275.22
2018	291.9	289.658	2.242	0.00768	291.9
2019	335.56	328.34	7.22	0.02152	335.56
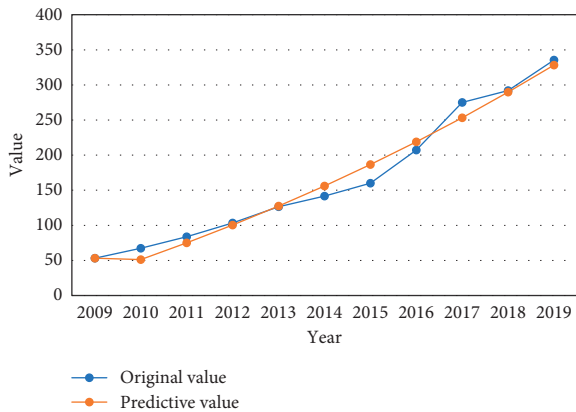
FIGURE 1: Comparison of predicted value with the original value.

FIGURE 2: Comparison of prediction accuracy of three algorithms.

FIGURE 3: Comparison of modeling error rates of different algorithms.
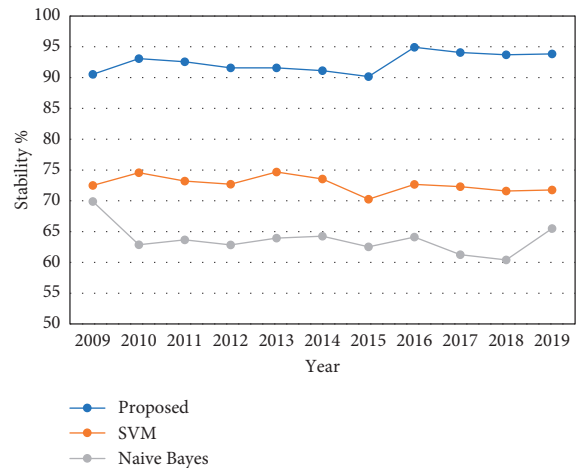
FIGURE 4: Comparison of stability of different algorithms.

compared. The results of the comparison measure the comprehensive effectiveness of the two different algorithms to establish the tourism revenue optimization forecast model. The results of the comparison are shown in Figures 3, 4, and 5.

From the analysis in Figures 3–5, it can be concluded that the comprehensive effectiveness of the establishment of the tourism revenue optimization forecast model established by the algorithm of this paper is better than the effectiveness of the SVM algorithm and the Naive Bayesian model. This is because when using the algorithm in this paper to establish a

tourism revenue optimization forecast accuracy model, a vector autoregressive model based on tourism revenue and local economy two variables is constructed, and the relationship between the two variables is tested by the Granger method, and the proportion of tourism industry is calculated. On this basis, the correlation effect of local tourism income is calculated, and a tourism income optimization forecast accuracy model is established.
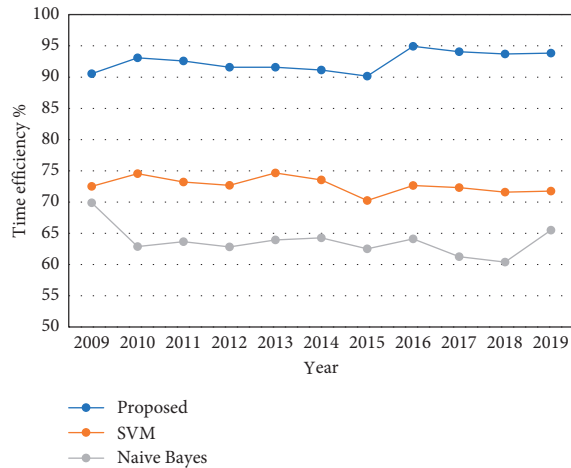
FIGURE 5: Comparison of time efficiency of different algorithms.

5. Conclusion

The development of modern tourism is closely related to the development of modern transportation. Transportation construction is an important condition for the development of tourism resources and the construction of tourist destinations. Related research shows that areas with obvious advantages in transportation accessibility have a relatively high level of regional tourism economic development. It is necessary to grasp the geographical position of the northwestern region to the east, seize the opportunity of the "Belt and Road" and the construction of the economic belt along the Yellow River, improve transportation services, speed up the construction of transportation networks, and continuously improve transportation accessibility. The development and construction of tourism resources are very important to the development of the tourism economy. In the development process, we must pay attention to the investigation of the tourist attraction radius of the scenic spot and the excavation of the characteristics of the scenic spot, avoid the appearance of homogeneous scenic spots, and establish a unique tourism brand image system. Furthermore, due to the low location and environmental carrying capacity of arid and semiarid areas, the stereotype that tourism is equal to a "smoke-free industry" must be changed, and the coupled and coordinated development of the environment and tourism must be emphasized to create a good environment for cultural tourism and natural tourism. The tourism service industry is a related industry generated by tourism activities. The more mature it is, the more prosperous the tourism economy will be. In order to promote the maturity of the tourism service industry, it is first necessary to establish a multichannel financing mechanism, increase capital investment in the tourism industry, and improve supporting service facilities; secondly, it is necessary to regulate the order of the tourism market and strengthen the supervision of the quality of tourism services; and finally, it is necessary to transform tourism services. The business philosophy is to establish a professional team of talents, pay attention to feedback from tourists, and advocate "refined" and "individualized" services.

Data Availability

The experimental data used to support the findings of this study are available from the author upon request.

Conflicts of Interest

The author declares no conflicts of interest.

References

[1] S. Milne and I. Ateljevic, "Tourism, economic development and the global-local nexus: theory embracing complexity," *Tourism Geographies*, vol. 3, no. 4, pp. 369–393, 2001.

[2] G. Shaw and A. M. Williams, "Tourism, economic development and the role of entrepreneurial activity," *Progress in tourism, recreation and hospitality management*, vol. 2, pp. 67–81, 1990.

[3] W. Z. Danish and Z. Wang, "Dynamic relationship between tourism, economic growth, and environmental quality," *Journal of Sustainable Tourism*, vol. 26, no. 11, pp. 1928–1943, 2018.

[4] W. Edwards, "The theory of decision making," *Psychological Bulletin*, vol. 51, no. 4, pp. 380–417, 1954.

[5] K. L. Milkman, D. Chugh, and M. H. Bazerman, "How can decision making be improved?" *Perspectives on Psychological Science*, vol. 4, no. 4, pp. 379–383, 2009.

[6] X. Xiang, Q. Li, S. Khan, and O. I. Khalaf, "Urban water resource management for sustainable environment planning using artificial intelligence techniques," *Environmental Impact Assessment Review*, vol. 86, Article ID 106515, 2021.

[7] J. R. Lund and J. Wiese, "Less is more: exploring support for time management planning," in *Proceedings of the Designing Interactive Systems Conference*, pp. 392–405, Diego, CA, USA, June 2021.

[8] Y. Pan, G. Weng, C. Li, and J. Li, "Coupling coordination and influencing factors among tourism carbon emission, tourism economic and tourism innovation," *International Journal of Environmental Research and Public Health*, vol. 18, no. 4, p. 1601, 2021.

[9] Y. Huan, T. Liang, and H. Li, "A systematic method for assessing progress of achieving sustainable development goals: a case study of 15 countries," *The Science of the Total Environment*, vol. 752, Article ID 141875, 2021.

[10] E. Childs, C. D. Remein, R. M. Bhasin et al., "How to launch and continually enhance an effective medical campus faculty development program: steps for implementation and lessons learned," *Journal of Healthcare Leadership*, vol. 13, pp. 147–156, 2021.

[11] S. V. Bryukhovetskaya, K. A. Artamonova, A. A. Gibadullin, S. A. Ilminskaya, and Z. M. Kurbonova, "Management of digital technology development in the national economy," in *Proceedings of the IOP Conference Series: Earth and Environmental Science*, Krasnoyarsk, Russia, November 2020.

[12] X. Li, S. Zhou, and K. Yin, "Measurement of the high-quality development level of China's marine economy," *Marine Economics and Management*, vol. 4, 2021.

[13] D. A. Shepherd and M. Gruber, "The lean startup framework: closing the academic-practitioner divide," *Entrepreneurship: Theory and Practice*, vol. 45, Article ID 1042258719899415, 2020.

[14] N. Mou, Y. Zheng, T. Makkonen, T. Yang, J. Tang, and Y. Song, "Tourists' digital footprint: the spatial patterns of

tourist flows in Qingdao, China," *Tourism Management*, vol. 81, Article ID 104151, 2020.

[15] W. M. Lim and W.-M. To, "The economic impact of a global pandemic on the tourism economy: the case of COVID-19 and Macao's destination- and gambling-dependent economy," *Current Issues in Tourism*, pp. 1–12, 2021.

[16] V. K. Nguyen, R. Natoli, and S. Divisekera, "Innovation and productivity in tourism small and medium enterprises: a longitudinal study," *Tourism Management Perspectives*, vol. 38, Article ID 100804, 2021.

[17] A. Nieto Masot and N. Ríos Rodríguez, "Rural tourism as a development strategy in low-density areas: case study in northern Extremadura (Spain)," *Sustainability*, vol. 13, no. 1, p. 239, 2021.

[18] K. Wang, Y. Pang, C. He et al., "Three Gorges Reservoir construction induced dissolved organic matter chemistry variation between the reservoir and non-reservoir areas along the Xiangxi tributary," *The Science of the Total Environment*, vol. 784, Article ID 147095, 2021.

[19] K. Zhong, Y. Wang, J. Pei, S. Tang, and Z. Han, "Super efficiency SBM-DEA and neural network for performance evaluation," *Information Processing & Management*, vol. 58, no. 6, Article ID 102728, 2021.

[20] J. Pei, "Solving the problem of charging and discharging of electric vehicles based on particle swarm algorithm," in *Proceedings of the 2019 2nd International Conference on Information Systems and Computer-Aided Education (ICISCAE)*, pp. 534–538, IEEE, Dalian, China, September 2019.

[21] J. Su, X. Zhao, D. Hong, Z. Luo, and H. Chen, "Q-value fine-grained adjustment based RFID anti-collision algorithm," *IEICE-Transactions on Communications*, vol. E99, no. 7, pp. 1593–1598, 2016.

[22] Z. Kuang, L. Zhang, L. Zhao, and L. Zhao, "Energy- and spectral-efficiency tradeoff with α-Fairness in energy harvesting D2D communication," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 9, pp. 9972–9983, 2020.

[23] X. Liu, J. Zhang, and S. Jiang, "Accurate localization of tagged objects using mobile RFID-augmented robots," *IEEE Transactions on Mobile Computing*, vol. 20, 2019.

[24] J. Su, Y. Chen, Z. Sheng, Z. Huang, and A. X. Liu, "From M-ary query to bit query: a new strategy for efficient large-scale rfid identification," *IEEE Transactions on Communications*, vol. 68, no. 4, pp. 2381–2393, 2020.

[25] J. Pei, J. Li, and B. Zhou, "A recommendation algorithm about choosing travel means for urban residents in intelligent traffic system," in *Proceedings of the 2021 IEEE 5th Advanced Information Technology, Electronic and Automation Control Conference (IAEA)*, pp. 2553–2556, IEEE, Chongqing, China, March 2021.

[26] D. Hong, L. Gao, and J. Yao, "Graph convolutional networks for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, 2020.

[27] D. Hong, N. Yokoya, J. Chanussot, and X. X. Zhu, "An augmented linear mixing model to address spectral variability for hyperspectral unmixing," *IEEE Transactions on Image Processing*, vol. 28, no. 4, pp. 1923–1938, 2019.

[28] D. Hong, L. Gao, N. Yokoya et al., "More diverse means better: multimodal deep learning meets remote-sensing imagery classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 5, pp. 4340–4354, 2021.

[29] E. Brynjolfsson, C. Wang, and X. Zhang, "The economics of IT and digitization: eight questions for research," *MIS Quarterly*, vol. 45, no. 1, pp. 473–477, 2021.

[30] K. A. Fachrudin, "Financial Health Assessment Model for Listed Companies in Indonesia," in *Proceedings of the 2nd Economics and Business International Conference*, Vienna, Austria, June 2021.

[31] A. Sutantoputra, "Do stakeholders' demands matter in environmental disclosure practices? Evidence from Australia," *Journal of Management & Governance*, pp. 1–30, 2021.

[32] J.-H. Chen, T.-S. Chou, J.-P. Wang, H.-H. Wei, and T.-H. Yang, "Sustainable corporate governance: the impact factors for top consulting engineering companies in taiwan," *Sustainability*, vol. 13, no. 14, p. 7604, 2021.

[33] X. Sun and Y. Lei, "Research on financial early warning of mining listed companies based on BP neural network model," *Resources Policy*, vol. 73, Article ID 102223, 2021.

[34] R. Sparapani, C. Spanbauer, and R. McCulloch, "Nonparametric machine learning and efficient computation with bayesian additive regression trees: the BART R package[J]," *Journal of Statistical Software*, vol. 97, no. 1, pp. 1–66, 2021.

[35] A. Kammoun and M.-S. AlouiniFellow, "On the precise error analysis of support vector machines," *IEEE Open Journal of Signal Processing*, vol. 2, pp. 99–118, 2021.

[36] K. K. Sharma and A. Seal, "Spectral embedded generalized mean based k-nearest neighbors clustering with S-distance," *Expert Systems with Applications*, vol. 169, Article ID 114326, 2021.

[37] Z. Zheng, H. Sun, and Y. Zhou, "Multiple discriminant analysis for collaborative representation-based classification," *Pattern Recognition*, vol. 112, Article ID 107819, 2021.

[38] R. Xiao, X. Cui, H. Qiao et al., "Early diagnosis model of Alzheimer's disease based on sparse logistic regression with the generalized elastic net," *Biomedical Signal Processing and Control*, vol. 66, Article ID 102362, 2021.

[39] C. Nicolaidis, G. Schnider, J. Lee et al., "Development and psychometric testing of the AASPIRE adult autism healthcare provider self-efficacy scale," *Autism*, vol. 25, no. 3, pp. 767–773, 2021.

[40] B. T. Pham, M. D. Nguyen, T. Nguyen-Thoi et al., "A novel approach for classification of soils based on laboratory tests using Adaboost, Tree and ANN modeling," *Transportation Geotechnics*, vol. 27, Article ID 100508, 2021.

[41] M. Nasir, S. Simsek, E. Cornelsen, S. Ragothaman, and A. Dag, "Developing a decision support system to detect material weaknesses in internal control," *Decision Support Systems*, Article ID 113631, 2021.

[42] K. G. Al-Hashedi and P. Magalingam, "Financial fraud detection applying data mining techniques: a comprehensive review from 2009 to 2019," *Computer Science Review*, vol. 40, Article ID 100402, 2021.

[43] B. D. Liengaard, P. N. Sharma, G. T. M. Hult et al., "Prediction: coveted, yet forsaken? Introducing a cross-validated predictive ability test in partial least squares path modeling," *Decision Sciences*, vol. 52, no. 2, pp. 362–392, 2021.

[44] J. Sun, H. Fujita, Y. Zheng, and W. Ai, "Multi-class financial distress prediction based on support vector machines integrated with the decomposition and fusion methods," *Information Sciences*, vol. 559, pp. 153–170, 2021.

[45] M. Durica, L. Svabova, and J. Fonda, "Financial distress prediction in Slovakia: an application of the CART algorithm," *Journal of International Studies*, vol. 14, no. 1, 2021.

Retraction

Retracted: Construction of the Open Oral Evaluation Model Based on the Neural Network

Scientific Programming

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope

(2) Discrepancies in the description of the research reported

(3) Discrepancies between the availability of data and the research described

(4) Inappropriate citations

(5) Incoherent, meaningless and/or irrelevant content included in the article

(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] Z. Chen, X. Zhang, Z. Li, and A. Li, "Construction of the Open Oral Evaluation Model Based on the Neural Network," *Scientific Programming*, vol. 2021, Article ID 3928246, 11 pages, 2021.

Research Article

Construction of the Open Oral Evaluation Model Based on the Neural Network

Zhixin Chen, Xu Zhang [iD], Zhiyuan Li, and Anchu Li

Northeast Electric Power University, Jilin City, Jilin 132012, China

Correspondence should be addressed to Xu Zhang; 20132487@neepu.edu.cn

According to the problem of low efficiency and low scoring accuracy of the traditional oral language scoring system, this study builds an open oral language evaluation model based on the basic principles of deep learning technology. Firstly, the basic methods of the convolutional neural network (CNN) and long short-term memory (LSTM) neural network are introduced. Then, we combine the convolutional neural network (CNN) and long short-term memory (LSTM) neural network to design an open oral scoring model based on CNN + LSTM, which divides the oral evaluation model into the speech scoring model and text scoring model and makes a specific implementation of two scoring models, respectively. An experimental environment is then built to preprocess the data, and finally, the model built in this study is trained and simulated. The experimental results show that the CNN + LSTM network evaluation model has a better comprehensive scoring performance, higher scoring efficiency, and higher accuracy and has feasibility and practicability.

1. Introduction

The rapid development of the internet information level has promoted the continuous progress of education. In traditional oral English learning, teachers mostly adopt the manual correction method to students' oral test and training results, which is tedious, time-consuming, and inefficient and is not conducive to the long-term development of students' oral English learning. In recent years, computer-assisted learning system has been widely used in the field of education, and more and more students and teachers in China begin to use computer-assisted oral learning. However, the traditional computer-assisted oral learning system has some shortcomings in the open oral evaluation, such as the lack of intelligence and low performance. It is an inevitable trend for the development of education to design and implement an intelligent and open oral scoring system.

How to improve the performance and efficiency of the open oral evaluation model is still the focus of many scholars. According to the demand of diagnostic oral English proficiency assessment, this paper puts forward corresponding solutions and strategies for the demand of the oral English assessment

model. Based on the GESE oral test method, a formative hierarchical evaluation model of oral English teaching in higher vocational colleges is constructed, and its positive "backwash effect" law is obtained. Many of the research tasks proposed in [1] focus on English, and it is difficult to explore a richer variety of languages within the framework of the SLU task. Minor languages need to be updated in the training process of the SLU model. This problem is explained from three aspects: the input of the neural network, the comparison between the French version and the best setting in different ways, and the comparison with the most advanced methods. Tsai et al. [2] aimed at the use of mobile phones, mailboxes, and other tools in daily life, which are realized by recurrent neural networks (RNNs). The character language model based on LSTM can not only capture the Boltzmann statistics of the system but also reproduce the dynamics. It can be applied to capture the time evolution of typical trajectories in chemistry and biophysics. An acoustic speech method for SLD proposed in [3] is a new method based on the original SLD. It uses an attention-based neural network model to capture the language and a Gaussian smoothing method to locate language changes. It is more effective in dealing with code-switching of monolingual

segments, but its performance will be affected by the duration of monolingual segments, which needs further study. Peng [4] proposed that learning assessment can let students know whether they have achieved the ideal academic achievements and goals through computer English teaching, so as to further improve their shortcomings. Based on the neural network and artificial intelligence technology, this model can realize this evaluation. It is based on hearing, introduces wavelet entropy features, and applies it to the adaptive model. Combined with the control experiment, the performance of the model is analyzed. Mathematical statistics can directly show the effect. Research proves that the model can meet the expectations. With the deepening development of higher education, English teaching has been paid more and more attention. While improving the quality of education [5], learning evaluation is the most fundamental measure. By studying the existing problems and characteristics of the learning evaluation system, analyzing the defects of the traditional system, then summing up the shortcomings, and developing the advantages, this paper puts forward a college English teaching quality evaluation system based on information fusion and optimized RBF neural network decision-making algorithm. Jiang et al. [6] proposed an improved fuzzy RBF neural network model based on back-propagation learning, which combines a large number of problems in English teaching before, such as large quantity and complexity, and then optimizes with the development limitations and existing shortcomings of neural networks. This model is an objective reference to a university teaching case and puts forward an improved quality evaluation method, which is subjective and random, so that the evaluation results are more in line with the actual situation. Mohammed et al. [7] used Arabic, English, and French to evaluate the recognition performance of linear predictive coding (LPC) and/or mel-frequency cepstrum coefficient (MFCC) and artificial neural network (ANN) and also tested LPC and MFCC, hidden layer, different neurons in the hidden layer, and different transfer functions to illustrate their usability. Li [8] proposed a teaching evaluation model based on the improved BP neural network, which is helpful to improve English teaching quality. It summarizes and analyzes several teaching elements, designs the evaluation system, and improves the shortcomings in calculation by improving nonmonotone linear search and adaptive step change. Finally, through practical verification, the model can better evaluate English teaching and improve the effect. Hermanto et al. [9] used two models, namely, recurrent neural network (RNN) and statistical network, using the n-gram model. Compared with the statistical language model, the neural language model achieves better results in the field of machine translation. The latter two are deficient in accuracy and calculation speed, while the RNN has improved its evaluation scores in these aspects. Esan et al. [10] proposed a machine translation method based on the recurrent neural network model. By testing the model with manual and automatic evaluation techniques, combined with real manual evaluation, it shows that the model is smooth and usable, basically consistent with human judgment, and relevant and has high translation effect and accuracy.

However, the above research has not solved the intelligent problem of open oral learning, and the research direction needs to be further studied. The above literature research also puts forward many oral evaluation methods and has achieved good results. However, many solutions are based on machine learning or algorithms, which have low effect, do not realize intelligent processing, and cannot meet the requirements of oral evaluation. Although there are some neural network solutions, the translation efficiency and accuracy are relatively low, and this paper combined with the CNN + LSTM model can solve such problems. Based on this, this study combines the learning characteristics of the neural network, according to the basic principles and structural characteristics of the convolution neural network (CNN) and long short-term memory (LSTM) neural network, and designs and implements an open oral scoring model based on CNN + LSTM. The final results show that this model can improve the intelligent level of open oral learning, and the evaluation model has relatively high scoring efficiency and accuracy, which further shows that the evaluation model is feasible and practical.

This paper proposes that the neural network can solve the translation problem from the perspective of intelligent analysis of oral English, which can improve the accuracy of oral English and build a standard evaluation model. The CNN-LSTM model proposed in this paper has the highest evaluation accuracy, which can reach 82%, and the effect is the best. The key points will be to score the oral pronunciation and oral content separately, and finally, the total score after summation will be taken as the final score result of the examinee. When constructing the scoring model, this paper uses different neural networks to implement and test. Finally, the model with the highest relevance of the man-machine score will be applied to the actual correction task to help teachers reduce work pressure. At the same time, it can also be used as a tool for students' oral self-evaluation.

2. Basic Method of the Neural Network

2.1. Convolution Neural Network (CNN)

2.1.1. Fundamentals of the CNN. Convolutional neural network (CNN) is a representative algorithm of deep learning [11]. It is a feedforward neural network with convolution calculation and depth structure, which is very suitable for processing one-dimensional time-domain sequence data and image data. Compared with other neural networks, CNN can learn the original data efficiently and quickly, so as to extract the specific features of the data; that is, it has the ability of representation learning. CNN is widely used in computer vision, natural language processing, and other fields. The network has four remarkable characteristics, which are local perception, weight sharing (convolution operation), pooling processing, and multiconvolution kernel operation [12]. Among them, convolution operation is an operation that defines two integrable functions for convolution operation. It mainly includes two operation modes, namely, continuous convolution operation and discrete convolution operation.

Let both $f(x)$ and $g(x)$ belong to integrable functions in the real number field, and the new function

$J(x) = (f * g)(x)$ is the convolution of functions $f(x)$ and $g(x)$, in which $(f * g)(x) = (g * f)(x)$ and $(f * g)(x)$ belong to integrable functions.

The continuous convolution operation is expressed as

$$J(x) = (f * g)(x) = \int _ f(a)g(x - a)\mathrm{d}a. \quad (1)$$

The expression of the discrete convolution operation is

$$J(t) = (f * g)(t) = \sum_{n=-\infty}^{\infty} f(n)g(t - n). \quad (2)$$

Among them, the convolution operation mode in the CNN is discrete convolution, which can easily deal with discrete data problems in actual measurement.

2.1.2. Central Neural Network Structure. As shown in Figure 1, the network structure of the convolution neural network (CNN) is mainly divided into five network layers, namely, input layer, convolution layer, pooling layer, full connection layer, and output layer, belonging to multilayer perceptron (MLP) [13]. The following five layers are described in detail.

(1) Input Layer. The input layer of the CNN can process multidimensional data. Like other neural network algorithms for deep learning, CNN's learning method is calculated by the gradient descent algorithm, and the input data features of the CNN need to be standardized. Standardizing the features of the input data can greatly improve the learning efficiency of the CNN.

(2) Convolution Layer. Convolution layer, pooling layer, and full connection layer are all hidden layers in the CNN. The convolution layer has the function of feature extraction from the input data. There are many convolution kernels in the convolution layer, and the composition of convolution kernels is similar to that of neurons in the feedforward neural network [14], such as

$$Z^{l+1}(i, j) = \left[Z^l \otimes \omega^{l+1}\right](i, j) + b$$

$$= \sum_{k=1}^{K_l} \sum_{x=1}^{f} \sum_{y=1}^{f} \left[Z_k^l \left(s_0 i + x, s_0 j + y\right)\omega_k^{l+1}\right], \quad (3)$$

$$(i, j) \in \{0, 1, \ldots, L_{l+1}\} L_{l+1} = \frac{L_l + 2p - f}{s_0} + 1.$$

In formula (3), the summation part is equivalent to solving a cross-correlation, B represents the deviation, Z^l and Z^{l+1} represent the input and output of the first layer, respectively, which can also be called the feature map, L_{l+1} represents the size of Z_{l+1}, if the length and width of the feature map are the same, $Z(i, j)$ represents the pixels of the feature map, f, s_0, p represents the number of channels of the feature map, and GG is the parameters of the convolution layer, which are convolution kernel size, convolution step size, and filling layer, respectively.

Under special circumstances, if the convolution kernel size $f = 1$ and convolution step size $S_0 = 1$ do not include filling, the cross-correlation calculation in the convolution layer is similar to matrix multiplication, thus constructing a fully connected network between convolution layers, as shown in the following:

$$Z^{l+1} = \sum_{k=1}^{K_l} \sum_{i=1}^{L} \sum_{j=1}^{L} \left(Z_{i,j,k}^l \omega_k^{l+1}\right) + b = \omega_{l+1}^T Z_{l+1} + b,$$

$$L^{l+1} = L. \quad (4)$$

(3) Pooling Layer. After the convolution layer extracts the features, the pooling layer selects the features and filters the information. The pooled area selected by the pooled layer is similar to the step of scanning the feature map by the convolution kernel and is also determined by the pooled size, the pooled step size, and the number of filled layers [15].

Among them, the pooling layer is divided into Lp pooling and random pooling or mixed pooling. Lp pooling is mainly inspired by the hierarchical structure in the visual cortex, and its expression is as follows:

$$A_k^l(i, j) = \left[\sum_{x=1}^{f} \sum_{y=1}^{f} A_k^l \left(S_0 i + x, S_0 j + y\right)^p\right]^{1/p}. \quad (5)$$

In equation (5), S_0 is the pooling step size, (i, j) is the pixel, and p is the prespecified parameter. If $p = 1$ and Lp pooling takes the average value in the pooling area, it is mean pooling; if $p \longrightarrow \infty$ and Lp pooling takes the maximum value in the region, it is maximum pooling. These two pooling methods are the most commonly used methods in the CNN.

Hybrid pooling and random pooling are mainly extended from the concept of Lp pooling. Random pooling is randomly selected according to a specific probability in the pooled area, which can ensure that a part of nonmaximum excitation signals enter the next construction. Hybrid pooling can represent a linear combination of mean pooling and maximum pooling, as shown in the following:

$$A_k^l = \lambda L_1\left(A_k^l\right) + L_\infty\left(A_k^l\right); \quad \lambda \in [0, 1]. \quad (6)$$

It is found that hybrid pooling and random pooling have a regularization function compared with mean pooling and maximum pooling, which can avoid overfitting of the CNN.
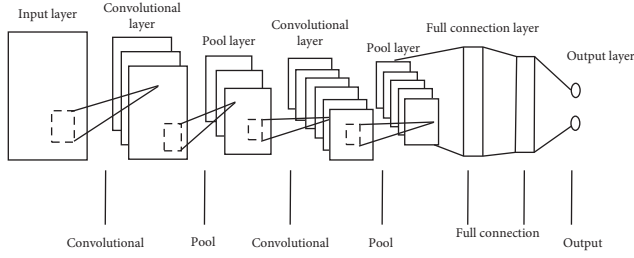
Figure 1: Central neural network structure.

(4) Fully Connected Layer. The fully connected layer belongs to the hidden layer in the traditional feedforward neural network. It is mainly located at the end of the hidden layer of the CNN, which only transmits signals to other fully connected layers. The main function of the full connection layer is to combine the extracted data features nonlinearly and output them.

(5) Output Layer. The upper part of the output layer is mostly a fully connected layer, and its structure and working principle are the same as those of the output layer in the traditional feedforward neural network. The output layer only outputs the image classification label through the logic function or normalized exponential function. The output layer recognizes objects according to the center coordinates, size, and classification of output objects. The output layer directly outputs the classification results of each pixel in the image to complete the semantic segmentation of the image.

2.2. Long Short-Term Memory (LSTM) Neural Network

2.2.1. Fundamentals of LSTM. Long short-term memory (LSTM) is a kind of time cycle neural network, which is designed to solve the long-term dependence of the RNN of the cycle neural network [16]. The main working mechanisms are mainly divided into the gating mechanism, forgetting mechanism, and circular memory mechanism. The three working mechanisms can effectively guarantee the normal operation of LSTM and make it play its long-term and circular memory functions, thus avoiding gradient disappearance and explosion. The following three mechanisms are analyzed in detail. LSTM output dimension setting is 300, return_sequences parameter is set to true, and the activation parameter is set to sigmoid.

(1) Gating Mechanism. In the LSTM network model, the key element of the gating mechanism is the activation function, which is usually sigmoid function and tan*h* function. *S* function is a commonly used activation function in neural networks. The characteristic of this function is that it can control the output real value within the range of [0, 1], as shown in the following:

$$\sigma(x) = \frac{1}{\left(1 + e^{-x}\right)}. \tag{7}$$

Tan*h* function, that is, hyperbolic tangent function: this function can control the output real value in the range of [−1,

1], with 0 bit as the center, which is just complementary to the *S* function without 0 as the center. The function is shown in the following:

$$\tan h(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}. \tag{8}$$

The gating mechanism can control the storage, memory, and update of information, and the basic principle is shown in formulas (9) and (10):

$$f(x) = \sigma(Wx + b), \tag{9}$$

$$\sigma(x) = \frac{1}{\left(1 + e^{-x}\right)}. \tag{10}$$

(2) Forgetting Mechanism. In the LSTM network model, forgetting mechanism must be realized based on the gating mechanism. By analyzing the influence of historical information on the memory unit, it is determined whether to retain or forget some information [17].

(3) Cyclic Memory Mechanism. Among them, in the LSTM network model, a new state *ct*, namely, memory unit, is added in the LSTM network model, and the main function of this memory unit is to transmit information circularly.

2.2.2. Remote Access Service Network Architecture. As shown in Figure 2, LSTM network structure is mainly divided into four parts: input gate, forgetting gate, memory unit, and output gate. The basic principles and methods of the four parts are analyzed in detail.

(1) Input Gate. The input gate mainly controls the filtering input data through the sigmoid function, wherein the input is mainly divided into two inputs, namely, the data directly input are controlled as shown in equation (11), and the other control input is stored in the candidate memory unit as shown in equation (12):

$$Z_i = \sigma\left(\sum_j K_{i,j} x_j^i + \sum_j W_{i,j} h_j^{(i-1)} + b_i, \tag{11}\right.$$

$$Z = \tan h\left(\sum_j K_{i,j} x_j^i + \sum_j W_{i,j} h_j^{(i-1)} + b. \tag{12}\right.$$

In the above formula, x_j^i represents the input data at time t, $K_{i,j}$ represents the input weight, $W_{i,j}$ represents the cycle weight of the forgetting gate, $h_j^{(i-1)}$ represents the output value of LSTM cells at the previous time, b_i represents the bias of the input gate, and b represents the bias of candidate memory cells.

(2) Forgetting the Door. Forgetting gate Z_f mainly uses the sigmoid function to map the input data x_j^i at time t in the range of [0, 1]. When the value is 0, it means that all information cannot pass this gate; when the value is 1, it means that all information is allowed to pass, as in the following:
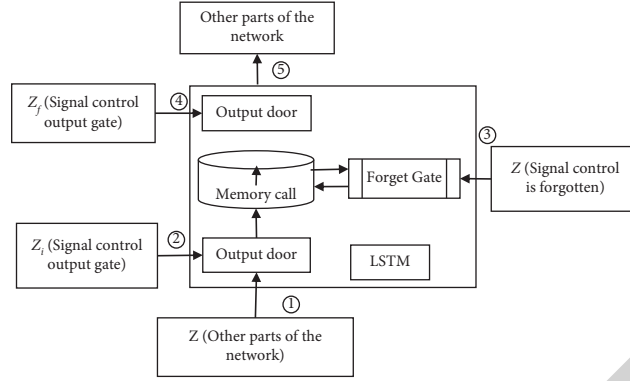
Figure 2: LSTM network architecture.

$$Z_f = \sigma\left(\sum_j K_{i,j}^f x_f^j + \sum_j W_{i,j}^f h_j^{(i-1)} + b_f\right). \tag{13}$$

In equation (13), x_j^i represents the data input at time t; $K_{i,j}^f$ represents the input weight; $W_{i,j}^f$ represents the cyclic weight of the forgetting gate; $h_j^{(i-1)}$ represents the cell output value of LSTM at last time; b^f stands for the offset of the forgetting gate [18].

(3) Memory Unit. The memory cell C_i is cyclically updated mainly through the candidate memory cell Z and the memory cell C_{i-1} at the previous time and is adjusted by the input gate Z_i and the forgetting gate Z_f, as shown in the following:

$$C_i = Z_f \cdot C_{i-1} + Z_i \cdot Z. \tag{14}$$

(4) Output Gate. The output gate Z_o mainly controls whether the memory unit C_i outputs through the sigmoid function. When the value is 0, it means that all information cannot pass this gate [19]; when the value is 1, it means that all information is allowed to pass, as shown in equation (15). The final output value H_i of LSTM cells is controlled by the $\tan h$ activation function, as shown in equation (16):

$$Z_o = \sigma\left(\sum_i K_{i,j}^o x_j^i + \sum_j W_{i,j}^o h_j^{(i-1)} + b_o\right), \tag{15}$$

$$h_i = Z_o \cdot \tan h(C_i). \tag{16}$$

In the above formula, x_j^i represents the data input at time t; $K_{i,j}^o$ represents the input weight; $W_{i,j}^o$ represents the cyclic weight of the output gate; $h_j^{(i-1)}$ represents the cell output value of LSTM at last time; b^o represents the offset of the output gate.

3. Construction of the Open Oral Scoring Model Based on CNN + LSTM

Combined with the basic principles and network structure of the CNN and LSTM neural network models, this study proposes an open oral scoring model based on the CNN + LSTM neural network. Because the model data trained by the traditional machine learning model and BP neural network model have low accurate fitting degree, the correlation between the manual feature extraction and manual score is low. This study breaks through the limitations of artificial extraction through the one-dimensional convolution neural network and long short-term memory network to integrate and process data more effectively, combined with the CNN + LSTM neural network to build the speech scoring model and text scoring model. Among them, the cyclic neural model has a large computational workload when dealing with lengthy sequence data. In this study, before the LSTM neural network, the one-dimensional convolution neural network is used to preprocess the data to shorten the sequence data, so as to improve the computational efficiency, extract more accurate features, and input them to the LSTM layer for processing.

As shown in Figure 3, the designed speech scoring model is mainly composed of MFCC feature vectors; 2 convolution blocks; bidirectional LSTM layer, namely, BLSTM, MeanOverTime layer, and full connection layer. The specific steps are as follows:

(1) Input the MFCC eigenvector firstly

(2) The filters in convolution block 1 and convolution block 2 convolve the input MFCC feature vectors in a translation way, and after the calculation is completed, the data are pooled by using the maximum pooling layer in the convolution block [20]

(3) After the data are processed by the convolution layer, they are input to the bidirectional LSTM network in the lower layer, that is, BLSTM, and the speech features are better extracted and fitted by the bidirectional LSTM network

(4) After feature extraction and fitting, the MeanOverTime layer is used to sum and average the corresponding positions of N vectors with equal lengths output by the upper LSTM network, and finally, a one-dimensional vector is output

(5) Finally, the one-dimensional vector output by the MeanOverTime layer is fitted through the full connection layer, so as to output the corresponding speech score results
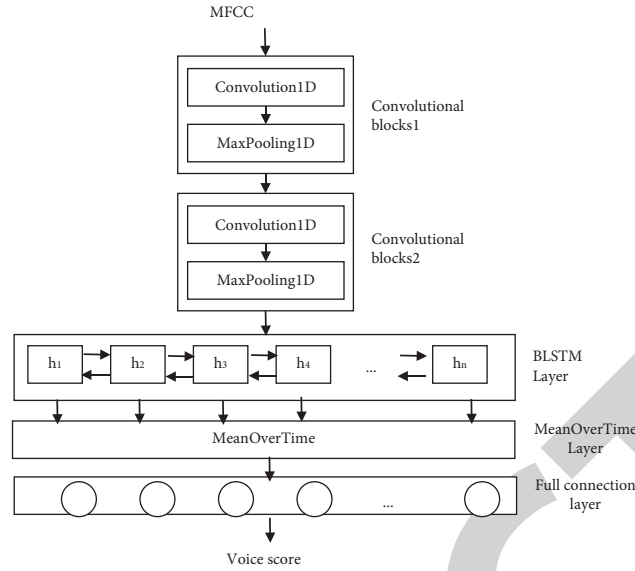
Figure 3: Speech scoring model based on CNN + LSTM.

As shown in Figure 4, the text scoring model based on CNN + LSTM is mainly composed of five layers: word embedding layer, one-dimensional convolution layer, LSTM layer, MeanOverTime layer, and full connection layer. The main function of the MeanOverTime layer is to sum and average the corresponding positions of N equal length vectors (including all intermediate state outputs of the LSTM network) output by the LSTM layer and finally output a one-dimensional vector. Finally, the full connection layer combines the one-dimensional vector and outputs the corresponding speech scoring results.

The specific steps of text scoring design are as follows:

(1) Firstly, the word embedding matrix based on the GloVe model is designed, and the dimension size of the word embedding layer is set to 50. If words cannot be found in the matrix, the word embedding layer vector is set to 0.

(2) The main function of the one-dimensional convolution layer is to shorten the length of the network input sequence and better extract speech features.

(3) The convolution layer is followed by the LSTM layer, which uses the function of selecting "memory" and "forget" information of the LSTM network to better extract and fit text features.

(4) MeanOverTime layer in the text scoring model has the same function as the MeanOverTime layer in the speech scoring model.

(5) Finally, the one-dimensional vector output by the MeanOverTime layer is fitted by the full connection layer, so as to output the corresponding text scoring results.

4. Experimental Verification

4.1. Environmental Construction. To verify the feasibility of the speech scoring model and text scoring model based on

the CNN + LSTM network, the experimental data are the recordings of 600 students' open oral test in a university in Hebei province. The recording files are all in the MP3 form, and the audio attributes are 16 bits and 16 kHz sampling rates. The spoken pronunciation and content are scored by manual scoring. Before training and testing, the FFmpeg tool is used to convert the MP3 format of the recording to the PCM format. Finally, the dataset is divided into two parts: one is the training dataset and the other is the test dataset. Among them, the training set is 480 pieces of data, and the test set is 120 pieces of data.

In the above, this study constructs a speech scoring model and a text scoring model based on the CNN + LSTM network. For these two models, this study uses the RMSProp optimizer and mean square error loss function to train. The training iteration times are set to epoch 50, the batch value is set to 10, and the learning rates are set to 0.001 and 0.01, respectively. On the contrary, because the activation function is used in the output layer of the two scoring models, only the scoring results in the range of 0–1 can be obtained. Therefore, this study needs to normalize the manual scoring results before training the two models. The normalization process is shown in the following:

$$x^* = \frac{x - x_{\min}}{x_{\max} - x_{\min}}. \tag{17}$$

4.2. Data Preprocessing. In the scoring model based on CNN + LSTM constructed in this study, the working form of the scoring model is to transform spoken recording and speech recognition text into a numerical vector representation. In this study, too many datasets are selected, which will lead to jumbled experimental process and increase the difficulty of the experiment. Therefore, it is necessary to preprocess the data in order to achieve better experimental results.
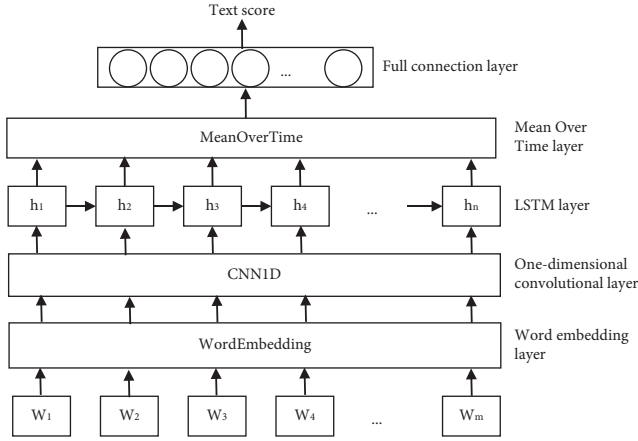
Figure 4: Text scoring model based on CNN + LSTM.

4.2.1. Data Cleaning. In the above dataset, due to the nonstandard spoken language of most students and the recognition defects of the speech recognition system itself, there are large errors in feature extraction of the tested speech recognition text, and the accuracy rate is relatively low, thus affecting the final recognition results. Therefore, this experiment cleans the experimental data, filters and removes some useless spoken words, which make the speech text structure clearer, and increases the effectiveness of the generated model. In this paper, adaptive filter (LMS) is used for data cleaning. Voice data formats include MP3 format, VOC format, and Au format. In the system, it is expressed as binary format coding. The system samples audio data in KB/s.

4.2.2. Feature Extraction. Feature extraction is very important in the scoring model, and the reliability and accuracy of the scoring model are determined by feature extraction. In the neural network scoring model based on CNN + LSTM constructed in this study, two kinds of features, speech class and text class, are mainly extracted. Specific feature extraction is shown in Table 1. Speech class feature extraction is directly based on the speech signal, and text class feature extraction is based on the output of the speech recognition engine.

According to the summary of feature extraction categories in Table 1, this study selects four phonetic features to evaluate students' oral pronunciation quality, fluency, and content richness. Among them, the specific function of the rate of speech (ROS) is to describe oral fluency, and the expression is

$$ROS = \frac{N_{words}}{t - t_s}. \tag{18}$$

In equation (18), N_{words} is the total number of words in students' spoken English, T is the total duration of oral recording, and t_s is the mute duration in recording.

In this study, five text features are selected to evaluate the oral content of candidates. Text feature extraction can better reflect the richness of the oral content and effectively deal with special data. In text feature extraction, grammar is

regarded as the standard to judge students' oral English. Through part-of-speech tags, we can judge whether there are grammatical problems in the spoken content in vocabulary. The row label is shown in Table 2.

In the text extraction of speech recognition, grammatical errors exist in sentences and are difficult to be found. In this study, the correct rate of text grammar is calculated by

$$correct\_ratio = \frac{N_g}{N_s}. \tag{19}$$

In equation (19), N_S represents the total number of ternary or quaternary tag combinations in the text, and N_g represents the total number of correct tag combinations.

4.2.3. Data Conversion. In this study, a pretrained word embedding model is used to transform the speech recognition text into a vector representation, as shown in Figure 5.

For spoken speech recording data, the extracted mel-frequency cepstrum coefficient (MFCC) is used as the input of the speech scoring model, as shown in Figure 6.

4.3. Empirical Results

4.3.1. Indicators for Evaluating System Performance. In the scoring system constructed in this study, the speech and text scoring model is based on the CNN + LSTM neural network. For the comprehensive evaluation and comparison of the scoring performance of the two models, this study sets up three evaluation indicators to quantitatively analyze the experimental results.

(1) Pearson's correlation coefficient is the most widely used performance test index in the field of automatic scoring. The main function of this coefficient is to reflect the linear correlation between two sequences, as shown in the following:

$$\rho_{X,Y} = \frac{cov(X, Y)}{\sigma_X \sigma_Y}. \tag{20}$$

In formula (20), $cov(X, Y)$ is denoted as the oblique variance of X and Y, and σ_X and σ_Y are both standard deviations. X is the system score, Y is the teacher score, and the range of $\rho_{X,Y}$ is [−1, 1]. If the result is positive, it indicates that X and Y are positively correlated, and if it is negative, it indicates that X and Y are negatively correlated. The closer the absolute value of this coefficient is to 1, the higher the correlation between X and Y.

(2) Average difference of the man-machine score: it mainly describes the deviation degree between the machine score and manual score. The indicator is represented by the letter d, as shown in the following:

$$d = E|S_{Machine} - S_{Human}|. \tag{21}$$

(3) Accuracy: this experiment uses the method of the maximum error of man-machine scoring to judge

TABLE 1: Summary of the feature extraction of the speech class and text class.

Feature category	Name to be signed	Brief description of characteristics
Phonetic class	Articulation rate	Speed of speech
	Num silence	Number of voice pauses
	Posterior score	Postpronunciation verification probability score
	Speaking ratio	Ratio of effective speaking time to total recording time
Text class	Content length	Total number of words in the text
	Unique words	Number of nonrepeating words in the text
	Syntactic tree depth	Sum of all syntactic tree depths in the text
	Semantic similarity	Semantic similarity between the text and topic
	Good grammar ratio	Correct rate of text grammar

TABLE 2: Part-of-speech tags.

Part-of-speech tags	Description
NN	Noun (singular)
NNS	Noun (plural)
VB	By bar (prototype)
VBD	Verb (past tense)
VUN	Moving river (past participle)
JJ	Adjectives
RB	Adverb
IN	Subordinate conjunctions
CC	Conjunctions
PRP	Personal pronoun
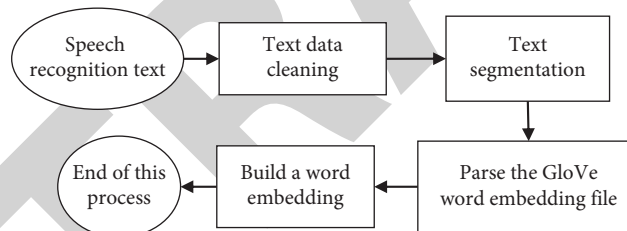
FIGURE 5: Text vector flow.
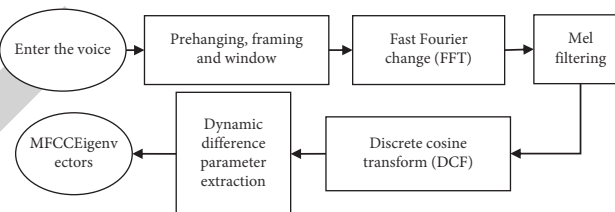
FIGURE 6: MFCC feature extraction flow.

whether the scoring result is accurate or not. If the difference between the manual score and machine score is less than the maximum error, it indicates that the score result is accurate.

Finally, the ratio of the number of accurate scoring results in the test set to the total number of samples is calculated, which is used as the accuracy index of the scoring model. In practical applications, the error of man-machine scoring is less than 1 point, which shows that the machine prediction result is accurate.

4.3.2. Analysis and Evaluation of the Scoring Model. In the CNN + LSTM neural network model, 9 features are extracted as the input of speech and text scoring models, and Pearson's correlation coefficients between each feature and manual scoring are calculated, respectively. The calculated results are shown in Tables 3 and 4.

It can be seen from Tables 3 and 4 that the Pearson correlation coefficients of num silence and speaking ratio are 0.44 and 0.42, respectively, which are highly correlated with manual scores. It shows that teachers pay more attention to

TABLE 3: Correlation between phonetic features and manual scores.

Phonetic features	Pearson's correlation coefficient
Articulation rate	0.37
Num silence	0.44
Posterior score	0.31
Speaking ratio	0.42

TABLE 4: Correlation between text class features and manual scoring.

Text class feature	Pearson's correlation coefficient
Content length	0.57
Unique words	0.59
Syntactic tree depth	0.27
Semantic similarity	0.33
Good grammar ratio	0.24
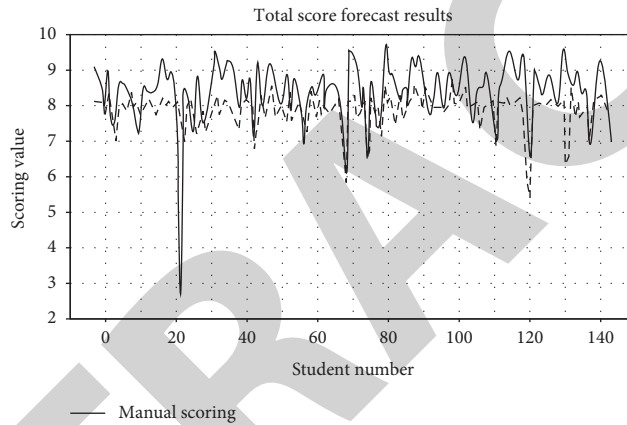
FIGURE 7: Scoring results of the BP scoring model.
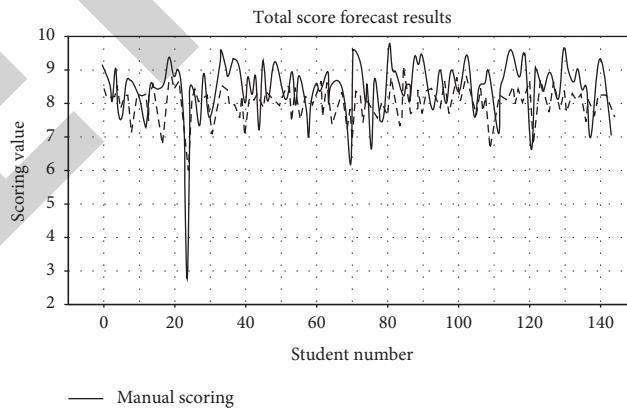
FIGURE 8: Scoring results of the CNN + LSTM neural network scoring model.

TABLE 5: Performance evaluation of the scoring model.

	Pearson's correlation	Average difference	Accuracy (%)
Manual scoring model	0.764	0.485	—
CNN + LSTM model	0.694	0.639	82.3
BP model	0.533	0.601	80.0

students' oral fluency and effective oral duration when grading oral English. The Pearson correlation coefficients of content length and unique words are 0.57 and 0.59, respectively, which have high correlation with manual scoring. It shows that teachers pay more attention to students' mastery of vocabulary and the richness of the oral content.

4.3.3. Test Results. As shown in Figures 7 and 8, 120 test data were selected to test the scoring models of the BP neural network and CNN + LSTM neural network, and the performance of these two scoring models was comprehensively evaluated through the three evaluation indexes set above.

It can be seen from Figures 7 and 8 that the CNN + LSTM neural network scoring model has a better fitting effect than the BP neural network scoring model, and the CNN + LSTM neural network scoring model has stronger adaptability. In Figures 7 and 8, the dotted line represents the standard actual test sample data. In the BP scoring model in Figure 7, it is quite different from the standard sample data, indicating that the evaluation effect is poor. In Figure 8, the difference between the CNN + LSTM model and the standard sample data is small, indicating that the evaluation effect is good.

It can be seen from Table 5 that the CNN + LSTM neural network model is better than the BP neural network model in the Pearson correlation coefficient and accuracy, and the difference between them in man-machine score correlation is obvious. CNN + LSTM is slightly better than the BP model in the average difference index.

5. Conclusion

To sum up, the open oral English scoring model based on CNN + LSTM designed in this study has better fitting effect and adaptability and can improve the intelligent level of open oral English learning. Compared with the traditional artificial scoring and BP neural network scoring model, the CNN + LSTM neural network scoring model has higher scoring accuracy and efficiency. Experimental results show that the CNN + LSTM neural network model has better scoring performance and has certain practicability when the training dataset is small.

Data Availability

The experimental data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest regarding this work.

References

[1] S. Ghannay, C. Servan, and S. Rosset, "Neural networks approaches focused on french spoken language understanding: application to the MEDIA evaluation task," in *Proceedings of the 28th International Conference on Computational Linguistics*, pp. 2722–2727, Barcelona, Spain, 2020.

[2] S. T. Tsai, E. J. Kuo, and P. Tiwary, "Learning molecular dynamics with simple language model built upon long short-term memory neural network," *Nature Communications*, vol. 11, no. 1, p. 5115, 2020.

[3] S. E. Chazan, S. Gannot, and J. Goldberger, "Attention-based neural network for joint diarization and speaker extraction," in *Proceedings of the 2018 16th International Workshop on Acoustic Signal Enhancement*, pp. 301–305, Tokyo, Japan, 2018.

[4] N. Peng, "Performance evaluation of English learning through computer mode using neural network and AI techniques," *Journal of Intelligent and Fuzzy Systems*, vol. 40, no. 3, pp. 1–11, 2020.

[5] Y. Chen, "College English teaching quality evaluation system based on information fusion and optimized RBF neural network decision algorithm," *Journal of Sensors*, vol. 2021, Article ID 6178569, 9 pages, 2021.

[6] Y. Jiang, J. Zhang, and C. Chen, "Research on a new teaching quality evaluation method based on improved fuzzy neural network for college English," *International Journal of Continuing Engineering Education and Life-Long Learning*, vol. 28, no. 3-4, p. 293, 2018.

[7] E. M. Mohammed, M. S. Sayed, A. M. Moselhy, and A. A. Abdelnaiem, "LPC and MFCC performance evaluation with artificial neural network for spoken language identification," *International Journal of Signal Processing Image Processing & Pattern Recognition*, vol. 6, no. 3, pp. 55–66, 2013.

[8] H. Li, "Application research of BP neural network in English teaching evaluation," *Telkomnika Indonesian Journal of Electrical Engineering*, vol. 11, no. 8, pp. 4602–4608, 2013.

[9] A. Hermanto, T. B. Adji, and N. A. Setiawan, "Recurrent neural network language model for English-Indonesian machine translation: experimental study," in *Proceedings of the International Conference on Science in Information Technology*, pp. 132–136, Yogyakarta, Indonesia, 2016.

[10] A. Esan, J. Oladosu, C. Oyeleye et al., "Development of a recurrent neural network model for English to Yorùbá machine translation," *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 5, pp. 602–609, 2020.

[11] T. Maximilian, L. J. Lennart, and E. Nicole, "Deep learning-based detection and classification of geographic atrophy using a deep convolutional neural network classifier," *Graefes Archive for Clinical & Experimental Ophthalmology*, vol. 256, pp. 2053–2060, 2018.

[12] T. Liang, G. Yang, F. Lv, J. Zhang, Z. Cao, and Q. Li, "Convolutional neural networks for text classification with multi-size convolution and multi-type pooling," *Database Systems for Advanced Applications*, Springer, vol. 10829, pp. 3–12, Berlin, Germany, 2018.

[13] A. H. Fath, F. Madanifar, and M. Abbasi, "Implementation of multilayer perceptron (MLP) and radial basis function (RBF) neural networks to predict solution gas-oil ratio of crude oil systems," *Petroleum*, vol. 6, no. 1, pp. 80–91, 2020.

[14] Y. Gao, G. Cai, H. Li et al., "Diagnosis and prognosis prediction of ovarian cancer with feedforward neural network by mining real-world laboratory tests," *Gynecologic Oncology*, vol. 159, pp. 338-339, 2020.

[15] L. Liu, C. Shen, and A. Hengel, "Cross-convolutional-layer pooling for image recognition," *IEEE Transactions on Pattern*

Research Article

SAR Target Recognition Based on Joint Representation of Multimode Representations

Youchun Qiu 🆔

Luzhou Vocational & Technical College, Luzhou 646000, China

Correspondence should be addressed to Youchun Qiu; qiuyouchun@lzy.edu.cn

For the problems of feature extraction and decision making in synthetic aperture radar (SAR) image target recognition, a method based on multimode clustering and decision fusion is proposed. The bidimensional variational mode decomposition (BVMD) is used to decompose the SAR image to obtain multiple modes, which provide multilevel descriptions of the target characteristics. Clustering is performed based on the intrinsic correlation of multiple modes, and several subsets with different modes are selected. Based on the joint sparse representation (JSR), each mode subset is classified, and the corresponding reconstruction error vector is obtained. The linear weighted fusion is employed to fuse the results from different mode subsets. Finally, a decision is made based on the fused results. Experiments are carried out based on the MSTAR dataset. The results show the effectiveness of the method under the standard operating condition (SOC) and robustness under extended operating conditions (EOCs).

1. Introduction

rSynthetic aperture radar (SAR) can work in all weather conditions to obtain high-resolution images that can be used for interpretation [1]. The existing SAR target recognition methods are mainly improved or innovated from the two key steps of feature extraction and classification in order to improve the final recognition performance. Feature extraction aims to obtain de-redundant, low-dimensional representations of the original SAR images. In [2–7], the geometric features were extracted such as target area, shadow, contour, and so on to design a target recognition method. In [2], the Zernike moments were adopted as regional features for SAR target recognition. Ding et al. applied binary morphological operations in region matching and defined a robust similarity measure. The target outlines were used as basic features to evaluate the similarities between different SAR images with appellation to target recognition in [6, 7]. The principal component analysis (PCA), kernel PCA (KPCA), monogenic signal, bidimensional empirical mode decomposition (BEMD), and multiresolution representations were employed to develop SAR target recognition algorithms [8–15]. In [11], the monogenic signal was introduced into SAR target recognition with good performance. BEMD was employed by Chang et al. to enhance the discrimination of extracted features. Taking into account the electromagnetic scattering characteristics of the target, the scattering center parameters of the target can be estimated by analyzing the pixel distribution of the SAR image. In [16–18], several matching schemes based on scattering centers were developed and applied. In the classification stage, the corresponding decision-making mechanism is mainly designed by using mature classifiers or according to the characteristics of the features, including K-nearest neighbor (KNN) [8], support vector machine (SVM) [19–21], sparse representation-based classification (SRC) [21–26], and convolutional neural network (CNN) [27–40].

This paper proposes a SAR target recognition method based on multimode clustering and decision fusion. The bidimensional variational mode decomposition (BVMD) [41, 42] is used to obtain multimode representations of the SAR image, which can more effectively characterize the global, detailed, and bidimensional time-frequency characteristics of the target. In [13], these multilevel decomposition features were proved effectively to improve the performance of SAR target recognition. Therefore, BMVD

can provide effective features for SAR target recognition. However, this method only selected several decomposition modes empirically and did not fully analyze the contribution of each mode to the final recognition result. As a result, part of the modes involved in the decision making may have adverse effects on the final recognition. Therefore, this paper first clusters the multiple modes obtained by BVMD by investigating their intrinsic correlation and obtains several mode subsets. In each subset, multiple modes obtained by clustering have strong internal correlation. In this paper, the joint sparse representation (JSR) is used for classification of each mode subset [11–13], and the reconstruction error vector corresponding to each category is obtained. As reported, JSR is an extended version of traditional SRC, which could be directly used to handle several sparse representation problems especially when they shared some correlations. For the case of the BVMD features in this paper, they are generated from the same inherently related object. Therefore, JSR is a suitable classifier for them. For the reconstruction error vectors of different mode subsets, the linear weighted fusion is used to perform fusion analysis on them. Finally, the target label of the test sample is determined according to the final fusion error vector. Some experiments are designed and performed based on the MSTAR dataset. According to the experimental results, the performance of the proposed can be quantitively validated.

2. Basics of BVMD

First developed by Dragomiretskiy and Zosso [41], the variational mode decomposition (VMD) provides an adaptive signal processing tool to decompose the input as components with specified frequencies. Compared with wavelet analysis, empirical mode decomposition (EMD), etc., it is validated that VMD has better effectiveness and robustness. As an extension of VMD, BVMD was developed to process 2D signals like images [42]. The basic problem in BVMD is stated as follows:

$$\begin{cases} \min_{u_k, \omega_k} & \left\{ \sum_{k=1,\dots K} \alpha_k \left\| \nabla \left[u_{AS,k}(x) e^{-j(\omega_k, x)} \right] \right\|_2^2 \right\} \\ \text{s.t.} & \forall x: \sum_{k=1,\dots K} u_k(x) = f(x), \end{cases} \tag{1}$$

where $f(x)$ is the input; $u_{AS,k}(x)$ represents the 2D analytic signal corresponding to the kth decomposition; and ω_k provides a reference direction in the frequency domain.

The problem in equation (1) can be reformulated by the Lagrangian multiplier as follows:

$$L(\{u_k\}, \{\omega_k\}, \lambda) = \sum_{k=1,\dots K} \alpha_k \left\| \nabla \left[u_{AS,k}(x) e^{-j(\omega_k, x)} \right] \right\|_2^2 + \left\| f(x) - \sum_{k=1,\dots K} u_k(x) \right\|_2^2 + \langle \lambda(x), f(x) - \sum_{k=1,\dots K} u_k(x) \rangle. \tag{2}$$

Afterwards, an unconstrained optimization problem is obtained as follows:

$$\min_{u_k, \omega_k} \max_{\lambda} L(\{u_k\}, \{\omega_k\}, \lambda), \tag{3}$$

where λ and α_k correspond to the Lagrangian multiplier and balance parameter, respectively. $\{u_k\}$ contain the K decompositions, and $\{\omega_k\}$ include the corresponding center frequencies.

According to (3), the alternate direction method of multipliers (ADMM) can be employed to solve the above problem. The decomposition process is updated as follows:

$$\widehat{u}_k^{n+1}(\omega) = \frac{\left(\widehat{f}(\omega) - \sum_{i \neq k} \widehat{u}_i(\omega) + (\widehat{\lambda}(\omega)/2) \right)}{1 + 2\alpha_k |\omega - \omega_k|}, \tag{4}$$

where $\widehat{f}(\omega)$, $\widehat{u}(\omega)$, and $\widehat{\lambda}(\omega)$ correspond to the Fourier transforms of $f(x)$, $u(x)$, and $\lambda(x)$, respectively.

The center frequency ω_k is updated using a similar idea:

$$\omega_k^{n+1} = \frac{\int_{\Omega_k} \omega |\widehat{u}_k^{n+1}(\omega)|^2 d\omega}{\int_{\Omega_k} |\widehat{u}_k^{n+1}(\omega)|^2 d\omega}, \tag{5}$$

where $|\widehat{u}_k(\omega)|^2$ represents the power spectrum on the half-plane Ω_k.

A standard gradient ascent can be used to update the Lagrangian multiplier λ with a fixed time step τ:

$$\lambda^{n+1}(x) = \lambda^{n+1}(x) + \tau \left(f(x) - \sum_{k=1,\dots,K} u_k^{n+1}(x) \right). \tag{6}$$

This paper employs BVMD for feature extraction in SAR target recognition. The decomposed multimode representations could comprehensively describe the properties of the target, including the region, outline, etc. Therefore, different modes from BVMD could complement each other as for providing discriminative information for the original image. Therefore, it is promising that the joint use of the multimode representations could enhance the recognition performance.

3. Decision Fusion Based on Multimode Representations

3.1. Clustering of Multimode Representations. Based on the BVMD decompositions, the multimode representations of the same SAR image can be obtained, which share certain correlations. However, due to the impact of EOCs, these correlations may not be global. Therefore, there may be several subsets with strong correlations in multimode representations. This paper uses a correlation-based clustering

algorithm to achieve multimode subset division. The correlation metric is defined as follows:

$$\lambda = \max\left(\frac{\sum_k\sum_l[\mathbf{c}_1(k,l) - m_1][\mathbf{c}_2(k - \Delta k, l - \Delta l) - m_2]}{\left[\sum_k\sum_l[\mathbf{c}_1(k,l) - m_1]^2[\mathbf{c}_2(k - \Delta k, l - \Delta l) - m_2]^2\right]^{1/2}}\right),$$
(7)

where $\mathbf{c}_1$ and $\mathbf{c}_2$ are two different modes obtained by decompositions of the same SAR image; m_1 and m_2 represent the pixel mean; and Δk and Δl are the two-dimensional offset distances of the reference image along the mode image. The correlation coefficient obtains the maximum image correlation under different offset distances while taking the maximum value.

Denote the modes obtained by the decomposition of the same SAR image as $V = \{\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_N\}$, and the correlation coefficient between any two modes is calculated according to the similarity measure in equation (7). The results are shown in Table 1. In the table, $\lambda_{i,j}$ denotes the similarity between the ith and jth decompositions from BVMD. On this basis, all the modes are clustered using the correlation threshold T_c. When the correlation between a single two modalities is higher than T_c, they are considered to belong to a modal subset. After the clustering, several mode subsets can be obtained. In each subset, the modes in it share high correlations.

3.2. Joint Representation.

For the multiple mode subsets obtained by clustering, this paper uses the JSR to independently represent and classify them. For a certain subset $\mathbf{y}_i$, which contains K modes, i.e., $\left[\mathbf{g}^{(1)} \;\; \mathbf{g}^{(2)} \;\; \cdots \;\; \mathbf{g}^{(K)}\right]$, the basic form of JSR is as follows:

$$\min_{\Lambda}\left\{q(\Lambda) = \sum_{k=1}^{K}\left\|\mathbf{g}^{(k)} - \mathbf{D}^{(k)}\alpha^{(k)}\right\|\right\},$$
(8)

where $\mathbf{D}^{(k)}$ is the dictionary corresponding to the kth mode in the subset; $\alpha^{(k)}$ is the corresponding sparse coefficient vector; and $\Lambda = \left[\alpha^{(1)} \;\; \alpha^{(2)} \;\; \cdots \;\; \alpha^{(k)}\right]$ is the sparse coefficient matrix.

In order to make full use of the correlation of different modes in this subset, the JSR model further uses the ℓ_1/ℓ_2 norm to constrain the coefficient matrix Λ. The updated objective function is as follows:

$$\min_{\Lambda}\left(q(\Lambda) + \lambda\|\Lambda\|_{2,1}\right).$$
(9)

According to the sparse coefficient matrix obtained by solving equation (9), the reconstruction error of each mode in the subset can be calculated separately to obtain the sum of reconstruction errors, as shown below:

$$r_i(j) = \sum_{k=1}^{K}\left\|\mathbf{g}^{(k)} - \mathbf{D}_j^{(k)}\alpha_j^{(k)}\right\|, \quad j = 1, 2, \ldots, C,$$
(10)

where $\mathbf{D}_j^{(k)}$ and $\alpha_j^{(k)}$ correspond to the dictionary and sparse coefficients of the ith mode and the jth class, respectively, and $r_i(j)$ is the reconstruction error of the ith mode to the jth class.

The reconstruction error vector for the M mode subsets can be calculated according to the same idea mentioned above, denoted as $\mathbf{r}_i(j), i = 1, 2, \ldots, M$. Afterwards, the linear weighting can be employed to fuse the reconstruction error vectors of different mode subsets as follows:

$$\mathbf{r}(i) = \omega_1\mathbf{r}_1(i) + \omega_2\mathbf{r}_2(i) + \cdots + \omega_M\mathbf{r}_M(i),$$
(11)

where $\omega_i (i = 1, 2, \ldots, M)$ are the weights corresponding to different mode subsets, which are determined according to the number of modes in each subset, i.e., $\omega_i = M_i/M$ where M_i is the number of modalities in the ith subset. For the reconstructed error vector after fusion, the target label of the test sample can be determined according to the principle of the minimum error. According to the above analysis, the SAR target recognition process designed in this paper is shown in Figure 1.

4. Experiments

4.1. MSTAR Dataset.

The SAR image dataset released by the US DARPA/AFRL in the MSTAR program is used as the experimental data source. This dataset is obtained by the X-band airborne SAR platform, containing multiview SAR images of 10 types of ground stationary vehicles, with a range and azimuth resolution of 0.3 m. Figure 2 shows the optical images of these 10 targets. Based on this dataset, multiple types of experimental scenarios can be set to conduct a more comprehensive performance analysis of the proposed method.

In the process of testing the proposed method, it is simultaneously compared and analyzed with some methods in the existing literature, including the SRC method [22], monogenic signal method [11], BEMD method [13], and A-ConvNet method [30]. All these reference methods are implemented by the author according to the idea in the original literatures. They are tested and compared with the proposed method under the same conditions.

4.2. Results and Discussion

4.2.1. SOC.

Table 2 shows a typical SOC based on the MSTAR dataset. The threshold value in the mode clustering algorithm is set as 0.45, and the proposed method is used to classify 10-class test samples, which obtains the confusion matrix as shown in Figure 3. Among them, the horizontal and vertical coordinates correspond to the real class and the predicted result, respectively. The diagonal element corresponds to the correct recognition rate of the corresponding target. The average recognition rate of the proposed method for all 10 targets is calculated to reach 99.12%, showing its excellent performance under SOC. The four types of comparison methods are tested under the same condition, and their average recognition rates are compared with the proposed one as shown in Table 3. It can be seen that various methods can achieve good performance under SOC. Compared with the BEMD method, the proposed method performs multimode clustering and decision fusion based on the inner correlations to further

TABLE 1: Correlation coefficients among multiple modes.

	$\mathbf{c}_1$	$\mathbf{c}_2$	$\cdots$	$\mathbf{c}_N$
$\mathbf{c}_1$	λ_{11}	λ_{12}	$\cdots$	λ_{2N}
$\mathbf{c}_2$	λ_{21}	λ_{22}	$\cdots$	λ_{2N}
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
$\mathbf{c}_N$	λ_{N1}	λ_{N2}	$\cdots$	λ_{NN}
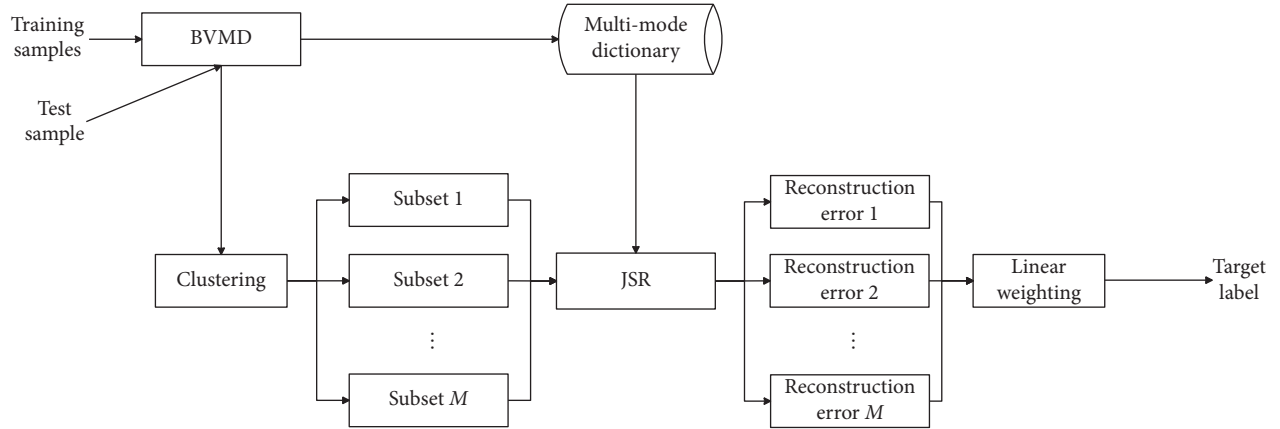
FIGURE 1: Procedure of SAR target recognition based on clustering of multiple modes and decision fusion.

(1) BMP2 (2) BTR70 (3) T72 (4) T62 (5) BRDM2

(6) BTR60 (7) ZSU23/4 (8) D7 (9) ZIL131 (10) 2S1

FIGURE 2: Images of targets in the MSTAR dataset used in the experiments [25].

TABLE 2: Training and test sets under SOC.

Target class	Training set		Test set	
	Configuration	Samples	Configuration	Samples
BMP2	9563	233	9563	195
			9566	196
			c21	196
BTR70	c71	233	c71	196
T72	132	232	132	196
			812	195
			s7	191
T62	A51	299	A51	273
BRDM2	E-71	298	E-71	274
BTR60	7532	256	7532	195
ZSU23/4	d08	299	d08	274
D7	13015	299	13015	274
ZIL131	E12	299	E12	274
2S1	B01	299	B01	274
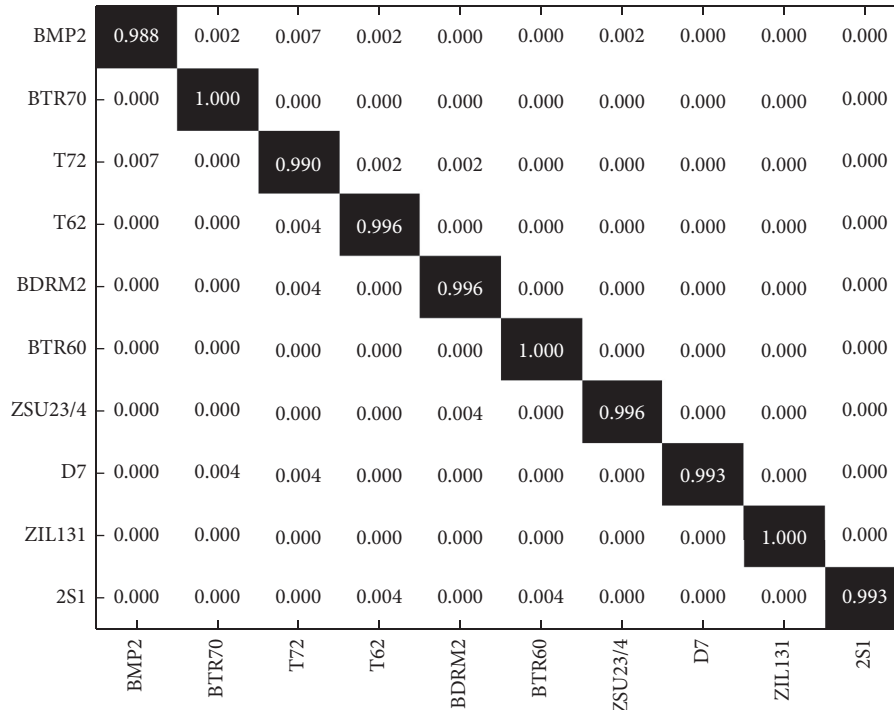
Figure 3: Results of 10-class classification.

Table 3: Average recognition rates under SOC.

Method type	Proposed	SRC	Monogenic	BEMD	ConvNet
Average recognition rate (%)	99.12	96.12	98.26	98.74	98.68

improve the recognition performance. As a method based on deep learning, ConvNet's classification ability is closely related to the scale of training samples. In the experimental settings of Table 2, there are differences in the configurations of some samples, which affect the overall classification performance of ConvNet to a certain extent.

The mode clustering threshold determines the final mode composition participating in the JSR classification, which has an important influence on the final classification performance. Therefore, we test the classification results of the proposed method under several typical threshold values for 10 types of targets. The results are shown in Table 4. When the threshold is small, the clustering algorithm has weaker constraints on the correlation between different modes. As a result, the algorithm degrades into the traditional JSR classification. When the threshold is large, the requirements for the interrelationship in the clustering algorithm are too strict. At this time, each mode tends to become a subset independently, which leads to insufficient investigation of the internal relevance of different modes. The comparison shows that the proposed method has the best performance at times, and subsequent tests and comparative analysis will be carried out under this threshold.

4.2.2. EOCs. Different from the SOC, the EOCs mainly refer to the large difference between the acquisition conditions of the test sample and the training sample, resulting in a low overall similarity. Under most non-cooperative conditions, SAR target recognition occurs under EOCs, typically including target configuration differences, depression angle differences, and noise interference. In this paper, the proposed method is tested under the above three types of EOCs, which are denoted as EOC-1, EOC-2, and EOC-3, respectively. Tables 5 and 6 show the training and test sets under the configuration difference and the depression angle difference. The former sets a completely different target model in the test set from the training set, and the latter sets the test set at 30° and 45° depression angles (the training set is from 17°). Tables 7 and 8 compare the average recognition rates of various methods under EOC-1 and EOC-2, respectively. For configuration differences, the proposed method effectively retains the correct decision adapted to the configuration difference samples through multimode clustering, and the final average recognition rate is also higher than that of the four types of comparison methods. When the depression angle is 30°, all methods can maintain a correct recognition rate higher than 90%. The difference between the test and training samples caused by the change of the depression angle at this time is relatively small. However, when the depression angle is 45°, the recognition performance of various methods decreases drastically. At this time, there is a big difference between the test sample and the training sample. The proposed method achieves the highest

TABLE 4: Average recognition rates at different clustering thresholds.

Threshold T_c	0.25	0.35	0.45	0.55	0.65
Average recognition rate (%)	98.56	98.78	99.12	98.96	98.52

TABLE 5: Training and test sets for EOC-1.

Target class	Training set		Test set	
	Configuration	Samples	Configuration	Samples
BMP2	9563	233	9566	196
			c21	196
T72	132	232	812	195
			s7	191
BTR60	7532	256	7532	195
T62	A51	299	A51	273

TABLE 6: Training and test sets for EOC-2.

Target class	Training set		Test set	
	Depression (°)	Samples	Depression (°)	Samples
2S1	17	299	30	288303
			45	
BDRM2	17	298	30	287303
			45	
ZSU23/4	17	299	30	288303
			45	

TABLE 7: Average recognition rates of different methods under EOC-1.

Method type	Proposed	SRC	Monogenic	BEMD	ConvNet
Average recognition rate (%)	98.06	94.62	97.01	97.27	96.12

TABLE 8: Average recognition rates of different methods under EOC-2.

Method type	Average recognition rate (%)	
	30°	45°
Proposed	97.90	73.60
SRC	94.16	68.07
Monogenic	96.96	69.86
BEMD	97.08	70.56
ConvNet	95.82	67.72

performance in both cases, showing its stronger robustness to depression angle differences.

According to [17], 10 types of target test samples in Table 2 are processed by adding noise to construct test sets at different signal-to-noise ratios (SNRs). This paper defines SNR as follows:

$$SNR (dB) = 10 lg \frac{\sum_{h=1}^{H} \sum_{w=1}^{W} |Im(i, l)|^2}{HW\sigma^2}, \quad (12)$$

where $Im(i, l)$ is the SAR pixel; σ^2 is the variance of the added additive white Gaussian noise; and the denominator term is the total energy of the added noise.

As the noise intensifies, the significance of the target characteristics gradually weakens, and the difficulty of recognition also increases. The noise samples are used to test various methods, and the comparison results are obtained, as shown in Figure 4. It can be seen that as the SNR decreases, the recognition performance of various methods decreases to varying degrees. The proposed method obtains stronger noise robustness by combining the advantages of clustering multiple modes and decision fusion. A comprehensive comparison of the recognition results under the three types of EOCs shows that the proposed method has stronger adaptability to EOCs and is beneficial to obtain more reliable recognition results.
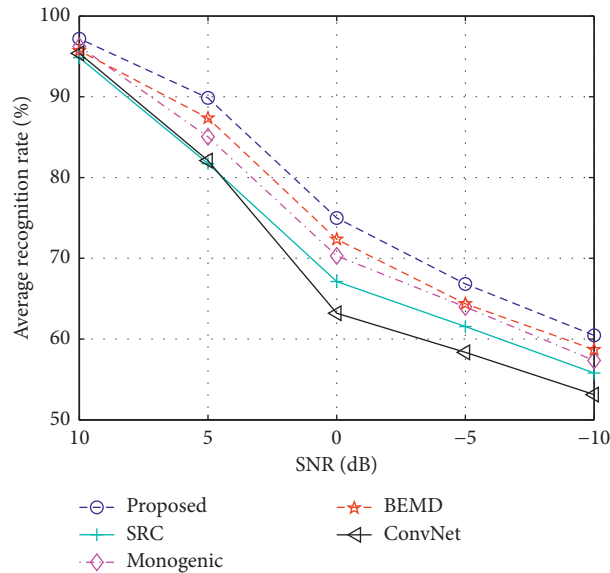
FIGURE 4: Average recognition rates of different methods under EOC-3.

5. Conclusion

This paper uses BVMD to decompose SAR images, thus obtaining multimode representations. In order to adaptively classify each test sample, all modes are clustered based on the principle of correlation, and multiple mode subsets with intrinsic correlation are obtained. The JSR is used to make decisions on each mode subset, and finally the decision is obtained through linear weighted fusion. Based on the MSTAR dataset, the proposed method is tested and compared under SOC and three types of EOCs. The experimental results show that the proposed method is more effective and robust than several existing methods.

Data Availability

The MSTAR dataset can be accessed upon request.

Conflicts of Interest

The author declares that there are no conflicts of interest.

References

[1] K. El-Darymli, E. W. Gill, D. Power, and C. Moloney, "Automatic target recognition in synthetic aperture radar imagery: a state-of-the-art review," *IEEE Access*, vol. 4, pp. 6014–6058, 2016.

[2] M. Amoon and G. a. Rezai-rad, "Automatic target recognition of synthetic aperture radar (SAR) images based on optimal selection of Zernike moments features," *IET Computer Vision*, vol. 8, no. 2, pp. 77–85, 2014.

[3] B. Ding, G. Wen, C. Ma, and X. Yang, "Target recognition in synthetic aperture radar images using binary morphological operations," *Journal of Applied Remote Sensing*, vol. 10, no. 4, Article ID 046006, 2016.

[4] C. Shan, B. Huang, and M. Li, "Binary morphological filtering of dominant scattering area residues for SAR target recognition," *Computational Intelligence and Neuroscience*, vol. 2018, Article ID 9680465, 15 pages, 2018.

[5] L. Jin, J. Chen, and X. Peng, "Synthetic aperture radar target classification via joint sparse representation of multi-level dominant scattering images," *Optik*, vol. 186, pp. 110–119, 2019.

[6] J. Tan, X. Fan, S. Wang et al., "Target recognition of SAR images by partially matching of target outlines," *Journal of Electromagnetic Waves and Applications*, vol. 33, no. 7, pp. 865–881, 2019.

[7] S. Papson and R. M. Narayanan, "Classification via the shadow region in SAR imagery," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 48, no. 2, pp. 969–980, 2012.

[8] A. K. Mishra, "Validation of PCA and LDA for SAR ATR," in *Proceedings of the IEEE TENCON*, pp. 1–6, Hyderabad, India, January 2008.

[9] Z. Cui, Z. Cao, J. Yang, J. Feng, and H. Ren, "Target recognition in synthetic aperture radar images via non-negative matrix factorisation," *IET Radar, Sonar & Navigation*, vol. 9, no. 9, pp. 1376–1385, 2015.

[10] W. Xiong, L. Cao, and Z. Hao, "Combining wavelet invariant moments and relevance vector machine for SAR target recognition," in *Proceedings of the IET International Radar Conference*, pp. 1–4, Guilin, China, December 2009.

[11] G. Dong, G. Kuang, N. Wang, L. Zhao, and J. Lu, "SAR target recognition via joint sparse representation of monogenic signal," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 8, no. 7, pp. 3316–3328, 2015.

[12] Y. Zhou, Y. Chen, R. Gao, J. Feng, P. Zhao, and L. Wang, "SAR target recognition via joint sparse representation of monogenic components with 2D canonical correlation analysis," *IEEE Access*, vol. 7, pp. 25815–25826, 2019.

[13] M. Chang, X. You, and Z. Cao, "Bidimensional empirical mode decomposition for SAR image feature extraction with application to target recognition," *IEEE Access*, vol. 7, pp. 135720–135731, 2019.

[14] M. Yu, G. Dong, H. Fan, and G. Kuang, "SAR target recognition via local sparse representation of multi-manifold

regularized low-rank approximation," *Remote Sensing*, vol. 10, no. 2, p. 211, 2018.

[15] Y. Huang, J. Peia, J. Yanga, B. Wang, and X. Liu, "Neighborhood geometric center scaling embedding for SAR ATR," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 50, no. 1, pp. 180–192, 2014.

[16] L. C. Potter and R. L. Moses, "Attributed scattering centers for SAR ATR," *IEEE Transactions on Image Processing*, vol. 6, no. 1, pp. 79–91, 1997.

[17] B. Ding, G. Wen, J. Zhong, C. Ma, and X. Yang, "A robust similarity measure for attributed scattering center sets with application to SAR ATR," *Neurocomputing*, vol. 219, pp. 130–143, 2017.

[18] B. Ding, G. Wen, X. Huang, C. Ma, and X. Yang, "Target recognition in synthetic aperture radar images via matching of attributed scattering centers," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 10, no. 7, pp. 3334–3347, 2017.

[19] Q. Zhao and J. C. Principe, "Support vector machines for SAR automatic target recognition," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 37, no. 2, pp. 643–654, 2001.

[20] M. E. Demirhan and Ö. Salor, "Classification of targets in SAR images using SVM and k-NN techniques," in *Proceedings of the 2016 24th Signal Processing and Communication Application Conference (SIU)*, pp. 1581–1584, Zonguldak, Turkey, June 2016.

[21] H. Liu and S. Li, "Decision fusion of sparse representation and support vector machine for SAR image target recognition," *Neurocomputing*, vol. 113, pp. 97–104, 2013.

[22] J. J. Thiagaraianm, K. N. Ramamurthy, P. Knee, A. Spanias, and V. Berisha, "Sparse representations for automatic target classification in SAR images," in *Proceedings of the 4th International Symposium Communication, Control Signal Process*, pp. 1–4, Limassol, Cyprus, May 2010.

[23] H. Song, K. Ji, Y. Zhang, X. Xing, and H. Zou, "Sparse representation-based SAR image target classification on the 10-class MSTAR data set," *Applied Sciences*, vol. 6, no. 26, 2016.

[24] B. Ding and G. Wen, "Sparsity constraint nearest subspace classifier for target recognition of SAR images," *Journal of Visual Communication and Image Representation*, vol. 52, pp. 170–176, 2018.

[25] W. Li, J. Yang, and Y. Ma, "Target recognition of synthetic aperture radar images based on two-phase sparse representation," *Journal of Sensors*, vol. 2020, Article ID 2032645, 12 pages, 2020.

[26] L. Yu, L. Wang, and Y. Xu, "Combination of joint representation and adaptive weighting for multiple features with application to sar target recognition," *Scientific Programming*, vol. 2021, Article ID 9063419, 9 pages, 2021.

[27] X. X. Zhu, D. Tuia, L. Mou et al., "Deep learning in remote sensing: a comprehensive review and list of resources," *IEEE Geoscience and Remote Sensing Magazine*, vol. 5, no. 4, pp. 8–36, 2017.

[28] M. Kang, K. Ji, X. Leng, X. Xing, and H. Zou, "Synthetic aperture radar target recognition with feature fusion based on a stacked autoencoder," *Sensors*, vol. 17, no. 1, p. 192, 2017.

[29] D. E. Morgan, "Deep convolutional neural networks for ATR from SAR imagery," in *Proceedings of the SPIE*, pp. 1–13, Baltimore, MA, USA, May 2015.

[30] S. Chen, H. Wang, F. Xu, and Y. Jin, "Target classification using the deep convolutional networks for SAR images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 6, pp. 1685–1697, 2016.

[31] J. Zhao, Z. Zhang, W. Yu, and T. Troung, "A cascade coupled convolutional neural network guided visual attention method for ship detection from SAR images," *IEEE Access*, vol. 6, pp. 50693–50708.

[32] R. Min, H. Lan, Z. Cao, and Z. Cai, "A gradually distilled CNN for SAR target recognition," *IEEE Access*, vol. 7, pp. 42190–42200.

[33] L. Wang, X. Bai, and F. Zhou, "SAR ATR of ground vehicles based on ESENet," *Remote Sensing*, vol. 11, no. 11, p. 1316, 2019.

[34] P. Zhao, K. Liu, H. Zou, and X. Zhen, "Multi-stream convolutional neural network for SAR automatic target recognition," *Remote Sensing*, vol. 10, no. 9, p. 1473, 2018.

[35] J. Ding, B. Chen, H. Liu, and M. Huang, "Convolutional neural network with data augmentation for SAR target recognition," *IEEE Geoscience and Remote Sensing Letters*, vol. 13, no. 3, pp. 364–368, 2016.

[36] Y. Yan, "Convolutional neural networks based on augmented training samples for synthetic aperture radar target recognition," *Journal of Electronic Imaging*, vol. 27, no. 2, Article ID 023024, 2018.

[37] D. Malmgren-Hansen, A. Kusk, J. Dall, A. A. Nielsen, R. Engholm, and H. Skriver, "Improving SAR automatic target recognition models with transfer learning from simulated data," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 9, pp. 1484–1488, 2017.

[38] S. A. Wagner, "SAR ATR by a combination of convolutional neural network and support vector machines," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 52, no. 6, pp. 2861–2872, 2016.

[39] O. Kechagias-Stamatis and N. Aouf, "Fusing deep learning and sparse coding for SAR ATR," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 55, no. 2, pp. 785–797, 2019.

[40] C. Jiang and Y. Zhou, "Hierarchical fusion of convolutional neural networks and attributed scattering centers with application to robust SAR ATR," *Remote Sensing*, vol. 10, no. 6, p. 819, 2018.

[41] K. Dragomiretskiy and D. Zosso, "Variational mode decomposition," *IEEE Transactions on Signal Processing*, vol. 62, no. 3, pp. 531–544, 2014.

[42] K. Dragomiretskiy and D. Zosso, "Two-dimensional variational mode decomposition," in *Proceedings of the Energy Minimization Methods in Computer Vision and Pattern Recognition 2015 EMMCVPR*, pp. 197–208, Hong Kong, China, January 2015.

Research Article

A SAR Target Recognition Method Based on Decision Fusion of Multiple Features and Classifiers

Zhengwu Lu, Guosong Jiang, Yurong Guan, Qingdong Wang, and Jianbo Wu [ID]

College of Computer Science, Huanggang Normal University, HuangGang 438000, China

Correspondence should be addressed to Jianbo Wu; wujianbo@hgnu.edu.cn

A synthetic aperture radar (SAR) target recognition method combining multiple features and multiple classifiers is proposed. The Zernike moments, kernel principal component analysis (KPCA), and monographic signals are used to describe SAR image features. The three types of features describe SAR target geometric shape features, projection features, and image decomposition features. Their combined use can effectively enhance the description of the target. In the classification stage, the support vector machine (SVM), sparse representation-based classification (SRC), and joint sparse representation (JSR) are used as the classifiers for the three types of features, respectively, and the corresponding decision variables are obtained. For the decision variables of the three types of features, multiple sets of weight vectors are used for weighted fusion to determine the target label of the test sample. In the experiment, based on the MSTAR dataset, experiments are performed under standard operating condition (SOC) and extended operating conditions (EOCs). The experimental results verify the effectiveness, robustness, and adaptability of the proposed method.

1. Introduction

Synthetic aperture radar (SAR) obtains effective ground observation data through two-dimensional high-resolution imaging, supporting related applications in the military and civilian fields. SAR target recognition uses the feature analysis and classification decision to determine the target class [1]. Feature extraction obtains the effective feature descriptions of the target in SAR images, including geometric shape, scattering centers, and projection transformation features. Literature [2–7] designed SAR target recognition methods based on geometric features such as target region, contour, and shadow. In [2, 3], the Zernike moments were used to describe the target area. In [4], a recognition method was proposed based on target region matching. In [6], the target contour distribution was modeled based on the elliptic Fourier descriptor. Papson proposed a SAR target recognition method based on shadow features [7]. The scattering center features describe the target's backscattering electromagnetic characteristics at the high-frequency region. In [8–10], SAR target recognition

methods were developed using the attribute scattering center as the basic feature. Projection transformation features can be further divided into projection ones and image decomposition ones. Projection features mainly use mathematical transformation algorithms, such as principal component analysis (PCA) or kernel PCA (KPCA) [11, 12] and nonnegative matrix factorization (NMF) [13]. Image decomposition methods include the wavelet analysis [14], monogenic signal [15], and bidimensional empirical mode decomposition (BEMD) [16]. The methods mentioned above are all based on a single feature to carry out target recognition. In fact, combining a variety of different features can effectively improve the performance of SAR target recognition. In [17], the multitask compressive sensing was employed to implement joint classification of multiple features of SAR images. In [18], a multifeature hierarchical decision fusion method was proposed. In [19], a multifeature and multirepresentation fusion strategy was proposed for target recognition. According to the extracted feature categories, the classifier has to analyze and make decisions accordingly to obtain the target label of the unknown

sample. In [20], a SAR target recognition method was developed based on K-nearest neighbor (KNN). The support vector machine (SVM) was employed in [19, 20] as a base classifier to design a SAR target recognition method. The sparse representation classification-based classification (SRC) was employed for SAR target recognition in [12–22]. With the development of deep learning in recent years, the convolutional neural network (CNN) has gradually become a hot tool in SAR target recognition, and a number of representative methods have emerged [23–30]. Similarly, classifier fusion is also used and verified in SAR target recognition. In [21], SVM and SRC were used for fused classification. In [24], the CNN and SVM were combined to further improve the classification performance.

This study proposes a SAR target recognition method based on multiple features and classifiers. Three types of features, i.e., Zernike moments, KPCA, and monogenic signals are used for feature extraction. Zernike moments describe the geometric shape of the target and have the advantages of invariable translation and rotation. The features have clear physical meaning and reflect the details of the target [2, 3, 19]. KPCA extracts the projection features of the original image, which provide a concise feature vector and have a certain nonlinear description ability [11, 12]. The monogenic signal can effectively decompose the SAR image and obtain the multilevel and multifrequency description characteristics [15]. Therefore, the three types of features have good complementarity and can provide more sufficient discriminative information for decision-making. In the classification stage, SVM, SRC, and joint sparse representation (JSR) are used as the classifiers for Zernike moments, KPCA feature vectors, and monogenic features to obtain the corresponding decision variables. On this basis, multiple sets of linear weights are designed to perform weighted fusion on the decision variables of the three types of features [31–36], which determine the target label of the test sample finally. In the experiments, the proposed method is tested under standard operating condition (SOC) and extended operating conditions (EOCs) [37–42] based on the MSTAR dataset. The recognition results and comparative analysis verify the effectiveness and robustness of the proposed method.

2. Extraction of Multiple Features

2.1. Zernike Moment. Zernike moments are widely used in the description of SAR target regions because of the translation and rotation invariance and noise robustness [2, 3, 19]. For the image $I(r, \theta)$ in the polar coordinates, the Zernike moment is calculated as follows:

$$Z_{nl} = \frac{n+1}{\pi} \int_0^{2\pi} \int_0^1 [V_{nl}(r, \theta)] * I(r, \theta) r \mathrm{d}r \mathrm{d}\theta, \quad (1)$$

where $n = 0, 1, \ldots, \infty$, $l = 0, \pm 1, \ldots$, $n - |l|$ is even, and $|l| \leq n$.

Zernike polynomials $V_{nl}(r, \theta) = R(r)e^{il\theta}$ are a set of orthogonal complete complex-valued functions on the unit circle $x^2 + y^2 \leq 1$, which satisfy the following condition:

$$\int_0^{2\pi} \int_0^1 [V_{nl}(r, \theta)] * V_{mk}(r, \theta) r \mathrm{d}r \mathrm{d}\theta = \frac{\pi}{n+1} \delta_{mn} \delta_{kl}. \quad (2)$$

On this basis, the rotation invariant feature is constructed as follows:

$$Z_{nl} = Z_{nl} \times Z_{n,-l} (l = 0, 1, 2, \ldots). \quad (3)$$

Based on the above equations, the Zernike moment at any order of the input image can be calculated. Among them, the high-order moments can effectively reflect the detailed information in the image, which is beneficial to improve the recognition performance. In this study, the 3–8th moments of Zernike moments are used to describe the SAR target region, which constitute a feature vector.

2.2. KPCA. PCA calculates the best projection direction by analyzing the data structure of a large number of samples to achieve data dimensionality reduction [11]. For the sample set $X = \{x_1, x_2, \ldots, x_n\}, X \in \mathrm{R}^{m \times n}$, their mean and covariance matrix are first calculated as follows:

$$\overline{X} = \frac{1}{n} \sum_{i=1}^{n} x_i, \quad (4)$$

$$Q = \sum_{i=1}^{n} (x_i - \overline{X})^T (x_i - \overline{X}). \quad (5)$$

The eigenvalues and eigenvectors of the covariance matrix are calculated as follows:

$$[V, D] = \mathrm{EIG}(Q). \quad (6)$$

In equation (6), the vector V stores the eigenvalues, and each eigenvalue corresponds to the eigenvector in the matrix D. By selecting feature vectors corresponding to several largest eigenvalues, a projection matrix can be constructed for feature extraction of the samples.

KPCA is an extension of PCA in nonlinear space, which can process datasets with nonlinear structure more efficiently [11, 12]. KPCA first processes the data by introducing a kernel function (typically polynomial and radial basis kernel) and then performs PCA operations in high dimensions. In this study, KPCA is used to process SAR images and obtain 80-dimensional feature vectors for subsequent classification.

2.3. Monogenic Signal. The monogenic signal is a two-dimensional signal decomposition algorithm, which can effectively analyze the multilevel spectrum characteristics of the original image [15]. For the input image $f(z)$, its Riesz transformation is calculated as $f_R(z)$, where $z = (x, y)^T$ represents the two-dimensional coordinates. The corresponding monogenic signal $f_M(z)$ is calculated as follows:

$$f_M(z) = f(z) - (i, j) f_R(z), \quad (7)$$

where i and j are the imaginary units. $f(z)$ and its Riesz transform, respectively, correspond to the real part and imaginary part of the monogenic signal, respectively.

Accordingly, the characteristics of the monogenic signal are defined as follows:

$$A(z) = \sqrt{f(z)^2 + |f_R(z)|^2},$$

$$\varphi(v) = a \tan 2\left(|f_R(z)|, f(z)\right) \in (-\pi, \pi), \quad (8)$$

$$\theta(z) = a \tan 2\left(f_y(z)/f_x(z)\right) \in \left(-\frac{\pi}{2}, \frac{\pi}{2}\right].$$

In the above equations, $f_x(z)$ and $f_y(z)$ correspond to the i-imaginary part and j-imaginary part of the monogenic signal, respectively; $A(z)$ represent the amplitude information; $\varphi(z)$ and $\theta(z)$ correspond to the local phase and azimuth information, respectively.

The three types of features obtained based on monogenic signal decomposition have different characteristics. Among them, $A(z)$ mainly reflects the gray distribution characteristics of the image. $\varphi(z)$ and $\theta(z)$ reflect the local detail information and shape characteristics of the image. Therefore, the joint use of the characteristics of the monogenic signal is conducive to constructing a more informative characterization. In this study, according to [9], the SAR image is decomposed by monogenic signal, and three corresponding feature vectors are obtained through downsampling and vector concatenation.

3. Multiple Classifiers and Decision Fusion

3.1. SVM for Zernike Moments.
For the classification problem of the two types of patterns, SVM obtains the best classification interface by minimizing structural risks [24]. For the sample x with unknown label, the decision hyperplane of SVM classification is as follows:

$$w^T \cdot \phi(x) + b = 0, \quad (9)$$

where w is the weight coefficient vector, which is used to describe the relevant parameters of the hyperplane, $\phi(\cdot)$ represents the kernel function, and b represents the bias.

At first, SVM was proposed for the recognition of two types of patterns, namely, the hyperplane of equation (9) was used to distinguish between the two types. In the later period, researchers extended it to the classification of multiclass models through strategies such as "one-to-one" and "one-to-many." Through the training of a large number of labeled training samples, a suitable classification surface can be obtained. At the same time, choosing a suitable kernel function can effectively enhance the nonlinear classification ability of SVM. When using SVM for multicategory classification, the (pseudo) posterior probability of each category is output to represent the possibility that the current sample belongs to a certain training class. The type of test sample can be determined by the principle of maximum posterior probability. In this study, SVM is used to classify Zernike moments [43–46].

3.2. SRC for KPCA Features.
SRC uses sparse representation as a basic algorithm to characterize test samples with unknown classes and then determines their category based on

the analysis of the reconstruction errors [14, 15, 23]. Dictionary construction is one of the key steps in SRC. Existing methods mostly use samples of all training classes to construct a global dictionary $A = [A^1, A^2, \ldots, A^C] \in R^{d \times N}$, in which $A^i \in R^{d \times N_i} (i = 1, 2, \ldots, C)$ contains all training samples from the ith training category. Accordingly, the sparse reconstruction of the test sample y is described as follows:

$$\hat{\alpha} = \arg \min \|\alpha\|_0,$$
$$\text{s.t.} \quad \|y - A\alpha\|_2^2 \le \varepsilon, \quad (10)$$

where α is the sparse representation coefficient vector to be solved, and ε is the reconstruction error threshold.

Under the constraint of ℓ_0 norm, it is very difficult to solve the sparse representation coefficient in equation (10). For this reason, researchers used ℓ_1 norm minimization [23] to approximate the problem in equation (10) and converted it into a convex optimization problem that is easy to solve. In addition, the orthogonal matching pursuit algorithm (OMP) [14], Bayesian compressive sensing (BCS) [15], and other algorithms can also be used to obtain the approximate solution of equation (10). Based on the solutions, the decision process is performed as follows:

$$r(i) = \|y - A_i \hat{\alpha}_i\|_2^2 (i = 1, 2, \ldots, C),$$
$$\text{identity}(y) = \arg \min_i (r(i)), \quad (11)$$

where $\hat{\alpha}_i$ are the coefficients corresponding to the ith training class from the extraction in $\hat{\alpha}$, and $r(i)$ $(i = 1, 2, \ldots, C)$ represent the reconstruction errors of different classes.

Studies have shown that SRC has good robustness against noise interference and occlusion [23], which can effectively supplement SVM. This study uses SRC to classify KPCA feature vectors.

3.3. JSR for Monogenic Features.
For the different types of features extracted from the same SAR image, they have a certain inherent correlation. For this reason, this study uses JSR to jointly represent them, thereby improving the overall accuracy [5, 15, 26]. The three monogenic feature vectors obtained from the test sample y are denoted as $[y^{(1)} y^{(2)} y^{(3)}]$. This study uses JSR for classification. The basic representation process is as follows:

$$\min_{\beta} \left\{ g(\beta) = \sum_{k=1}^{3} \left\| y^{(k)} - \Phi^{(k)} x^{(k)} \right\| \right\}, \quad (12)$$

where $\Phi^{(k)}$ is the global dictionary corresponding to the kth feature, $x^{(k)}$ is the corresponding coefficient vector, and $\beta = [x^{(1)} x^{(2)} x^{(3)}]$.

The objective function in equation (12) does not take into account the inherent relationship of the three types of features. This goal can be achieved by constraining the sparse matrix β. The updated objective function is as follows:

$$\min_{\beta} \left(g(\beta) + \lambda \|\beta\|_{2,1} \right). \quad (13)$$

In equation (13), the ℓ_1/ℓ_2 norm is used to constrain β, which can effectively use the internal relationship of the three types of features.

According to the obtained coefficient matrix β, the sum of the reconstruction errors of each class for the three types of features can be calculated, and then, the target label of the test sample can be decided:

$$\text{Identity}(y) = \min_i \sum_{k=1}^{3} \left\| y^{(k)} - \Phi_i^{(k)} x_i^{(k)} \right\|, \tag{14}$$

where $\Phi_i^{(k)}$ and $x_i^{(k)}$ are the part of the dictionary and the corresponding coefficient vector corresponding to the kth feature in the ith class.

3.4. Decision Fusion. For SRC, the output result is the reconstruction error vector $[r_1, r_2, \ldots, r_C]$. First, these reconstruction errors are converted into a probability vector according to the following:

$$P_i^S = 1 - \frac{r_i}{\sum_{j=1}^{C} r_j}. \tag{15}$$

This study uses multiple sets of weights for linear fusion, so as to obtain a more robust result. Denoted r_k^i as the decision variable of the kth feature of the ith class, first construct N weight vectors:

$$W = \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1N} \\ w_{21} & w_{22} & \cdots & w_{2N} \\ w_{31} & w_{32} & \cdots & w_{3N} \end{bmatrix}. \tag{16}$$

In the formula, each column in the matrix W represents a weight vector, which satisfies

$$\sum_{k=1}^{K} w_{ki} = 1, \quad w_{ki} \geq 0. \tag{17}$$

The weighting process under a set of weight vectors is as follows:

$$R_n^i = \begin{bmatrix} r_1^i & r_2^i & r_3^i \end{bmatrix} \begin{bmatrix} w_{n1} \\ w_{n2} \\ w_{n3} \end{bmatrix}. \tag{18}$$

Therefore, in the group of N random weight vectors, the ith training sample obtains a weighted result $R = \begin{bmatrix} R_1^i & R_2^i & \cdots & R_N^i \end{bmatrix}$, which is called the fusion decision vector. Finally, the N decision variable is averaged as the final decision value of the ith category, and the target category of the test sample is determined through various comparisons. It can be seen that under the action of multiple sets of weight vectors, the three types of features participating in the fusion can be fully analyzed.

4. Experiments

4.1. MSTAR Dataset. The MSTAR dataset is a representative dataset for testing and evaluation of current SAR target recognition methods. The dataset contains ten types of

military vehicle targets, as shown in Figure 1. The SAR images of various targets are acquired by X-band airborne radar with a resolution of 0.3 m. The MSTAR dataset has abundant samples, and a number of representative operating conditions can be set accordingly. The azimuth angles of various targets cover 0°–360°, which can be used for comprehensive training and testing. Some targets include several different configurations, which can be used to investigate the performance of the recognition methods under configuration variance. Some targets have multiple different depression angles, which can be used to investigate the performance of the recognition methods under large depression angle differences.

During the experiments, several types of existing methods are compared, which are mainly divided into two categories. The first category employs the three types of features used in this study, but uses a single feature, which are, respectively, denoted as Zernike [2], KPCA [12], and monogenic [15]. The second category is the multifeature fusion method, i.e., the methods in the [17, 18], which are, respectively, denoted as fusion 1 and fusion 2. The third category is the currently popular deep learning methods, using the A-ConvNet method in [23]. Subsequent experiments are first carried out under SOC and then under two EOCs of configuration variance and depression angle variance.

4.2. Results and Discussion

4.2.1. SOC. In the SAR target recognition problem, SOC generally refers to the high overall similarity between the test and the training samples, and the recognition difficulty is relatively low. Table 1 provides the training and test samples under SOC, which are from 17° and 15° depression angles, respectively. The test and training samples for various targets are from the same target confirmations. The proposed method is used to classify the 10 types of targets shown in Figure 1, and the confusion matrix shown in Figure 2 is obtained. Among them, the diagonal value marks the correct recognition rate of the corresponding target. In the experiment, the average recognition rate is defined as the proportion of correctly recognized samples in all test samples. The average recognition rate of the proposed method for 10 types of targets is 99.46%. Table 2 compares the average recognition rates of various methods under current experimental settings, which are all higher than 98%, reflecting the low difficulty of problems under SOC. Compared with the three types of methods based on single features, this study significantly improves the final recognition performance through their combined use. Compared with the other two types of multifeature fusion methods, the performance of this study is better, which shows that the designed decision fusion algorithm has stronger effectiveness. The CNN method can achieve high performance under SOC, but it is still lower than the proposed method. In summary, the proposed method can achieve superior performance under SOC, which verifies its effectiveness.
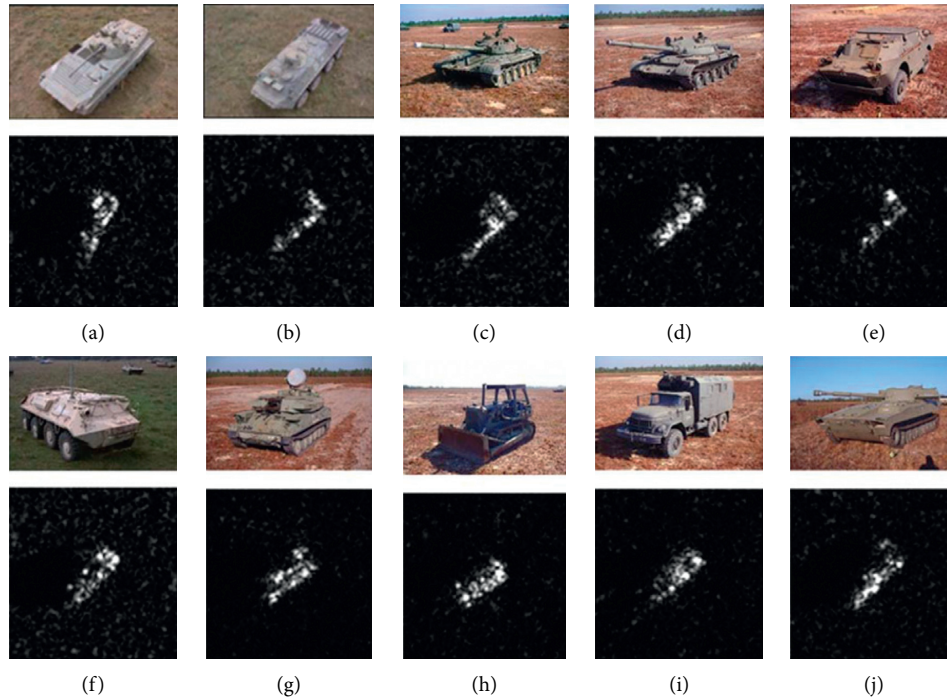
FIGURE 1: Images of targets in the MSTAR dataset used in the experiments.

TABLE 1: Training and test samples under SOC.

Class	Training (17°)		Test (15°)	
	Configuration	Samples	Configuration	Samples
BMP2	9563	233	9563	195
BTR70	c71	233	c71	196
T72	132	232	132	196
T62	A51	299	A51	273
BRDM2	E-71	298	E-71	274
BTR60	7532	256	7532	195
ZSU23/4	d08	299	d08	274
D7	13015	299	13015	274
ZIL131	E12	299	E12	274
2S1	B01	299	B01	274

4.2.2. EOCs. EOCs are defined with reference to SOC, which mainly examine the differences between the test and the training samples due to factors such as target, background, and sensor variations. The typical EOCs that can be set based on the MSTAR dataset mainly include configuration variance and depression angle variance, which are tested in the subsequent experiments.

Configuration Variance. The configuration variance is mainly due to the situation caused by the change of the target itself, which refers to the different configurations of the test sample and the training sample from the same target. Table 3 provides the current experimental scene. Among them, the test samples and training samples of BMP2 and T72 targets are from different configurations. The appearance similarity between BTR70 and these two types of targets is relatively high (as shown in Figure 1), and its introduction increases the overall recognition difficulty. Various methods are tested

under current conditions, and their average recognition rates are given in Table 4. Compared with the three types of single feature-based methods, the performance advantages of this method are very significant, indicating that their joint representation and weighted fusion can effectively improve the robustness of recognition. Compared with the two types of multifeature fusion methods, the recognition rate of the proposed method is higher, reflecting its stronger robustness. The performance degradation of the CNN method under current conditions is very obvious, mainly due to the weak coverage of the training samples to the test samples, and the corresponding training network adaptability also decreases.

Depression Angle Variance. As the relative viewing angle between the target and the sensor changes, the corresponding SAR images obtained will also have larger differences. In particular, when the test sample and the training
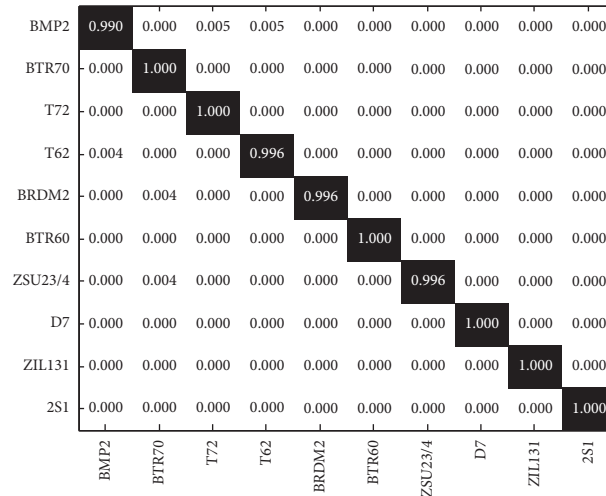
Figure 2: Recognition results under SOC.

Table 2: Average recognition rates under SOC.

Method	Average recognition rate (%)
Proposed	99.46
Zernike	98.02
KPCA	98.36
Monogenic	98.87
Fusion 1	99.02
Fusion 2	99.13
A-ConvNet	99.08

Table 3: Training and test samples under different configurations.

Class	Training (17°)		Test (15°)	
	Configuration	Samples	Configuration	Samples
BMP2	9563	233	9566	196
			c21	196
T72	132	232	812	195
			s7	191
BTR60	7532	256	7532	195
T62	A51	299	A51	273

Table 4: Classification results under configuration differences.

Method	Average recognition rate (%)
Proposed	98.78
Zernike	96.92
KPCA	96.53
Monogenic	97.46
Fusion 1	97.82
Fusion 2	98.20
A-ConvNet	97.24

sample come from a large depression angle, the difficulty of the recognition is greatly increased. Table 5 provides the training and test samples under the condition of depression angle variance. Among them, the training samples are all from a 17° depression angle; the test samples are divided into two subsets, corresponding to 30° and 45° depression angles,

respectively. Independent testing is performed on the test samples at two depression angles, and the average recognition rate of various methods is shown in Figure 3. It is intuitively obvious that the recognition result at the 45° depression angle is significantly lower than that of 30°. The proposed method obtains the highest average recognition

TABLE 5: Training and test samples under depression angle variance.

Class	Training		Test	
	Depression angle	Samples	Depression angle	Samples
2S1	17°	299	30°	288
			45°	303
BDRM2	17°	298	30°	287
			45°	303
ZSU23/4	17°	299	30°	288
			45°	303
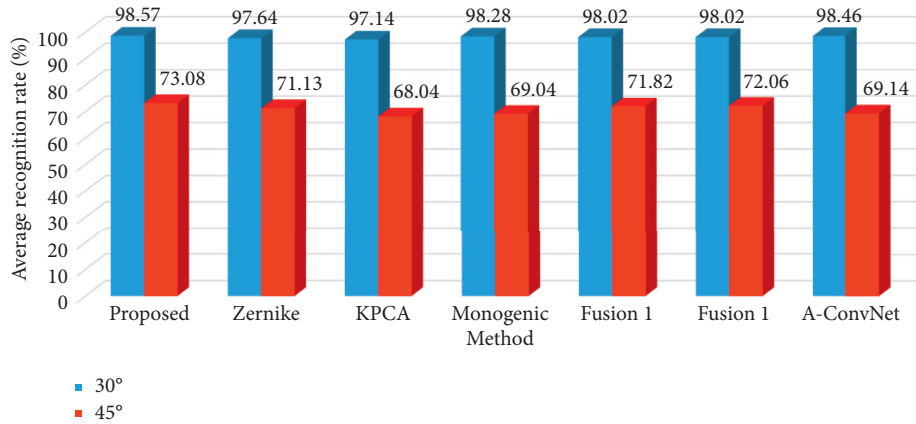
FIGURE 3: Recognition results of different methods at different depression angles.

rate at both depression angles, indicating that it has better robustness to depression angle variance. Through the effective fusion of the three types of features, the proposed method can more comprehensively investigate the image changes caused by the variances in the depression angle, so as to obtain more reliable recognition results.

5. Conclusion

In this study, a multifeature and multiclassifier SAR target recognition method is proposed. Zernike moments, KPCA, and monogenic signal are used to describe the characteristics of the original SAR image, and the corresponding feature vectors are obtained. In the classification stage, SVM, SRC, and JSR are used to make decisions on the three types of features, and then, their decision vectors are weighted and fused based on the multiple weight vector. Finally, the target label of the test sample is determined according to the fused decision variables. The three types of features and the three classifiers have good complementarity, so they can provide more effective information for target recognition. In the experiment, the proposed method is tested and verified under SOC, configuration variance, and depression angle variance based on the MSTAR dataset. The results show the performance advantages of the proposed method.

Data Availability

The dataset used to support the findings of this study is available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] K. El-Darymli, E. W. Gill, P. McGuire, D. Power, and C. Moloney, "Automatic target recognition in synthetic aperture radar imagery: a state-of-the-art review," *IEEE Access*, vol. 4, pp. 6014–6058, 2016.

[2] M. Amoon and G. a. Rezai-rad, "Automatic target recognition of synthetic aperture radar (SAR) images based on optimal selection of Zernike moments features," *IET Computer Vision*, vol. 8, no. 2, pp. 77–85, 2014.

[3] B. Ding, G. Wen, C. Ma, and X. Yang, "Target recognition in synthetic aperture radar images using binary morphological operations," *Journal of Applied Remote Sensing*, vol. 10, no. 4, Article ID 046006, 2016.

[4] C. Shan, B. Huang, and M. Li, "Binary morphological filtering of dominant scattering area residues for SAR target recognition," *Computational Intelligence and Neuroscience*, vol. 2018, Article ID 9680465, 15 pages, 2018.

[5] L. Jin, J. Chen, and X. Peng, "Synthetic aperture radar target classification via joint sparse representation of multi-level dominant scattering images," *Optik*, vol. 186, pp. 110–119, 2019.

[6] J. Tan, X. Fan, S. Wang et al., "Target recognition of SAR images by partially matching of target outlines," *Journal of Electromagnetic Waves and Applications*, vol. 33, no. 7, pp. 865–881, 2019.

[7] S. Papson and R. M. Narayanan, "Classification via the shadow region in SAR imagery," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 48, no. 2, pp. 969–980, 2012.

[8] L. C. Potter and R. L. Moses, "Attributed scattering centers for SAR ATR," *IEEE Transactions on Image Processing*, vol. 6, no. 1, pp. 79–91, 1997.

[9] B. Ding, G. Wen, J. Zhong, C. Ma, and X. Yang, "A robust similarity measure for attributed scattering center sets with application to SAR ATR," *Neurocomputing*, vol. 219, pp. 130–143, 2017.

[10] B. Ding, G. Wen, X. Huang, C. Ma, and X. Yang, "Target recognition in synthetic aperture radar images via matching of attributed scattering centers," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 10, no. 7, pp. 3334–3347, 2017.

[11] A. K. Mishra, "Validation of PCA and LDA for SAR ATR," in *Proceedings of the IEEE TENCON*, pp. 1–6, Hyderabad, India.

[12] Z. Cui, Z. Cao, J. Yang, J. Feng, and H. Ren, "Target recognition in synthetic aperture radar images via non-negative matrix factorisation," *IET Radar, Sonar & Navigation*, vol. 9, no. 9, pp. 1376–1385, 2015.

[13] W. Xiong, L. Cao, and Z. Hao, "Combining wavelet invariant moments and relevance vector machine for SAR target recognition," in *Proceedings of the IET International Radar Conference*, pp. 1–4, Guilin China, 22 April 2009.

[14] G. Dong, G. Kuang, N. Wang, and L. Zhao, J. Lu, "SAR target recognition via joint sparse representation of monogenic signal," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 8, no. 7, pp. 3316–3328, 2015.

[15] Y. Zhou, Y. Chen, R. Gao, J. Feng, P. Zhao, and L. Wang, "SAR target recognition via joint sparse representation of monogenic components with 2D canonical correlation analysis," *IEEE Access*, vol. 7, Article ID 25815, 2019.

[16] M. Chang, X. You, and Z. Cao, "Bidimensional empirical mode decomposition for SAR image feature extraction with application to target recognition," *IEEE Access*, vol. 7, Article ID 135720, 2019.

[17] M. Yu, G. Dong, H. Fan, and G. Kuang, "SAR target recognition via local sparse representation of multi-manifold regularized low-rank approximation," *Remote Sensing*, vol. 10, no. 2, p. 211, 2018.

[18] Y. Huang, J. Peia, J. Yanga, B. Wang, and X. Liu, "Neighborhood geometric center scaling embedding for SAR ATR," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 50, no. 1, pp. 180–192, 2014.

[19] Q. Zhao and J. C. Principe, "Support vector machines for SAR automatic target recognition," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 37, no. 2, pp. 643–654, 2001.

[20] C. Tison, N. Pourthie, and J. Souyris, "Target recognition in SAR images with support Vector Machines (SVM)," in *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium*, pp. 456–459, Barcelona Spain, July 2007.

[21] M. E. Demirhan and Ö. Salor, "Classification of targets in SAR images using SVM and k-NN techniques," in *Proceedings of the 24th Signal Processing and Communication Application Conference (SIU)*, pp. 1581–1584, Zonguldak Turkey, 2016.

[22] H. Liu and S. Li, "Decision fusion of sparse representation and support vector machine for SAR image target recognition," *Neurocomputing*, vol. 113, pp. 97–104, 2013.

[23] J. J. Thiagaraianm, K. N. Ramamurthy, P. Knee, A. Spanias, and V. Berisha, "Sparse representations for automatic target classification in SAR images," in *Proceedings of the 4th Int. Symp. Commun., Control Signal Process*, pp. 1–4, Limassol, Cyprus, 5 March 2010.

[24] H. Song, K. Ji, Y. Zhang, X. Xing, and H. Zou, "Sparse representation-based SAR image target classification on the 10-class MSTAR data set," *Applied Sciences*, vol. 6, no. 1, p. 26, 2016.

[25] B. Ding and G. Wen, "Sparsity constraint nearest subspace classifier for target recognition of SAR images," *Journal of Visual Communication and Image Representation*, vol. 52, pp. 170–176, 2018.

[26] W. Li, J. Yang, and Y. Ma, "Target recognition of synthetic aperture radar images based on two-phase sparse representation," *Journal of Sensors*, vol. 2020, Article ID 2032645, 12 pages, 2020.

[27] L. Yu, L. Wang, and Y. Xu, "Combination of Joint Representation and Adaptive Weighting for Multiple Features with Application to SAR Target Recognition," *Scientific Programming*, vol. 2021, Article ID 9063419, 9 pages, 2021.

[28] Y. Xin, L. Kuan, and L. Jiao, "SAR automatic target recognition based on classifiers fusion," in *Proceedings of the International Workshop on Multi-Platform/Multi-Sensor Remote Sensing and Mapping*, pp. 1–5, Xiamen China, 12 Jan. 2011.

[29] Z. Cui, Z. Cao, J. Yang, and J. Feng, "A hierarchical propelled fusion strategy for SAR automatic target recognition," *EURASIP Journal on Wireless Communications and Networking*, vol. 2013, pp. 1–8, 2013.

[30] U. Srinivas and V. Monga, "Meta-classifiers for exploiting feature dependence in automatic target recognition," in *Proceedings of the IEEE Radar Conference*, pp. 147–151, Kansas City MO USA, 27 May 2011.

[31] R. Huan and Y. Pan, "Decision fusion strategies for SAR image target recognition," *IET Radar, Sonar & Navigation*, vol. 5, no. 7, pp. 747–755, 2011.

[32] X. X. Zhu, D. Tuia, L. Mou et al., "Deep learning in remote sensing: a comprehensive review and list of resources," *IEEE Geoscience and Remote Sensing Magazine*, vol. 5, no. 4, pp. 8–36, 2017.

[33] M. Kang, K. Ji, X. Leng, X. Xing, and H. Zou, "Synthetic aperture radar target recognition with feature fusion based on a stacked autoencoder," *Sensors*, vol. 17, no. 1, p. 192, 2017.

[34] D. E. Morgan, "Deep convolutional neural networks for ATR from SAR imagery," in *Proceedings of the SPIE*, pp. 1–13, Maryland, United States, 13 May 2015.

[35] S. Chen, H. Wang, F. Xu, and Y. Jin, "Target classification using the deep convolutional networks for SAR images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 6, pp. 1685–1697, 2016.

[36] J. Zhao, Z. Zhang, W. Yu, and T.-K. Truong, "A cascade coupled convolutional neural network guided visual attention method for ship detection from SAR images," *IEEE Access*, vol. 6, Article ID 50693, 2018.

[37] R. Min, H. Lan, Z. Cao, and Z. Cui, "A gradually distilled CNN for SAR target recognition," *IEEE Access*, vol. 7, Article ID 42190, 2019.

[38] L. Wang, X. Bai, and F. Zhou, "SAR ATR of ground vehicles based on ESENet," *Remote Sensing*, vol. 11, no. 11, p. 1316, 2019.

[39] P. Zhao, K. Liu, H. Zou, and X. Zhen, "Multi-stream convolutional neural network for SAR automatic target recognition," *Remote Sensing*, vol. 10, no. 9, p. 1473, 2018.

[40] J. Ding, B. Chen, H. Liu, and M. Huang, "Convolutional neural network with data augmentation for SAR target recognition," *IEEE Geoscience and Remote Sensing Letters*, vol. 13, no. 3, pp. 364–368, 2016.

[41] Y. Yan, "Convolutional neural networks based on augmented training samples for synthetic aperture radar target recognition," *Journal of Electronic Imaging*, vol. 27, no. 2, Article ID 023024, 2018.

[42] D. Malmgren-Hansen, A. Kusk, J. Dall, A. A. Nielsen, R. Engholm, and H. Skriver, "Improving SAR automatic target recognition models with transfer learning from simulated data," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 9, pp. 1484–1488, 2017.

[43] S. A. Wagner, "SAR ATR by a combination of convolutional neural network and support vector machines," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 52, no. 6, pp. 2861–2872, 2016.

[44] O. Kechagias-Stamatis, N. Aouf, and Z. Liu, "Fusing deep learning and sparse coding for SAR ATR," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 55, no. 2, pp. 785–797, 2019.

[45] C. Jiang and Y. Zhou, "Hierarchical Fusion of Convolutional Neural Networks and Attributed Scattering Centers with Application to Robust SAR ATR," *Remote Sensing*, vol. 10, no. 6, p. 819, 2018.

[46] C. Chang and C. Lin, "LIBSVM: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, no. 3, pp. 296–389, 2011.

Research Article

Target Recognition of Synthetic Aperture Radar Images by Updated Classifiers

Jingyu Li [ID] **and Cungen Liu**

School of Information and Electrical Engineering, Shandong Jianzhu University, Jinan 250101, China

Correspondence should be addressed to Jingyu Li; lijingyu@sdjzu.edu.cn

For the problem of reliable decision in synthetic aperture radar (SAR) target recognition, a method based on updated classifiers is proposed. The convolutional neural network (CNN) and support vector machine (SVM) are used as basic classifiers to classify samples with unknown target labels. The two decisions are fused and the reliability of the fused decision is evaluated. The classified test samples with high reliabilities are added to the original training samples to update the classifiers. The updated classifiers have stronger classification abilities and the fused result of the two classifiers can obtain a more reliable decision. The proposed method is tested and verified based on the moving and stationary target acquisition and recognition (MSTAR) dataset. The experimental results verify the effectiveness and robustness of the proposed method.

1. Introduction

High-resolution synthetic aperture radar (SAR) can provide strong support for earth observation. In the military field, SAR images can be used to detect and identify targets of interest to obtain high-value intelligence. The problem of SAR target recognition is a specific application of traditional pattern recognition technology [1]. With the support of labeled training samples, the classification of unknown test samples can be performed. Feature extraction and classifier design are two important steps in the SAR target recognition methods. The former obtains high discriminative features through the analysis of SAR images, thereby improving the overall accuracy and efficiency of subsequent classification. Commonly used SAR image features mainly describe target geometry [2–7], electromagnetic scattering characteristics [8–10], or distributions of pixel values by projection or transformation [11–18]. The classifier learns a reliable decision-making surface based on a large number of training samples and then classifies the test samples. The classifiers commonly used for SAR target recognition include nearest neighbor (NN) [11], support vector machine (SVM) [19–22], and sparse representation-based classification (SRC) [23–27]. In addition, the multiclassifier fusion is also used in SAR target recognition [28–31]. In recent years, the deep learning algorithms have been widely applied and verified in various research fields, and a large number of SAR target methods based on deep learning models have emerged. A typical representative is the convolutional neural network (CNN) [32–45]. The deep learning methods integrate feature learning and classification, thus avoiding the separation of feature extraction and classifier design in traditional methods. However, the deep learning methods have a large demand for the training samples, and the final classification accuracy is often poor when there are fewer training samples. Different from the field of optical image processing, the data resources of SAR images are scarce. As a result, it is difficult to train a reliable deep learning classification model, which brings obstacles to the application of related methods. In [40, 41], more available samples were generated by means of data augmentation to improve the classification performance of the networks. In [42], multisource image data (such as optical images and electromagnetic simulation data) were processed through transfer learning to assist in training neural networks suitable for SAR target recognition.

This paper proposes a SAR target recognition method based on updated classifiers. The key idea is using those test samples whose categories can be reliably confirmed during the classification process to optimize the original classifier. In detail, a CNN is first designed as the dominant classifier for SAR target recognition. In addition, this paper selects SVM as an auxiliary classifier and confirms the classification result of the test sample together with CNN. The original training samples are used to train the CNN and the SVM classifiers. For a certain test sample to be recognized, it is first classified by CNN and SVM respectively to obtain the corresponding decision values. Then, the final decision value is obtained through the weighted fusion. Based on the fused decision values, the reliability of the final decision is calculated. When the current test sample has a higher reliability level than the preset threshold, it is furthered used as a training sample to optimize the CNN and also enhances the training samples for SVM. With the continuous increase of test samples with confirmed labels, the classification performance of CNN and SVM is continuously enhanced. So, the recognition results obtained by the fusion of the two classifiers are more reliable. The main contributions of this paper are as follows: first, the updated mechanism is introduced under the decision fusion framework of CNN and SVM classifiers. Decision fusion of different classifiers is a common way to improve decision accuracy. In the traditional methods, the classified test samples are not fully used, and the online update of the classifier is lacked. In the problem of SAR target recognition, the number of training samples is very limited. This paper confirms the test samples and updates the classifier, which can effectively improve the classification ability of the classifiers. Second, an effective criterion for decision-making reliability is developed and used for the selection of test samples. Although the accuracy of fusion decision is improved, there is still some probability of misclassification. The introduction of test samples with the wrong target labels will result in a decrease in the performance of the updated classifiers. Based on the fusion of probabilistic decision variables, this paper defines decision reliability levels to select test samples whose decision reliability is higher than the preset threshold to update the classifier. So, the effectiveness of the updated classifiers can be as ensured. In the experiments, the proposed method is evaluated on the moving and stationary target acquisition and recognition (MSTAR) dataset. The results show that the proposed method has advantages over a single classifier and traditional classifier decision fusion methods.

2. Basic Classifiers

2.1. CNN.
CNN is an extension of traditional neural networks, which can be used for two-dimensional signal (image) processing. It conducts deep mining of original data by setting multiple different convolutional layers. In each convolutional layer, different convolution kernels are used to extract two-dimensional features from the original data, so as to obtain multilevel features. Finally, the network builds an efficient classification framework through end-to-end training to realize the classification of test samples. At

present, CNN has been widely used in the field of image processing, and a series of CNN-based SAR target recognition methods have appeared [32–45]. It should be pointed out that the classification performance of CNN is closely related to the size and coverage of training samples. When the number of training samples is small, the finally trained network has poor adaptability and cannot handle the SAR target recognition task well.

Based on the existing research studies, this paper designs a CNN as shown in Figure 1, which includes three convolutional layers, three maximum pooling layers, and two fully connected layers. In each convolutional layer, the rectified linear unit (ReLU) is used as the activation function to enhance the nonlinear adaptability of the network. The maximum pooling layer is set after the convolutional layer to improve the overall training efficiency of the network. Finally, the conversion of input data to category labels is achieved through two fully connected layers (take 10-class recognition as an example). The end of the network uses Softmax as the basic classifier and outputs the possibility that the test sample belongs to each class in the form of posterior probability. In addition, the overall complexity of the network structure is relatively low, which is beneficial to improve the efficiency of overall SAR target recognition.

2.2. SVM.
SVM is chosen as another classifier in this paper. Since first proposed by Vapnik et al. in 1995, SVM has been one of the most popular classifiers in the field of pattern recognition. In Zhao and Pricipe [19], SVM was first introduced to SAR target recognition with good performance. Afterwards, many relevant researches were developed on SVM and improved the recognition performance [20–22]. Based on the principle of structural risk minimization, SVM aims to find a hyperplane to separate the patterns from two different classes in a two-class classification problem. For a test sample x, SVM generally makes the decision as follows:

$$f(x) = \sum_{i=1}^{M} w_i y_i K(x_i, x) + b \; \alpha_i \geq 0, \forall i. \tag{1}$$

In equation (1), $x_i (i = 1, \cdots, M)$ represents a support vector from the training samples; $y_i = \pm 1$ denote the labels of the two different classes. $w_i (i = 1, \cdots, M)$ are the weights and b is the bias, which are the parameters to be estimated during the training process. $K(\cdot)$ is the kernel function, which can be specifically designed to handle different types of classification problems. For example, the polynomial and radial basis function (RBF) kernels are two commonly used kernel functions in SVM.

The traditional two-class SVM can be generalized to multiclass ones via the strategies such as "one-versus-one" or "one-versus-rest". Then, SVM can be directly used to classify multiple classes. The famous LIBSVM [46] provided an excellent toolbox to use SVM for different applications, which can be smoothly used for multiclass recognition problems. In this paper, the multiclass SVM with RBF kernel is employed to perform the classification for SAR images.
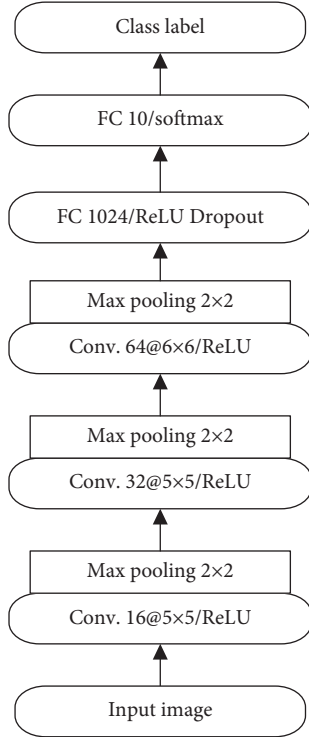
FIGURE 1: Structure of the CNN used in this paper.

3. Updated Classifiers for Target Recognition

3.1. Principle for Decision and Updating Classifiers. For the classification results from CNN and SVM, this paper uses a weighted fusion algorithm to obtain the final decision values. CNN uses Softmax as the classifier, and its output decision value is the posterior probability vector $[P_1, P_2, \cdots, P_C]$. For SVM, the output result is also in the form of probabilities, denoted as $[P_1^S, P_2^S, \cdots, P_C^S]$. Afterwards, the classical weighting (equal weight) algorithm [24, 25] is used to fuse the posterior probability vectors of CNN and SVM, as follows:

$$P_i^F = 0.5 * P_i + 0.5 * P_i^S, \, i = 1, 2, \cdots, C, \quad (2)$$

where $[P_1^F, P_2^F, \cdots, P_C^F]$ denotes the final decision variables after fusion. Accordingly, this paper defines the decision reliability as follows:

$$R = \min\left(\frac{P_K^F}{\max\left(P_i^F\right)}\right)(i \neq K), \quad (3)$$

where P_K^F is the maximum probability, so the reliability level is $R \geq 1$.

Correspondingly, the larger the value R, the more reliable the classification result. An appropriate decision threshold is set. When the decision reliability is higher than the threshold, the current decision is considered reliable. And the corresponding test sample is added to the original training samples to update CNN and SVM. Otherwise, the training set is not updated.

3.2. Procedure of Target Recognition. In this paper, the training set is dynamically updated by analyzing the reliability of the decisions of test samples. Therefore, the classification performance of CNN and SVM can be enhanced so their fused result will be more reliable. The key steps of the implementation of the proposed method are shown in Figure 2, which can be summarized as follows:

Step 1: the original training samples are used to train the CNN shown in Figure 1 and the SVM classifier.

Step 2: for the test sample with unknown target label, CNN and SVM are used to classify them respectively. Their results are fused and the reliability level of decision-making is calculated.

Step 3: if the decision-making reliability of the current test sample is higher than the preset threshold, it is added to the original training samples and dynamically updates the training set of CNN and SVM.

With the increasing number of test samples with confirmed target labels, the CNN and SVM classification capabilities have also been continuously enhanced. Therefore, the classification accuracy of subsequent test samples can also be improved. In general, the overall accuracy of target recognition can be improved.

4. Experiments

4.1. MSTAR Dataset. The MSTAR dataset is a representative dataset for testing SAR target recognition methods, which is employed for a long time. Figure 3 shows the optical and SAR images of the 10 targets in the dataset. Among them, the SAR images can cover all-round aspect angles and several depression angles, and the image resolution is 0.3 m. Through flexible processing and simulation on the original SAR images, a variety of operating conditions can be setup to test the proposed method including the standard operating condition (SOC) and extended operating condition (EOC).

Several types of reference methods are set up in the experiments, including the method of SVM-based classifier in [19] (denoted as SVM), the cascade coupled CNN designed in [36] (denoted as CSCNN), the CNN with data augmentation used in [41] (denoted as Aug-CNN), and the method based on the decision fusion of SVM and SRC in [22] (denoted as SVM + SRC).

In the follow-up experiments, the validation is first carried out under SOC to analyze and verify the basic performance of the proposed method. Then, several typical EOCs are established to verify the robustness of the proposed method under the conditions of depression angle variance, noise corruption, and reduced training set.

4.2. Results and Discussion

4.2.1. SOC. Table 1 shows a typical SOC based on the MSTAR dataset. The training and test sets contain the SAR images of the 10 types of targets acquired at the depression angles of 17° and 15°, respectively, and both cover 0°~360° azimuth angles. With the reliability threshold set to 1.4, the

Reliable samples
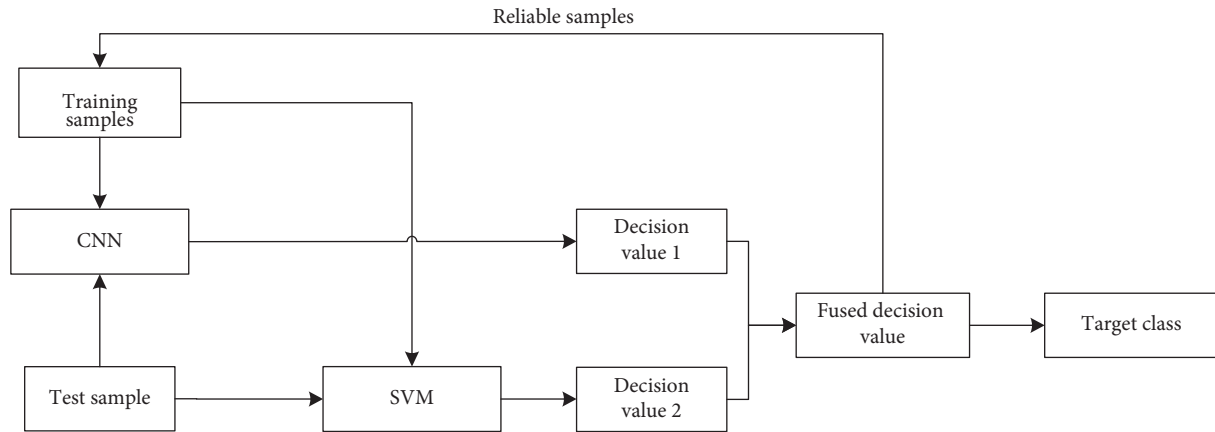
Figure 2: Implementation of target recognition in the proposed method.

(1) BMP2   (2) BTR70   (3) T72   (4) T62   (5) BRDM2

(6) BTR60   (7) ZSU23/4   (8) D7   (9) ZIL131   (10) 2S1
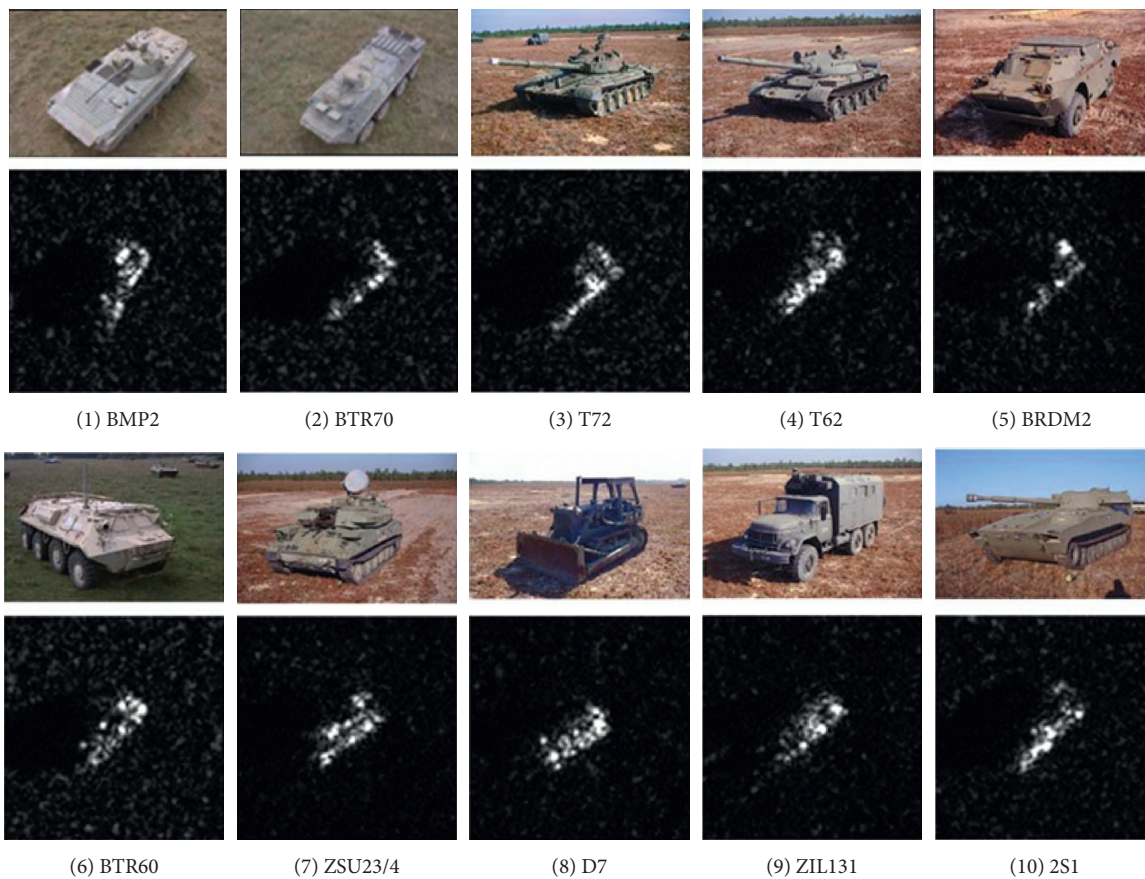
Figure 3: Optical and SAR images of targets in the MSTAR dataset for recognition.

proposed method classifies the test samples and obtains the corresponding results. The classification accuracy of different targets keeps higher than 98%. The recognition rates of BMP2 and T72 are relatively low due to the configuration differences between their training and test samples. Among the remaining 8 types of targets, the recognition rate of BTR60 is relatively low, mainly because its test samples share high similarity with the samples of BMP2 and T72 targets, which increases the probability of false classification. The four types of reference methods are also tested under the

same condition, and the average recognition rates of all methods for 10 types of targets are summered in Table 2. The comparison shows that the recognition performance of the proposed method under SOC is better than that of the reference methods, which verifies its stronger effectiveness. Compared with the methods using SVM or CNN alone, this paper confirms the test samples and enriches the training set and organically combines the classification results of the two classifiers to significantly improve the final classification accuracy. The Aug-CNN method improves the Performance

Table 1: Training and test sets under SOC.

Class	Training (17°)		Test (15°)	
	Configuration (SN)	Number of samples	Configuration (SN)	Number of samples
BMP2	9563	233	9563	195
			9566	196
			c21	196
BTR70	—	233	—	196
T72	132	232	132	196
			812	195
			S7	191
T62	—	299	—	273
BRDM2	—	298	—	274
BTR60	—	256	—	195
ZSU23/4	—	299	—	274
D7	—	299	—	274
ZIL131	—	299	—	274
2S1	—	299	—	274

Table 2: Results comparison under SOC.

Method	Average recognition rate (%)
Proposed	99.15
SVM	96.72
CSCNN	98.93
Aug-CNN	99.02
SVM + SRC	98.28

of traditional CNN by simulating training samples, but the overall amount of information in the original training set is still limited. As a result, the improvement is also very limited and it will also bring additional time consumption. Compared with the fusion method of SVM and SRC, the proposed method dynamically updates the two classifiers while fusing CNN and SVM, and the enhancement of the final recognition performance is also very obvious.

According to the basic principles of the proposed method, the selection of the reliability threshold has a direct impact on the final classifier update and recognition performance. In the previous experiment, the reliability threshold was set to be 1.4. This experiment mainly verifies this key parameter and investigates its influence on the final recognition performance. Table 3 lists the average recognition rates of the proposed method for 10 types of targets at different thresholds. It can be seen that the proposed method can maintain an average recognition rate over 98.5% at each threshold, but its performance is also directly related to the threshold selection. When the threshold is relatively low (for example, 1.1), the decision reliability is easy to meet the requirements, and more samples are used for updating the classifiers. However, some of the misclassified samples will also cause the performance of the classifier to decrease, resulting in a decrease in the overall average recognition rate. When the threshold is large (for example, 1.7), the decision reliability is very high at this time, resulting in only a few test samples that can be used for classifier update, which makes the proposed method degenerate into a traditional decision fusion method. After comparative analysis and repeated

experiments, this paper selects 1.4 as an appropriate threshold. This threshold can effectively ensure the rationality of test sample selection and the reliability of classifier update.

4.2.2. Depression Angle Variance. The difference in depression angle between the test sample and the training set will cause the characteristic difference of the SAR images. Table 4 shows the training and test sets under the condition of depression angle variance set using the MSTAR dataset, which includes 3 types of targets. Among them, the training set uses SAR images with a depression angle of 17°, and the training set contains samples with two depression angles of 30° and 45°. Experiments are carried out at the two depression angles, respectively. The average recognition rates of various methods are summarized as shown in Table 5. It is easy to find that the recognition performance of various methods at 45° depression angle is significantly lower than the results at 30°, indicating that large depression angle difference will seriously affect the correct decisions. Through the use of highly reliable test samples and the fusion of CNN and SVM, the proposed method achieves better recognition performance than the reference ones. Compared with the CSCNN, SVM and Aug-CNN methods using single classifiers, the performance superiority of the proposed method is very significant, which demonstrates the effectiveness of decision fusion and dynamic update of the classifiers. Compared with the SVM + SRC method, this paper introduces CNN with stronger classification performance. In addition, those samples after reliable classifications are used to update CNN and SVM, so the final decision result has higher reliability.

4.2.3. Noise Corruption. In order to test the performance of the proposed method under noise corruption, different degrees of noise are added to the original test samples in Table 1 to construct test sets at different signal-to-noise ratios (SNR). All the methods are tested at different SNRs, and the results are shown in Figure 4. It can be seen that the

TABLE 3: Results of the proposed method at different reliability thresholds.

Threshold	1.1	1.2	1.3	1.4	1.5	1.6	1.7
Average recognition rate (%)	98.74	98.98	99.04	99.15	99.07	98.86	98.57

TABLE 4: Experimental setup under depression angle variances.

| Class | Training | | Test | |
	Depression	Number of samples	Depression	Number of samples
2S1	17°	299	30°	288
			45°	303
BDRM2	17°	298	30°	287
			45°	303
ZSU23/4	17°	299	30°	288
			45°	303

TABLE 5: Average recognition rates of different methods under depression angle variances.

| Method | Depression | |
	30°	45°
Proposed	97.02	72.92
SVM	93.27	66.08
CSCNN	96.42	66.73
Aug-CNN	96.48	68.74
SVM + SRC	94.66	67.88

FIGURE 5: Comparison of results under reduced training set.

which improves the noise adaptability of the network. Since the training samples used in the CSCNN method are all from high SNRs, its classification accuracy under noise corruption, especially test samples from low SNRs, is significantly reduced. The proposed method uses the combination of CNN and SVM to dynamically update the training samples to improve the coverage of noise corruption situations, and the final decision can better handle the noisy samples.

4.2.4. Reduced Training Set. As mentioned above, the training samples available in SAR target recognition are often very limited. As a result, it is difficult to cover the possible situations of the test samples (such as view angles and noise levels). Therefore, it is very important to improve the robustness of the recognition method under the condition of limited training samples. Based on Table 1, 80%, 60%, 40%, and 20% of the original training set are randomly
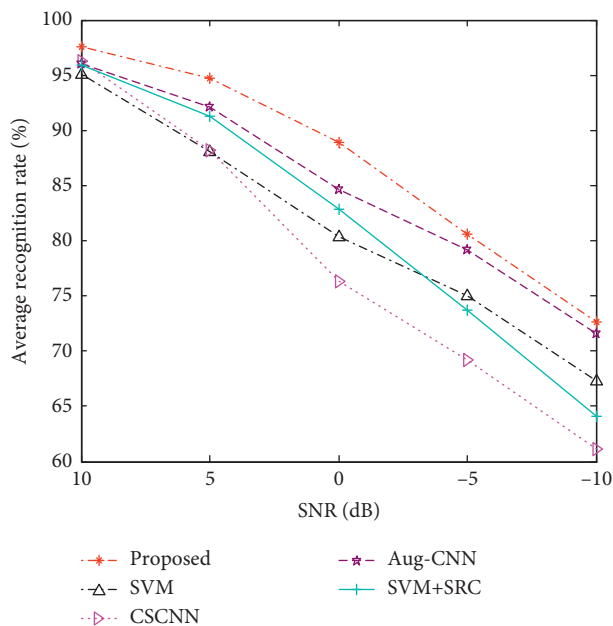

FIGURE 4: Comparison of results under noise corruption.

proposed method maintains the highest recognition rate at each SNR, showing its better noise robustness. The performance of the Aug-CNN method ranks second only to the proposed method, mainly because the method adds noisy samples generated by simulation to augment the training set,
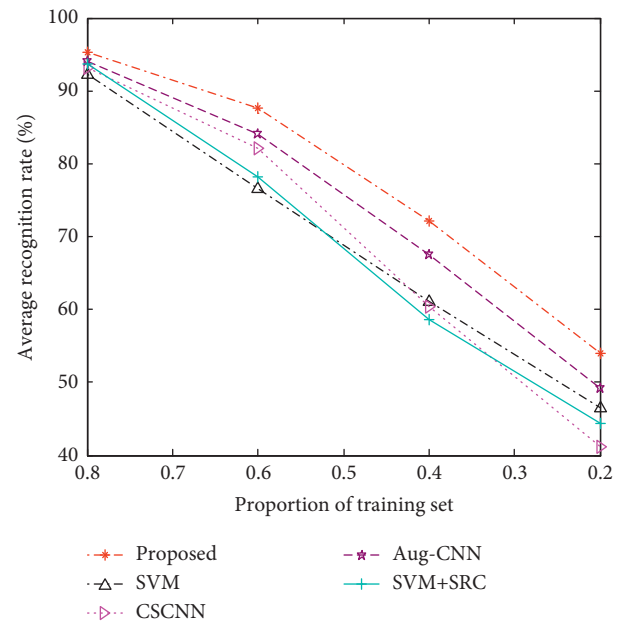
selected to construct reduced training sets and then the original test sets are classified. The results of different methods achieved under reduced training sets are shown in Figure 5. The proposed method has obvious performance advantages in the case of fewer training samples. On the one hand, the fusion of CNN and SVM at the decision-making layer improves the fault tolerance of the overall decision-making. On the other hand, with the increasing number of confirmed test samples, the training set has actually been effectively supplemented, so the updated CNN and SVM classification results are more reliable. The four types of reference methods are limited by the size of the training samples, so their overall recognition performance is also limited.

5. Conclusion

The paper proposes a SAR target recognition method based on updating the classifiers, which continuously updates the available training samples through the confirmation of the target label of the test sample, thereby improving the classification performance. CNN and SVM are used as basic classifiers to improve independent classification performance on the basis of updating training samples. At the same time, the two results are combined at the decision-making stage to obtain a more reliable recognition result. Experiments are carried out based on the MSTAR dataset, and the classification performance of the proposed method is tested under SOC, depression angle variance, noise corruption, and reduced training set, which is also compared with several existing methods. The experimental results show that the recognition performance of the proposed method is better than the existing methods under all conditions, verifying its effectiveness.

Data Availability

The MSTAR dataset is publicly available.

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] K. El-Darymli, E. W. Gill, D. Power, and C. Moloney, "Automatic target recognition in synthetic aperture radar imagery: a state-of-the-art review," *IEEE Access*, vol. 4, pp. 6014–6058, 2016.

[2] M. Amoon and G. Rezai-rad, "Automatic target recognition of synthetic aperture radar (SAR) images based on optimal selection of Zernike moments features," *IET Computer Vision*, vol. 8, no. 2, pp. 77–85, 2014.

[3] B. Ding, G. Wen, C. Ma, and X. Yang, "Target recognition in synthetic aperture radar images using binary morphological operations," *Journal of Applied Remote Sensing*, vol. 10, no. 4, Article ID 46006, 2016.

[4] C. Shan, B. Huang, and M. Li, "Binary morphological filtering of dominant scattering area residues for SAR target recognition," *Computational Intelligence and Neuroscience*, vol. 2018, Article ID 9680465, 2018.

[5] L. Jin, J. Chen, and X. Peng, "Synthetic aperture radar target classification via joint sparse representation of multi-level dominant scattering images," *Optik*, vol. 186, pp. 110–119, 2019.

[6] J. Tan, X. Fan, S. Wang et al., "Target recognition of SAR images by partially matching of target outlines," *Journal of Electromagnetic Waves and Applications*, vol. 33, no. 7, pp. 865–881, 2019.

[7] S. Papson and R. M. Narayanan, "Classification via the shadow region in SAR imagery," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 48, no. 2, pp. 969–980, 2012.

[8] L. C. Potter and R. L. Moses, "Attributed scattering centers for SAR ATR," *IEEE Transactions on Image Processing*, vol. 6, no. 1, pp. 79–91, 1997.

[9] B. Ding, G. Wen, J. Zhong, C. Ma, and X. Yang, "A robust similarity measure for attributed scattering center sets with application to SAR ATR," *Neurocomputing*, vol. 219, pp. 130–143, 2017.

[10] B. Ding, G. Wen, X. Huang, C. Ma, and X. Yang, "Target recognition in synthetic aperture radar images via matching of attributed scattering centers," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 10, no. 7, pp. 3334–3347, 2017.

[11] A. K. Mishra, "Validation of PCA and LDA for SAR ATR," in *Proceedings of the IEEE TENCON*, pp. 1–6, Hyderabad, India, November 2008.

[12] Z. Cui, Z. Cao, J. Yang, J. Feng, and H. Ren, "Target recognition in synthetic aperture radar images via non-negative matrix factorisation," *IET Radar, Sonar & Navigation*, vol. 9, no. 9, pp. 1376–1385, 2015.

[13] W. Xiong, L. Cao, and Z. Hao, "Combining wavelet invariant moments and relevance vector machine for SAR target recognition," in *Proceedings of the IET International Radar Conference*, pp. 1–4, Guilin, China, April 2009.

[14] G. Dong, G. Kuang, N. Wang, L. Zhao, and J. Lu, "SAR target recognition via joint sparse representation of monogenic signal," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 8, no. 7, pp. 3316–3328, 2015.

[15] Y. Zhou, Y. Chen, R. Gao, J. Feng, P. Zhao, and L. Wang, "SAR target recognition via joint sparse representation of monogenic components with 2D canonical correlation analysis," *IEEE Access*, vol. 7, pp. 25815–25826, 2019.

[16] M. Chang, X. You, and Z. Cao, "Bidimensional empirical mode decomposition for SAR image feature extraction with application to target recognition," *IEEE Access*, vol. 7, pp. 135720–135731, 2019.

[17] M. Yu, G. Dong, H. Fan, and G. Kuang, "SAR target recognition via local sparse representation of multi-manifold regularized low-rank approximation," *Remote Sensing*, vol. 10, no. 2, p. 211, 2018.

[18] Y. Huang, J. Peia, J. Yanga, B. Wang, and X. Liu, "Neighborhood geometric center scaling embedding for SAR ATR," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 50, no. 1, pp. 180–192, 2014.

[19] Q. Zhao and J. C. Principe, "Support vector machines for SAR automatic target recognition," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 37, no. 2, pp. 643–654, 2001.

[20] C. Tison, N. Pourthie, and J. Souyris, "Target recognition in SAR images with support vector machines (SVM)," in *Proceedings of the 2007 IEEE International Geoscience and Remote Sensing Symposium*, pp. 456–459, Barcelona, Spain, July 2007.

[21] M. E. Demirhan and Ö. Salor, "Classification of targets in SAR images using SVM and k-NN techniques," in *Proceedings of the 2016 24th Signal Processing and Communication Application Conference (SIU)*, pp. 1581–1584, Zonguldak, Turkey, May 2016.

[22] H. Liu and S. Li, "Decision fusion of sparse representation and support vector machine for SAR image target recognition," *Neurocomputing*, vol. 113, pp. 97–104, 2013.

[23] J. J. Thiagaraianm, K. N. Ramamurthy, P. Knee, A. Spanias, and V. Berisha, "Sparse representations for automatic target classification in SAR images," in *Proceedings of the 4th International Symposium on Communications, Control and Signal Processing*, pp. 1–4, Limassol, Cyprus, March 2010.

[24] H. Song, K. Ji, Y. Zhang, and X. Xing, "Sparse representation-based SAR image target classification on the 10-class MSTAR data set," *Applied Sciences*, vol. 6, no. 26, 2016.

[25] B. Ding and G. Wen, "Sparsity constraint nearest subspace classifier for target recognition of SAR images," *Journal of Visual Communication and Image Representation*, vol. 52, pp. 170–176, 2018.

[26] W. Li, J. Yang, and Y. Ma, "Target recognition of synthetic aperture radar images based on two-phase sparse representation," *Journal of Sensors*, vol. 2020, Article ID 2032645, 2020.

[27] L. Yu, L. Wang, and Y. Xu, "Combination of joint representation and adaptive weighting for multiple features with application to SAR target recognition," *Scientific Programming*, vol. 2021, Article ID 9063419, 2021.

[28] Y. Xin, L. Kuan, and L. Jiao, "SAR automatic target recognition based on classifiers fusion," in *Proceedings of the International Workshop on Multi-Platform/Multi-Sensor Remote Sensing and Mapping*, pp. 1–5, Xiamen, China, January 2011.

[29] Z. Cui, Z. Cao, J. Yang, and J. Feng, "A hierarchical propelled fusion strategy for SAR automatic target recognition," *EURASIP Journal on Wireless Communications and Networking*, vol. 39, pp. 1–8, 2013.

[30] U. Srinivas and V. Monga, "Meta-classifiers for exploiting feature dependence in automatic target recognition," in *Proceedings of the IEEE Radar Conference*, pp. 147–151, Kansas City, USA, May 2011.

[31] R. Huan and Y. Pan, "Decision fusion strategies for SAR image target recognition," *IET Radar, Sonar & Navigation*, vol. 5, no. 7, pp. 747–755, 2011.

[32] X. X. Zhu, D. Tuia, L. Mou et al., "Deep learning in remote sensing: a comprehensive review and list of resources," *IEEE Geoscience and Remote Sensing Magazine*, vol. 5, no. 4, pp. 8–36, 2017.

[33] M. Kang, K. Ji, X. Leng, X. Xing, and H. Zou, "Synthetic aperture radar target recognition with feature fusion based on a stacked autoencoder," *Sensors*, vol. 17, no. 1, p. 192, 2017.

[34] D. E. Morgan, "Deep convolutional neural networks for ATR from SAR imagery," in *Proceedings of the SPIE*, pp. 1–13, San Francisco, CA, USA, January 2015.

[35] S. Chen, H. Wang, F. Xu, and Y. Jin, "Target classification using the deep convolutional networks for SAR images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 6, pp. 1685–1697, 2016.

[36] J. Zhao, Z. Zhang, W. Yu, and T.-K. Truong, "A cascade coupled convolutional neural network guided visual attention method for ship detection from SAR images," *IEEE Access*, vol. 6, pp. 50693–50708, 2018.

[37] R. Min, H. Lan, Z. Cao, and Z. Cui, "A gradually distilled CNN for SAR target recognition," *IEEE Access*, vol. 7, pp. 42190–42200, 2019.

[38] L. Wang, X. Bai, and F. Zhou, "SAR ATR of ground vehicles based on ESENet," *Remote Sensing*, vol. 11, no. 11, p. 1316, 2019.

[39] P. Zhao, K. Liu, H. Zou, and X. Zhen, "Multi-stream convolutional neural network for SAR automatic target recognition," *Remote Sensing*, vol. 10, no. 9, p. 1473, 2018.

[40] J. Ding, B. Chen, H. Liu, and M. Huang, "Convolutional neural network with data augmentation for SAR target recognition," *IEEE Geoscience and Remote Sensing Letters*, vol. 13, no. 3, pp. 364–368, 2016.

[41] Y. Yan, "Convolutional neural networks based on augmented training samples for synthetic aperture radar target recognition," *Journal of Electronic Imaging*, vol. 27, no. 2, Article ID 023024, 2018.

[42] D. Malmgren-Hansen, A. Kusk, J. Dall, A. A. Nielsen, R. Engholm, and H. Skriver, "Improving SAR automatic target recognition models with transfer learning from simulated data," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 9, pp. 1484–1488, 2017.

[43] S. A. Wagner, "SAR ATR by a combination of convolutional neural network and support vector machines," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 52, no. 6, pp. 2861–2872, 2016.

[44] O. Kechagias-Stamatis, N. Aouf, and Z. Liu, "Fusing deep learning and sparse coding for SAR ATR," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 55, no. 2, pp. 785–797, 2019.

[45] C. Jiang and Y. Zhou, "Hierarchical fusion of convolutional neural networks and attributed scattering centers with application to robust SAR ATR," *Remote Sensing*, vol. 10, no. 6, p. 819, 2018.

[46] C. Chang and C. Lin, "LIBSVM: a library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, no. 3, pp. 296–389, 2011.

Retraction

Retracted: Research on Network Layer Recursive Reduction Model Compression for Image Recognition

Scientific Programming

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope

(2) Discrepancies in the description of the research reported

(3) Discrepancies between the availability of data and the research described

(4) Inappropriate citations

(5) Incoherent, meaningless and/or irrelevant content included in the article

(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] H. Ling, W. Zhang, Y. Tao, and M. Zhou, "Research on Network Layer Recursive Reduction Model Compression for Image Recognition," *Scientific Programming*, vol. 2021, Article ID 4054435, 11 pages, 2021.

Research Article

Research on Network Layer Recursive Reduction Model Compression for Image Recognition

Hongfei Ling [ID],[1,2] **Weiwei Zhang** [ID],[1,2] **Yingjie Tao,**[1,2] **and Mi Zhou**[1,2]

[1]*College of Engineering, Huaqiao University, Quanzhou 362021, Fujian, China*
[2]*Industrial Intelligence and System Fujian University Engineering Research Center, Huaqiao University, Quanzhou 362021, Fujian, China*

Correspondence should be addressed to Weiwei Zhang; weiweizh@hqu.edu.cn

ResNet has been widely used in the field of machine learning since it was proposed. This network model is successful in extracting features from input data by superimposing multiple layers of neural networks and thus achieves high accuracy in many applications. However, the superposition of multilayer neural networks increases their computational cost. For this reason, we propose a network model compression technique that removes multiple neural network layers from ResNet without decreasing the accuracy rate. The key idea is to provide a priority term to identify the importance of each neural network layer, and then select the unimportant layers to be removed during the training process based on the priority of the neural network layers. In addition, this paper also retrains the network model to avoid the accuracy degradation caused by the deletion of network layers. Experiments demonstrate that the network size can be reduced by 24.00%–42.86% of the number of layers without reducing the classification accuracy when classification is performed on CIFAR-10/100 and ImageNet.

1. Introduction

Convolutional neural network (CNN) is a commonly used neural network model in the field of computer vision as it can achieve high accuracy in various tasks in the field of image recognition [1, 2]. Network models can deepen the network structure by CNNs and thus improve the accuracy of tasks such as recognition or detection. For example, LeNet-5, proposed by LeCun et al. uses a 5-layer CNN to classify handwritten text. Later, VGGNet-19 utilized 22 layers to further improve the accuracy [2]. The residual network (ResNet) [3] even uses 152 layers of neural networks to achieve the optimal performance for the current task competition. As a result, ResNet is now commonly used as one of the models of standard CNNs in diverse fields, such as medical disease map classification, forestry pest, and disease classification [4].

ResNet is used to solve the problem of performance degradation caused by increasing depth. The biggest difference between DenseNet and ResNet is that in DenseNet we never combine features through summation before they are passed into a layer; instead, we provide them all as separate inputs. ResNeXt proposes aggregated transformations, using a parallel stack of blocks with the same topology to replace the original three-layer convolution block of ResNet, which improves the accuracy of the model without significantly increasing the parameter level. At the same time, due to the same topology, the number of hyperparameters is also reduced, which is convenient for model transplantation. In SE-ResNet and SE-ResNeXt, SENet can be regarded as a channel-wise attention [5, 6]. SENet adds a branch to calculate the channel-wise scale after the normal action and then multiplies the obtained value to the corresponding channel [7].

CNN improves accuracy through deep structure on the one hand, and on the other hand the computational cost required for learning and model inference increases as the number of layers increases. During model training, computational resources are enhanced by adding hardware devices, and computation time can be significantly reduced by distributed algorithms. However, with the

advent of the Internet of Things (IoT) era, models often need to be deployed again on end devices with limited computational resources, for example, image classification on embedded systems, text recognition on portable devices, and speech recognition on mobile devices. A higher task level requires a larger amount of hardware computational memory, and thus the significant problem that arises is the high operational and inference cost requirements of IoT end devices and realistic scenarios where end devices often struggle to meet the high demand for computational resources [8]. Therefore, how to effectively reduce the computational cost of CNN training and inference has received significant attention from researchers. For example, for the problem of computational cost of deep neural network models, Denton et al. [7] proposed to try to reduce the computational cost by cutting the number of layers, preserving all network layers of the residual network, and changing the number of network layers executed according to the input data. However, it is also necessary to save all the network layers in the scheme and not just consider cutting the cost of computational resource consumption. On the contrary, deciding which network layer is skipped also increases the memory consumption due to the additional modules required to determine it. For this reason, Chen et al. [8] proposed a method to decrease the number of residual network layers during learning, which can shorten the time of inference computation and reduce the memory consumption at the same time. However, this scheme removes the network layers completely statically, so it is sometimes difficult to maintain a high accuracy rate. Rastegari et al. [9] used the distillation maneuver to learn new models with fewer network layers from learned models with more layers, but their experiments showed a huge decrease in accuracy. Moreover, from the point of view of reducing the computational cost, to reduce the model inference time and computational resource cost, the use of static deletion of layers and distillation can be satisfied.

In this paper, we present a model compression method that uses layer deletion and retraining and can suppress accuracy degradation. The proposed method imports judgment values that determine the importance of each layer of the residual network. The judgment values are used to determine or remove unimportant layers after learning and preventing the accuracy degradation. This paper is to retrain the residual network after removing the layers. The experimental results show that layer deletion and retraining in such a way are applicable for overall model compression and reduce the cost of computational resources. Maintaining accuracy in the deleted layers requires retraining by removing individual parameters and then retraining with different hyperparameter settings. Experiments using the CIFAR-10/100 image dataset for the image classification task cut the number of network layers by 24.00% to 42.86%. Accordingly, the computation time for model inference decreased by 60.23% to 76.69%, and the number of parameters of the model was reduced to 69.82% to 93.15%.

1.1. Related Work.
Present-day model compression schemes for ResNet fall into three broad categories.

In the first category, Jaderberg et al. [10] dynamically decide whether to execute or skip the next layer in the middle of the inference calculation of the residual network. Han et al. [11] decide whether or not to execute a layer by adding a gate function to each layer. The signal from the previous layer is input to the gate function; if it outputs 1, the next layer is executed, and if it outputs 0, it is not executed. Liu et al. [12] take action based on reinforcement learning to decide which layer is executed by the residual network, which attempts to reduce the number of layers executed by rewarding the accuracy of the actual execution while suppressing the reduction in accuracy. However, while these approaches reduce the average inference time, they require additional gate functions and neural networks and thus suffer from increased memory consumption.

In the second category, Huang et al. [13] used a model that multiplies the reasonable judgment values by the output of the network layers. et al. [14] set many judgment values to 0 in learning by adding L1 regularization on this judgment value, while being able to remove such layers completely since the scalar is the same as the layer corresponding to the value 0 for the unexecuted state. Wu et al. [15] set a threshold value in the output of each layer and removes the layers below the threshold value. These methods differ from the methods in the first category in that the time and memory consumption of the model inference computation can be reduced simultaneously, except that the layers can be cut completely. However, it is difficult to maintain the accuracy rate in order to completely remove layers. Although experiments show that these methods can actually maintain accuracy in data such as CIFAR-10, it is difficult to maintain accuracy in actual data such as specific real-time images. In addition, these methods require adjustment of hyperparameters of continuous values such as regularization strength and threshold. That is, our method is used to obtain the Resnet network with the required number of layers by continuously adjusting the hyperparameters, which does not cause degradation in the accuracy of the image data, and additional cost is spent.

In the third category, Wen et al. [16] used the technique of distillation, which is a framework for efficiently training another neural network (student) using information from an already trained neural network (teacher). Specifically, it aids student learning by imposing a constraint that the probability distribution of the output of the teacher model and the output of the student model should be parsimonious. In addition, it is easy to use as the number of layers required can be set directly for teacher-student learning in the case of hardware memory constraints. Since then, [17–19] distillation methods have been applied to the compression of various models. These schemes are different from the layer parameter removal involved in this study.

2. ResNet

ResNet is a CNN neural network model widely used in the field of image recognition [3]. ResNet achieves the deep structure of the model by stacking multiple residual blocks, and each residual block is composed of residual units constructed from multiple convolutional layers. Residual units in the residual network model are calculated as follows:

$$x_{i+1} = x_i + F(x_i), \tag{1}$$

where x is the input signal to the residual unit and $F(x_i)$ is a module consisting of a convolution layer, batch normalization, and ReLU activation function [20]. Thus, the residual unit takes the input signal through constant mapping and nonlinear mapping and adds its result as a new kind of computational unit.

Multiple residual units of the same size in different dimensions are superimposed in each residual block. When changing the residual block, downsampling, or increasing the number of channels is performed, as shown in Figure 1, the dimensionality of the residual unit is changed.

3. The Proposed Method

The proposed method statically eliminates the number of ResNet layers while minimizing the loss of its accuracy. This method gradually reduces the number of layers of the residual network by iterative layer deletion and retraining.

3.1. Deleting Residual Blocks. Previous studies have shown that schemes that directly remove the residual blocks significantly reduce the model accuracy [14, 21, 22]. However, if these residual blocks are removed, it is difficult to recover the accuracy by retraining. Therefore, it is necessary to remove the residual blocks that have little impact on accuracy by model compression techniques.

For this problem, our approach introduces a variable that represents the importance of the residual units. This variable is a judgment value that can be learned from the training data as well as from the model parameters, introducing a variable for each residual unit. Specifically, the importance of $F(x_i)$ in equation (1) is learned. Identifying the unimportant $F(x_i)$ removes it, and the input of the next residual block becomes $x_i + 1 = x_i$, which has the same effect as removing the residual block itself. By introducing the variables indicating the importance into the residual unit in formula (1), we can obtain

$$x_{i+1} = x_i + \omega_i F(x_i), \tag{2}$$

where ω_i is a judgment-valued variable that can be learned by error backpropagation. ω_i can be viewed as a judgment-valued layer overlaid on top of $F(x_i)$, so calculations such as error backpropagation can be easily implemented using a deep learning framework. When the absolute value $|\omega_i|$ of ω_i becomes small, the output size of $F(x_i)$ is considered small, such that $F(x_i)$ is considered to have little effect on the output. Therefore, in this method, ω_i is used as the judgment
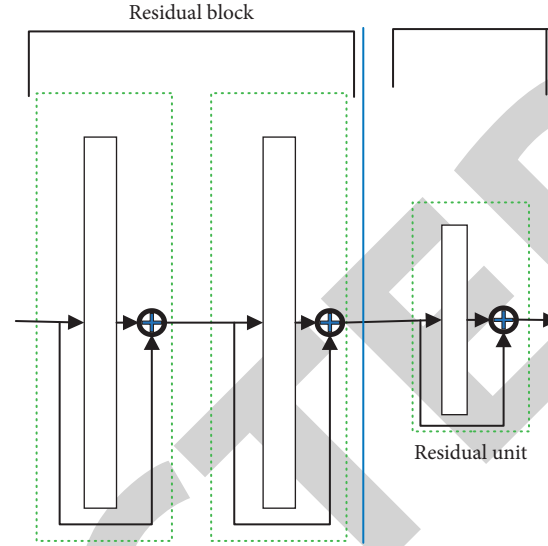
FIGURE 1: The change of residual block and residual unit in ResNet.

value for the importance of the residual block, and $|\omega_i|$ is used as the target to remove the residual block.

In the residual network, there are residual blocks that cannot be removed from the beginning to the end. According to [23], removing residual blocks immediately after a residual block change significantly reduces the model accuracy because a new intermediate representation is obtained by downsampling and increasing the number of channels when the residual block changes (Figure 1). Since such a residual block is important to maintain accuracy, instead of removing the residual block immediately after a residual block change, the usual formula (1) is used in this paper. This solution requires an additional parameter (importance), which is not a vector or tensor, but a scalar, so it has little impact on the size of the model [24].

3.2. Retraining. The authors in [14, 23] and others showed that the accuracy has a tendency to decrease if multiple residual blocks are removed from the residual network at a time. Based on the present method of alternating the removal and retraining of residual blocks, instead of removing multiple residual blocks at a time, the residual blocks are removed in a gradual manner. Baker et al. [25] categorized the method of removing parameters from the model as elemental variational kernel/group filtering level, and this method removes parameters of the layer level, which is not included in the above category. Therefore, the amount of model modification by retraining is also considered to be large. In retraining, this paper sets the learning rate initial value larger in the stochastic gradient descent optimizer to significantly update the parameters [11, 26]. Specifically, this paper applies the same learning rate for retraining as in the first learning. Such a learning rate setting contrasts with the small learning rate used in [24, 27] for retraining in single parameter removal.

3.3. Overall Algorithm. Algorithm 1 is the pseudocode of our method, where the learning rate of an optimization algorithm such as stochastic gradient descent is denoted as η, the total

number of residual blocks in the residual network is denoted as L, the number of residual blocks deleted at one time is denoted as K, the total number of target residual blocks is denoted as L', and the number of retraining iterations is n.

In Algorithm 1, indicating the importance of residual blocks is first initialized for all residual blocks (line 1). The network model weights are initialized with uniform random numbers, and the residual network is trained using stochastic gradient descent optimization (line 2). This algorithm gradually removes the residual blocks in the future and imports L (line 3) indicating the current number of residual blocks. The residual blocks are retrained in a loop (lines 4 to 10). Residual blocks are removed during the algorithm loop with a target of the preset number of residual blocks, but residual block removal is no longer performed when the accuracy of the model decreases. The set of I that indicates the index of the residual unit that becomes the object of deletion is set within the loop (line 5). The indexes of the residual blocks judged as unimportant according to the weights are appended to I (line 6). $\omega_i F(x_i)$ corresponding to the index contained therein is deleted from (2), and the deletion of the layer is performed (line 7). Along with the deletion of residual blocks, the current number of residual blocks $L\prime$ is updated (line 8). For the residual network after the deletion of residual blocks, n retraining iterations are performed (line 9). At this point, the initial value of the learning rate η for retraining is the same as the initial learning rate used in line 2.

The algorithm in this paper is shown in the flowchart in Figure 2. Firstly, the variable of residual block importance is introduced to identify whether it is an inherent residual block or an unimportant residual block to delete it, to update the number of residual blocks. If the set number of residual blocks is not reached, then return to the residual block importance judgment for layer deletion. If it is reached, then retraining is performed. If the accuracy rate does not drop, then return to the residual block importance judgment for layer delete. If the accuracy rate decreases, then the whole layer deletion process is ended.

4. Experimental Results and Analysis

4.1. Experimental Settings

4.1.1. Dataset.
The CIFAR-10/100 [28] and ImageNet datasets [29] are used as the experimental validation datasets. CIFAR-10/100 consists of image with 10/100 classifications per dataset. The image size is $32 \times 32 \times 3$. ImageNet is a dataset with 1000 classifications, and the image size is $224 \times 224 \times 3$. In this experiment, a $224 \times 224 \times 3$ single-center crop is applied to the images during training and testing with reference to the literature [30]. Furthermore, as in the literature [31, 32], in this paper, color and proportional aspect ratio enhancements are applied as data enhancements during the training process.

4.1.2. Model.
In this evaluation, the residual network model experimented refers to [30] with three convolutional layers for each residual block and combines batch normalization

and activation function ReLU. The number of residual blocks is assumed to be 3 for CIFAR-10/100 and 4 for ImageNet. Thus, as proposed by He [3], the number of layers is 56 for CIFAR-10/100 and 50 for ImageNet, and the dimensionality of the residual units changes using a projection scheme [32, 33] at the time of the change of residual blocks.

4.1.3. Hyperparameters.
For hyperparameters, according to the settings used in the standard image classification [22], the optimizer uses SGD with modulus of 0.9, with an initial learning rate of 0.1 and 200 iterations. The learning rate in CIFAR-10/100 is 0.81. The batch setting is 128 in CIEAR10/100 and 512 in ImageNet.

4.1.4. Baseline.
The baseline solution is model compression using residual networks without model compression and distillation [14, 22]. As described in a related study, model distillation dynamically skips layers compared to the static removal of layers to reduce computational cost while maintaining accuracy. In addition, model distillation, unlike other methods, can directly specify the number of layers and is therefore easy to use in situations such as hardware memory limitations, and its model use is consistent with the structure of the residual network model used in this paper.

4.1.5. Implementation Platform.
In this paper, the Keras and TensorFlow [13] are used to implement the designed model and baseline. In addition, CUDA and CUDNN libraries are used for GPU accelerated training to implement the stacked capsule network model. The platform was trained using an Intel i7-10500U Processor, 2.7 GHz, 3M processor speed, 8 GB RAM, 1 TB hard disk, Nvidia GeForce GPU for the system.

4.2. Experimental Results

4.2.1. Results of the Proposed Method.
The residual block of (2) is employed in the residual network of our method; however, the representation of equation (1) is used in the residual unit transformation. The performance comparison between the deleted residual blocks and the original residual network is given in Table 1. A residual block has 3 convolutional layers, and it can be found that even after some residual blocks are deleted, the accuracy on each dataset does not abate but increases; thus, for model compression, the deletion of neural network layers can be performed, which can effectively avoid overfitting. The number of retraining iterations n is set to be the same as the original training iterations, but in this experiment, even if n is set to be smaller than the original learning iterations, the accuracy can be maintained to some extent. Table 1 compares the number of remaining network layers in each residual block before and after removing the network layers, and it can be seen that the first residual block retain the least number of network layers. In turn, this allows the model to be compressed substantially and effectively.

Initialization: learning rate η, number of residual units L, number of residual units removed at one time is k, number of retraining iterations is n.

(l) ResNet's parameters and weights $\{\omega_i : L \in [L]\}$ are initialized.

(2) Set the initial learning rate η to train ResNet.

(3) $L' = L$

(4) while $L' > L$ do

(5) $I = \theta$

(6) Select k in ascending order of the absolute value of ω_i and add that index to I

(7) Delete $\omega_i F(x_i)$ of $i \in I$ from equation (2)

(8) $L' \leftarrow L' - k$

(9) Retraining in SGD with initial learning rate for n iterations

(10) End

ALGORITHM 1: Our method.
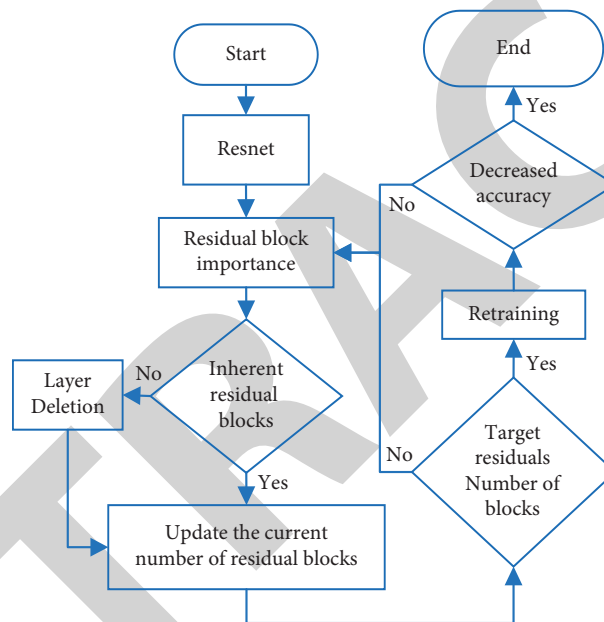
FIGURE 2: The flowchart of our algorithm.

Table 1 represents the number of remaining layers in each residual block of the minimum residual network where no accuracy degradation occurs, and according to Table 1, it can be seen that the proposed method can remove different layers of network layers in each residual block. According to the results of this paper, the first residual block, i.e., the residual block closest to the input, has the least number of layers; i.e., the most layers can be removed, and for the latter residuals blocks, some of the network layers can be also removed separately without affecting the final accuracy. Moreover, according to [13, 14], the first residual can be said to have little impact on the accuracy even if the network layers are reduced.

4.2.2. Accuracy. This subsection evaluates the image classification accuracy of the test data when the number of layers of the residual network is changed; however, the accuracy of the validation data is evaluated following the customary

evaluation since the ImageNet dataset does not exist for the test data. The comparison method is a residual network varying the number of network layers from 56 to 11 in CIFAR-10/100. Six residual blocks are employed in each dataset, and each residual block has 18, 15, 12, 9, 6, and 3 layers respectively. In ImageNet classification training in this paper, the method is changed from 50 to 17 layers for the residual network and distillation method using 50, 34, and 18 layers of the model as a comparison object.

The experimental results are visualized in Figure 3 with the graphs of the number of network layers and the accuracy of image classification. The blue dashed line indicates the accuracy of the residual network with different number of layers, the green dashed line indicates the accuracy of the residual network with different number of layers learned by distillation, and the red dashed line shows the initial accuracy of the proposed method. The proposed method maintains the initial accuracy indicated by the red dashed line while the layers are deleted. Eventually, the proposed

TABLE 1: The number of remaining layers of each residual block and accuracy before and after layer deletion.

Dataset	Layer number	Accuracy (%)	1st residual block Layer number	2nd residual block Layer number	3rd residual block Layer number	4th residual block Layer number
CIFAR-10	56	92.88	18	18	18	None
	32	93.05	3	15	12	None
CIFAR-100	56	71.83	18	18	18	None
	35	71.99	3	12	18	None
ImageNet	50	75.89	9	12	18	9
	38	76.12	6	6	15	9

(a)

(b)

(c)
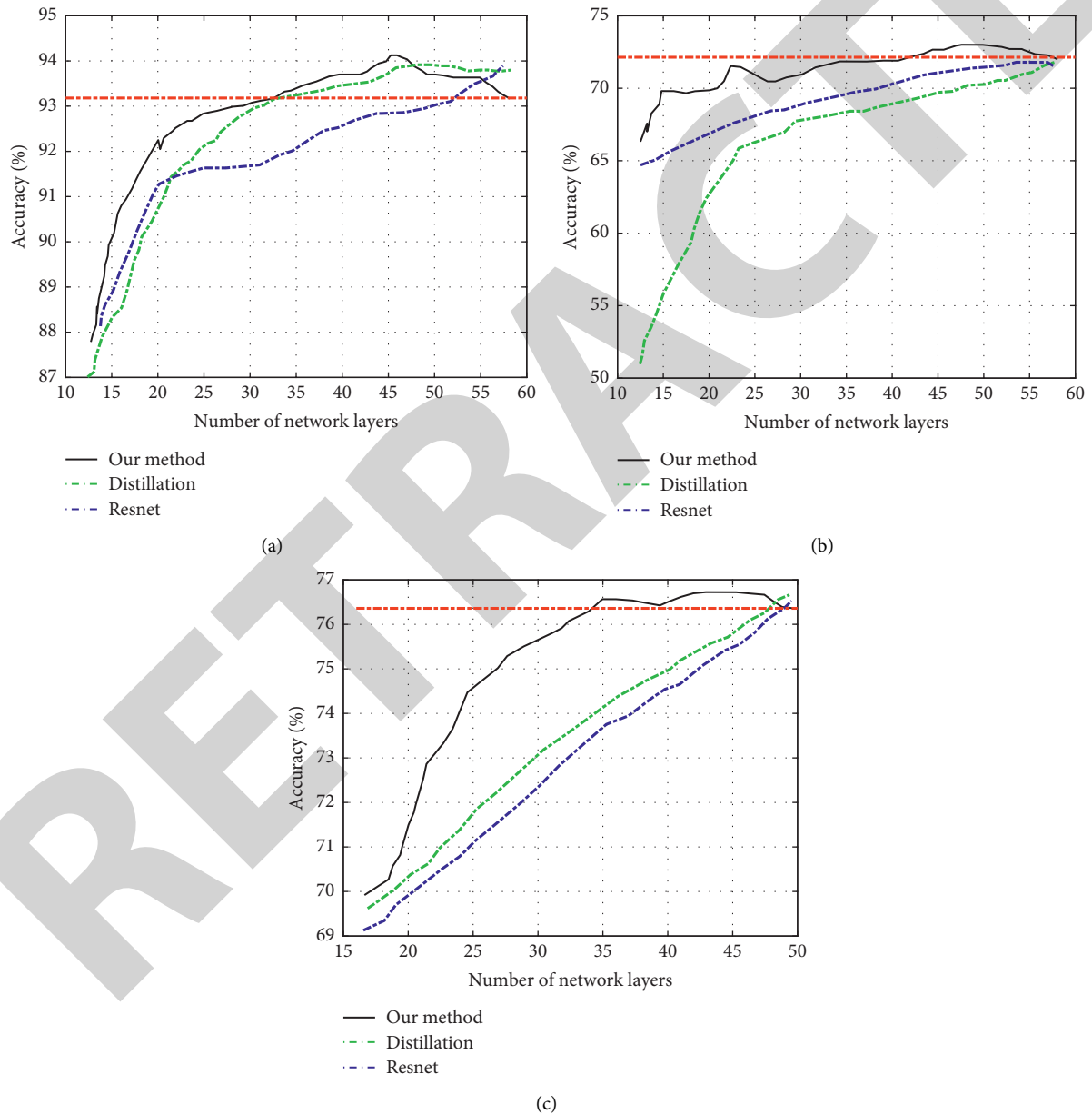
FIGURE 3: Accuracy of different network layers in different methods: (a) CIFAR-10; (b) CIFAR-100; (c) ImageNet.

method maintains the accuracy of 32, 35, and 38 layers in CIFAR-10/100 and ImageNet, respectively, which are the numbers of layers corresponding to the intersection of the

black and red dashed lines. That is, 42.86%, 37.50%, and 24.0% of the network layers were removed in CIFAR-10/100 and ImageNet, respectively. Comparing the blue dashed line

and the black line, compared with the usual case of the residual network with layer change learning, the model distillation achieves essentially the same accuracy as the method in this paper in CIFAR-10 but cannot maintain the accuracy in CIFAR-10/images; on the other hand, the method in this paper can also maintain the accuracy while cutting the number of layers in CIFAR-100 and ImageNet. The reason why the method in this paper maintains the original accuracy can be attributed to the fact that a priority term is introduced in this paper to identify the importance of each neural network layer, and then the unimportant layers are selected to be removed during the training process based on the priority of the neural network layers. In addition, the network model is also retrained to avoid accuracy degradation when the network layers are removed.

4.2.3. Calculating Cost.

This section evaluates the computational cost of the minimum residual network for which the method in this paper is able to maintain accuracy, as an evaluation metric, the execution time of sequential propagation used in the MAC model [13, 18] inference calculations using the representation of the product and the number of operations, the execution time of backpropagation, and the number of parameters of the model. MAC is calculated using the convolutional layer and the mean of all combined layers. The execution time is evaluated as the average of 100 execution times. The experimental setup in the study is the same as in the previous section. The proposed method can reduce the number of layers while maintaining high accuracy to 32, 35, and 38 layers in CIFAR-10/100 and ImageNet, respectively, and therefore these models are evaluated above utilizing price metrics.

Table 2 indicates the evaluation results. Regarding the number of MACs, it was cut to 60.93%, 62.89%, and 78.59% in CIFAR-10/100 and ImageNet, respectively. The execution time of sequential propagation was cut to 60.23%, 70.13%, and 76.69%, respectively. The execution time of reverse propagation was cut to 59.71%, 60.44%, and 78.9%. In terms of quantity, without accuracy degradation, the number of parameters can be cut to 69.82%, 90.50%, and 93.15%, respectively. These experimental results demonstrate that the proposed method speeds up the model inference computation without increasing the memory consumption and without accuracy degradation. In Section 1.1, it is elaborated that the dynamic layer skipping maneuver tends to maintain accuracy but increases memory consumption; conversely, the static layer deletion maneuver cuts memory consumption but decreases accuracy. When compared in a minimal residual network that does not cause the accuracy degradation caused by the proposed method, the proposed method improves all computational cost metrics without accuracy degradation.

4.2.4. Hyperparameter Dependence.

This section explores the relationship between the number of iterations of retraining and the accuracy of the number of residual blocks removed for the hyperparameters of the methods in this paper.

For 56-layer residual network learned in CIFAR-10, the accuracy versus the number of network layers when the number of iterations to be retrained is 30, 60, and 120 is illustrated in Figure 4. It can be known that the accuracy is easily maintained when the number of zones is high, but it can also be maintained to some extent when the number of iterations is low. In this method, even in the case of 30 iterations of retraining, the accuracy can be maintained to some extent while deleting layers in order to be used as the initial value for retraining, and retraining can be started from the model with a certain degree of high accuracy from the beginning. Figure 4 shows that the accuracy is still good when the number of iterations of retraining is fixed at 60 and the number of deleted residual blocks can be 1, 2, or 4. Notice that when $n = 4, 44, 32, 20$, the smaller the $n = 4$, the greater the improvement in accuracy through retraining. As shown in [3], the smaller the number of residual blocks removed, the smaller the decrease in accuracy, which can be considered as the reason why retraining can be started from the beginning with a certain degree of high accuracy. In addition, the minimum residual network that was able to maintain the initial accuracy was 32 layers with $n = 2$. The residual network with 44 layers with $n = 4$ had an accuracy of 92.84%, achieving essentially the same accuracy as the preliminary accuracy of 92.88%.

4.2.5. Weighted Regularization Effect.

In the proposed method, the absolute value of ω_i is used as an indicator of the importance of the residual blocks; i.e., if the absolute value of ω_i is small, the output of $F(x_i)$ is reduced and the result $\omega_i F(x_i)$ becomes smaller and has a smaller impact on the overall output. The average relationship between the absolute value of ω_i and the size (L2) on the validation data for the 56-layer residual network learned in CIFAR-10 can be visualized in each residual block, as shown in Figure 5. The smaller the importance in Figure 5, the smaller the weight of the output. It can be known that the smaller the absolute value of ω_i, the smaller the L2 parametric number of $\omega_i F(x_i)$, so it can be said that the absolute value of $\omega_i F(x_i)$ can be used as an indicator of the importance of the residual unit according to ω_i reduction $F(x_i)$.

4.2.6. Weight Distribution.

In this paper, the residual network is trained on the CIFAR-10 dataset, and the weight distribution of the residual blocks is visualized to investigate the nature of the "importance" introduced in the proposed method. Figure 6 demonstrates the importance histogram of the residual blocks for the residual network trained with CIFAR-10. Initialized with uniform random numbers, the trained shape becomes a mixture of two Gaussian distributions. In element-level parameter removal, the histogram of known learning parameters is monotonously normally distributed with a mean around 0 [12]. Since the absolute value of the parameter is used as importance in element-level parameter deletion, the most frequent values are around 0. Many parameters are judged to be unimportant, and parameters close to 0 are considered to have little impact on the output even if they are deleted, and can be retrained with

TABLE 2: Calculation cost in model reasoning.

Dataset	Layers number	Accuracy (%)	MAC	Sequential propagation (msec)	Reverse propagation (msec)	Parameters
CIFAR-10	56	92.88	8.19×10^7	6.584	12.93	585.9K
	32	93.05	4.99×10^7	3.970	7.721	409.1K
CIFAR-100	56	71.83	8.65×10^7	6.203	13.36	613.6K
	35	71.99	5.44×10^7	4.350	8.075	555.3K
ImageNet	50	75.89	4.11×10^9	29.95	59.51	25.55M
	38	76.12	3.23×10^9	22.97	46.53	23.80M
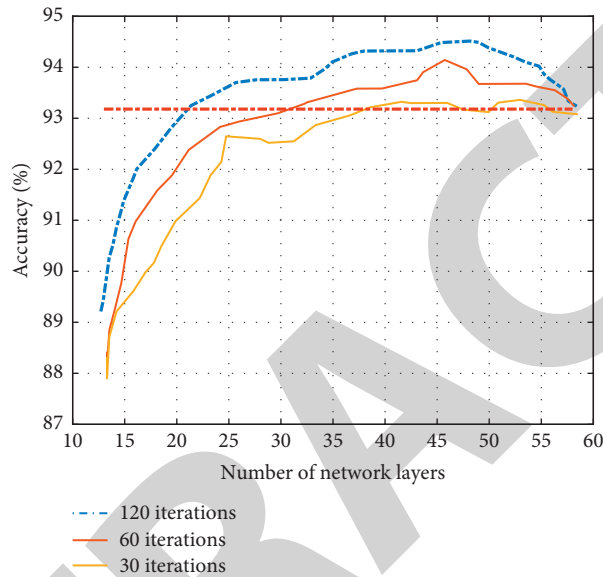
FIGURE 4: The change of the accuracy of deletion and retraining of ResNet layer in CIFAR-10 training.

(a)   (b)

FIGURE 5: Continued.

(c)
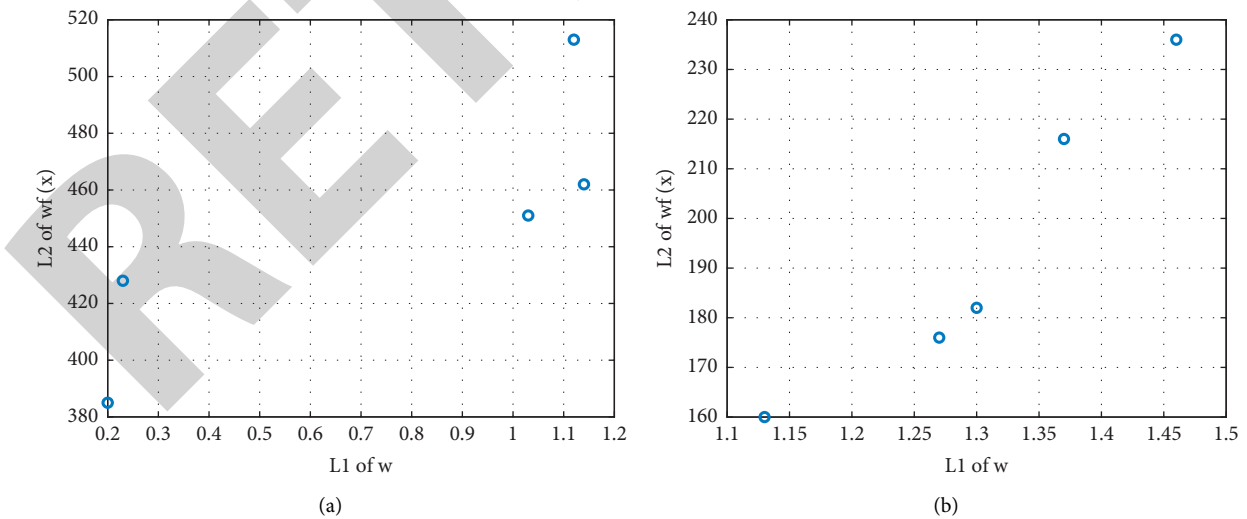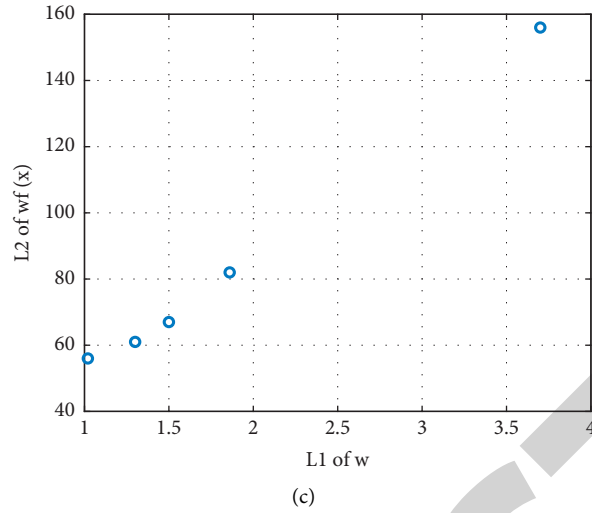
Figure 5: The average distribution of the weight of each residual block of ResNet in CIFAR-10: (a) CIFAR-10; (b) CIFAR-100; (c) ImageNet.
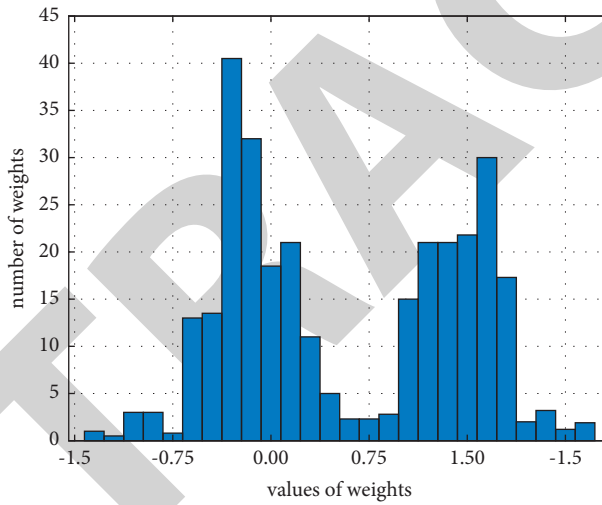
Figure 6: Weight histogram of residual unit in CIFAR-10.

(a)

(b)
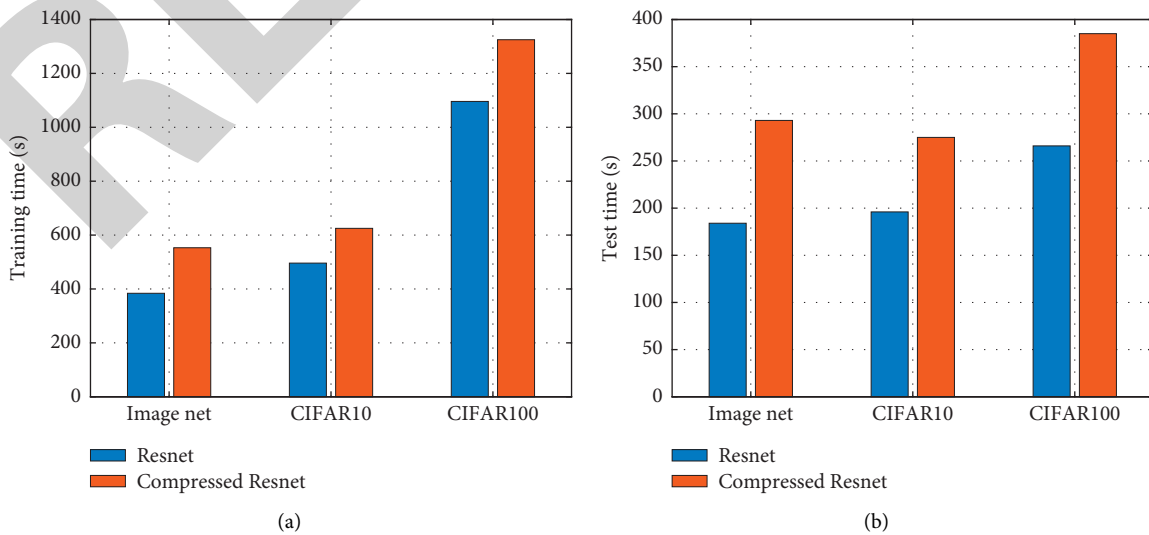
Figure 7: Time spent on (a) training stage and (b) testing stage.

Table 3: The number of model parameters before and after compression.

Models/model parameters	CIFAR-10	CIFAR-100	ImageNet
ResNet	585.9K	613.6K	25.55M
Compressed ResNet	409.1K	555.3K	23.80M

small learning rates to maintain accuracy as in [12]. On the other hand, the histogram of cascading parameter deletion is like two normal distributions symmetrical around 0, as shown in Figure 6. If we take the absolute value as the importance in this paper, the most common value is not 0, but around 0.7. In other words, when layer deletion is performed with the proposed method, there are cases when some network layers with higher importance must be deleted. In this case, the model needs to be modified substantially to maintain the accuracy of the model, so a larger learning rate needs to be set during the retraining process.

4.2.7. Model Complexity Analysis. Figure 7(a) indicates the time spent in the training phase for the original ResNet and the compressed ResNet of this paper. It can be clearly seen that the training time of the compressed ResNet in this paper is the least on all three different datasets, which indicates that the model complexity is lower than that of the original ResNet. In the testing phase indicated in Figure 7(b), the compressed ResNet in this paper takes even less time. These results show that the compressed ResNet in this paper has faster recognition speed for both image samples at the end of the training phase. The numbers of parameters of the original ResNet and the compressed ResNet in this paper are shown in Table 3, and the number of parameters of the compressed ResNet in this paper is very much smaller than that of the original ResNet, which indicates that the model overhead cost of the compressed ResNet in this paper is lower than that of the original ResNet. This indicates that the compressed ResNet in this paper can be flexibly and quickly deployed on existing hardware.

5. Conclusions

In order to reduce the computational cost of inference, this study proposes a method to reduce the number of residual network layers without reducing the accuracy. The proposed method has a scalar parameter to identify the insignificant residual units, based on which the residual units are selected and removed. In the image classification task of CIFAR-10/100 and ImageNet, the number of network layers is effectively removed for model compression while maintaining accuracy compared to existing methods.

Data Availability

The dataset used in this paper are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest regarding this work.

Acknowledgments

References

[1] C. Zhang, T. Xie, and K. Yang, "Positioning optimisation based on particle quality prediction in wireless sensor networks," *Networks, IET*, vol. 8, no. 2, pp. 107–113, 2019.

[2] Y. X. Zou, J. S. Yu, Z. Chen, J. Chen, and Y. Wang, "Convolution neural networks model compression BasedOn feature selection for image classification," *Control Theory & Applications*, vol. 34, no. 6, pp. 746–752, 2017.

[3] K. He, X. Zhang, S. Ren, and J. Sun, "Residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, Hilton Head, SC, USA, 15-15 June 2000.

[4] Z. Ye and S. Xiao, "Compression of convolutional neural network applied to image classification," *Journal of Beijing Information Science and TechnologyUniversity*, vol. 33, no. 3, pp. 52–56, 2018.

[5] S. Han, J. Pool, J. Tran, and W. J. Dally, "Learning both weights and connections for efficient neural network," in *Proceedings of the Advances in Neural Information Processing Systems (NIPS)*, Montreal, Canada, December 2015.

[6] M. Müller, J. Finke, L. Stahl, Y. Tong, H. Fricke, and T. Vallee, "Development and validation of a compression flow model ofnon-Newtonian adhesives," *The Journal of Adhesion*, vol. 2021, no. 4, pp. 1–38, 2021.

[7] E. L. Denton, W. Zaremba, J. Bruna, Y. Lecun, and R. Fergus, "Exploiting linear structure within convolutional networks for efficient evaluationp," in *Advances in Neural Information Processing Systems*, Montreal, Canada, December 2014.

[8] W. Chen, J. T. Wilson, S. Tyree, K. Q. Weiberger, and Y. Chen, "Compressing neural networks with the hashing trick,"vol. 37, pp. 2285–2294, in *Proceedings of the 32nd International Conference on Machine Learning (ICML)*, vol. 37, pp. 2285–2294, PMLR, Lille, France, July 2015.

[9] M. Rastegari, V. Ordonez, J. Redmon, and A. Farhadi, "Xnornet: Imagenet classifification using binary convolutional neural networks," in *Proceedings of the European Conference on Computer Vision*, pp. 525–542, (ECCV), Amsterdam, Netherlands, October 2016.

[10] M. Jaderberg, A. Vedaldi, and A. Zisserman, "Speeding up convolutional neural networks with low rank expansions," in *Proceedings of British Machine Vision Conference (BMVC)*, Nottingham, UK, September 2014.

[11] S. Han, J. Pool, J. Tran, and W. J. Dally, "Mcdnn: an approximation-based execution framework for deep stream processing under resource constraints," in *Proceedings of the 14th Annual International Conference on Mobile Systems, Applications, and Services*, pp. 123–136, Association for Computing Machinery, New York, NY, USA, 2016.

Research Article

Parameter Estimation of Attribute Scattering Center Based on Water Wave Optimization Algorithm

Zhangkai Zhou and Yihan Li [ORCID]

School of Information Science and Engineering, Lanzhou University, Lanzhou 730000, China

Correspondence should be addressed to Yihan Li; yhli18@lzu.edu.cn

For the problem of attribute scattering center parameter estimation in synthetic aperture radar (SAR) image, a method based on the water wave optimization (WWO) algorithm is proposed. First, the segmentation and decoupling of high-energy regions in SAR image are performed in the image domain to obtain the representation of a single scattering center. Afterwards, based on the parameterized model of the attribute scattering center, an optimization problem is constructed to search for the optimal parameters of the separated single scattering center. In this phase, the WWO algorithm is introduced to optimize the parameters. The algorithm has powerfully global and local searching capabilities and avoids falling into local optimum while ensuring the optimization accuracy. Therefore, the WWO algorithm could ensure the reliability of scattering center parameter estimation. The single scattering center after solution is eliminated from the original image and the residual image is segmented into high-energy regions, so the parameters of the next scattering center are estimated sequentially. Finally, the parameter set of all scattering centers in the input SAR image can be obtained. In the experiments, firstly, the parameter estimation verification is performed based on the SAR images in the MSTAR dataset. The comparison of the parameter estimation results with the original image and the reconstruction based on the estimated parameter set reflect the effectiveness of the proposed method. In addition, the experiment is also conducted using the SAR target recognition algorithms based on the estimated attribute parameters. By comparing the recognition performance with other parameter estimation algorithms under the same conditions, the performance superiority of the proposed method in attribute scattering center parameter estimation is further demonstrated.

1. Introduction

Synthetic aperture radar (SAR) has all-day and all-weather high-resolution imaging capabilities, which is widely used in military and civilian fields [1–4]. Unlike clear and intuitive optical images, SAR images reflect the electromagnetic scattering characteristics of the target, in which the appearance of the target is not obvious, which brings obstacles to image interpretation. In order to quantitatively describe the distribution characteristics of SAR images, the researchers model the local scattering phenomenon by means of the scattering center model. The most representative one is the attribute scattering center model [5–7]. The attribute scattering center model contains attribute parameters related to the local structures of the target, such as intensity distribution, positions, and lengths. SAR target recognition methods based on the attribute scattering centers also fully

verified the effectiveness of attribute parameters for distinguishing different targets [8–11]. However, the form of the attribute scattering center model is complicated, so the parameter estimation problem is a hard one. In the early stage, researchers used the "divide and conquer" strategy to realize the sequential estimation of each scattering center in the image domain. In the parameter estimation of a single scattering center, many classical optimization algorithms are usually used, such as the genetic algorithm, simulated annealing, particle swarm algorithm, and wolf swarm algorithm [12–16]. According to the theory of compressive sensing, the researchers also designed the attribute scattering center estimation algorithms based on sparse representation [17], assuming that the attribute scattering centers in a single SAR image are sparsely distributed in the entire parameter space. The sparse representation methods avoid the local decoupling in the image domain, but the consistency of the

estimated results with the original image is often difficult to guarantee. In addition, it has strong randomness. For this aspect, in order to ensure the physical meanings of the scattering center parameters, the scattering center estimation algorithm in the image domain has stronger reliability.

In this paper, based on the traditional image domain attribute scattering center estimation methods, the water wave optimization (WWO) algorithm is introduced to realize the parameter optimization of a single scattering center. For a single scattering center image area decoupled from the image domain, the best attribute parameters are obtained through the WWO algorithm. Inspired by the theory of shallow water waves, researchers proposed a new heuristic algorithm in 2015, i.e., the WWO algorithm [18]. The algorithm is inspired by the theoretical model of shallow water waves. The individual water waves use propagation, refraction, and wave breaking to effectively search in the solution space. It has the advantages of concise principle, fewer control parameters, and easy implementation. At present, the WWO algorithm has been widely used and verified in the fields of task allocation, image processing, etc., with higher accuracy and robustness [19–23]. For this

reason, this paper introduces the WWO algorithm into the parameter estimation of the single attribute scattering center to obtain more reliable estimation results. In the experiment, the MSTAR dataset is used to verify the proposed method, including the contributions to the parameter estimation of SAR images and the target recognition based on attribute scattering center matching, respectively. The experimental results prove the effectiveness of the proposed method and its superiority compared with the existing methods.

2. Attributed Scattering Center Model

As a parameterized model to describe the local scattering characteristics of radar targets in the high-frequency area [24], the attribute scattering center model has stronger description capabilities than the previous geometric theory of diffraction (GTD) model [25, 26] in both the physical and signal layers. In [8–11], SAR target methods were developed based on the attribute scattering center, which verified its significant advantages in improving the robustness of target recognition. Based on the attribute scattering center model, the scattering characteristics of a single scattering center are described as follows:

$$
E_i(f, \phi; \theta_i) = A_i \cdot \left(j \frac{f}{f_c} \right)^{\alpha_i} \cdot \exp\left(\frac{-j4\pi f}{c} (x_i \cos \phi + y_i \sin \phi) \right)
$$
$$
\cdot \sin c\left(\frac{2\pi f}{c} L_i \sin(\phi - \overline{\phi_i}) \right) \cdot \exp(-2\pi f \gamma_i \sin \phi).
\tag{1}
$$

In equation (1), f represents the frequency of the incident wave and ϕ is the azimuth angle. $\theta_i = [A_i, \alpha_i, x_i, y_i, L_i, \overline{\phi_i}, \gamma_i]$ is the attribute parameter set. Specifically, A_i represents the relative amplitude of different scattering centers; x_i, y_i record the positions of the scattering center; α_i is a frequency-dependent factor whose value is directly related to the local structure; for the distributed scattering centers, L_i and $\overline{\phi_i}$ respectively represent the length and direction angle; for the localized scattering centers, γ_i refers to the position-dependent parameter.

For a SAR image containing an interested target, there are several scattering centers. Then, based on the attribute scattering center model, the overall scattering characteristics of the target can be described as follows:

$$
E(f, \phi; \theta) = \sum_{i=1}^{P} E_i(f, \phi; \theta_i),
\tag{2}
$$

where $\theta = \{\theta_i\}$ $(i = 1, 2, \ldots, P)$ is the attribute scattering center set of the target; P represents the number of the scattering centers.

3. Method Description

3.1. Parameter Estimation in the Image Domain. Although the description accuracy of the attribute scattering center model is high, its complex form makes the parameter

estimation problem a complicated and hard one. In general, the SAR image data to be analyzed can be expressed as follows:

$$
D(f, \phi) = E(f, \phi; \theta) + N(f, \phi),
\tag{3}
$$

where $E(f, \phi; \theta)$ is the model component described in equation (2); the noise component $N(f, \phi)$ in the image is generally modeled as a Gaussian distribution. Under the framework of the maximum likelihood estimation, the joint estimation process of the parameter set is described as follows:

$$
\widehat{\theta}_{ML} = \arg \min_{\theta} \| \| D - E(\theta) \|^2.
\tag{4}
$$

Equation (4) gives the ideal form of the parameter estimation of the attribute scattering centers. However, due to the complexity of the model itself and the large number of parameters, the optimization process is nonconvex, which is easy to fall into the local optimum, resulting in unstable and inaccurate estimation results. As a remedy, the researchers proposed a "divide and conquer" strategy in the image domain, which separates the image area of a single scattering center according to the significant characteristics of the effective scattering center [8–11]. Afterwards, an optimization algorithm can be used to estimate the parameters of a single scattering center. Then, the parameters of all scattering centers can be obtained sequentially. According to

previous works, the process of the image domain parameter estimation algorithm can be generally divided into three steps: decoupling, single scattering center parameter estimation, and sequential estimation. First, the input SAR image is analyzed, and the image segmentation algorithm is used to obtain the region with the largest energy as the image domain representation of a single scattering center. Then, for the decoupled single scattering center data, the optimization is performed under the constraints of the attribute scattering center model to obtain the best attribute parameters. Here, it is necessary to introduce an appropriate optimization algorithm to ensure the precision of the parameter estimation. Finally, the estimated components are removed from the original image (the reconstruction result of the scattering center is subtracted from the original image) to update the input image, and the parameters of the next scattering center are obtained. Through sequential iterative updates, the attribute parameter set of all scattering centers of the target can be output when the termination condition is met.

3.2. Parameter Estimation Based on WWO Algorithm.

In the WWO algorithm, the problem space is analogous to the seabed, and each individual in the population is analogous to a "water wave" object [18]. Each water wave has two attributes: initial wave height and wavelength. The core of the WWO algorithm is to assign a wavelength inversely proportional to its fitness for each solution and to make each solution propagation (search) range proportional to its wavelength: the closer the point to the sea level, the better the corresponding solution. The WWO algorithm mainly conducts efficient search in the problem space by simulating the three operations of water wave propagation, refraction, and wave breaking. Without loss of generality, the following takes solving the maximum value as an example to introduce the WWO algorithm.

Step 1: the dissemination stage. In each iteration, all individual water waves need to enter the propagation stage. The expression of the propagation process in each dimension is

$$s\prime(d) = s(d) + N_r \lambda L(d), \qquad (5)$$

where N_r is a random number in the interval $[-1,1]$; λ is the wavelength of the water wave; $L(d)$ is the length of the solution space in the dimension d; $s(d)$ is the position of the individual s in the dimension d; $s'(d)$ is the position of the water wave in the dimension d after the propagation process. If the new individual exceeds the boundary, the new individual is given a random position in the dimensional search space. After the propagation process is completed, calculate the new individual fitness value generated by the propagation; if $\text{fit}(s\prime) > \text{fit}(s)$, then s' is replaced to s and the initial wave height of the wave height dimension h_{max} is updated; otherwise, the wave height h is reduced by 1. If the new individual is better than the original individual and better than the best individual in the group, it

enters the stage of breaking waves. After each iteration is completed, all water waves update their wavelengths according to the following equation:

$$\lambda = \lambda \theta^{-[\text{fit}(s) - F_{min} + \varepsilon]/(\text{fit}_{max} - \text{fit}_{min} + \varepsilon)}, \qquad (6)$$

where fit_{max} and fit_{min} are the maximum fitness value and the minimum fitness value of the group, respectively; θ is the wavelength attenuation factor; ε is a constant greater than zero to prevent the denominator from being 0.

Step 2: refraction stage. The individual wave height h of the water wave is determined whether it is 0. If the wave height of the water wave h is reduced to 0, the next generation of individuals will be produced through the following refraction operation; if the wave height h is not 0, the individual will enter the next round of iteration.

$$s'(d) = N\left(\frac{s^*(d) + s(d)}{2}, \frac{|s^*(d) - s(d)|}{2}\right), \qquad (7)$$

where s^* is the best individual in the population; $N(\mu, \sigma)$ is a Gaussian random number with the mean μ and standard deviation σ. After the refraction process is completed, the wave height is updated to h_{max}, and the individual's wavelength $\lambda\prime$ is updated. If the fitness is negative, the wavelength is updated according to the following equation to take the reciprocal:

$$\lambda' = \lambda \frac{\text{fit}(s)}{\text{fit}(s')}. \qquad (8)$$

Step 3: wave breaking stage. When the group finds that a new individual has a better fitness value than the best individual s^* in the current group, then break the waves. The main process of breaking waves is to randomly select search k individuals whose dimensions are around the individual s^*, and search for nearby areas s^*. The new wave individual is updated according to the following equation in the selected dimension to obtain the new individual:

$$s'(d) = s(d) + \beta N(0, 1)L(d), \qquad (9)$$

where β is the wave breaking factor, generally taken in the interval [0.001, 0.01]. If the individual fitness value of the generated water wave is not better than that of the individual s^*, then keep it; otherwise, produce the best individual replacement s^*.

Owing to the characteristics and advantages of the WWO algorithm, this paper introduces it into the parameter estimation of the attribute scattering centers. For the single scattering center data obtained by decoupling the image domain, the WWO algorithm is used to solve the best attribute parameters. According to the strategy of order inertial estimation, the parameter set of all attribute scattering centers in the SAR image is obtained.

(a) (b) (c) (d) (e)

(f) (g) (h) (i) (j)

Figure 1: Illustration of targets in the MSTAR dataset.

4. Experiments

4.1. Preparation. This paper tests the proposed method based on the SAR image samples in the MSTAR dataset. This dataset contains SAR images of 10 types of vehicle targets shown in Figure 1. The resolution of these images reaches 0.3 m, and the local scattering phenomenon of the target is obvious. In this experiment, the proposed method is used to estimate the scattering center parameters of MSTAR SAR images. Since the true value of the scattering center parameters of these SAR images is unknown, the validity of the estimation results can only be preliminary determined by comparing the correspondence between the estimated parameters and the strong local scattering of the target (mainly the position parameters) and the consistency between the reconstructed image and the original image. In addition, since the attribute scattering center has been widely used in the SAR target recognition, this paper further evaluates the effectiveness of the proposed method by comparing the parameter estimation results with other ones in the sense of contributions to the recognition performance.

4.2. Parameter Estimation Performance. The proposed method is used to verify a typical SAR image from the MSTAR dataset, and the upper limit of the number of scattering centers is chosen to be 15. Figure 2 shows the projection result of the position parameters on the original
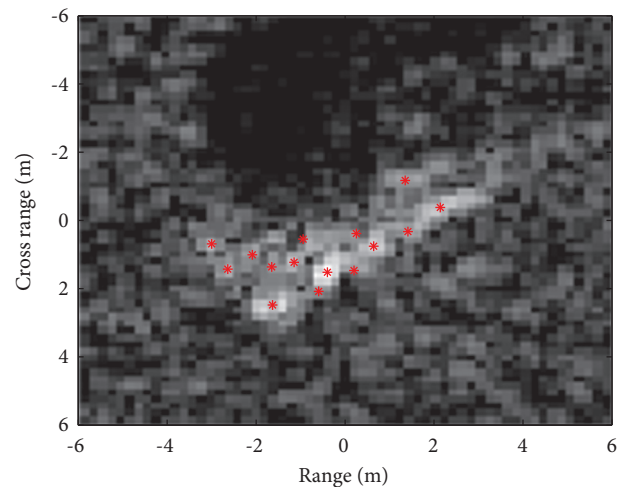
Figure 2: Illustration of locations of estimated scattering centers.

SAR image. It can be seen that there is a strong agreement with the local strong scattering center of the target, indicating that the accuracy of the position parameter in the estimation result has a higher accuracy. Figure 3(a) shows the target image reconstructed based on the estimated scattering centers. The classical image correlation is used as the measurement degree, and the similarity between it and the original SAR image is calculated to be 0.932, indicating that the reconstructed result is very similar to the original

(a)

(b)
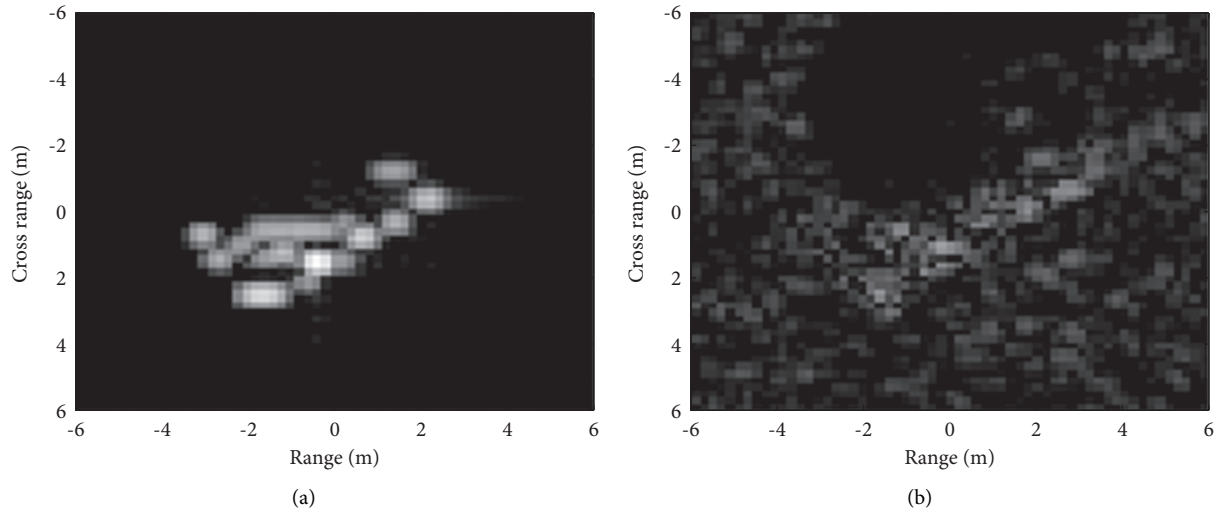
FIGURE 3: (a) Reconstructed target and (b) residuals from reconstruction based on the estimated scattering centers.

TABLE 1: Training and test samples of the ten targets.

Class	BMP2	BTR70	T72	T62	BDRM2	BTR60	ZSU23/4	D7	ZIL131	2S1
Training (17°)	232	233	231	299	298	256	299	299	299	299
Test (15°)	195	196	196	273	274	195	274	274	274	274

image. The high correlation indicates the validity of the estimation results of various attribute parameters of the scattering center. Figure 3(b) shows the residual component after the subtraction between the original image and the reconstructed image. Intuitively, the reconstruction residual is mainly from the background component, which is redundant due to clutters or noises. Through the estimation of the target scattering center in the original SAR image and reconstruction, the background factors existing in the original image are effectively eliminated, which has strong support for the analysis of target characteristics. The above results all show the effectiveness of the proposed method for estimating the parameters of the scattering centers in SAR images.

4.3. Target Recognition Performance. In [8–11], SAR target recognition methods are designed based on the attribute scattering centers. One of the typical scenarios is to classify the 10 types of targets shown in Figure 1 under the standard operating condition, under which the training and test sets are shown as Table 1. In order to further verify the effectiveness of the proposed scattering center parameter estimation method, the results of different estimation algorithms can be used for the same SAR target recognition method based on the attribute scattering centers. Then, the recognition performance can be compared. Accordingly, this paper employs different attribute scattering center estimation algorithms, i.e., the simulated annealing, the particle swarm, and sparse representation, to extract the attribute scattering centers. Then, the estimated scattering centers are matched using the method in [9] to classify the 10-class test samples in Table 1. Under different algorithms,

TABLE 2: Average recognition rates of the scattering center matching method using different parameter estimation algorithms.

Parameter estimation algorithm	Average recognition rate (%)
Simulated annealing	98.02
Particle swarm	98.36
Sparse representation	98.84
Proposed	99.08

the average recognition rates obtained by the recognition method are listed in Table 2. The comparison shows that the attribute scattering center estimated based on the proposed method can achieve the highest average recognition rate, indicating that it has a greater contribution to the recognition performance, reflecting its relatively higher estimation accuracy. Therefore, the contribution of different scattering center extraction algorithms to target recognition can indirectly reflect the stronger effectiveness of the proposed method.

5. Conclusion

To handle the problem of attributed scattering center parameter estimation in SAR images, this paper proposes a method based on the WWO algorithm. On the basis of the traditional image domain "divide and conquer" strategy, the WWO algorithm is introduced to estimate the parameters of each attribute scattering center in a sequential way. In the experiments, the performance of the proposed method is tested based on the MSTAR dataset. First, a typical demonstration of parameter estimation is carried out based on MSTAR SAR images. The effectiveness of the proposed method is verified by comparing the estimated position

parameters with the position of strong scattering points in the image and the similarity between the reconstructed image and the original one. Afterwards, the contribution of different parameter estimation algorithms to SAR target recognition is compared, which further reflects the superior performance of the proposed method.

Data Availability

The MSTAR dataset is publicly available.

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] K. El-Darymli, E. W. Gill, P. Mcguire, D. Power, and C. Moloney, "Automatic target recognition in synthetic aperture radar imagery: a state-of-the-art review," *IEEE Access*, vol. 4, pp. 6014–6058, 2016.

[2] M. Amoon and G. Rezai-rad, "Automatic target recognition of synthetic aperture radar (SAR) images based on optimal selection of Zernike moment features," *IET Computer Vision*, vol. 8, no. 2, pp. 77–85, 2014.

[3] B. Ding, G. Wen, C. Ma, and X. Yang, "Target recognition in synthetic aperture radar images using binary morphological operations," *Journal of Applied Remote Sensing*, vol. 10, no. 4, Article ID 046006, 2016.

[4] C. Shan, B. Huang, and M. Li, "Binary morphological filtering of dominant scattering area residues for SAR target recognition," *Computational Intelligence and Neuroscience*, vol. 2018, Article ID 9680465, 15 pages, 2018.

[5] L. C. Potter and R. L. Moses, "Attributed scattering centers for SAR ATR," *IEEE Transactions on Image Processing*, vol. 6, no. 1, pp. 79–91, Jan. 1997.

[6] M. J. Gerry, L. C. Potter, I. J. Gupta, and A. Merwe, "A parametric model for synthetic aperture radar measurements," *IEEE Transactions on Antennas and Propagation*, vol. 47, no. 7, pp. 1179–1188, 1999.

[7] Z. Li, K. Jin, B. Xu, W. Zhou, and J. Yang, "An improved attributed scattering model optimized by incremental sparse Bayesian learning," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 5, pp. 2973–2987, 2016.

[8] B. Ding, G. Wen, J. Zhong, C. Ma, and X. Yang, "A robust similarity measure for attributed scattering center sets with application to SAR ATR," *Neurocomputing*, vol. 219, pp. 130–143, 2017a.

[9] B. Ding, G. Wen, X. Huang, C. Ma, and X. Yang, "Target classification in synthetic aperture radar images via matching of attributed scattering centers," *IEEE Journal of Selected Topics in Applied Earth Observation and Remote Sening*, vol. 10, no. 7, pp. 3334–3347, 2017.

[10] X. Zhang, "Noise-robust target recognition of SAR images based on attribute scattering center matching," *Remote Sensing Letters*, vol. 10, no. 2, pp. 186–194, 2019.

[11] K. E. Dungan and L. C. Potter, "Classifying vehicles in wide-angle radar using pyramid match hashing," *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 3, pp. 577–591, 2011.

[12] M. A. Koets and R. L. Moses, "Image domain feature extraction from synthetic aperture imagery," in *Proceedings of the 1999 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings. ICASSP99*, vol. 4, pp. 2319–2322, Phoenix, AZ, USA, March 1999.

[13] S. Lee and S. B. Kim, "Parallel simulated annealing with a greedy algorithm for bayesian network structure learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 32, no. 6, pp. 1157–1166, 2020.

[14] D. Cao, A. Modiri, G. Sureka, and K. Kiasaleh, "DSP implementation of the particle swarm and genetic algorithms for real-time design of thinned array antennas," *IEEE Antennas and Wireless Propagation Letters*, vol. 11, pp. 1170–1173, 2012.

[15] B. Zhang, X. Sun, L. Gao, and L. Yang, "Endmember extraction of hyperspectral remote sensing images based on the discrete particle swarm optimization algorithm," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 11, pp. 4173–4176, 2011.

[16] Q. Al-Tashi, S. J. Abdul Kadir, H. M. Rais, S. Mirjalili, and H. Alhussian, "Binary optimization using hybrid grey wolf optimization for feature selection," *IEEE Access*, vol. 7, pp. 39496–39508, 2019.

[17] H. Liu, B. Jiu, F. Li, and Y. Wang, "Attributed scattering center extraction algorithm based on sparse representation with dictionary refinement," *IEEE Transactions on Antennas and Propagation*, vol. 65, no. 3, pp. 2604–2614, 2017.

[18] Y. Zheng, "Water wave optimization: a new nature inspired metaheuristic," *Computers & Operations Research*, vol. 55, pp. 1–11, 2015.

[19] Y. Lu, J. Lu, and T. Jiang, "Energy-conscious scheduling problem in a flexible job shop using a discrete water wave optimization algorithm," *IEEE Access*, vol. 7, pp. 101561–101574, 2019.

[20] Z. Ren, C. Wang, X. Fan, and Z. Ye, "Fuzzy clustering based on water wave optimization," in *Proceeedings of the 13th International Conference on Computer Science & Education (ICCSE)*, pp. 1–5, Colombo, Sri Lanka, August 2018.

[21] J.-Y. Wu, M.-X. Zhang, X. Wu, and Y.-J. Zheng, "A water wave optimization algorithm for order selection and delivery path optimization for takeaway deliverymen," in *Proceeedings of the 2021 11th International Conference on Information Science and Technology (ICIST)*, pp. 620–662, Chengdu, China, May 2021.

[22] H. Chen, Y. Hou, Q. Luo, Z. Hu, and L. Yan, "Text feature selection based on water wave optimization algorithm," in *Proceeedings of the 2018 Tenth International Conference on Advanced Computational Intelligence (ICACI)*, pp. 546–551, Xiamen, China, March 2018.

[23] S. Zhao, Z. Li, X. Yun, K. Wang, X. Lyu, and B. Liu, "IIR filters designing by water wave optimization," in *Proceeedings of the 2017 13th IEEE International Conference on Control & Automation (ICCA)*, pp. 347–352, Ohrid, Macedonia, July 2017.

[24] K. El-Darymli, P. Mcguire, E. W. Gill, D. Power, and C. Moloney, "Characterization and statistical modeling of phase in single-channel synthetic aperture radar imagery," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 51, no. 3, pp. 2071–2092, 2015.

[25] L. C. Potter, D.-M. Chiang, R. Carriere, and M. J. Gerry, "A GTD-based parametric model for radar scattering," *IEEE Transactions on Antennas and Propagation*, vol. 43, no. 10, pp. 1058–1067, 1995.

[26] A. Uzer, "Improvement of the diffraction coefficient of GTD by using multiplicative calculus," *IEEE Transactions on Antennas and Propagation*, vol. 62, no. 7, pp. 3741–3748, 2014.

Research Article

Construction of Corporate Investment Decision Support Model Based on Deep Learning

Jian-tao Song [ORCID]

School of Finance and Accounting, Yellow River Conservancy Technical Institute, Kaifeng 475004, China

Correspondence should be addressed to Jian-tao Song; songjiantao@yrcti.edu.cn

Aiming at the problem of corporate investment decision support, this paper proposes and constructs a stock quality evaluation model based on deep learning and applies it to the stock quality evaluation of e-commerce enterprises. Firstly, LSTM neural network is used to construct the evaluation and prediction model. Secondly, the evaluation index system is constructed. Finally, the structure and parameters of the model are designed, and the prediction model is tested and evaluated through simulation experiments. The experiments prove that the model is reasonable and feasible, which can provide a reference for investors to make decisions.

1. Introduction

In recent years, with the development of the Internet, the e-commerce industry has developed rapidly, but the development of e-commerce is uncertain, and the industry has great competitiveness and bears great risks. The emergence of stocks has a certain impact on enterprise investment, and the stock market is also uncertain and risky. Therefore, it is of great significance for enterprises and investors to study stocks in the field of e-commerce. If a stock quality evaluation model is established to evaluate e-commerce enterprises reasonably, it is beneficial for the state to supervise the stock market and provide decision support reference for investors. With the development of computer technology, deep neural network is widely used in stock model research, and its structural characteristics are suitable for stock problem analysis. Yang et al. [1] proposed that deep learning should be used to improve the momentum strategy of stock market, and its improvement effect is good; Zeng and Nie [2] proposed a stock recommendation system based on deep learning, which is helpful for stock investment; Zhang et al. [3] proposed using deep learning algorithm to predict stock prices and provide effective reference for investors. On the basis of the above research, this paper puts forward a company investment decision support model based on deep

learning and applies it to the evaluation of e-commerce stock quality, so as to correctly analyze the stock development of e-commerce enterprises and provide reference for investors. Literature [4] aimed at the rapid development of e-commerce today; whether the decision-making is accurate or not directly affects the operating effect of e-commerce companies. Investment decision-making system is an important way for investors to earn profits, and it is a model to support companies. By studying the data of e-commerce and Internet, combined with relevant theories, this paper puts forward a prediction model of corporate decision quality evaluation based on deep learning algorithm. It effectively provides support for investors in decision-making. According to the development direction and current state of domestic and foreign e-commerce investment market, literature [5] puts forward the modeling idea of investment decision quality evaluation index system and completes the investment decision support model by using modeling and simulation tools. Portfolio replication is realized by extracting the features of portfolio model by automatic encoder. Python programming language and library have been applied and implemented in the classification model of deep neural network (DNN) in literature [6]. Compared with traditional classification methods, deep learning classification model has better effect, because its accuracy can

change with the size of data, thus increasing. There are many shortcomings in the traditional portfolio theory in literature [7], such as the construction of optimal socially responsible investment. Nowadays, people are more interested in how to invest in the environment, governance, and other aspects of the company, that is, socially responsible portfolio. Deep Responsibility Portfolio (DRIP) model uses reinforcement learning technology to balance the portfolio repeatedly, so as to achieve socially responsible portfolio. According to the data results, the latter is more competitive and has better social impact than the traditional portfolio model. Literature [8] aims at connecting suppliers and customers and uses the intelligent functions provided by cloud manufacturing platform and various processing schemes generated to effectively help customers choose manufacturing services. Its characteristic is that the more suppliers, the more schemes to choose from and the more complicated the customer's decision. Literature [9] aimed at the unstoppable trend of economic globalization; investors need to examine the reliability of solutions in the process of sound financial strategy, which leads to the intelligent decision support system. It uses mathematical tools of game theory and can consider many factors when making decisions on the same and similar problems. The rapid development of this kind of algorithms and components promotes the development of models based on the properties of bilinear differential games with multiple terminal planes. In literature [10], for real estate development companies must select high-performance and competitive cities as development projects, real estate asset evaluation is the top priority of regional strategy. The primary driving force for real estate investors is whether the region has urban competitiveness to attract all kinds of investment. The application of the comprehensive complex evaluation model to practical cases shows that it can map the multidimensional factor spectrum of the attractiveness of alternative real estate choices. According to literature [11], every investor expects investment to be proportional to income. Clear investment is a longer-term plan, with low or medium risks and low or medium returns. Investors' decisions will be influenced by market and psychology. The concept of investment decision support technology establishes a time frame for each goal, allows people to better choose the tools to achieve the goal, and defines the expected return or risk, investment portfolio, and the steps in the decision-making process of investment analysis. In view of the entertainment and commerce matched with the development of tourism economy, literature [12] analyzes the current situation of enterprises from the three-dimensional perspective of enterprise strategic portfolio and uses deep learning to analyze the impact of different portfolio characteristics on the operation and development of enterprises. It has been proved by practice that there are obvious differences between combination characteristics and single characteristics in enterprise decision-making and income. This paper puts forward the analysis of the impact of different combination policies on macroeconomy, and the forecast preparation rate of combination is as high as 98.57%. In view of the complexity of investment decision analysis application fields and expert

systems and the inability to provide accurate evaluation, literature [13] proposes a large-scale project investment decision system based on natural language evaluation and applies it to cloud model. Furthermore, a multiprocessing and multigranularity language decision-making system is proposed to support investment decision-making. The current situation of the research is explained, only the neural network is applied to various scenarios, and no industrialization decision is studied. The applications of these algorithms are based on a single neural network and have poor prediction results. Deep learning model is of important research significance for corporate investment, which can reduce investment risk and improve economic efficiency. This paper first introduces the LSTM algorithm, explains the application of deep network model in stock quality evaluation, and finally studies the stock forecast to analyze the investment decision behavior.

2. LSTM Neural Network Algorithm

BP neural network algorithm, RNN model, and LSTM algorithm can all be used in the training of evaluation and prediction model, in which BP neural network algorithm can only input the original data network one by one and cannot show the time series relationship between data during training and can only carry out error back-propagation layer by layer and cannot minimize the convergence of the network when dealing with gradient descent [14]. Although RNN model can input multiple original data at the same time and consider the time series relationship, it cannot solve the problem of gradient disappearance; LSTM algorithm is a special recurrent neural network. The information of its hidden layer can be transmitted to the output end or the hidden layer at the next time. LSTM algorithm can also update the weight continuously through input, output, and forgetting operations to avoid gradient expansion and disappearance. Compared with BP neural network algorithm, RNN model, and LSTM algorithm, the advantages of LSTM are that it has long-time memory and is simple to implement; the problems of gradient vanishing and gradient explosion existing during long sequence training are solved. This paper chooses LSTM algorithm to build the model, and its model structure is shown in Figures 1 and 2.

The training steps of LSTM algorithm are divided into four steps:

> Step 1: Calculate the output of LSTM hidden layer. Let w, o_t, c, and f denote output gate, cell state, forgetting gate, and input gate, respectively, let b denote offset, let w denote weight coefficient matrix, and T (tan$h$) and S (sigmoid) are activation functions; then the expressions of forgetting gate and input layer are shown in the two following equations (1) and (2) [15]:

$$f_t = \sigma\left(w_f\left[h_{t-1}, x_t\right] + b\right), \quad f_t \in [0,1]. \tag{1}$$

$$\begin{aligned} i_t &= \sigma\left(w_i\left[h_{t-1}, x_t\right] + b_i\right), \\ c1_t &= \tan h\left(w_c\left[h_{t-1}, x_t\right] + b_c\right). \end{aligned} \tag{2}$$

Update the cell state:
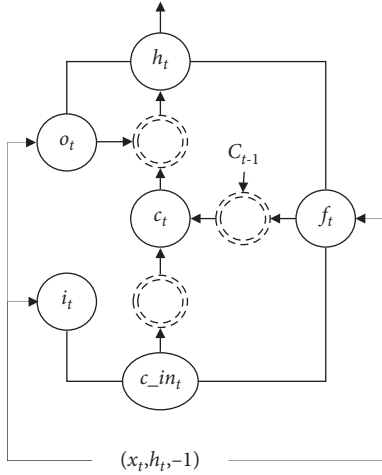
FIGURE 1: LSTM hidden layer structure diagram.
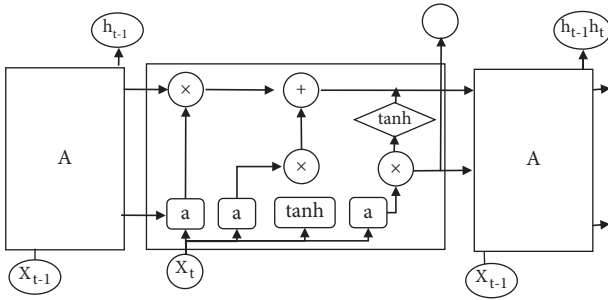
FIGURE 2: LSTM algorithm structure diagram.

$$c_t = f_t * c_{t-1} + i_t * c1_t. \tag{3}$$

The model output is obtained by activating the function:

$$o_t = \sigma\left(w_o\left[h_{t-1}, x_t\right] + b\right),$$
$$h_t = o_t * \tan h\left(c_t\right). \tag{4}$$

Step 2: The error δ of each LSTM cell is calculated by back-propagation, including the error terms in time and network level. Let $\delta\, t$ denote the error term at time t, let $\delta\, t - 1$ denote the error term at time $t - 1$, and let $\delta\, k$ denote the error term at time k; then the delay reverse transmission of the error term is expressed as [16]

$$\delta_t^T = \frac{\partial E}{\partial h_t},$$

$$\delta_{t-1}^T = \frac{\partial E}{\partial h_{t-1}} \tag{5}$$

$$\delta_k^T = \prod_{j=k}^{t-1} \delta_{o,j}^T w_{oh} + \delta_{f,j}^T w_{fh} + \delta_{i,j}^T w_{ch}.$$

Error terms are reversely transmitted to the upper layer along the network:

$$\delta = \left(\delta_{f,t}^T w_{fx} + \delta_{i,t}^T w_{ix} + \delta_{o,t}^T w_{ox}\right) \cdot f'\left(net_t^{l-1}\right) \tag{6}$$

Step 3: Calculate the weight gradient according to the error [17].

$$\frac{\partial E}{\partial w_{oh}} = \sum_{j=1}^{t} \delta_{o,j} h_{j-1}^T,$$

$$\frac{\partial E}{\partial w_{fh}} = \sum_{j=1}^{t} \delta_{f,j} h_{j-1}^T,$$

$$\frac{\partial E}{\partial w_{ih}} = \sum_{j=t}^{t} \delta_{i,j} h_{j-1}^T,$$

$$\frac{\partial E}{\partial w_{ch}} = \sum_{j=1}^{t} \delta_{c,j} h_{j-1}^T, \tag{7}$$

$$\frac{\partial E}{\partial w_{o}} = \sum_{j=1}^{t} \delta_{o,j},$$

$$\frac{\partial E}{\partial w_{i}} = \sum_{j=1}^{t} \delta_{i,j},$$

$$\frac{\partial E}{\partial w_{f}} = \sum_{j=1}^{t} \delta_{f,j},$$

w is the error with the fastest value at the point. In the decision of the company in the multi-index decision for error calculation, it finds out the index with the fastest change.

Step 4: Update the weights based on the previous step.

3. Application of Deep Network Model in Stock Quality Evaluation

3.1. Data Sources. The data of this paper comes from an information website. The stocks of 500 e-commerce companies are selected as experimental samples, and the sample stocks are divided into excellent stocks, common stocks, and loss stocks, with a ratio of 1 : 1 : 1. The price-earnings ratio on the last day of 16 years is taken as the basis for differentiation. The sample stock annual reports from 2014 to 2016 were downloaded, and 30 indicators and 12,946 pieces of data were found in these annual reports. Excel 2017, Python data processing software, and TensorFlow system framework were adopted.

3.2. Indicator Selection. At present, there are many stock evaluation indexes in the market, such as net asset income and price-earnings ratio. Because stocks have risks and volatility, they can only be used as reference. In this paper, the price-earnings ratio is used as a reference to classify the sample data. In order to find the hidden connection of the company and grasp the development status of the enterprise,

the existing indicators are not added to the evaluation system. When selecting indicators, we should follow compliance, scientificity, quantification, and feasibility so as to correctly analyze the stock quality of enterprises. The evaluation indicators selected in this paper include the following: asset-liability ratio, earnings per share, growth rate of operating income, proportion of R&D investment, total share capital, number of circulating shares, issue price, number of senior executives, and 30 other indicators [18].

3.3. Sample Data Processing

3.3.1. Quantification of Nonnumerical Indicators. There are nonnumerical indexes in the index system, which must be transformed into numerical indexes if they are to be identified and processed by computer. In statistics, Likert scale is often used to transform text description into digital expression. There is no unified quantitative form in other disciplines, and the quantitative method is usually determined according to the characteristics and experience of indicators. Some indicators in this paper are expressed in the form of numbers and words, and there are three dimensions of expression. After consulting relevant data, this paper chooses the maximum-minimum standardization method to convert these indicators into numerical indicators. The specific operation is as follows: all nonnumerical indicators are standardized by maximum-minimum, and three dimensions are added to obtain one-dimensional index data.

3.3.2. Dimension Reduction Processing of Index Data. The correlation of all evaluation indexes is strong and weak. Dimension reduction is applied to the indexes; that is, the indexes with weak correlation are removed, which can improve the aggregation degree of the model and help to accurately evaluate the stock situation of enterprises. Let r indicate the correlation of indexes; when $r < 0.1$, it is not correlated, $0.1 < r < 0.3$ indicates weak correlation, $0.3 < r < 0.5$ indicates medium correlation, and $0.5 < r < 1$ indicates strong correlation. In this paper, PCA method is used to reduce the dimension of index data, and its steps are divided into three steps. First, after processing the sample data, it is converted into matrix form; the second is to calculate the index correlation coefficient matrix; the third is to show the index $r > 0.1$, and calculate the variance percentage [19].

From this, we can get 18 of the 30 indexes with $R > 0.3$, and then remove the indexes with $R < 0.3$. By calculating the variance percentage of these 18 indexes, all of them can reach more than 95%, which shows that PCA dimension reduction is feasible; see Figure 3.

3.3.3. Standardization of Indicator Data. Because the basic flow of the network model is to set the initial weight and input variables as the input of the hidden layer, get the output of the hidden layer through the activation function, and take it as the input of the next hidden layer, and operate repeatedly until the final network output. If the input data is
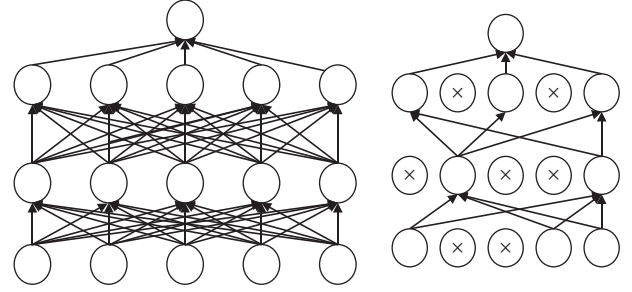
FIGURE 3: Schematic diagram of neural network after dimensionality reduction.

too large, under the action of activation function, the output data of network nodes may fall into the saturation interval, and the output of neurons will show the maximum or minimum value, resulting in the derivative gradually tending to 0. In the process of error back-propagation, the modified parameter value is very small, the learning rate is slow, and the network cannot reach the convergence state. Therefore, this paper normalizes the data to prevent the paralysis of neural network caused by too large sample data. There are three commonly used methods for data normalization:

(1) Logarithmic function transformation method, the formula of which is [20]

$$X_i' = \log x_i. \tag{8}$$

(2) Inverse cotangent function transformation method, the formula of which is

$$X_i' = \frac{2 \tan^{-1} x_i}{\pi}. \tag{9}$$

(3) Linear function transformation method, the formula of which is

$$X_i' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{10}$$

According to the characteristics of samples, this paper adopts the linear function transformation method. Its principle is to calculate the maximum and minimum values of the original data and subtract the original data x from the minimum value to correspond to [0, 1].

3.3.4. Discretization of Indicator Data. The purpose of discretization of index data is to evenly distribute the values of each piece of index data in the data samples, thus reducing the risk of overfitting and making the model have strong stability. Discretization is robust and can deal with abnormal data in sample data. Therefore, discretization method is usually used to deal with multiple continuous data in sample data. There are three main methods of discretization, namely, optimization and isometric and equal frequency discretization, among which optimization includes information gain and chi-square test, isometric means that the sample data is divided into N equal parts and the interval of

each sample data is the same, and equal frequency means that the frequency of divided sample data is the same.

By observing and analyzing the original data, It is found that the original data is too large, and there are many continuous data and abnormal data, which will lead to the inability of the training model to fit the actual results. Therefore, it is necessary to discretize and standardize the sample data and train the processed data at the same time, so as to get a better model.

In this paper, the continuous data of 15 pieces of index data are discretized, and the stock quality types are divided into three grades by equidistance dispersion method, with values of 0, 1, and 2, respectively. After the data is standardized and discretized, the data is input into the network and then trained several times according to the training steps of LSTM algorithm, so that the number of hidden layer nodes is determined to be 50, and the whole processing process is realized by Python language.

When evaluating the stock quality of enterprises, 18 indicators are selected as input variables: dividend, asset liability ratio, three-year earnings per share from 2014 to 2016, three-year reorganization, issue price, shareholding ratio of shareholders, listing time, number of legal proceedings, key pollutant discharge units, social responsibility report, total share capital, and number of circulating shares. Select 300 groups of training sample data, 200 groups of test sample data, and stock category as the evaluation output.

3.4. Establishment of LSTM Neural Network Evaluation and Prediction Model

3.4.1. Structural Design of Prediction Model.
The input layer and output layer of neural network model are set according to data characteristics and targets, while other parameters are set according to experience or experiment, without fixed procedures and steps, which reflects the freedom of network model structure design. The hierarchy of neural network is generally set according to experience, and the research shows that the hierarchy is small, and its network popularization is strong, which is convenient for extracting features. The hierarchy is large, the network error is small, and the accuracy is high, but, at the same time, it also increases the complexity, the training speed, and time of the model, which may lead to overfitting phenomenon. Therefore, the number of hierarchies can be set according to the effect that the model needs to achieve. Generally speaking, in order to make the model have higher accuracy, we choose more than three layers of networks, and this paper chooses 10 layers of LSTM network to build the stock quality evaluation prediction model.

The sample data contains 500 stocks, and each stock has 18 pieces of data. According to the parameter setting basis of the input layer, the output layer, and the hidden layer, the input layer node number m is set to 18, the output layer point number n is set to 3, and the trial-and-error method is adopted to calculate the initial value of the hidden layer node number, assuming that s represents the hidden layer node number and a represents a constant, and the trial-and-error method expression is shown in formula (11), and the hidden layer node number is calculated to be 50.

$$
\begin{aligned}
s &= \log_2 m, \\
s &= \sqrt{mn}; \\
s &= \sqrt{m+n} + a.
\end{aligned}
\tag{11}
$$

3.4.2. Selection of Activation Function.
Tanh, ReLU, and sigmoid are all active functions. When the neural network is small, Tanh has the best expressive force, while ReLU is the most widely used active function at present. Compared with sigmoid function and Tanh function, the gradient of ReLU function has no saturation force, but it can quickly adjust parameters and alleviate the gradient disappearance process, and the calculation speed is faster. Therefore, ReLU function is selected as the activation function, and its expression is as follows (Figure 4):

$$
f(x) = \max(0, x).
\tag{12}
$$

4. Simulation Experiment

4.1. Original Data Processing Results.
The original data are standardized and discretized, and the results are given in Tables 1 and 2.

The stock quality type is divided into three grades with the values of 0, 1, and 2. The value range is evenly divided into 3 equal parts, with equal spacing. It mainly helps LSTM to realize its process and can submit the prediction effect.

Take the above data as network input, and train it repeatedly according to LSTM training steps, and get the number of hidden layer nodes as 50. The whole simulation experiment is completed by Python language.

4.2. The Results are Implemented in Python.
Python is commonly used open-source software, which can combine languages C, C++, Java, and so forth and put packages not written by Python language into Python programming platform, which can identify and read packages. Python itself also has packages and extension libraries, including various functions inside, which can perform fast operations on numeric arrays.

4.2.1. Simulation Experiment Steps.
There are seven steps in the simulation experiment. The first step is to define the file path of storage model, training set and test set; the second step is to define an LSTM single layer with a "switch"; the third step is to set up a multilayer LSTM network; the fourth step is to use dynamic_rnn function to make tf build a multilayer LSTM network and define the output; the fifth step is to define the loss function and train the neural network through the optimizer of tf.train; the sixth step is to solve the problem of overfitting; and the seventh step is to train LSTM neural network according to parameters.
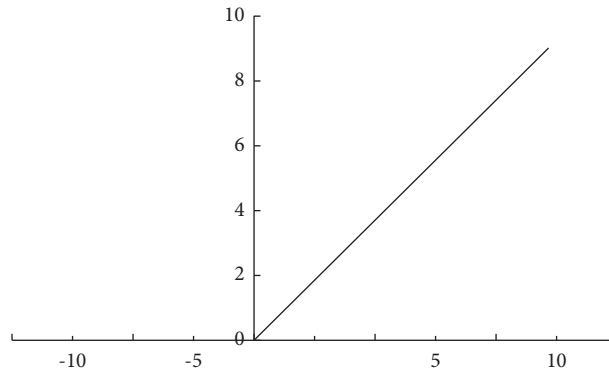
FIGURE 4: ReLU activation function.

TABLE 1: Partial data after standardization.

Code	Industry	Capital	Ating cap	Property	Oclcholder	To market	Price	Add stock	Profit
1	0.667	0.032	0.035	1	0.302	0.333	0.052	1	0.129
2	0.33	0.001	0.001	1	0.329	0.250	0.055	0	0.005
3	1.000	0.001	0.002	0	0.110	0.917	0.193	1	0.046
4	0.667	0.010	0.012	0	0.410	0.750	0.177	1	0.355
5	0.333	0.001	0.001	0	0.327	0.917	0.225	0	0.074
6	0.333	0.088	0.088	1	0.521	0.417	0.029	1	0.027
7	0.333	0.000	0.000	1	0.224	0.250	0.072	0	0.074
8	0.333	0.010	0.012	0	0.390	0.833	0.128	1	0.355
9	0.333	0.005	0.006	1	0.342	0.792	0.541	0	0.428
10	0.333	0.023	0.024	1	0.182	0.250	0.014	0	0.775
11	0.333	0.002	0.002	0	0.426	0.542	0.030	0	0.022
12	0.667	0.005	0.007	1	0.350	0.583	0.018	1	0.077
13	0.667	0.186	0.186	1	0.252	0.833	0.019	0	0.027
14	0.667	0.002	0.003	1	0.503	0.250	0.036	1	0.059
15	0.667	0.006	0.010	0	0.415	0.792	0.528	0	0.01
16	0.667	0.002	0.003	1	0.382	0.167	0.064	1	0.062
17	0.333	0.008	0.008	0	0.490	0.917	0.051	0	0.034
18	0.333	0.000	0.001	0	0.384	1.000	0.191	1	0.607

TABLE 2: Partial data after discretization.

General_capital	Floating_capital	Stock_holder	Time_to_market	Price	2014_profit	2015_profit
0	0	0	0	0	0	0
0	0	1	0	0	0	1
0	0	0	2	0	0	0
0	0	1	2	0	1	0
0	0	1	2	0	0	0

4.2.2. The Running Result of the Data Model after Standardization and Discretization. After 5 times of standardization, the prediction accuracy of the model is only 46%, and when the training is increased to 15 times, the prediction accuracy can reach 70%. Then the test set data is input into the model, and the accuracy rate is 77%, which shows that after the data is standardized, the simulation accuracy is not high. The fitting degree of the model after discretization can reach 92%, which is higher than that after standardization, which shows that the training model has high reliability after discretization of the data.

4.2.3. Actual Forecast Results. Selecting the stocks of 25 companies, the index data as the network input into the model, the results show that the prediction accuracy of the model is about 75%, indicating that about 19 stocks have the accuracy of quality prediction. Compared with other related models, the stock evaluation prediction model in this paper improves the evaluation accuracy, and the model is feasible.

Through the above analysis, it is known that there are 18 indicators that have strong correlation with stock quality evaluation; these 18 pieces of index data are input into the stock quality evaluation model, and, after repeated training

of the model, the stock quality evaluation and prediction results are obtained, and the results are accurate. According to the prediction results, the enterprise is comprehensively analyzed to evaluate the development status and investment potential of the enterprise, thus providing a reference for investors and deciding whether to give decision support.

5. Conclusion

This paper designs and constructs a company investment decision support model based on deep learning and applies it to the evaluation of enterprise stock quality. In this study, LSTM multilayer neural network is used to establish the model, and the parameters in the input layer, output layer, and hidden layer of the network are set reasonably. At the same time, the model is trained repeatedly through simulation experiments, and the trained model is used to evaluate the test sample data, and the evaluation effect is good, which verifies the feasibility of the model. Due to my limited knowledge and technology, this paper needs further in-depth study. The existing problems in the present stage are as follows: First, when large-scale data analysis is applied to the company's investment decision support model, it takes a certain amount of time to verify its economic feasibility. Second, deep learning requires large-scale server operation, especially for real-time data acquisition and time effectiveness of analysis. The focus of future research work can solve the timeliness of data analysis, analyze all kinds of data in parallel, and then conduct joint analysis.

Data Availability

The experimental data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The author declares that there are no conflicts of interest regarding this work.

References

[1] Q. Yang, X. Wang, and Q. Tang, "Research on momentum strategy of Chinese a-share market based on deep learning method," *Journal of Guangxi University of Finance and Economics*, vol. 32, no. 3, pp. 36–55, 2019.

[2] A. Zeng and W. Nie, "Stock recommendation system based on deep bidirectional LSTM," *Computer Science*, vol. 46, no. 10, pp. 84–89, 2019.

[3] Q. Zhang, D. Yan, and J. Han, "Research on stock price forecasting based on deep learning and decomposition algorithm," *Computer Engineering and Application*, vol. 57, no. 5, pp. 56–64, 2021.

[4] Z. Lei, "Research and analysis of deep learning algorithms for investment decision support model in electronic commerce," *Electronic Commerce Research*, vol. 20, no. 2, pp. 275–295, 2020.

[5] L. Hou, "Decision support model of e-commerce enterprise investment based on deep learning algorithm," *Information Systems and E-Business Management*, 2021.

[6] S. Bayraci and O. Susuz, "A deep neural network (DNN) based classification model in application to loan default prediction," *Theoretical and Applied Economics*, vol. 621, no. 4, pp. 75–84, 2019.

[7] N. N. Y. Vo, X. He, S. Liu, and G. Xu, "Deep learning for decision making and the optimization of socially responsible investments and portfolio," *Decision Support Systems*, vol. 124, Article ID 113097, 2019.

[8] A. Simeone, A. Caggiano, B. Deng, and L. Boun, "A deep learning based-decision support tool for solution recommendation in cloud manufacturing platforms," *Procedia CIRP*, vol. 86, pp. 68–73, 2019.

[9] V. Lakhno, V. Malyukov, B. Akhmetov, N. Gerasymchuk, H. Mohylnyi, and P. Kravchuk, "Decision support model for assessing projects by a group of investors with regards of multi-factors," in *Artificial Intelligence and Bioinspired Computational Methods. CSOC 2020*, R. Silhavy, Ed., vol. 1225, pp. 1–10, Springer, Cham, 2020.

[10] V. D. Giudice, P. D. Paola, T. Francesca, P. J. Nijkamp, and A. Shapira, "Real estate investment choices and decision support systems," *Sustainability*, vol. 11, no. 11, pp. 1–18, 2019.

[11] S. Gupta and L. N. Das, "Optimal investment decision model based on simplex algorithm with variable optimal value evaluation process," *ICACM 2018. Springer Proceedings in Mathematics & Statistics*, vol. 320, pp. 433–443, Springer, Singapore, 2020.

[12] B. Li, "Construction of business strategic planning structure model based on deep learning algorithm," *Information Systems and E-Business Management*, vol. 25, pp. 1–18, 2021.

[13] X. Mi, H. Liao, and X. J. Zeng, "Investment decision analysis of international megaprojects based on cognitive linguistic cloud models," *International Journal of Strategic Property Management*, vol. 24, no. 6, pp. 414–427, 2020.

[14] y. Fu, "Intelligent financial decision making: integration of soft computing and machine learning," *Financial Research*, no. 6, pp. 31–36, 2019.

[15] Q. Yang, G. Zhang, and B. Ma, "Management semantics and the risk of future stock price crash: an analysis of mediating effect based on investor sentiment," *Journal of Zhongnan University of Economics and Law*, no. 1, pp. 26–36, 2020.

[16] Q. Han, G. Cai, and Z. Zhao, "Research on the growth evaluation method of small and medium-sized enterprises based on big data technology," *Statistics and Information Forum*, vol. 34, no. 7, pp. 123–128, 2019.

[17] X. Tang, M. Tan, and S. Li, "Research on the model construction and implementation of enterprise bankruptcy prediction system," *Journal of Information Science*, vol. 38, no. 10, pp. 1051–1065, 2019.

[18] H. Zhang, H. Shen, and Y. Liu, "Research on multi factor quantitative stock selection based on self attention neural network," *Mathematical Statistics and Management*, vol. 39, no. 3, pp. 556–570, 2020.

[19] X. Niu and W. Zonghao, "Research on multi factor strategy of Chinese stock market based on Integrated Learning," *Journal of Guizhou Party School*, vol. 20, no. 4, pp. 22–37, 2020.

[20] T. Lin, Q. Zhang, and X. Qi, "Securities big data analysis and research," *Computer Technology and Development*, vol. 30, no. 10, pp. 179–186, 2020.

Research Article

Semantic Extraction of Basketball Game Video Combining Domain Knowledge and In-Depth Features

Yufeng Du,[1] **Quan Zhao** ⓘ **,**[2] **and Xiaochun Lu**[1]

[1]*Ministry of Sports, Xi'an Academy of Fine Arts, Xi'an 710065, Shaanxi, China*
[2]*School of Sciences, Xi'an Technological University, Xi'an 710021, Shaanxi, China*

Correspondence should be addressed to Quan Zhao; zhaoquan@xatu.edu.cn

The team sports game video features complex background, fast target movement, and mutual occlusion between targets, which poses great challenges to multiperson collaborative video analysis. This paper proposes a video semantic extraction method that integrates domain knowledge and in-depth features, which can be applied to the analysis of a multiperson collaborative basketball game video, where the semantic event is modeled as an adversarial relationship between two teams of players. We first designed a scheme that combines a dual-stream network and learnable spatiotemporal feature aggregation, which can be used for end-to-end training of video semantic extraction to bridge the gap between low-level features and high-level semantic events. Then, an algorithm based on the knowledge from different video sources is proposed to extract the action semantics. The algorithm gathers local convolutional features in the entire space-time range, which can be used to track the ball/shooter/hoop to realize automatic semantic extraction of basketball game videos. Experiments show that the scheme proposed in this paper can effectively identify the four categories of short, medium, long, free throw, and scoring events and the semantics of athletes' actions based on the video footage of the basketball game.

1. Introduction

In recent years, the spectacle and attention of various sports events have been increasing, and a large number of sports events are broadcasted and shared on the Internet in the form of videos, and sports videos have become an efficient and irreplaceable force for information dissemination on the Internet [1]. With the explosive growth of sport video data on the Internet platform, how to manage this information scientifically has become a major challenge in the current era of big data.

The traditional method based on the manual annotation to categorize and integrate the events occurring in a video not only requires a lot of human resources but also has a high error rate because manual annotation is easily influenced by human subjective factors and cannot meet the various needs of different people [2]. So a reasonable semantic analysis of video content has important theoretical significance and broad application prospects. There are multiplayer

collaborative confrontation videos in sports such as basketball, where sports videos have difficulties such as serious obscuration, fast movement speed, and complex action changes [3].

On the other hand, the current sports video dataset is quite large, containing tens of thousands of videos and hundreds of classes, and there is an urgent need to address how to use them effectively. Moreover, these classes may be specific to certain domains, such as sport, and the dataset may contain noisy labels [4]. Another key open question is as follows: what is an appropriate spatiotemporal representation pattern of the video? Recent video representations for action recognition are mainly based on two different CNN architectures: (1) 3D spatiotemporal convolution, which may learn complex spatiotemporal dependencies but is difficult to scale in terms of recognition performance; (2) dual-stream architecture, which decomposes the video into motion and appearance streams and trains separate CNNs for each stream, eventually fusing the outputs [5]. While

both approaches have made rapid progress, both stream structures typically outperform spatiotemporal convolution because they can easily exploit the new ultra-deep structure and pretrained models for still image classification.

However, the dual-stream structure largely ignores the long-term temporal structure of the video and essentially learns classifiers that operate on single frames or short blocks of a few (up to 10) frames [6], potentially forcing consistency in classification scores over different segments of the video [7]. During testing, T uniformly sampled frames are independently classified, and the classification scores are averaged to obtain a final prediction. This approach raises the issue of whether such temporal averaging can model the complex spatiotemporal structure of human behavior, which is exaggerated when multiple action classes share the same subactions. For example, consider the intricate combination of "basketball shooting" shown in Figure 1. Since there are only a few consecutive video frames, it can be easily confused with other actions, such as "running," "dribbling," "jumping," and "throwing." "Using postfusion or averaging is not the right best solution, because it requires that frames belonging to the same subaction are assigned to multiple classes. What we need is a worldwide feature descriptor for the video that collects evidence about the appearance of the scene and the movement of the person throughout the video without having to uniquely assign each frame to a single action.

In order to address this problem, this paper analyzes basketball game videos in multiplayer collaborative sports videos, studies the semantic event classification problem of basketball game videos, and aims to propose a deep learning-based event analysis method for basketball sports videos to realize the automatic analysis of basketball videos, so as to assist coaches in tactics formulation, assist players in action analysis, and assist video viewers in rapid search for interesting video segments, and promote the rapid development of basketball and even sports. We develop an end-to-end trainable video-level representation that aggregates convolutional scripts to different parts of the imaged scene and the entire time span of the video.

In this paper, we take a basketball game video in a collaborative multiplayer sports video as an example for semantic event analysis of sports video. The global motion features, group motion features, and individual pose features of the video are extracted separately to realize the eventual analysis of basketball game video with multi-feature fusion. Firstly, basketball video events are classified by combining sports domain knowledge, and secondly, global and group motion patterns of players in basketball games are expressed, in which the spatiotemporal extension of NetVLAD [8] aggregation layer in deep learning [9] is used, and NetVLAD can be used well in still images for instance-level recognition tasks. For our extension of NetVLAD to videos, we also address the following issues. First, aggregation of frame-level features across time into video-level representations selects the best method for aggregation at different levels of the network from output probabilities to separate layers of convolutional descriptors
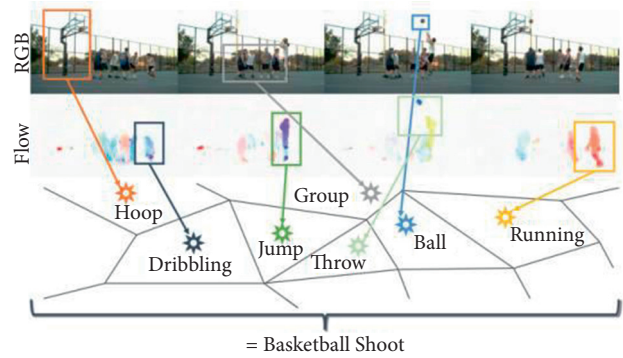
FIGURE 1: Schematic representation of basketball video actions (the spatiotemporal aggregation of a set of action primitives on the appearance and motion flow of the video in our method. For example, basketball shoot can be represented as a collection of appearance features corresponding to "player," "ball," and "basketball hoop" and motion features corresponding to "run," "jump," and "shoot").

[10] and shows that aggregation picks through the last layer of convolutional descriptors to perform best. Second, different strategies for aggregating features from spatial and temporal streams for best-combining signals from different streams in a multistream architecture show some surprising results as the best results can be obtained by aggregating spatial and temporal streams into their separate individual video-level representations. We support our study with quantitative experimental results while providing intuition for obtaining the results.

The contributions of this paper are as follows:

(1) We first designed a scheme that combines dual-stream network and learnable spatiotemporal feature aggregation, which can be used for end-to-end training of video semantic extraction to bridge the gap between low-level features and high-level semantic events.

(2) Then, an algorithm based on the knowledge from different video sources is proposed to extract the action semantics. The algorithm gathers local convolutional features in the entire space-time range, which can be used to track the ball/shooter/hoop to realize automatic semantic extraction of basketball game videos.

(3) Experiments show that the scheme proposed in this paper can effectively identify the four categories of short, medium, long, free throw, and scoring events and the semantics of athletes' actions based on the video footage of the basketball game.

2. Related Work

The purpose of basketball sports video analysis is to extract key information from video content for play summaries, computer-aided recommendations, and content insertion, among others. In recent years, the content oriented to basketball sports video analysis mainly includes player detection, action recognition, and event detection [11].

2.1. Player Detection. The task of player detection is to automatically detect the key target of the basketball video and give its location or area in the form of a candidate box search, as shown in Figure 1. In basketball sports videos, the foreground target plays a crucial role in the event scene, and the location and behavior of this target are critical in high-level semantic event analysis or sports game and tactics analysis. For example, the work in [12] uses low-level feature information while combining domain knowledge of basketball to identify specific events and implement segmentation of a given video into wide-angle and close-up shots. The study in [13] proposes a hybrid approach of manual observation and machine learning to integrate statistics in a logical rule model for highlight clip detection, where clips of player breaks are cleverly utilized to aid in event recognition. The work in [14] is based on the motion characteristics of basketball to achieve content understanding, first based on a specific region enhancement method to achieve the segmentation of the ball in the video and then based on matching key points to achieve the reconstruction of the 3D scene to establish the motion trajectory of basketball. The work in [15] focused on analyzing free-throw events in basketball videos by detecting the principal locations of free throws and calculating the hit rate of free throws by analyzing the state of players. The methods mentioned above are based on low-level information, including visual information from human observation, etc. Traditional methods are limited to matching and classifying in the low-level feature space of video frames, which is not only inefficient but also of far lower quality and far less differentiation of intraclass features.

2.2. Action Recognition. Action recognition in video means recognizing the motion of the person in the video. Extracting features of human motion from video sequences is also a crucial part of video analysis. In the telephoto case, only the trajectory of the motion target as a whole is generally extracted for analysis, but in a close-up activity like sports video, a specific expression of the target's limbs or torso needs to be extracted from the video sequence frames, i.e., action recognition [16].

Traditional action recognition methods have static features based on shape and contour, and dynamic features based on optical flow and motion information. [17] are to ensure action recognition by doing shape matching of key video frames extracted from the video with the saved original action. The study in [18] uses fuzzy optical flow features to calculate the optical flow field of the target in the video and achieves online action recognition by template matching. The study in [19] proposed expressing the action recognition of the target by the features of key point motion history. The features extracted by traditional methods are manually designed features, and thus, the performance of action recognition is not high.

In recent years, with the development of deep learning, some new solutions have emerged in the field of basketball video. The work in [20] developed a novel recursive convolutional neural network for large-scale visual learning that is not specific to basketball video analysis but includes basketball videos in the dataset. The work in [21]

analyzed a basketball game with a single shooter in life based on deep learning. The method extracts the spatial features of sampled frames as well as the motion information in the video by VGG-16 [22] and subsequently achieves the fusion of spatial and temporal features by a layer called the aggregation layer pooling in this paper, and the whole network is end-to-end structured. The work in [23] implemented basketball tracking using deep learning, which is a method to determine whether a 3-point shot is successful or not by RNN, while the model can predict the trajectory of the basketball very well. The work in [24] implemented NBA tactical analysis based on deep learning, firstly by locating the players as well as the relative positions of the ball, and secondly by implementing tactical classification based on convolutional networks. Extraction of key player regions often causes greater interference due to the existence of certain occlusions between players, and the algorithm has a complex framework and high computational complexity. The player occlusion problem is very serious during basketball games, which tends to reduce the ability to track key players, and coupled with the complex background of basketball videos, this will lead to serious target confusion while the computational effort will increase dramatically [25, 26].

2.3. Event Detection. Event detection means giving the start and endpoints of semantic events in the video. Starting from when a player is ready to shoot from the free-throw line to when the ball goes in the box and land is a free throw event in a basketball game video. The study in [27] implements event detection based on CNN extended network by introducing cascaded CNN networks to express local motion information while incorporating trajectory information to achieve event analysis of surveillance videos; similar to action recognition, event analysis can be implemented based on a dual-stream CNN model, which introduces both spatial information and temporal features from a global perspective. The work in [28] tracks all players while learning the key players in basketball events based on the attention mechanism and achieves the learning of player sequence information by Bi-LSTM, the learning of global video frame sequence information by LSTM, and finally the combined expression of local and global information to achieve event detection. In [29], firstly, player tracking is achieved based on keyframes, secondly, player state changes are drawn by LSTM, and player information is fused by different kinds of team division methods to achieve event representation.

3. The Proposed Dual-Stream Network

We just try to learn a video representation, which is trained for end-to-end action recognition. To achieve this, we introduce architecture as shown in Figure 2. In detail, we sample frames from the entire video and convert aggregated features from the appearance (RGB) and motion (flow) streams into a fixed-length vector at the exclusive video level using the vocabulary "action words" [30]. This representation is subsequently passed to the classifier, which outputs a final classification score. The
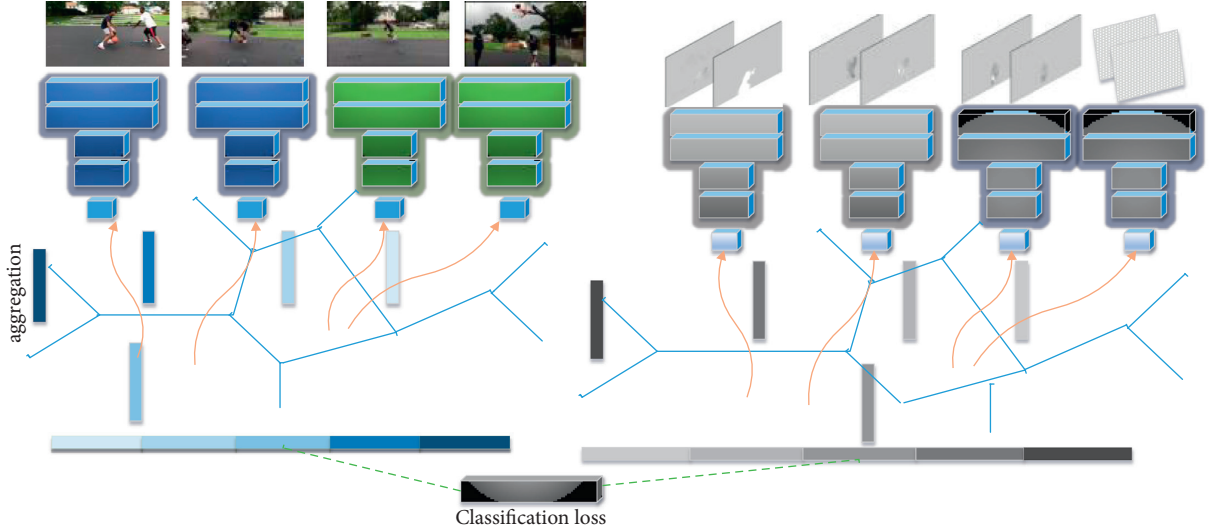
FIGURE 2: Our network architecture. (We use a standard CNN architecture (VGG-16) to extract features from sampled appearance and motion frames in the video. These features are then brought together across space and time using a pooling layer of this paper's aggregation layer, which can be trained end-to-end and has a classification loss.)

parameters of the aggregation layer—a set of "action words"—are learned in a differentiated manner together with the feature extractor to accomplish the ultimate task of action classification.

In the following, we first describe the reasonable spatiotemporal aggregation layer [31]. We then discuss the possible positions of the aggregation layer in the overall architecture and strategy for combining the emergent and motion streams. Finally, we give implementation details. We more specifically describe our improvements to ActionVLAD. (1) We have improved the fusion method, as described in the introduction of the paper. (2) We changed the network structure. For example, we added CNN and FC to the network and shared their parameters based on the idea of ResNet. (3) The application scenarios of the designed model are different.

3.1. Dual-Stream Aggregation Network Model.

Consider $x_{i,t} \in R^D$, the D-dimensional local descriptors extracted from the spatial locations $i \in \{1 \ldots N\}$ and frames $t \in \{1 \ldots T\}$ of the video. We want to aggregate these descriptors spatially and temporally throughout the video while preserving their information content. This is achieved by first dividing the descriptor space R^D into K cells using a vocabulary of K "action words" represented by anchor points $\{c_k\}$ (Figure 2, note that different lines represent different aggregation states). Each video descriptor x_i, T is then assigned to a cell and the difference between descriptors and anchor points is recorded using a residual vector $x_{it} - c_k$. The difference vectors [32] in the whole video are then summed as

$$V[j,k] = \sum_{t=1}^{T} \sum_{i=1}^{N} \underbrace{\sum_{k_l} e^{-\alpha \|x_{it} - c_k\|^2}}_{\text{Soft-assignment}} \underbrace{(x_{it}[j] - c_k[j])}_{\text{Residual}} . \quad (1)$$

Here, $x_{it}[j]$ and $c_k[j]$ are the j-th component of the descriptor vector x_{it} and anchor c_k, respectively, and α is a

tunable hyperparameter. The first term in (1) represents the soft assignment of descriptor x_{it} to unit k. The second term represents $x_{it}[j] - c_k[j]$, the residual between the descriptor and the anchor point of unit k. The two summation operators denote aggregation in time and space, respectively. The output is a matrix V where the kth column $V[-,k]$ represents the aggregation descriptor in the k-th cell. The columns of the matrix are then internally regularized [33], stacked, and L2 regularized [5] for a single descriptor of the entire video $v \in R^{\text{KD}}$.

Clearly known residual vectors record the differences between the extracted descriptors and the "typical actions" (or subactions) represented by the anchors. The residual vectors are then aggregated across the video by computing the sum of the residual vectors within each cell. Critically, all parameters, including the feature extractor, action words, and classifier, are learned jointly from the data in an end-to-end manner to better distinguish the target actions. This is because the spatiotemporal aggregation described in (1) is microscopic and allows backpropagation of error gradients to lower layers of the network. The aggregation used in this paper is a spatiotemporal extension of the NetVLAD [8] aggregation, while we introduce a sum across time T.

3.2. Aggregation Layer.

Theoretically, the spatiotemporal aggregation layers described above can be placed at any level of the network to bring the corresponding feature maps. In this section, we describe the different possible choices that will guide our experimental study.

In detail, we are based on the dual-stream architecture introduced by [34] on the VGG-16 network. Here, we consider only appearance streams but discuss different ways of combining appearance and motion streams with our aggregation in Section 3.1.

The two-stream model first trains a frame-level classifier that uses all frames from all videos and averages the

predictions of T uniformly sampled frames at test time. Choosing this basic network (pretrained at the frame-level) as a feature generator feeds the input from different frames to our trainable aggregation layer [4, 15, 35]. But which layer's activation should we share? We consider two main options. Consider pooling the outputs of the fully connected (fc) layers. These are represented as 1×1 spatial feature map with 4096-dimensional outputs for each T-frame of the video. In other words, we pool a 4096-dimensional descriptor from each T-frame of the video. Second, considering pooling features from the con volume layer (we consider VAN4 3 and VAN5 3), as shown in Section 3.1, we obtain the best performance by pooling features at the highest convolutional layer (VGG-16 for conv5 and 3) [35].

3.3. Dual-Stream Aggregation Mode. The aggregation layer in this paper can be employed to pool functions across different input mode streams. In our example, we consider appearance and motion streams [36], but any number of other data streams, such as warped streams or RGB differences [37], can be spanned in pooling. There are several possible ways to combine appearance and motion streams to obtain a trainable joint representation. We will explore the most salient ones in this section and outline them in Figure 3.

A single paper aggregation layer covers appearance and motion features in tandem (concat fusion). In this case, we connect the corresponding output feature mappings from appearance and motion, essentially assuming their spatial correspondence. We place a paper aggregation layer on top of this connected feature map, as shown in Figure 3(a). This allows the codebook to be constructed using the correlation between appearance and flow features.

A single paper aggregation layer covers all appearance and motion features (early fusion). We also experimented with a single paper aggregation layer for all features from the appearance and motion streams, as shown in Figure 3(b). This encourages the model to learn a single script space x_{ij} for both appearance and motion features, thus exploiting redundancy in the features.

Late Fusion. This strategy, shown in Figure 3(c), follows the standard test practice of appearance and motion weighted averaging of the last layer of features. Thus, we have two separate layers of this paper aggregation layer, one for each stream [15]. This allows both layers of this paper aggregation layer to learn a specialized representation of each input modality.

4. The Overall Framework of Integration Domain Knowledge

The overall framework of the proposed basketball event classification method incorporating domain knowledge and deep features is illustrated in Figure 4. Our method

framework is inspired by the idea of decomposing complex tasks into basic tasks and combining the work of multiple elementary channels to generate a unified network, and each part of the framework is described in detail in this subsection. First, we propose a 2-stage event classification method, which includes 5 types of coarse classification for 3 pointers, free throws, 2 pointers (including other 2 pointers and layups), dunks, and steals based on event-occ video clips with GCMP_DF_SVF information [30] and 2 types of coarse classification for layups and other 2 pointers based on preevent video clips with GCMP_DF_SVF information. 2 categories of coarse classification are based on the preevent video segment using GCMP_DF_SVF information for layups and other 2 pointers. Both event classification stages are achieved by our model to extract and express the spatial as well as sequential features of GCMPs. Secondly, we implement the prediction of event state success/failure in the video segment of the postevent video segment based on RGB_DF_VF features [15, 21] and finally combine the results of event type classification with the event state success\failure to achieve 11 classes of event classification, which are a 3-point shot success, 3-point shot failure, freethrow success, free-throw failure, other 2-point shot success, other 2 pointers, layup success, layup failure, slam dunk success, slam dunk failure, and steal [28].

The specific process is described as follows:

(1) 2-stage event classification method: firstly, we implement the event preclassification based on the depth feature basketball events in the event video segment; this stage does not consider the success and failure states only; it predicts the event type. Combined with the model described in part 3, in the event video segment, the features of other 2 pointers and layup events are relatively similar, so in this stage, we will first implement a 5-class classifier; that is, other 2 pointers and layup events are fused into one class; then, the 5 classes are 3 pointers, free throws, other 2 pointers and layups, dunks, and steals. The deep feature basketball event expression of basketball video players is extracted by optical flow graph, and then, the spatial features of basketball video are extracted by aggregation layer in this paper's model, and the sequence feature expression of deep feature basketball events is realized by dual flow network. Next is the implementation of event subclassification within the preevent video segment based on depth feature basketball events. This stage requires training a 2-class classifier similar to the 5-class classifier for additional 2 pointers and layups. As further illustrated in Figures 3–6, for a video segment of an unknown class, the video sequence of the event video segment is first input to the 5-class classifier and can be output directly if the output is a 3-point, free throw,

(a)   (b)   (c)
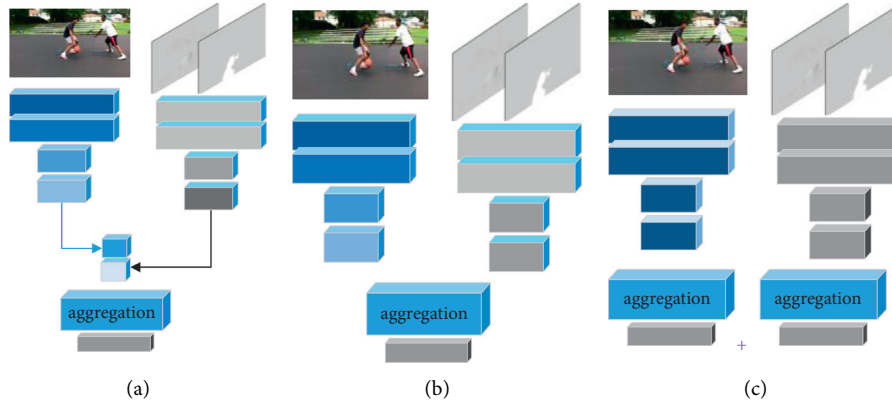
FIGURE 3: Aggregate patterns that combine appearance and motion flow. (a) Early fusion. (b) Intermediate fusion. (c) Late fusion.
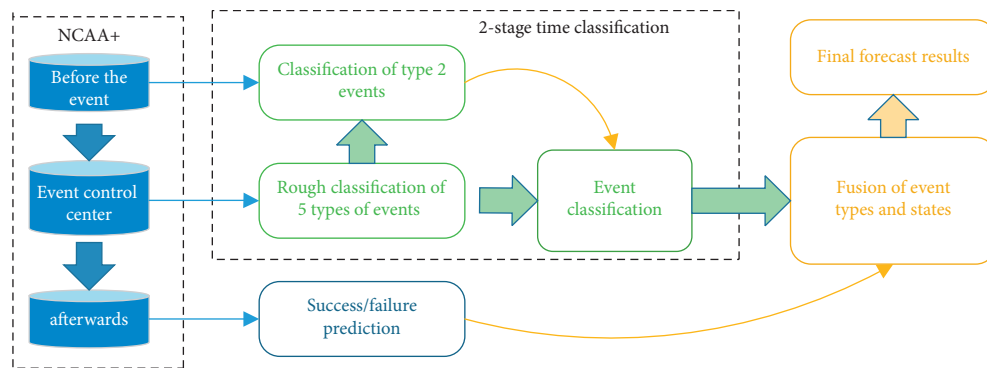
FIGURE 4: Basketball event classification based on domain knowledge and depth features.

dunk, or steal event. However, if the output is the category after the fusion of layups and other 2 pointers, the preevent video segment of the video segment needs to be input to the 2-class classifier for further judgment, and the output of the last two classes of classifiers is the discriminative type of the event.

(2) Classification method for success and failure event states: the success and failure state prediction is deployed in the postevent video segment based on the original video frame RGB_DF_VF. Combined with the analysis of different video segments of the same event in chapter 3 on the event classification ability, we extract RGB_DF_VF features for the success and failure classification of events, except for the stealing events (here only the success and failure states are set for the pitching events, without considering stealing). Finally, the entire event state analysis is achieved by voting the results of each frame in the postevent.

The final classification of all the 11 types of events is done by fusing the analysis of stages (1) and (2) above. Based on the analysis of (1) above, the prediction vectors of 6 types of events can be obtained through the 2-stage event classification, which can be expressed as points, free throws, layups, other points, dunks, and steals, and the prediction vectors of success and failure of the event state stages can be expressed

as success and failure based on the analysis of (2) above. It should be noted here that since steals cannot be determined whether their status is success or failure, the five categories of events that need to be analyzed in this paper are all shooting events, whose expression vectors are points, free throws, layups, other points, and dunks. The vectors of these events are all binary vectors, and each element of the predicted event is either "1" or "0" [20, 30].

5. Simulation Experiment and Result Analysis

Our experimental video data were selected from the video scenes captured by the main camera. First, five video sequences of basketball games from the Nth NCTU tournament were segmented into many video shots [7, 10], from which we selected about 157 video shots from these scenes.

In this thesis, a neural network model is constructed using the PyTorch framework [5] with a server operating system of Linux release (core). 7 Two Tesla V100 chips with 32 G of video memory each and a CPU frequency of 2.20 GHz are used.

5.1. Ball Detection and Tracking. The ball detection is built on a coarse to the fine method. Coarse processing is to search for possible locations of the ball in a larger spatial area, while exceptional processing is to precisely determine the shooter/ball location in a smaller search area. The position of the ball

in the current frame is invoked as the initial guess for shooter identification and ball finding in the next frame.

In the next frame, we apply pixel-by-pixel ball tracking by checking the model score of the adjacent candidates of the identified bounding circles in this paper. The ball tracking process in this frame is based on adequate processing. Figure 5 shows the distribution of this paper model score for ball tracking. The knowledge fusion architecture uses color histogram features and luminance features to calculate this paper model score. Figure 5(a) represents the ball tracking from the initial position (dashed surrounds) of the previous frame to the best position (solid surrounds) of the current frame.

The intensity of each block in Figure 5(b) indicates a paper model score. Blocks with lower intensity indicate lower paper model scores, and the position is measured in pixels. The arrows indicate the tracking trajectory of the current search from the initial position in the previous frame to the best position of the current frame and the local maxima of this paper's model score. We search for the brightest pixel, where the position indicates the predicted position of the ball.

5.2. Shooter Recognition. In basketball game videos, players from two different teams always wear two different colors. Here, we trained two different models of this paper model for the shooters in the game. Using the color histogram of the shooter's clothing as a feature vector, two distinct models of this paper model are pretrained. Once the knowledge fusion architecture locates the ball, we can search the area around the ball and search for the approximate location of the shooter. To determine the size of the shooter, we assume that the image size of the shooter can be quantified into 10 different scales (i.e., the block length can be between 50, 60, ..., 140) with an aspect ratio equal to 0.3. Here, we use this paper model tracking to identify the size of the shooter. Results of the shooter identification experiments are shown in Figure 6.

5.3. Shooting Detection and Tracking. We find that the edge information of the basket is significantly different from background regions such as spectators, players, and venues. For the input image, we find the corresponding edge mapping and then rescale the bounding box to normal size as the input feature for the knowledge fusion architecture. Since baskets often appear in the upper part of the image box, therefore, we only use coarse processing to check the candidate positions in the upper half of the image frame. We check the model score of this paper for the candidate baskets in the upper half of the image frame once every five pixels to determine the best coarse positioning of the baskets. However, the position of the basket is not the same in subsequent frames due to camera panning or scaling. Therefore, we need to find the best basket position by refinement, so as to find the basket position around the bounding box of the previous frame that has the best present paper model score in the current frame. The tracking results are presented in Figure 7.

5.4. Pitch Detection. The estimation of ball release timing is further complicated by the fact that players have different heights and arm lengths. Taller players may also have longer arms. To detect whether the ball was released, we compared the product of the pitcher's distance (h_1) with the pitcher's height and ratio (r_1). The pitcher's distance in the vertical direction (i.e., h_1) can be obtained from the distance between the ball tracking and the upper boundary of the pitcher's torso.

Here, we find a certain relationship between h_1 and h_2 (i.e., $r_1 = h_1/h_2$) when the player is about to release the ball. If the thrower's distance is greater than the product, the thrower is releasing the ball and a shot event may occur; the relationship between h_1, h_2, and r_1 is shown in Figure 8. Here, ratio 1 (i.e., r_1) can be achieved by proper training. By training a sequence of approximately 70 ball release examples, we find that the average ratio $r_1 = 0.655$.

5.5. Ball and Ring Contact Detection. Ball ring contact is important information for identifying shot put events. We identify ball-loop contact by calculating the ball-loop distance between the ball and the basket. Ball tracking and basket tracking steps can be used to get the position of the ball and the basket. We can find the ball ring contact when the positions of the ball's envelope and the basket's envelope box are very similar to each other. As shown in Figure 9, the distance between the ball and the basket is obtained by calculating the following:

(1) The upper boundary of the basket and the lower boundary of the ball as the distance Y of the ball ring

(2) The left/right boundary of the basket and the right/left boundary of the ball as the distance X of the ball ring

5.6. Ball Direction Recognition. The direction of the ball bouncing after the ball ring contact detection is usually unpredictable. We must establish the direction of the bouncing ball's motion relative to the basket. First, the ball position and basket position in the image are generated using ball tracking and basket detection techniques. When the direction of motion of the bouncing ball relative to the basket is downward, we can deduce that the video shot is a scoring shot. Figure 10(a) shows the scoring shots on the left and right sides, respectively. If the direction of motion of the bouncing ball is not downward, it may move upward or sideways, as illustrated in Figure 10(b).

5.7. Lens-Based Event Recognition. This subsection enables the retrieval of video footage by semantic understanding and classification of a large amount of input basketball video footage. In this section, we illustrate how to test the performance of our system and show experimental results in some aspects. We emphasize that the recognition rate metric is based on video shot units, while the updated probability of the model in this paper is based on frame units [7, 15].

(a)   (b)
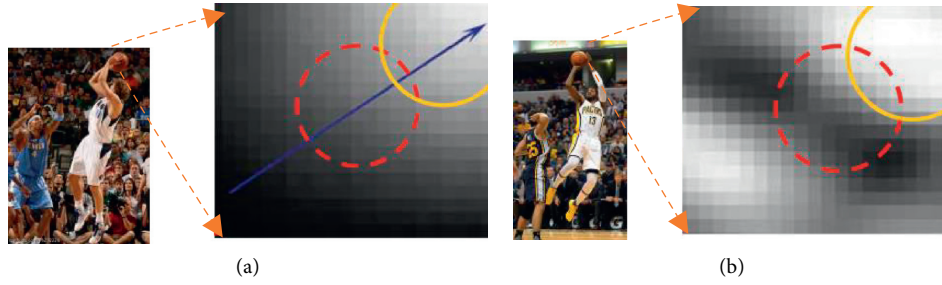
FIGURE 5: Score distribution of our model. (a) The color histogram as the input feature vector. (b) The grayscale value as the input feature vector.
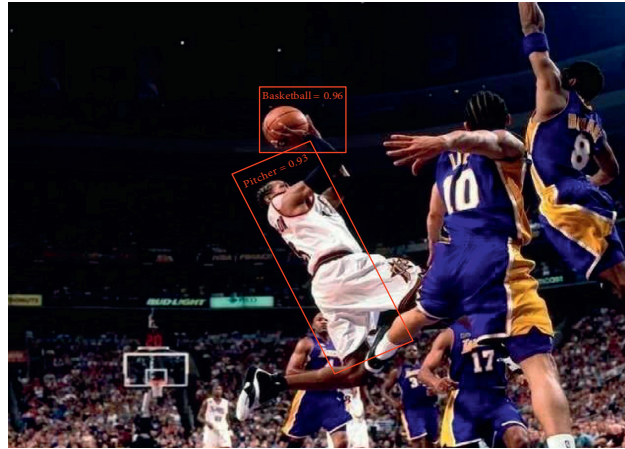
FIGURE 6: The result of shooter recognition.

Similarly, experimental video data are five basketball games consisting of two women's basketball games and three men's basketball games selected from the basketball video scenario described in the previous section. For the bottom node, preprocessing and feature extraction are utilized to obtain the shooter position, ball position, and basket position in each frame.

The interpretation of basketball video shots is achieved by classifying the video shots. The root node of the shot event consists of four states, indicating four different shot event categories. The input video shots are now classified according to the maximum posterior probability of the corresponding state. Accuracy and recall are calculated to measure the performance of shot event classification and score event detection indicated in Tables 1 and 2. It is clear that the performance of the schemes in this paper is very good; for example, both recall and precision in the short shot can reach more than 92%, which is a rare performance result in video data, and these are attributed to our dual-stream network setup and the integration of video knowledge.

5.8. Network Operations at the Aggregation Layer. Here, we evaluate the different locations in the network where the layers of this paper's aggregation layer can be inserted. Specifically, we compare the placement of this paper's aggregation layer after the last two evolutionary layers (conv4 3 and conv5 3) and the

last fully connected layer (fc7), embedding subsampled conv and fc features from a set of videos in tSNE as shown in Figure 11. In each case, we trained to the block of layers before this paper's aggregation layer; thus, in the case of conv4, conv4 1 indicates a loss, and in the case of fc7, fc7 indicates a loss. The results are shown in Figure 11 and clearly show that the best performance is obtained by aggregating the last evolutionary layer (conv5 3). fc6 characteristics obtain similar performance to fc7.

We believe that this is due to two reasons. First, pooling on fully connected layers prevents the aggregation layers in this paper from modeling spatial information, since these layers already compress a lot of information. Second, fc7 features are more semantic, so features from different frames are already comparable to each other and do not take advantage of the modeling capabilities of the aggregated layers in this paper; i.e., they would normally fall in the same cell. To test this hypothesis, we visualize the conv5 3 and fc7 appearance features from the same frame using public tSNE [22, 30] embedding in Figure 11. These plots clearly show that fc7 features from the same video (displayed in the same color) already resemble each other (Figure 11(b)), while conv5 3 features are more diverse and can benefit from the ability of this paper to aggregate layers to capture complex distributions in the feature space (Figure 11(a)), as described in Section 3.1.

(a)

(b)

(c)
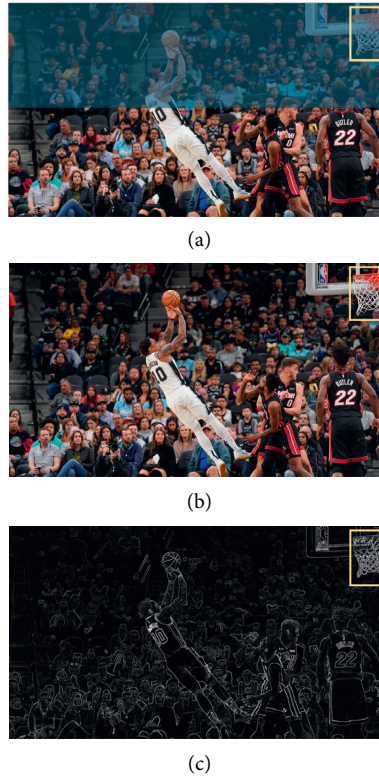
FIGURE 7: (a) Search range of basket detection. (b) Detection of the basket. (c) Edge information of the object.

FIGURE 8: Relationship between h_1, h_2, and r_1. h_1 is the pitcher distance, r_1 is the ratio 1, and $\overline{h_2}$ is the pitcher clothing height.

5.9. Different Fusion Effects. It can be clearly seen from Table 3 that late fusion has the best accuracy under different basketball skills. For example, in terms of basic facts, late fusion is 0.4 higher than intermediate fusion.

Similarly, the other two aspects are also better than other hybrid model schemes. This is also the reason why this paper selects late fusion as the fusion method of our model.

Figure 9: Ball ring distance.
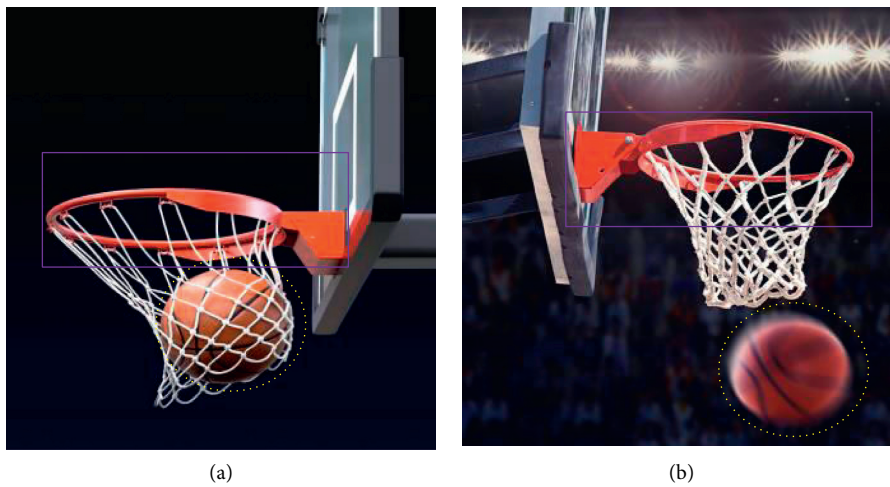
(a)

(b)

Figure 10: Bouncing balls can move in different directions. (a) A bouncing ball moving sideways (b) A bouncing ball moving downward.

Table 1: Shooting event classification results.

Video snapshot	Basic facts	Correct	Miss	False alarm	Recall	Accuracy
Close-up	34	35	3	1	0.9142	0.9652
Medium view	20	11	5	5	0.8362	0.7542
Far view	13	13	2	3	0.8324	0.8154
Penalty shot	17	17	2	6	0.8865	0.8423
Total	84	76	12	15	0.8722	0.8664

TABLE 2: Score event detection results.

Snapshot	Basic facts	Correct	Miss	False alarm	Recall	Accuracy
Score	35	35	3	3	0.8842	0.9152
No score	52	41	4	4	0.9142	0.9242
Total	87	76	7	7	0.9234	0.9258
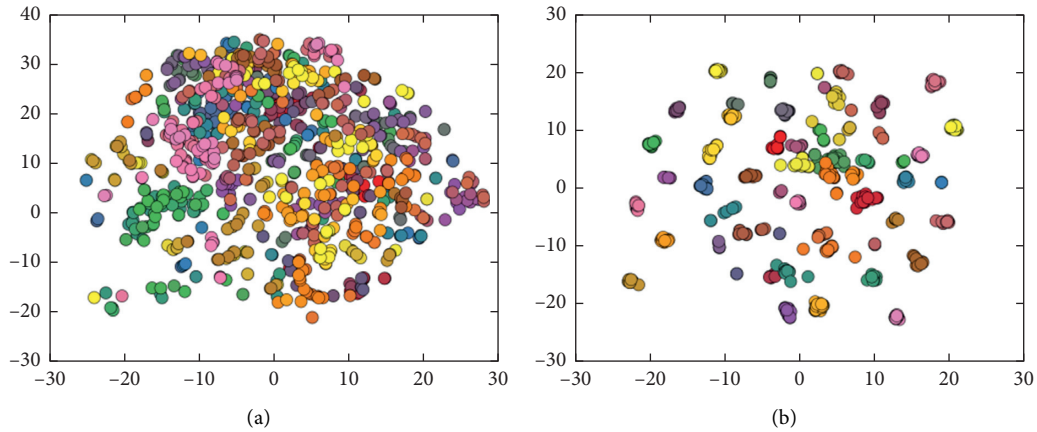
FIGURE 11: Embedding subsampled conv and fc features from a set of videos in tSNE. (a) conv5 3. (b) fc7.

TABLE 3: Accuracy of different hybrid models in different basketball skills.

	Basic facts	Correct	Miss
Early fusion	0.8652	0.7512	0.8743
Intermediate fusion	0.8832	0.7486	0.8859
Late fusion	0.9231	0.8159	0.8959

6. Conclusions

In this paper, we propose a successful spatiotemporal video feature aggregation method for basketball video semantic extraction. Our approach is end-to-end trainable, and our method is general enough to be applied as an aggregation layer-CNN layer in this paper for future video architecture, which may be useful for related tasks such as spatiotemporal localization of human behavior in long videos. We designed a dual-stream network combined with a reasonable spatiotemporal feature aggregation scheme trying to bridge the gap between low-level features and high-level semantic events, which can fuse knowledge from different video sources to design an automatic semantic system for basketball games videos. Experiments show that the model in this paper can accurately identify different shooting items or player actions in video footage of basketball games.

Data Availability

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Conflicts of Interest

The authors declare that they have no conflicts of interest regarding this work.

Acknowledgments

References

[1] L. Havasi and Z. Szlavik, "Using location and motion statistics for the localization of moving objects in multiple camera surveillance videos," in *Proceedings of the IEEE International Conference on Computer Vision Workshops*, IEEE, Barcelona, Spain, 2010.

[2] H. Kerr, B. Bakken, and G. House, "Future directions in sports-related concussion management," *Clinics in Sports Medicine*, vol. 40, no. 1, pp. 199–211, 2021.

[3] J. A. Blázquez López, S. García Martínez, A. Ferriz Valero, and J. Olaya Cuartero, "Cuantificación de la carga de entrenamiento y competición: análisis comparativo por posiciones en un equipo de la Liga Española de Baloncesto Oro (Quantification of training and competition load: comparative analysis by position in a Spanish Golden Baske," *Retos*, vol. 42, pp. 882–890, 2021.

[4] U. Avni, H. Greenspan, E. Konen, M. Sharon, and J. Goldberger, "X-ray categorization and retrieval on the organ and pathology level, using patch-based visual words," *IEEE Transactions on Medical Imaging*, vol. 30, no. 3, pp. 733–746, 2011.

[5] R. Li, Q. Liu, J. Gui et al., "Indoor relocalization in challenging environments with dual-stream convolutional neural networks," *IEEE Transactions on Automation Science and Engineering*, vol. 15, pp. 651–662, 2017.

[6] Z. Zhao, D. A. Klindt, A. Maia Chagas et al., "The temporal structure of the inner retina at a single glance," *Scientific Reports*, vol. 10, no. 1, p. 4399, 2020.

[7] A. Chadha, A. Abbas, and Y. Andreopoulos, "Video classification with cnns: using the codec as a spatio-temporal activity sensor," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 29, no. 2, pp. 475–485, 2017.

[8] R. Arandjelovic, P. Gronat, A. Torii et al., "NetVLAD: CNN architecture for weakly supervised place recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 5297–5307, Las Vegas, NV, USA, June 2016.

[9] A. Kamilaris and F. X. Prenafeta-Boldú, "Deep learning in agriculture: a survey," *Computers and Electronics in Agriculture*, vol. 147, pp. 70–90, 2018.

[10] L. Wang, Y. Qiao, and X. Tang, "Action recognition with trajectory-pooled deep-convolutional descriptors," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4305–4314, Boston, MA, USA, June 2015.

[11] C. M. Chen and L. H. Chen, "Novel framework for sports video analysis: a basketball case study," in *Proceedings of the 2014 IEEE International Conference on Image Processing (ICIP)*, pp. 961–965, IEEE, Paris, France, October 2014.

[12] G. Nagarajan and R. I. Minu, "Multimodal fuzzy ontology creation and knowledge information retrieval," *Proceedings of the International Conference on Soft Computing Systems*, Springer, New Delhi, India, pp. 697–706, 2016.

[13] D. W. Tjondronegoro and Y.-P. P. Chen, "Knowledge-discounted event detection in sports video," *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, vol. 40, no. 5, pp. 1009–1024, 2010.

[14] Z. Gao, P. Wang, H. Wang, M. Xu, and W. Li, "A review of dynamic maps for 3D human motion recognition using ConvNets and its improvement," *Neural Processing Letters*, vol. 52, no. 2, pp. 1501–1515, 2020.

[15] X.-B. Fu, S.-L. Yue, and D.-Y. Pan, "Camera-based basketball scoring detection using convolutional neural network," *International Journal of Automation and Computing*, vol. 18, no. 2, pp. 266–276, 2021.

[16] B. Li and X. Xu, "Application of artificial intelligence in basketball sport," *Journal of Education, Health and Sport*, vol. 11, no. 7, pp. 54–67, 2021.

[17] R. A. Minhas, A. Javed, A. Irtaza, M. T. Mahmood, and Y. B. Joo, "Shot classification of field sports videos using AlexNet Convolutional Neural Network," *Applied Sciences*, vol. 9, no. 3, p. 483, 2019.

[18] M. Rafiq, G. Rafiq, R. Agyeman, G. S. Choi, and S.-I. Jin, "Scene classification for sports video summarization using transfer learning," *Sensors*, vol. 20, no. 6, p. 1702, 2020.

[19] S. Su, J. Pyo Hong, J. Shi et al., "Predicting behaviors of basketball players from first person videos," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1501–1510, Honolulu, HI, USA, July 2017.

[20] K. Lu, J. Chen, J. J. Little, and H. He, "Lightweight convolutional neural networks for player detection and classification," *Computer Vision and Image Understanding*, vol. 172, pp. 77–87, 2018.

[21] Y. Yang, "Research on sports action training method based on generative confrontation network model and artificial intelligence," *Journal of Intelligent and Fuzzy Systems*, vol. 4, pp. 1–11, 2021.

[22] N. A. Rahmad, N. A. J. Sufri, N. H. Muzamil, and M. A. As'ari, "Badminton player detection using faster region convolutional neural network," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 14, no. 3, pp. 1330–1335, 2019.

[23] W. Liu, C. C. Yan, J. Liu, and H. Ma, "Deep learning based basketball video analysis for intelligent arena application," *Multimedia Tools and Applications*, vol. 76, no. 23, pp. 24983–25001, 2017.

[24] J. Žemgulys, V. Raudonis, R. Maskeliūnas et al., "Recognition of basketball referee signals from real-time videos," *Journal of Ambient Intelligence and Humanized Computing*, vol. 11, no. 3, pp. 979–991, 2020.

[25] D. Acuna, "Towards real-time detection and tracking of basketball players using deep neural networks," in *Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS 2017)*, pp. 4–9, Long Beach, CA, USA, January 2017.

[26] G. Junjun, "Basketball action recognition based on FPGA and particle image," *Microprocessors and Microsystems*, vol. 80, Article ID 103334, 2021.

[27] L. Wu, Z. Yang, Q. Wang et al., "Fusing motion patterns and key visual information for semantic event recognition in basketball videos," *Neurocomputing*, vol. 413, pp. 217–229, 2020.

[28] X. Fu, K. Zhang, C. Wang, and C. Fan, "Multiple player tracking in basketball court videos," *Journal of Real-Time Image Processing*, vol. 17, no. 6, pp. 1811–1828, 2020.

[29] L. Chen and W. Wang, "Analysis of technical features in basketball video based on deep learning algorithm," *Signal Processing: Image Communication*, vol. 83, Article ID 115786, 2020.

[30] L. Wu, Z. Yang, J. He et al., "Ontology-based global and collective motion patterns for event classification in basketball videos," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 7, pp. 2178–2190, 2019.

[31] L. Liu, "Objects detection toward complicated high remote basketball sports by leveraging deep CNN architecture," *Future Generation Computer Systems*, vol. 119, pp. 31–36, 2021.

[32] J. Quiroga, H. Carrillo, E. Maldonado et al., "As seen on tv: automatic basketball video production using Gaussian-based actionness and game states recognition," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pp. 894-895, Seattle, WA, USA, June 2020.

[33] B. Mahaseni, E. R. M. Faizal, and R. G. Raj, "Spotting football events using two-stream convolutional neural network and dilated recurrent neural network," *IEEE Access*, vol. 9, Article ID 61929, 2021.

[34] K. Rangasamy, M. A. As'ari, N. A. Rahmad, and N. F. Ghazali, "Hockey activity recognition using pre-trained deep learning model," *ICT Express*, vol. 6, no. 3, pp. 170–174, 2020.

[35] R. Girdhar, D. Ramanan, A. Gupta et al., "Actionvlad: learning spatio-temporal aggregation for action classification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 971–980, Honolulu, HI, USA, July 2017.

[36] V. Choutas, P. Weinzaepfel, J. Revaud et al., "Potion: pose motion representation for action recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7024–7033, Salt Lake City, UT, USA, June 2018.

[37] M. Trinidad-Fernández, D. Beckwée, A. Cuesta-Vargas et al., "Differences in movement limitations in different low back pain severity in functional tests using an RGB-D camera," *Journal of Biomechanics*, vol. 116, Article ID 110212, 2021.

Research Article

Digital Currency Illegal Behavior Detection Based on Mutual Information Prior Loss

Feng Yang [iD],[1,2] Guixin Dong,[1,2] Chaoran Cui,[1] Xiaojie Li,[1] Yaxi Su,[1] and Yilong Yin [iD][3]

[1]*School of Computer Science and Technology, Shandong University of Finance and Economics, Jinan 250014, China*
[2]*School of Software, Shandong University, Jinan 250101, China*
[3]*Shandong Key Laboratory of Blockchain Finance, Jinan 250014, China*

Correspondence should be addressed to Feng Yang; 20077869@sdufe.edu.cn

In recent years, with the rapid development of digital currency, digital currency brings us convenience and wealth, but also breeds some illegal and criminal behaviors. Different from traditional currencies, digital currency provides concealment to criminals while also exposing their behavior. The analysis of their behavior can be used to detect whether the current digital currency transaction is legal. There is a problem that most digital currency transactions are in compliance with laws and regulations, and only a small part of them uses digital currency to conduct illegal activities. It belongs to the problem of sample imbalance. It is quite challenging to accurately distinguish which transactions are legal and which are illegal in the massive digital currency transactions. For this reason, this study combines the mutual information and the traditional cross-entropy loss function and obtains the loss function based on the mutual information prior. The loss function based on the mutual information prior is that the bias of the category prior distribution is added after the output of the model (before the softmax), which makes the model consider category prior information to a certain extent when predicting. The experimental results show that the use of the loss function based on mutual information prior to the detection of digital currency illegal behavior has a good effect in SVM, DNN, GCN, and GAT methods.

1. Introduction

As Bitcoin exceeded the $50,000 mark, individuals' interests in digital currency increased on February 16, 2020. Bitcoin is one of many peer-to-peer virtual cryptocurrencies. The biggest feature of digital currency is the peer-to-peer transmission, which shows that it is a decentralized payment system. Unlike traditional currencies, digital currencies cannot depend on specific organizations to issue them. It is in line with proprietary algorithms, which is generated by computer calculations over a long period of time. Digital currencies which use a distributed database of nodes throughout the P2P network can record and confirm all transactions, such as Bitcoin, using cryptography design to ensure the security of all aspects of currency circulation. Generally speaking, digital currency can be seen as a set of quasicurrency issuing system, as well as a set of currency circulation and settlement network. It can not only solve the problem of overissuance of existing credit currency but also realize low-cost, quasi-real-time currency settlement.

Money laundering is regarded as a major illegal activity in digital currency. According to statistics, there are 1.6 trillion dollars of money laundering in the global market every year. The money that needs to be laundered is often the illegal income from terrorist activities, fraud, drug trade, corruption, and other criminal action. As one of the internationally recognized illegal economic activities, the behavior of money laundering will not only impede our country's economic development but also affect our country's political security and social stability. The traditional way of money laundering is to avoid the supervision of the domestic anti-money laundering department. In general, criminals would transfer a large amount of illegal money to

countries whose supervision is less strict or take advantage of policy loopholes between different countries so as to evade supervision.

Criminals are looking for a new type of vector-financial currency, mostly Bitcoin, while fiat money has become increasingly debased. Digital currency has been a major force in the money-laundering market out of its anonymity, decentralization, strong cross-border liquidity, lack of effective regulations, and opaque transaction information. Furthermore, it is also used in the dark net illegal trading and network extortion and other criminal activities. In 2017, a computer virus called WannaCry began to wreak havoc on computers all over the world, which can encrypt victims' computer files and threat victims of computer virus attacks to pay $600 in Bitcoin to WannaCry's Bitcoin address. The antiterrorist group discovered that about $140,000 worth of Bitcoins had been collected from the address WannaCry which is provided for the redemption money on August 3, 2017. Therefore, illegal money laundering has been a cancer and bomb in the process of economic development and social progress. Hence, it is extremely urgent to crack down on illegal money laundering.

The analysis of illegal detection of digital currency is a detection of abnormal phenomena. The majority of digital currency holders are using digital currency legally, while only a small number of criminals are using digital currency for illegal transactions, such as extortion and money laundering [1–5]. As a result, it is difficult to accurately sort out a small amount of illegal transactions in the vast data set of digital currency transactions. It can be seen as a typical problem of data imbalance. Previous research considered using classical methods to deal with data imbalance, such as modifying the weight of legal use and illegal use categories and taking focal loss [6] to replace the original loss function.

As for related materials of machine learning, mutual information is often used to measure the correlation between two things, which is regarded as a measure of the interdependence of variables. Different from the correlation coefficient, mutual information is not limited to the actual value of the random variable, while it determines the similarity between the product of two edge distributions and the joint distribution. It can be said that mutual information essentially reflects the relationship between input information and output information. As a result, if mutual information can be directly fitted, effective knowledge will be studied by individuals. In this article, cross-entropy fitting mutual information is considered to enhance the loss function, so that the problem of data sample imbalance can be relieved.

Digital currency transactions are ubiquitous. Graph Neural Network, which is very popular in recent years, can be seen as the key tool for processing graph data. There are two methods that are commonly used in Graph Neural Network. One is Graph Convolution Network in accordance with space, which represents graph convolution by aggregating the information of adjacent nodes. The other is spectral on the basis of Graph Convolution Network, which interprets graph convolution as removing noise from the image signal.

Because the trading network of digital currency is a graph structure, this article will take Graph Neural Network and traditional machine learning algorithm (DNN and support vector machine) as models to test the ordinary cross-entropy loss function, focal loss function, and mutual information priority-based loss function, respectively. The results show that the loss function based on mutual information prior proposed in this article has better property than the cross-entropy loss function and focal loss function.

Innovation:

(1) Use mutual information-based loss to alleviate the problem of data imbalance

(2) Use Graph Neural Network method

(3) Experiment on real data sets

2. Related Work

Illegal behavior detection towards digital currency has great significance, playing an important role in keeping the security of network finance, governing the chaos of network finance, and preventing the risk of network finance. At present, the illegal behavior detection of digital currency has been involved in many fields such as finance, economy, law, and even politics. Illegal behavior detection towards digital currency has wide application and important research significance no matter in financial background or technology background. However, there are serious data imbalance problems in illegal transactions of digital currency, and how to effectively alleviate the data imbalance problem has been the major work goal of researchers. Therefore, this article will briefly introduce the existing digital currency illegal behavior detection, so as to improve the loss function to alleviate the data imbalance.

2.1. Detection of Digital Currency Violations. Akcora [7] put forward a solution in accordance with topology data analysis to divide the Bitcoin network into 24 hour long windows. We need to extract the following six characteristics for the address in each window: income, neighbor, weight, length, count, and loop. The new address of known ransomware family is detected, and the appearance of unknown ransomware family address is predicted. New tools on the basis of TDA have largely improved the accuracy of ransomware detection. Chen et al. [8] put forward a method of cascading feature extraction based on transaction graph for phishing account identification and a dual sampling integration algorithm on the basis of lightGBM, so as to solve the problem of class imbalance in phishing fraud identification. It targets the entire blockchain ecosystem, providing users with early warning against phishing scams.

Weber et al. [9] put forward some cryptocurrency transaction authentication laws (especially for Bitcoin) to combat criminal activities, providing the anti-money laundering community with a data-rich, tagged set of transaction data that has never been publicly available before. It also provided a prototype for the visualization of such data and a model for enhancing human analysis and

interpretation. The goal is to inspire others to tackle the big challenge of anti-money laundering and make the world's financial system safer and more inclusive. These methods do not use the neural network method for detection. Neural network can be used for complex nonlinear modeling.

2.2. Improvement of the Loss Function to Alleviate the Data Imbalance.

Cao et al. [10] put forward a theoretically principled label distribution perception of marginal LDAM losses, which was motivated by minimization of edge-based generalization limits. This loss replaces the standard cross-entropy goal during training and can be applied to previous class-imbalance training strategies, such as reweighting or resampling. Then, a simple but effective training program is put forward. This plan postpones reweighting until after the initial phase, which allows the model to learn the initial representation while avoiding some of the complexities associated with reweighting or resampling.

Buda et al. [11] attached great importance to the influence of sample imbalance on the classification neural network and studied the influence of sample imbalance on the classification network in three general data sets of different sizes: MNIST/CIFAR-10/ImageNet as well as relevant ideas to solve the sample imbalance problems: oversampling/subsampling/two-stage training/threshold method. It is concluded that the sample imbalance is not conducive to the classification network. At present, the main method is oversampling. It is concluded that oversampling in general networks does not lead to overfitting. Focal loss was put forward by Lin et al. [6], and category imbalance was regarded as the reason that prevented the one-stage method from surpassing the two-stage method of top-performing. We need to focus on hard examples and reduce the weight of a large number of easy negatives, so as to use focal loss as well as an adjustment item in the cross entropy.

The loss function method based on mutual information prior which is put forward in this article plays an important role in promoting the research on the detection of digital currency illegal behaviors, effectively improving the serious imbalance of samples in the Bitcoin data set and improving the accuracy and recall rate of illegal categories.

3. Method

3.1. Question Raising.

To detect illegal digital currency transactions, we need to formalize digital currency transactions into a typical machine learning problem. We can regard detecting illegal digital currency transactions as a binary classification. The data set that is used in this article is provided by researchers from MIT, IBM, and Elliptic, a blockchain analytics company. It contains a data set of more than 200,000 Bitcoin transactions with a small number of tokens. It is by far the most cryptocurrency-related and tagged data set [9]. There are two categories in this data set. One is the legal category, which includes a series of legal services such as exchanges, wallet providers, and miners. The other category is illegal, which includes fraud, malware,

terrorist groups, ransomware, Ponzi schemes, and other illegal transactions with only a few classified as illegal transactions. More important for detecting illegal transactions of digital currency is the few categories (i.e., illegal categories).

This article formally describes the problem as follows:

Input: $X \in R^{|V|*F}$, where $|V|$ is the number of nodes and F is the characteristic dimension

Output: $Y \in \{0, 1\}$, where 0 is the legal category and 1 is the illegal category

Because there is serious data imbalance between legal category and illegal category, in order to solve this problem, this article uses the loss function in terms of the mutual information prior for mitigation.

3.2. Illegal Transaction Monitoring of Digital Currency Based on Graph Neural Network.

In general, the transactions of digital currency are generally modeled with a graph structure. During recent years, Graph Neural Network (GNN) [12], which is very popular in the field of deep learning, has excellent effect on the processing of graph structure. Therefore, this article considers Graph Convolutional Neural Network (GCN) [13–16] and Graph Attention Network (GAT) [17], the two pillars of the Graph Neural Network, as models for monitoring illegal transactions of digital currency.

3.2.1. Graph Convolutional Network (GCN).

Graph Convolutional Network (GCN) is similar to Cognition Evolutionary Algorithm. It usually uses spectral convolution to drive neighborhood aggregation, which consists of multi-layer graph convolution. In this article, the digital currency transaction graph is assumed from the data set $G = (N, E)$, where N represents the features of nodes and E represents the edge features of the graph. In the Graph Convolutional Network (GCN), the input of each layer is the node feature matrix H and the adjacency matrix A, and then the node feature matrix H is updated as the output through the weight matrix W. The node characteristic matrix H and the weight matrix W are not necessarily the same in each layer, while the adjacency matrix A is always the same. Equation (1) defines the mathematical formula of Graph Convolutional Network:

$$H^{(l+1)} = \sigma\left(\widetilde{D}^{-(1/2)}\widetilde{A}\widetilde{D}^{-(1/2)}H^{(l)}W^{(l)}\right), \quad (1)$$

where l represents the current number of layers and σ is usually taken as a ReLU activation function. The input feature matrix comes from the features of the graph nodes, such as $H^{(0)} = X$.

Formula (2) defines $\widetilde{A}$ and $\widetilde{D}$:

$$\widetilde{A} = A + I,$$
$$\widetilde{D} = \text{diag}\left(\sum_j \widetilde{A_{ij}}\right), \quad (2)$$

where $\widetilde{A}$ represents adjacency matrix plus self-linking and $\widetilde{D}$ represents degree matrix.

We can assume that the Graph Neural Network is used for node classification, the graph convolution layer is set, the activation function is ReLU, and the final output layer is softmax output. First, the characteristic $H^{(0)}$ and adjacency matrix A of the node are input, convolved through the layer graph l, and finally output $H^{(l)}$ through softmax, which is composed of the predicted probability. The graph convolution layer of each layer differs from its feed forward layer only by the front multiplication $\widehat{A}$, which can be understood as a set of transformation embedding of adjacent nodes. It is driven by a spectral filter and is the result of a linear function of the Laplace matrix. The parameter to be updated is $W^{(l)}$. Equation (3) defines the mathematical formula of a 2-layer Graph Convolutional Network (GCN):

$$H^{(2)} = \text{softmax}\left(\widehat{A} \cdot \text{ReLU}\left(\widehat{A}XW^{(0)}\right) \cdot W^{(1)}\right). \tag{3}$$

3.2.2. Graph Attention Network (GAT).
As another representative Graph Convolutional Network, Graph Attention Network (GAT) adds attention mechanism to achieve better neighbor aggregation compared with the GCN. By learning the weight of neighbors, the Graph Attention Network can achieve the weighted aggregation of neighbors. Therefore, GAT is not only robust to noisy neighbors but the use of the attention mechanism also makes the model more interpretable.

Equation (4) gives a linear transformation formula for node embedding of the l layer, where $W^{(l)}$ is the parameter to be trained in the network and H is the node feature matrix:

$$z_i^{(l)} = W^{(l)} H_i^{(l)}. \tag{4}$$

Equation (5) gives the formula for calculating the original attention score between pairs of nodes. First, the nodes z_i and z_j are embedded and spliced, where $||$ represents splicing, and then the dot product of the spliced embed and weight vector $\overrightarrow{a}$ is carried out, where $\overrightarrow{a}$ is the learnable weight vector and finally uses LeakyReLU as the activation function. This form of attention mechanism is often referred to as additive attention.

$$e_{ij}^{(l)} = \text{LeakyReLU}\left(\overrightarrow{a}^{(l)^T}\left(z_i^{(l)} \big\| z_j^{(l)}\right)\right). \tag{5}$$

Equation (6) gives the formula of attention weight. Softmax is used to calculate the original attention score of node entry edge, where $\mathcal{N}(i)$ represents the neighbor node.

$$a_{ij}^{(l)} = \frac{\exp\left(e_{ij}^{(l)}\right)}{\sum_{k \in \mathcal{N}(i)} exp\left(e_{ik}^{(l)}\right)}. \tag{6}$$

Equation (7) gives the node feature update formula and carries out the weighted summation based on attention on the features of all neighbor nodes:

$$H_i^{(l+1)} = \sigma\left(\sum_{j \in \mathcal{N}(i)} a_{ij}^{(l)} z_j^{(l)}\right). \tag{7}$$

In essence, GAT transforms the normalized constant used in GCN for first-order neighbor nodes into a neighbor node feature aggregation function using attention weights.

3.3. Mutual Information Prior Loss.
In statistics, the loss function is a measure of the degree of system error. In the supervised learning model, the loss function measures the difference between the predicted value of the sample produced by the model and the true label of the sample. In this section, we mainly introduce how to improve the loss function through the idea of mutual information, so as to alleviate the problem of class imbalance.

3.3.1. Cross Entropy.
The cross entropy loss function is the most widely used loss function in the field of machine learning today. Because the weight gradient of the last layer of the cross entropy loss function is not related to the derivative of the activation function but is only proportional to the difference between the real value and the output value, it converges quickly. Also, the back propagation is in the form of multiplication, so the whole weight matrix is updated quickly. Equation (8) gives the formula of cross entropy loss function:

$$H(p, q) = -\sum_i p(c_i) \log q(c_i), \tag{8}$$

where $q(c_i)$ is the output of the network, which is equivalent to the output result after the sample is input to the neural network, and $p(c_i)$ is the expected sample distribution, equivalent to the label of real data.

In this article, we mainly focus on the classification of labels. It is assumed that there are k categories, the training data are $(x, y) \sim D$, and the distribution of the modeling is $p_\theta(y|x)$. So, the optimization goal is to maximize likelihood estimation or to minimize the cross entropy. Equation (9) gives the formula for minimizing cross entropy:

$$\arg \min_\theta \mathbb{E}_{(x,y) \sim D}\left[-\log p_\theta(y|x)\right]. \tag{9}$$

3.3.2. Mutual Information.
Mutual information [18] measures the dependence between two random variables, that is, how much less uncertainty y is given an x. If there is no relationship between x and y, then the value of mutual information is the minimum of 0. If given x can completely eliminate the uncertainty of y, then the value of mutual information is the maximum entropy of x.

For any discrete random variable, its entropy is defined as formula (8).

For the joint entropy of a group of random variables, (x, y) is similar to the entropy definition of a single discrete random variable. Equation (10) gives the joint entropy $H(X; Y)$ formula:

$$H(X;Y) = -\sum_{y \in Y} \sum_{x \in X} p(x, y) \log(q(x, y)), \tag{10}$$

where $p(x, y)$ is the joint probability distribution function of X and Y and $p(x)$ and $p(y)$ are the edge probability distribution function of X and Y, respectively. Equation (11) gives the mutual information formula:

$$I(X;Y) = H(X) + H(Y) - H(X, Y), \tag{11}$$

where $H(X), H(Y),$ and $H(X, Y) > 0$ and $I(X;Y) > 0$.

Mutual information can essentially express the relationship between input and output. Compared with fitting conditional probability, if the model can directly fit mutual information, more essential knowledge will be learned because mutual information is the indicator that reveals the core relationship. However, the training of mutual information fitting is very difficult, and the conditional probability is generally trained. The most commonly used method is to use cross entropy $-\log p_\theta(y|x)$ directly for training. Therefore, how to make the model use cross entropy loss to fit the mutual information has become the focus of research [19].

Softmax is often used as the activation function of the last layer of neural network in binary classification or multiple classification problems and is used to represent the predicted output of neural network due to its normalization function and easy derivation. Equation (12) gives the formula of softmax generalization:

$$q(c_j) = \frac{e^{z_j}}{\sum_{i=1}^{n} e^{z_i}}. \tag{12}$$

In a neural network, z_j is the output of the upper layer, $q(c_j)$ is the distribution form of the output of this layer, and e^{z_j} is a sum of e^{z_i} within a batch.

3.3.3. Improved Loss Function Based on Mutual Information. First, this study assumes that logits is $f(x; \theta)$, the output result of the network model. Substituting it into softmax, equation (13) can be obtained:

$$p_\theta(y|x) = \frac{e^{f_y(x;\theta)}}{\sum_{i=1}^{K} e^{f_i(x;\theta)}}. \tag{13}$$

Equation (14) gives the loss function form of equation (13):

$$-\log p_\theta(y|x) = -\log \frac{e^{f_y(x;\theta)}}{\sum_{i=1}^{K} e^{f_i(x;\theta)}}$$
$$= \log\left[1 + \sum_{i \neq y} e^{f_i(x;\theta) - f_y(x;\theta)}\right]. \tag{14}$$

Equation (14) is the conventional softmax cross entropy. Equation (15) is the modeling of mutual information:

$$\log \frac{p_\theta(y|x)}{p(y)} \sim f_y(x; \theta) \Leftrightarrow \log p_\theta(y|x) \sim f_y(x; \theta) + \log p(y). \tag{15}$$

Equation (16) is the result after the softmax normalization is repeated in the form of the right end:

$$p_\theta(y|x) = \frac{e^{f_y(x;\theta) + \log p(y)}}{\sum_{i=1}^{K} e^{f_i(x;\theta) + \log p(i)}}. \tag{16}$$

Equation (17) is the loss function form of equation (16):

$$-\log p_\theta(y|x) = -\log \frac{e^{f_y(x;\theta) + \log p(y)}}{\sum_{i=1}^{K} e^{f_i(x;\theta) + \log p(i)}}$$
$$= \log\left[1 + \sum_{i \neq y} \frac{p(i)}{p(y)} e^{f_i(x;\theta) - f_y(x;\theta)}\right]. \tag{17}$$

If it is more general, the formula of equation (18) after adding regulatory factor τ is as follows:

$$-\log p_\theta(y|x) = -\log \frac{e^{f_y(x;\theta) + \log p(y)}}{\sum_{i=1}^{K} e^{f_i(x;\theta) + \log p(i)}}$$
$$= \log\left[1 + \sum_{i \neq y} \left(\frac{p(i)}{p(y)}\right)^\tau e^{f_i(x;\theta) - f_y(x;\theta)}\right]. \tag{18}$$

In this way, the model still uses cross entropy as the loss function, but in essence it is fitting mutual information. An offset associated with the label prior is applied to each logarithmic output (i.e., the result before softmax activation) [20]. The training process combined with a neural network is shown in Figure 1.

According to the above process, a neural network is first determined, data x are input into the neural network, and a coded logits is obtained through the neural network, and then logits is subtracted from one Inp(y), in which $p(y)$ is the prior information of the label. This is then fed into the softmax activation layer, resulting in the predicted output y.

4. Experiments and Results

4.1. Data Set. The data set used in this article was provided by Elliptic, Inc., with 203,769 node transactions and 234,355 edges. About 21 percent (42,019) of the transactions were marked as legal, 2 percent (4,545) were marked as illegal, and the rest were marked as unknown but with other characteristics. There are 49 different graphs, each of which is unrelated to each other. A node represents a transaction, and an edge represents the flow of Bitcoins.

There are two problems with this data set at present: a large number of unmarked nodes and a serious data imbalance due to the fact that illegal categories only account for 10% of the nodes with tags compared to legal categories (Figure 2).

4.2. Measurement Standard. In order to verify the advantages of the method in this article more objectively, we need to evaluate the effect of the model after the machine learning system completes the modeling. The evaluation indexes were Precision, Recall, and F1-Score for comparative analysis.
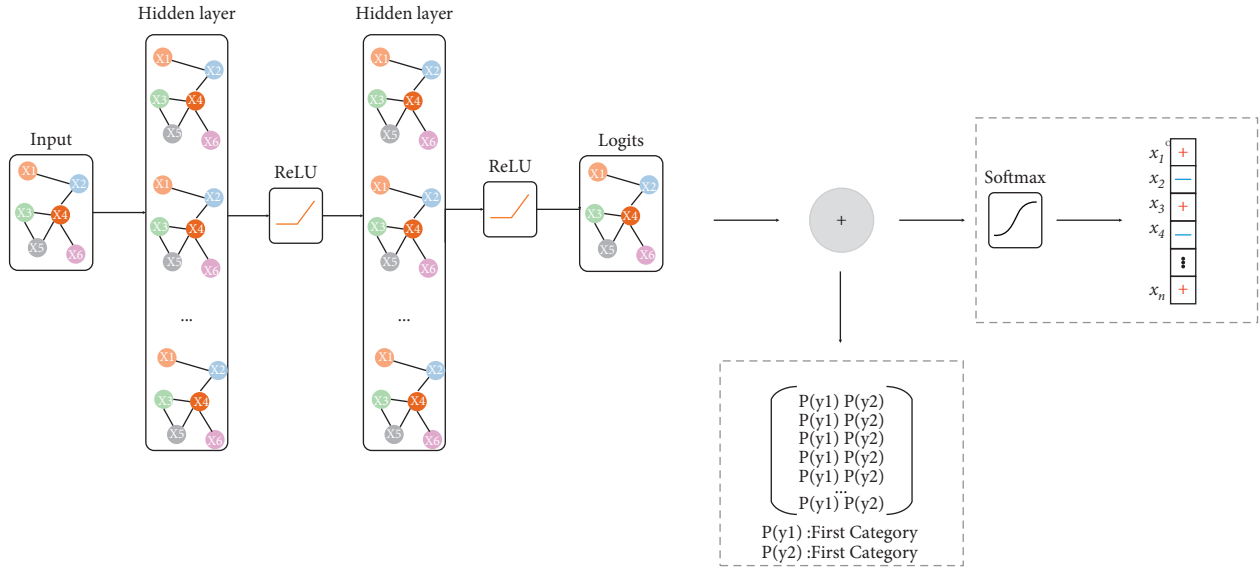
FIGURE 1: The process of training in a neural network.
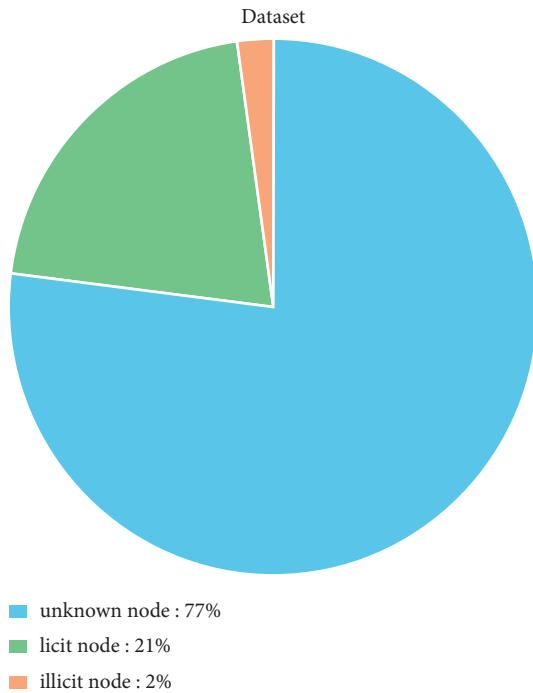
FIGURE 2: Proportion of node labels in the data set.

This is a dichotomous task, with a serious imbalance between legal and illegal categories. For the detection of illegal digital currency, it is more important for a few categories (i.e., illegal categories), so this article will focus on the F1-Score of illegal categories. We can take illegal as positive and legal as negative.

The calculation formula of Precision is as follows:

$$Precision = \frac{TP}{TP + FP}. \tag{19}$$

The Recall formula is as follows:

$$Recall = \frac{TP}{TP + FN}. \tag{20}$$

The F1-Score formula is as follows:

$$F1 - Score = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall}. \tag{21}$$

TP is truly indicating that the number is correctly divided into illegal. FP is false positive, indicating that legal numbers are classified as illegal. TN is true negative, which means the number of valid ones. FN is false negative, which means illegal is classified as a legal number.

4.3. Experiment and Result Analysis. The hardware environment of the experiment in this article is as follows: operating system Ubuntu 18.04.5LTS, CPU model Intel Xeon E5-2643, memory 318G, graphics card NVIDIA GeForce RTX 1080 8G *3, and GPU acceleration library CUDA10.0. The deep learning framework is PyTorch.

In this article, the Elliptic data set is divided into training set and test set according to the ratio of 6 : 1 : 3. That is to say, the first 31 graphs are used for training the model, and the 5 graphs from 31 to 36 are verified, and the 13 graphs are used for testing the model. The following are the experimental results obtained on the Elliptic data set in this article.

Since all the models lost their effectiveness after graph 43, the survey found that the United States had a serious crack down on digital currency crimes during that period, with only 1 or 2 of each chart flagged as illegal activities from graphs 43 to 49. Therefore, this article extracted data of different models and different loss functions from graphs 37 to 42 for comparison.

In this article, the Adam optimizer was used to set the learning rate to 0.001 to train the GCN and GAT models with 1000 epoch. Two hidden layers were used in the experiment, and the number of nodes in the hidden layer was set as 100. The prediction results and F1-Score of GCN, GAT, and two
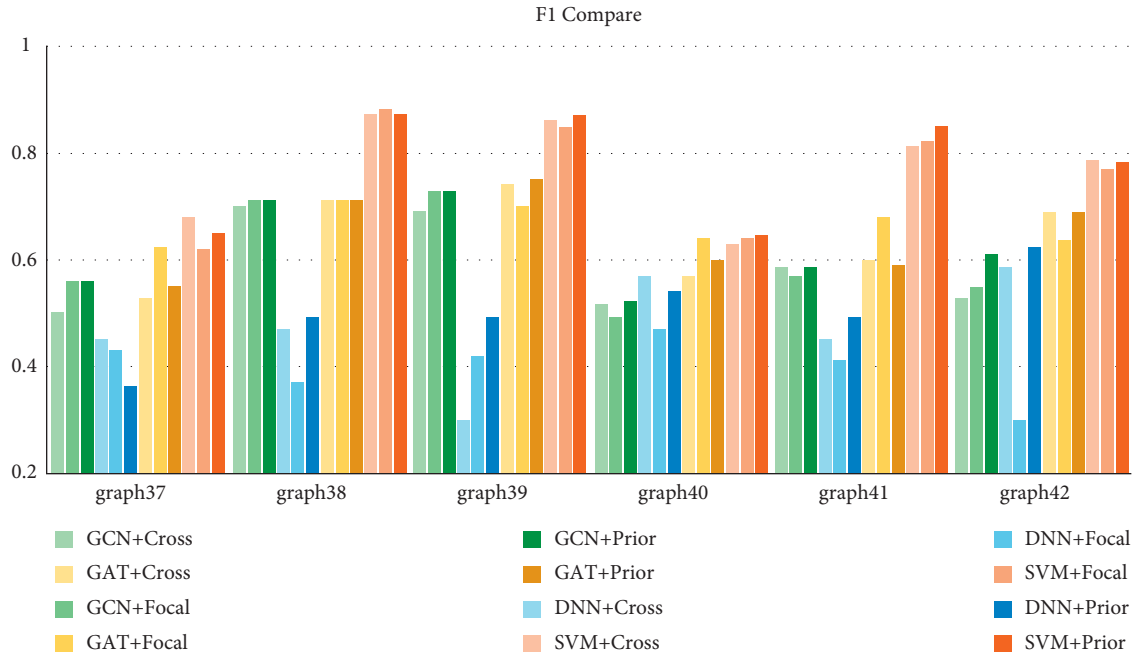
Figure 3: Experimental results of different methods on different loss.

Table 1: Graph 37–42 experimental results of different methods on different loss.

Model	Loss	Precision	Recall	F1
SVM	Cross	0.77	**0.77**	0.77
	Focal	0.84	0.70	0.76
	Prior	**0.86**	0.72	**0.78**
DNN	Cross	**0.43**	0.52	0.47
	Focal	0.38	0.43	0.40
	Prior	0.41	**0.67**	**0.50**
GCN	Cross	**0.70**	0.52	0.58
	Focal	0.68	0.55	0.60
	Prior	0.69	**0.58**	**0.62**
GAT	Cross	0.58	**0.71**	0.63
	Focal	**0.75**	0.59	**0.66**
	Prior	**0.75**	0.58	0.65

The bold values in the loss column are the methods we proposed, and the bold values in the prior are the best ones under the same evaluation index.

benchmark methods (DNN [21] (PyTorch implementation) and SVM [22] (PyTorch implementation)) are summarized in Figure 3.

According to the data in Figure 3, focal loss function on DNN and SVM did not improve but decreased compared with the ordinary cross-entropy loss function, but the loss function based on mutual information prior had a large improvement compared with it. On GCN, focal loss has a certain improvement compared with ordinary cross-entropy loss, and the loss based on mutual information prior has a certain improvement compared with focal loss. In GAT, it is found that although the loss based on mutual information prior is improved compared with the traditional cross-entropy loss, it is not as good as focal loss.

The average results of the indicators from graphs 37 to 42 are summarized in Table 1.

The following conclusions can be drawn from Table 1:

(i) Graph neural network is less effective than SVM; the reason may be that the node in the data set is not an account but a transaction. In normal graphs, nodes represent addresses and edges represent transactions, so the information learned using Graph Neural Network may be wrong.

(ii) The loss function based on mutual information prior in SVM improved by 1% compared with the traditional cross entropy loss F1-Score and 2% compared with focal loss F1-Score; in DNN, the cross entropy loss F1-Score increased by 3% compared with the traditional cross entropy loss F1-Score, and the F1-Score increased by 10% compared with focal loss; compared with the traditional cross entropy loss F1-Score in GCN, the F1-Score improved by 4%; compared with the focal loss F1-Score, the F1-Score improved by 2%. Compared

with the traditional cross entropy loss F1-Score, GAT increased by 2%, but compared with focal loss F1-Score, it decreased by 1%. The reason may be the conflict between the attention mechanism of GAT and the priori based on mutual information.

(iii) GAT has an overall improvement over GCN, indicating that introducing the attention mechanism into the graphics network is an effective operation.

It can be intuitively seen from the above chart that the loss function based on mutual information prior can effectively improve the problem of sample imbalance in the detection of digital currency illegal behaviors.

5. Conclusion

As for the serious imbalance of positive and negative samples in the detection of digital currency illegal behaviors, this study integrates mutual information and cross entropy loss function and obtains a loss function based on mutual information prior. On the Elliptic data set, a comparative experiment will be conducted whether to use the method of mutual information prior loss function. The results show that, using this loss function, the F1 value of each method is improved to varying degrees compared with the original cross-entropy loss function and focal loss function. Among them, GCN and DNN were significantly improved, and SVM also had a certain improvement. Although the improvement in GAT is not as good as focal loss, it also has a large improvement compared with the ordinary cross entropy loss. Therefore, the loss function based on mutual information prior proposed in this article plays a significant role in the detection of digital currency illegal behaviors.

Data Availability

The data sets used in this paper are publicly available.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] N. Abdelhamid, A. Ayesh, and F. Thabtah, "Phishing detection based Associative Classification data mining," *Expert Systems with Applications*, vol. 41, no. 13, pp. 5948–5959, 2014.

[2] A. A. Andryukhin, "Phishing attacks and preventions in blockchain based projects," in *Proceedings of the 2019 international conference on engineering technologies and computer science (EnT)*, pp. 15–19, IEEE, Moscow, Russia, March 2019.

[3] M. Bartoletti, S. Carta, T. Cimoli, and R. Saia, "Dissecting Ponzi schemes on e: identification, analysis, and impact," *Future Generation Computer Systems*, vol. 102, pp. 259–277, 2020.

[4] S. Meiklejohn, M. Pomarole, G. Jordan et al., "A fistful of bitcoins: characterizing payments among men with no names," in *Proceedings of the 2013 Conference on Internet Measurement Conference*, pp. 127–140, Torre Telefonica, in Barcelona, Spain, October 2013.

[5] M. Clouston, B. Haslhofer, and B. Dupont, "Ransomware payments in the bitcoin ecosystem," *Journal of Cybersecurity*, vol. 5, no. 1, 2019.

[6] T. Y. Lin, P. Goyal, R. Girshick et al., "Focal loss for dense object detection," in *Proceedings of the IEEE international conference on computer vision*, pp. 2980–2988, Venice, Italy, October 2017.

[7] C. Akcora, "Bitcoinheist: Topological data analysis for ransomware prediction on the bitcoin blockchain," in *Proceedings of the 2020 International Joint Conference on Artificial Intelligence (IJCAI)*, Yokohama, Japan, July 2020.

[8] W. Chen, X. Guo, Z. Chen et al., "Phishing scam detection on ethereum: towards financial security for blockchain ecosystem," in *Proceedings of the International Joint Conferences on Artificial Intelligence Organization*, pp. 4506–4512, Yokohama, Japan, July 2020.

[9] M. Weber, G. Domeniconi, J. Chen et al., "Anti-money laundering in bitcoin: experimenting with graph convolutional networks for financial forensics," 2019, https://arxiv.org/abs/1908.02591.

[10] K. Cao, C. Wei, A. Gaidon et al., "Learning imbalanced datasets with label-distribution-aware margin loss," 2019, https://arxiv.org/abs/1906.07413.

[11] M. Buda, A. Maki, and M. A. Mazurowski, "A systematic study of the class imbalance problem in convolutional neural networks," *Neural Networks*, vol. 106, pp. 249–259, 2018.

[12] F. Scarselli, M. Gori, A. C. Tsoi et al., "The graph neural network model," *IEEE Transactions on Neural Networks*, vol. 20, no. 1, pp. 61–80, 2008.

[13] J. Bruna, W. Zaremba, A. Szlam et al., "Spectral networks and locally connected networks on graphs," 2013, https://arxiv.org/abs/1312.6203.

[14] M. Defferrard, X. Bresson, and P. Vandergheynst, "Convolutional neural networks on graphs with fast localized spectral filtering," *Advances in Neural Information Processing Systems*, vol. 29, pp. 3844–3852, 2016.

[15] J. Chen, T. Ma, and C. Xiao, "Fastgcn: fast learning with graph convolutional networks via importance sampling," 2018, https://arxiv.org/abs/1801.10247.

[16] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," 2016, https://arxiv.org/abs/1609.02907.

[17] P. Veličković, G. Cucurull, A. Casanova et al., "Graph attention networks," 2017, https://arxiv.org/abs/1710.10903.

[18] A. Kraskov, H. Stögbauer, and P. Grassberger, "Estimating mutual information," *Physical Review*, vol. 69, no. 6, 2004.

[19] J. Su, "《Mutual Information to Alleviate the Problem of Category Imbalance》," 2020, https://kexue.fm/archives/7615.

[20] A. K. Menon, S. Jayasumana, A. S. Rawat et al., "Long-tail learning via logit adjustment," 2020, https://arxiv.org/abs/2007.07314.

[21] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006.

[22] C. Acortes and V. Vapnik, "Support vector networks," *Machine Learning*, vol. 20, no. 1, pp. 273–297, 1995.

Research Article

Detection, Integration, and Optimization of Acoustic Field Simulation in the Closed Space

Siming Meng [ID],[1] **Ge Lin,**[2] **and Xiaoyan Liang**[2]

[1]*Information Engineering Institute, Guangzhou Railway Polytechnic, Guangzhou 510430, China*
[2]*Sun Yet-sen University, Guangzhou 510006, China*

Correspondence should be addressed to Siming Meng; mengsiming@gtxy.edu.cn

Noise pollution in the closed space such as railway carriage is an important problem because the noise pollution seriously affects comfort and health of people in the closed space. We propose the method to detection, integration, and optimization of acoustic field simulation in the closed space. First, we analyze the acoustic field distribution in the virtual 3D close space. We use harmonic sound wave propagation in the closed space and present the distribution according to geometric analysis. Second, we introduce Delaunay triangulation and k-means clustering into visualization to form the quiet zone and show it in 3D perspective. Our method used acoustic simulation to develop the sound barrier system. The simulation results show that our method can improve the analysis of the noise problem in the closed space.

1. Introduction

The problem of noise in the closed space such as railway carriage is an important problem. It not only seriously affects the passengers' comfort but also damages their health. The researches study various methods to deal with the problem of noise pollution in the close space. The virtual sound barrier system (VSB-S) method [1–3] is the interesting method to improve the acoustic environment. VSB-S uses the acoustic interference method to eliminate noise pollution in the close space, so it will not disturb the space at all. In VSB-S, the noise sensor array [4] is used to obtain spatial noise distribution, and the sound waves of the same amplitude and opposite phase to the noise are emitted to counteract the noise. We call the area in the space of sound counteract is the quiet zone.

In general, wave-based methods and geometric methods [5] are used for sound propagation in the close space. The wave-based methods built solve the acoustic wave equations [6–8]. However, it is very slow to solve the problem of the high frequency sound source or open sound field. This is because the computational complexity of this method is proportional to the volume of the sound field space and the

fourth power of the maximum sound frequency. The geometric method uses straight line ray to describe sound propagation approximately [9]. We assume that noise is always along the ray propagation; the tangent direction of every point on ray is the direction of transmission of sound waves. The geometric acoustic methods usually cannot be described and are solved accurately by explicit equations, but they can efficiently simulate the sound transmission if ignored the reflected sound waves. So, we introduce the geometric acoustic approach for analysis of the acoustic field in a closed space.

On the basis of acoustic simulation, we use the computer simulation model to establish visual description of the sound field [10, 11]. Our method considers direct sound, reflected sound, and sound attenuation. We use the Delaunay triangulation model to get the silent area of VSB-S and use three-dimensional perspective to display the triangulation results. This study' contributions include two parts. First, we research the problem of sound comfort. In the past, the research on acoustics of the enclosed space mainly focused on hall acoustics and architectural acoustics. Then, the Delaunay triangulation method is preferred into wave-based acoustics to form the quiet zones in the enclosed space such

as railway carriage. We obtain a better subdivision result because of the application of the k-means clustering algorithm.

2. Related Work

Various methods have been proposed for the construction of the VSB system and visualization of sound field distribution. Due to the complexity of different types of sound waves and the multilateral nature of sound field environment, there is no simulation model that can adapt to all situations [12].

2.1. VSB System (VSB-S). Qiu et al. [1] used a group of loudspeaker arrays to build a virtual acoustic barrier applied to the driver's seat. They solve the problem of conflict with human's head using virtual sensor technology. Seyedin and Abedi [13] calculate sound propagation with complete reflection of sound waves from the walls by the FDTD method and estimate optimal amplitude to build the VSB system.

2.2. Sound Field Distribution Simulation. The technology of sound field [14–17] prediction in the closed space based on geometric acoustics is relatively mature and has been applied in many scenes. In a study [9], Cai et al. proposed a ray tracing method based on space partition to simulate the sound field in a closed space. In his method, the whole closed space is divided into finite convex polyhedral subspaces, so as to reduce the times of ray and wall intersection. In order to simplify the calculation, his method omits the validation of the collision point between the ray and the wall.

Based on the numerical method of sound wave propagation, equation is discretized, and then, the numerical calculation is carried out to approximate the solution [6]. There are many feasible methods to solve the wave equation, which are suitable for different application scenarios. The method of time domain solution (TDS) [7] depends on spatially invariant speed of sound. The method of an equivalent staggered grid scheme (ESG) [8] relies on variable density media. The adaptive rectangular decomposition method (ARD) [18] can combine the analytical solution of the rectangular subdomain wave equation with the finite difference template of interface processing between subdomains, so as to achieve high performance simulation with low error. A method of sound field simulation for spatial convolution circular array based on discrete Fourier transform (DFT) is proposed by Haneda et al. [19]. Tkamisinski [20] mentioned a visual simulation technology to analyze the influence of sound wave diffusion on music.

2.3. Delaunay Triangulation. Delaunay triangulation is a commonly used optimal triangulation method, which is widely used in mesh generation and surface reconstruction of a 3D scattered point set. It is one of the important research contents in CAGD, CGM, and CG. Delaunay triangulation has many advantages, such as maximizing the minimum angle and ensuring convergence. So, it can build the low complexity and high-quality mesh in the 3D space. Delaunay

triangulation is the most popular surface generation method and has a widespread application. Wang and Wu [21] and Turnbull et al. [22] used Delaunay triangulation to describe the water surface, for visible analyzing the interactions of water waves or others obstacle interfere. Yu et al. [23] showed a 3D geological modeling of Nanjing City used the Delaunay triangulation method.

2.4. Clustering Algorithm. Cluster analysis is an important method of data division or grouping. Common clustering methods include k-means, density-based clustering, maximum expectation clustering, hierarchical agglomerative clustering, and graph detection clustering. These cluster algorithms are proposed to solve different practical problems. Likas et al. [24] proposed a global clustering method. Through the deterministic global search process composed of dataset size, a cluster center is dynamically added in one search process to achieve global optimization. Accordingly, we use the k-means cluster algorithm [25] based on Euclidean distance in this study, with the number of clusters to ensure the precise zones.

We demonstrate the sound field distribution in close space slice as shown in Figure 1. The first line is with two sound sources while the second line is with three. In Figure 1, we use yellow point sign point noise source.

3. The Modeling of the Sound Field

In this study, we use a virtual closed cuboid space to simulate the sound field space [26]. In order to simplify the simulation modeling, we assume that there are no other obstacles and other interference sound sources in the space.

Our study uses a $7\,m \times 5\,m \times 3\,m$ cuboid to describe our virtual sound field space. Assuming that all the boundaries in the closed space are smooth rigid bodies, there is no diffuse reflection because it is so complex in our model [27]. At the same time, no other object in this virtual space should absorb or reflect the sound waves. Without considering any absorption and reflection, the sound propagation path between two points in the close space is a straight line.

The sound wave of the spherical point source propagates in the virtual space, and each point on the sound wave propagates on the uniform spherical surface in a certain period of time until it collides with the boundary of the space and reflects. Considering the complexity of the simulation model, when the noise reaches the boundary and reflects, we only consider the influence of the first radiated sound wave and ignore other sound waves.

3.1. Wave Equation of Sound. In this study, the sound field generated by a stable point source propagating in a closed space is studied. Through Fourier analysis [28], we can use some simple harmonic functions with different frequencies to express the acoustic vibration function. Therefore, the variation of the harmonic field with time is the basis of analyzing the variation of complex field with time. At any points in this closed space, the vibration information of the selected sound source [29, 30] is as follows:
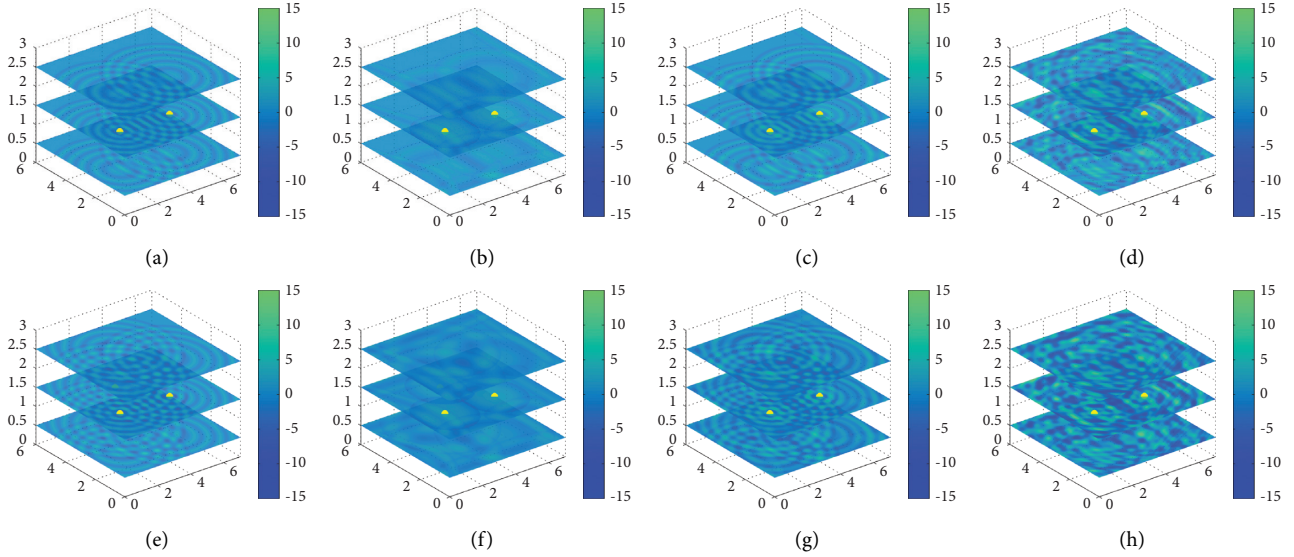
FIGURE 1: Examples of sound field distribution. The first line is with two sound sources while the second line is with three. Yellow points are sound source locations.

$$P(x, y, z; t) = e^{-2\alpha\gamma} A_0 \cos[w(t - t_0) + \varnothing_0], \quad (1)$$

where $t_0 = r/v$ is the time of sound wave transmission in the close space, r is the distance, v is the speed of sound in air, $t - t_0$ means that the vibration of the sound wave is beginning from the moment it spread to here, $\varnothing_0$ is the initial phase of the sound wave, and α is an attenuation factor.

3.2. Sound Wave Reflection.

It will get many reflected sound waves when reaching boundary in space. Figure 2 shows intersection of spherical sound waves from multiple point sources. Figure 2(a) shows spherical sound waves intersection from 2 point sources, and Figure 2(b) shows spherical sound waves intersection from 3 point sources. There are many factors affecting the sound field distribution. The sound field distribution after impact reflection is affected not only by the propagation direction of the point source but also by the propagation attenuation.

In our model, all the boundaries are smooth rigid bodies. After the point source sound waves collide and reflect on the boundary surface, they will be reflected at the same angle as the incident angle. Therefore, at any point in the close space, the value of acoustic intensity will be equal to the direct sound at that point and all the values reflected by all boundaries. In this study, the closed space environment is simulated as a virtual cuboid, and the boundary is regarded as a rigid reflector. In this way, we can ignore all kinds of attenuation [31–33] and only consider the boundary reflection.

3.3. Sound Wave Interference in the Closed Space.

A single proton vibrates longitudinally in a spherical acoustic wave. When sound waves with the same vibration frequency and amplitude collide, a superimposed energy will be produced. If the two protons vibrate in the opposite direction, the energy will cancel each other. We use p_1 and p_2 that express

sound pressures of two sound waves, and the synthetic sound field pressure is p. It can be calculated by the following formulas:

$$\nabla^2 (p_1 + p_2) = \frac{1}{c_0^2} p_1 \frac{\partial^2 (p_1 + p_2)}{\partial t^2}, \quad (2)$$

$$\nabla^2 (p_i) = \frac{1}{c_0^2} \frac{\partial^2 (p_i)}{\partial t^2}. \quad (3)$$

4. Quiet Area Integration

We detect the collision point by the feature of sparse and dense in the propagation of the point source. Then, we introduce Delaunay triangulation to represent the final quiet region due to the regularity and uniqueness of the surface formed by collision points. Finally, we use Euclidean distance for k-means clustering [34] to improve the final results.

4.1. Detection of Impact Points.

We assume that there are three sound source points in the closed space, and the frequency and amplitude of the sound source are the same. According to the stability of direct sound and the change of reflected sound, the collision position of sound wave in the space can be calculated [35, 36]. We consider the area of 1.0-2.0 meters in the vertical direction is our "sweet area." We discretize the sound field space and get the acoustic impact point through the sparse and dense characteristics of the discrete points.

4.2. Quiet Area Integration.

It is difficult to triangulate discrete points in space. The constancy of coherent wave provides a strong guarantee for our work. By controlling the wave length, we can get a good triangulation dataset.

(a)   (b)
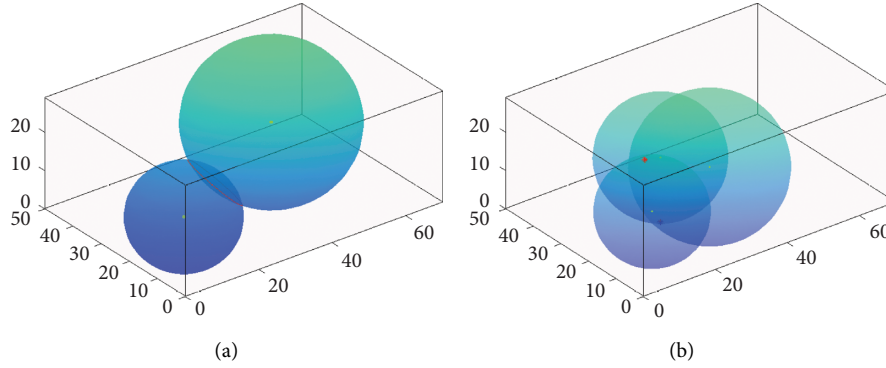
FIGURE 2: Intersection of spherical sound waves from multiple point sources. (a) Spherical sound waves intersection from 2 point sources, yellow points represent the locations of sound sources. (b) Spherical sound waves intersection from 3 point sources.

First, a super triangle is created to contain all the scattered points in the triangle list. Second, insert points one by one, find the triangle whose circumscribed circle contains the insertion point from the triangle list, and delete the common edge that has influence on it. Connect the insertion point with all the vertices of its related triangle to complete the insertion point into the Delaunay triangle list. Then, the new local triangles are optimized according to the optimization rules, and the generated triangles are put into the Delaunay triangle table. Loop the insertion until all points are inserted. Two triangular subdivision surfaces in the 3D space can be on the same plane. Let us first draw in two dimensions. During this process, some points may be deleted because they are on a line that is not referenced by triangulation.

The Delaunay triangulation result is shown in Figure 2.

4.3. Subzone Dividing. The impact point set we get from section discrete point detection may be distributing across the space sparsely [37, 38] or forming several dense parts distinctly. If we generate the quiet zones on the impact point set directly, we may get a surface that almost covers the whole space, which make our study become meaningless. In order to avoid this situation, we introduce the clustering algorithm before Delaunay triangulation to create space subsets. We implement surface subdivision from the subsets. With this method, we may get several precise quiet zones. We employ Euclidean distance to be the unique attribute work on the clustering algorithm. Any clustering algorithm can be used to obtain the final grouping of elements. We use the k-means algorithm to achieve data partition.

Iterating according to the following formulas could get labeling the initial dataset u.

$$C^{(i)} = \arg\min_j \left| d^{(i)} - u_j \right|^2, \tag{4}$$

where $d^{(i)}$ and u_j are the spatial coordinates vector. Its main point is to apply the distance between the original dataset and the initialization random data. Then, find out the data with the nearest distance to the initial data:

$$u_j = \frac{\sum_{i-1}^m l\{c^i = j\} d^i}{\sum_{i-1}^m l\{c^i = j\}}. \tag{5}$$

This formula aims at getting the average distance between all the original data and one initial data. Iterate the two formulas until the member in set u do not change anymore, and we get the labeling group dataset.

5. Experimental Results

5.1. Sound Field Distribution. Our approach obtains the visualization through the discrete points detection. When we calculate sound pressure of one point, we only consider the direct sound and the primary reflection. All boundaries we simulate are idealized rigid edge, and we do not put the fact into consideration that sound wave reflect from a wall would product a series of wavelet, which may cause standing wave in the whole space and diminish the reflection value [39–43]. Therefore, each time we calculate sound pressure of one point, and we should put the three direct sound and eighteen reflections into consideration. The two and three spherical waves intersection is shown in Figure 2.

We collect a data point per cubic decimeter; then, we have 118059 data to deal with. Figure 3 shows the results of visualization simulation and Delaunay triangulation [44, 45]. The first line shows the situation about two sound sources while the second line is about three. The first three columns are 0.001 s when the sound is not spreading in the whole room yet. The last column is 0.005 s when the room is full of sound waves. The third column is the sum of the first two columns. We could see that the first column is almost exactly the same with the third column because the value of reflection is so small to affect the entire waveform. But we could see that the waveform was not irregular anymore when time is at 0.005 s because of the strong reflection.

In our experiment, sound field distribution of a whole space in Figures 3(a) and 3(d) will cost about 1 and a half minutes, computing on the 64 bits system of Intel CPU, 16G RAM per time. Illustrated in Figures 3(a) and 3(d), through simulation, the distribution of sound field in space is visualized. Different sound intensities are marked with different colors.
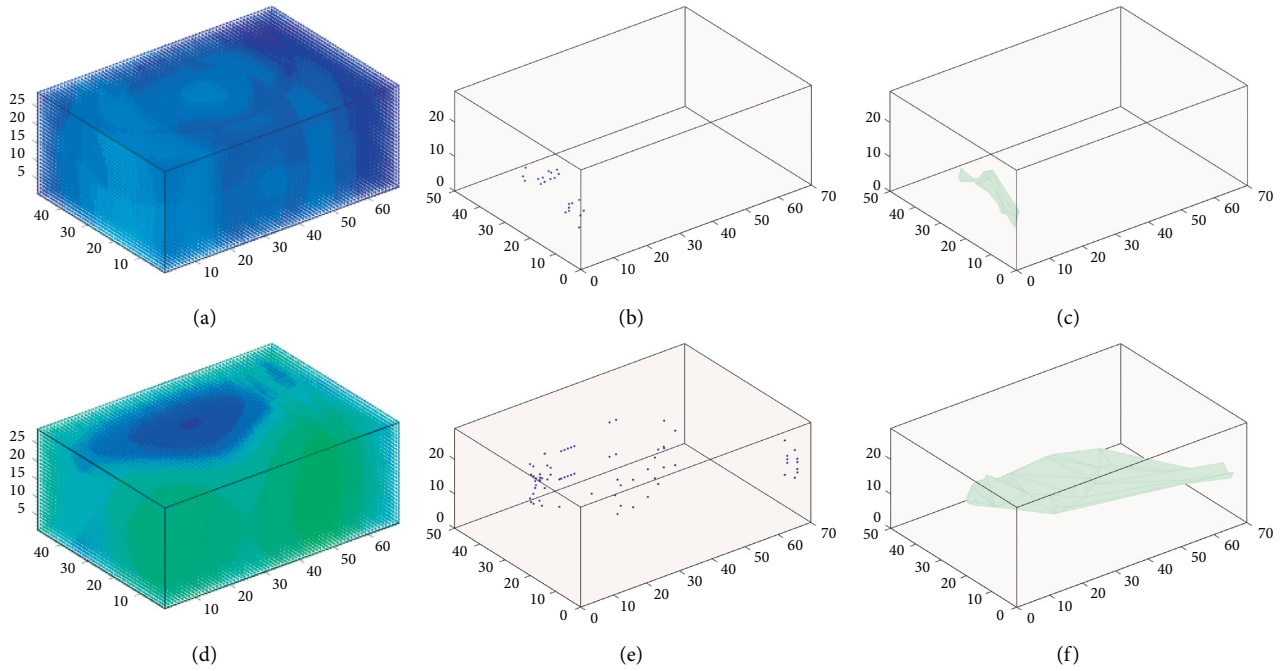
FIGURE 3: The results of visualization simulation and Delaunay triangulation. (a), (d) The distribution in $t = 0.01$ s and $t = 0.05$ s. (b), (e) The dataset of impact points in $t = 0.01$ s and $t = 0.05$ s. (c), (f) The Delaunay triangle surface in $t = 0.01$ s and $t = 0.05$ s.
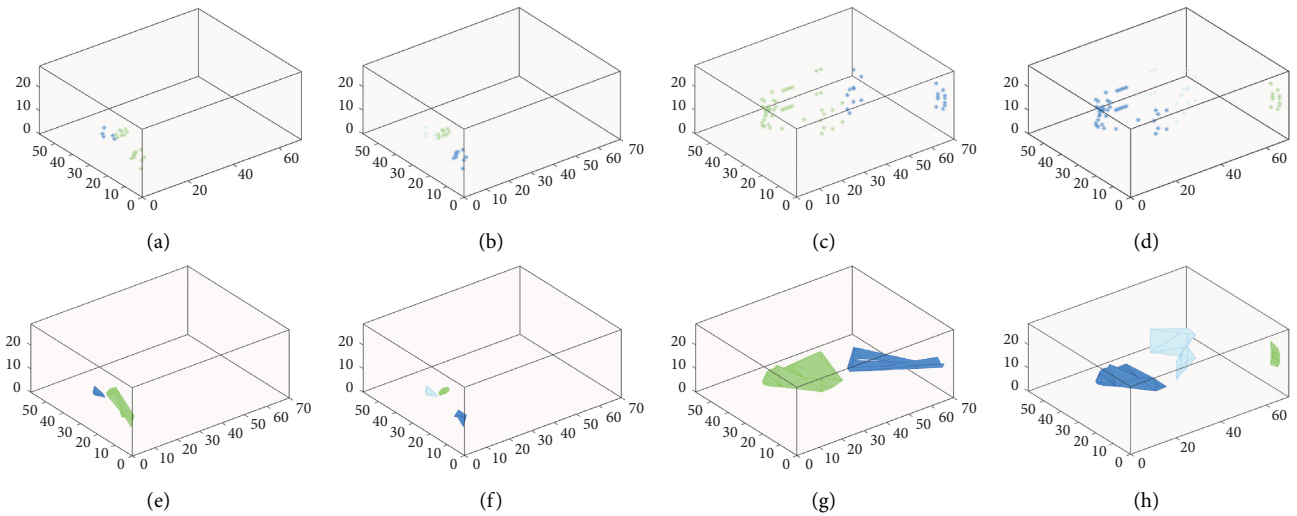
FIGURE 4: The results of cluster. (a)–(d) Cluster results of impact points at 0.01 s and 0.05 s. (e)–(h) Their cluster respective subdivision results.

5.2. Impact Points Detection.

We could simplify data detection. As shown in Figures 3(b) and 3(e), the amplitude of a point in vibration always approximate the point nearby because of the continuity of volatility. Because of this, there may be many impact points in a small region, which provides a powerful guarantee for Delaunay triangulation.

5.3. Delaunay Triangulation.

We map the set into the 2D space primarily. As a result, not all the impact points would be the vertex of the Delaunay triangle. Figures 3(c) and 3(f) show the final result of subdivision, and it is obvious that

some points have been removed. We can see the surface is so rough that it does not look like a zone.

5.4. Subzone Dividing.

The k-means clustering algorithm is used to distinguish the collision point set. The initial class number K, namely, the desired number of clusters, must be confirmed before clustering, and it is the user to specify parameters [46]. When we call the k-means function to do with the impact point set, it need to iterate based on the initial value. Iteration will not stop until the cluster does not change anymore.

As shown in Figure 4, we could get different subdivision results with different K values. As for the situation about Figure 3(f), we could improve it in Figures 4(g) and 4(h). The quiet zone we get could not cover large areas of points with high sound pressure values, and get the quiet zones in a close space. Figures 4(a) and 4(b) are cluster results of impact points at 0.01 s, while Figures 4(e) and 4(f) are their respective subdivision results. The others are time at 0.05 s.

6. Conclusion

With the problem of noise in closed a three-dimensional space, a simulation model of spherical wave propagation and interaction formed by multiple point sound sources is proposed and used to calculate the quiet area of the closed space. First, we analyze the sound field in a closed space with the geometric acoustics method and describe the intensity of the sound field with a color map. Second, Delaunay triangulation is introduced into the simulation, and finally, the quiet area is optimized by clustering.

First, we use the geometric wave acoustic analysis method to calculate the propagation of spherical sound wave, mainly considering the boundary reflection of smooth rigid body, analyzing the sound field distribution of a multipoint sound source and using the discrete method to detect the collision point. Then, Delaunay triangulation is used to form a continuous surface in the quiet area. Finally, the k-means clustering algorithm is used to optimize the quiet area. In the experiment, we give the visual description of different time and sound sources. The simulation results show that geometric analysis can directly and effectively analyze sound field distribution.

In the future research, we will learn from the near space physical signal acoustic suppression processing system and modify the experimental parameters according to the actual measurement data, so that the system has a practical application value. At the same time, we will verify the effectiveness of the model in complex noise scenes such as train compartments.

Data Availability

The dataset used to support the findings of this study is publicly available.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] X. Qiu, N. Li, and G. Chen, "Feasibility study of developing practical virtual sound barrier system," in *Proceedings of the 12th International Congress on Sound and Vibration*, Lisbon, Portugal, July 2005.

[2] H. Zou, X. Qiu, J. Lu et al., "A preliminary experimental study on virtual sound barrier system," *Journal of Sound and Vibration*, vol. 307, no. 1, pp. 379–385, 2007.

[3] H. Zou and X. Qiu, "Performance analysis of the virtual sound barrier system with a diffracting sphere," *Applied Acoustics*, vol. 69, no. 10, pp. 875–883, 2008.

[4] H. Zou, X. Qiu, F. Niu, and J. Lu, "A numerical and experimental study on virtual sound barrier," *Acta Acoustica*, vol. 32, no. 1, pp. 26–33, 2007.

[5] L. Antani, A. Chandak, and M. Taylor, "Direct-to-Indirect acoustic radiance transfer," *Visualization and Computer Graphics*, vol. 18, no. 2, pp. 261–269, 2011.

[6] N. Raghuvanshi, N. Galoppo, and M. C. Lin, "Accelerated wave-based acoustics simulation," in *Proceedings of the 2008 ACM symposium on Solid and physical modeling*, pp. 91–102, ACM, Stony Brook, NY, USA, June 2008.

[7] R. Mehra, N. Raghuvanshi, L. Savioja, M. C. Lin, and D. Manocha, "An efficient GPU-based time domain solver for the acoustic wave equation," *Applied Acoustics*, vol. 73, no. 2, pp. 83–94, 2012.

[8] L. Di Bartolo, C. Dors, and W. J. Mansur, "A new family of finite-difference schemes to solve the het-erogeneous acoustic wave equation," *Geophysics*, vol. 77, no. 5, pp. 187–199, 2012.

[9] M. Cai, Z. Chen, and J. Zou, "A ray tracing method based on indoor space partition-ing," *Applied Acoustics*, vol. 32, no. 1, pp. 34–38, 2013.

[10] X. Wang, X. Liu, Q. Zou, and L. Ge, "Multichannel LMS-based large area sound inhibition in home environment: preliminary simulation and experiments," in *Proceedings of the International Conference on Digital Home*, pp. 401–405, Guangzhou, China, November 2014.

[11] X. Liu, Q. Zou, X. Wang, and L. Ge, "Acoustic feedback effect reduction based on path compensation in multichannel FXLMS sound in-hibition systems," in *Proceedings of the International Conference on Digital Home*, pp. 390–394, Guangzhou, China, November 2014.

[12] Q. Q. Li, C. Ripamonti, and R. Caccialanza, "Simulation of deterministic tyre noise based on a monopole substitution model," *Applied Acoustics*, vol. 178, 2021.

[13] S. A. Seyedin and M. H. Abedi, *Optimizing Secondary Source Location in Acoustic ANC Systems*, IEEE, Manhattan, NY, USA, 2006.

[14] T. Jasa and N. Xiang, "Nested sampling applied in Bayesian room-acoustics decay analysis," *Journal of the Acoustical Society of America*, vol. 132, no. 5, pp. 3251–3262, 2012.

[15] K. Kowalczyk and M. Van Walstijn, "Room acoustics simulation using 3-D compact explicit FDTD schemes," *IEEE Transactions on Audio Speech and Language Processing*, vol. 19, no. 1, pp. 34–46, 2011.

[16] C. J. Webb and S. Bilbao, "Computing room acoustics with CUDA-3D FDTD schemes with boundary losses and viscosity," in *Proceedings of the Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*, pp. 317–320, Prague, Czech Republic, May 2011.

[17] D. T. Murphy, A. Southern, and L. Savioja, "Source excitation strategies for obtaining impulse responses in finite difference time domain room acoustics simulation," *Applied Acoustics*, vol. 82, pp. 6–14, 2014.

[18] L. Antani, A. Chandak, M. Wilkinson et al., "Validation of adaptive rectangular decomposition for three-dimensional wave-based acoustic simulation in architectural models,"

Proceedings of Meetings on Acoustics, vol. 19, no. 1, pp. 015–099, 2013.

[19] Y. Haneda, K. i. Furuya, S. Koyama, K. Niwa, and K. Kobayashi, "Sound field simulation for circular array based on spatial circular convolution," *Acoustical Science and Technology*, vol. 35, no. 2, pp. 99–107, 2014.

[20] T. Kamisinski, "Acoustic simulation and experimental studies of theatres and concert Hall-s," *Acoustic and Biomedical Engineering*, vol. 118, no. 1, pp. 78–82, 2010.

[21] C. Z. Wang and G. X. Wu, "Interactions between fully nonlinear water waves and cylinder arrays in a wave tank," *Ocean Engineering*, vol. 37, no. 4, pp. 400–417, 2010.

[22] M. S. Turnbull, A. G. L. Borthwick, and R. Eatock Taylor, "Wave-structure interaction using coupled structured-unstructured finite element meshes," *Applied Ocean Research*, vol. 25, no. 2, pp. 63–77, 2003.

[23] S.-J. Yu, F. Guo, L. Xiang, and F. Xu, "Application of constrained Delaunay tetrahedralization in 3D geological modeling," *Geography and Geo-Information Science*, vol. 29, no. 1, pp. 41–44, 2013.

[24] A. Likas, N. Vlassis, and J. Verbeek, "The global k-means clustering algorithm," *Pattern Recognition*, vol. 36, no. 2, pp. 451–461, 2003.

[25] G. H. Du, Z. M. Zhu, and X. Gong, *Fundamentals of acoustics*, Press of Nanjing University, Nanjing, China, 2012.

[26] E. Deines, M. Bertram, J. Mohring et al., "Comparative visualization for wave based and geometric acoustics," *Visualization and Computer Graphics*, vol. 14, no. 6, pp. 1707–1722, 2008.

[27] S. Siltanen, T. Lokki, S. Tervo, and L. Savioja, "Modeling incoherent reflections from rough room surfaces with image sources," *Journal of the Acoustical Society of America*, vol. 131, no. 6, pp. 4606–4614, 2012.

[28] A. Torras-Rosell, S. Barrera-Figueroa, and F. Jacobsen, "Sound field reconstruction using acousto-optic tomography," *Journal of the Acoustical Society of America*, vol. 131, no. 5, pp. 3786–3793, 2012.

[29] J. Yu, H. Pu, J. Chen, and S. Lu, "The measurement of sound speed and the study of sound waves," *College Physics*, vol. 26, no. 1, pp. 58–61, 2007.

[30] Y. Ji, H. Pu, J. Chen, and S. Lu, "The measurement of sound speed and the study of sound waves," *College Physics*, vol. 26, no. 1, pp. 58–61, 2007.

[31] W. Rao, "A study of effects of boundary and directional sound sources on virtual sound v-barrier system," Master's thesis, Master's thesis of Nanjing university, Nanjing, China, 2011.

[32] J. Peng, "Finite difference time domain method and its application in room acoustical simulation," *Technical Acoustics*, vol. 28, no. 1, pp. 53–57, 2009.

[33] G. Yan and J. Lin, "Continuous element for 1D problems," *Partial Differential Equations and the Finite Element Method*, vol. 21, no. 1, pp. 45–102, 2005.

[34] A. Likas, N. Vlassis, and J. Verbeek, "The global k-means clustering algorithm," *Pattern Recognition*, vol. 36, no. 2, pp. 451–461, 2003.

[35] S. Yu, F. Guo, X. Li, and F. Xu, "Application of constrained Delaunay tetrahedralization in 3D geological modeling," *Geography and Geo-Information Science*, vol. 29, no. 1, pp. 41–44, 2013.

[36] A. Bassuet, D. Rife, and L. Dellatorre, "Computational and optimization design in geometric acoustics," *Building Acoustics*, vol. 21, no. 1, pp. 75–86, 2014.

[37] W. Han, X. Bu, and M. Xu, "Yunpu Zhu Model of a surface acoustic wave sensing system based on received signal strength indication detection," *Measurement Science and Technology*, vol. 32, no. 8, pp. 85–103, 2021.

[38] E. Vevek and W. L. Chan, "Adomain decomposition technique for small amplitude wave interactions with shock waves," *Journal of Computational Physics*, vol. 437, 2021.

[39] C. Chen, T. Chen, and Y. Wang, "Observation of topological locally resonate and Bragg edge modes in a two-dimensional slit-typed sonic crystal," *Applied Physics Express*, vol. 12, no. 9, Article ID 097001, 2019.

[40] J. Lu, C. Qiu, L. Ye et al., "Observation of topological valley transport of sound in sonic crystals," *Nature Physics*, vol. 13, no. 4, pp. 369–374, 2017.

[41] S. Kang and J. Hwang, "Tuning the characteristics of photoacoustic pressure in a laser-induced photoacoustic generator: a numerical study," *Applied Mathematical Modelling*, vol. 94, pp. 98–116, 2021.

[42] J. Liu, J. Du, J. Wang, and J. Yang, "Effects of surface impedance on current density in a piezoelectric resonator for impedance distribution sensing," *Applied Mathematics and Mechanics*, vol. 42, no. 5, pp. 677–688, 2021.

[43] J. Liu, J. Du, J. Wang, and J. Yang, "Thin film bulk acoustic wave piezoelectric resonators with circular ring driving electrodes for mass sensing," *Integrated Ferroelectrics*, vol. 192, no. 1, pp. 57–66, 2018.

[44] P. Caday, M. V. de Hoop, V. Katsnelson, and G. Uhlmann, "Recovery of discontinuous Lame parameters from exterior Cauchy data," *Communications in Partial Differential Equations*, vol. 46, no. 4, pp. 680–715, 2020.

[45] S. Kugunavar and C. J. Prabhakar, "Content-based medical image retrieval using Delaunay triangulation segmentation technique," *Journal of Information Technology Research*, vol. 14, no. 2, pp. 48–66, 2021.

[46] K. Trojachanec, I. Kitanovski, I. Dimitrovski, and S. Loshkovska, "Longitudinal brain MRI retrieval for alzheimer's disease using different temporal information," *IEEE Access*, vol. 6, pp. 9703–9712, 2018.

Research Article

Personality Trait Detection Based on ASM Localization and Deep Learning

JinFeng Fu[1] and Hongli Zhang [2]

[1]*School of Marxism, Heilongjiang University of Science and Technology, Harbin 150022, China*
[2]*Department of Mathematics and Computer Science, Tongling University, Anhui, Tongling 244061, China*

Correspondence should be addressed to Hongli Zhang; zhl0@tlu.edu.cn

Global competition is the competition of human resources, the social demand for high-quality talents is increasing, and the demand for all kinds of talents is increasing. Therefore, how to scientifically and efficiently complete the preliminary screening of college students' mental health, so as to provide services for them, has become an important task. In order to solve the above problems, by combining the relevant professional knowledge of psychology, statistics, image processing, and artificial intelligence technology, a personality trait detection method based on active shape model (ASM) localization and deep learning is proposed. Firstly, the traditional ASM algorithm is improved and applied to facial feature point location, which provides training basis for further deep learning. It mainly includes three aspects of improvement: (1) 2D texture model based on Gabor wavelet and gradient features; (2) new multiresolution pyramid decomposition method; and (3) improved multiresolution pyramid search strategy. Secondly, the deep belief network model is used to train and classify the students' four personality traits and facial features, so as to dig out the relationship between the four personality traits and facial features. The experimental results show that the localization effect of the improved ASM algorithm is obviously better than that of the traditional algorithm, and the classifier after learning and training has a good effect in analyzing the four personality traits.

1. Introduction

In essence, global competition is the competition of human resources. Therefore, the social demand for high-quality talents is increasing, and the demand for all kinds of talents is increasing. Therefore, how to scientifically and efficiently complete the preliminary screening of college students' mental health, so as to provide services for them, has become an important task.

As a cross-discipline, management science and engineering include many research fields [1–5]. In order to classify the personality characteristics of talents accurately and quickly, management science and engineering need to combine relevant professional knowledge such as psychology, statistics, image processing, and artificial intelligence technology to establish a personality classification tree based on facial features to achieve efficient personality trait

classification. Thus, we can understand the potential psychological problems of college students at the first time, effectively improve their mental health, and help quickly find suitable roles according to their personality.

Face can best reflect a person's immediate mental outlook; face can express emotion and can better reflect a person's character and life [6–8]. The results show that, just like adults, even 3-year-olds tend to judge their personality traits, such as credibility and ability, by their faces, and children show surprising consistency with adults in judgment. With the increase of age, the tendency to judge people by their appearance also develops, and the older children are, the more similar their facial judgments are to adults. These psychological studies show that people's personality and even in a certain aspect of the ability and social behavior. The character of a character refers to the relatively stable and core psychological characteristics of his attitude and behavior in

the face of real things. It is a kind of personality characteristic with the highest correlation with society, which is closely related to facial features [9–12]. The features of human face can be regarded as the distance between facial contours, eyebrows, eyes, nose, mouth, and other feature points, including the distance between eyebrows and eyes, eyes and nose, nose and mouth, even the distance between the corner of the eye and the inner canthus, the left and right pupil, the corner of the left and right mouth, and so on. How to associate human character with complex facial features and find the relationship between them is the primary problem to be solved in this study.

In psychology, there are many ways to measure a person's personality characteristics, including psychological counseling, psychological scale testing, and other means [13, 14]. Different from the existing methods, on the basis of traditional personality test, this study combines face recognition technology and data mining technology to quickly and accurately find the relationship between facial features and personality traits and establish a student personality classification tree based on facial features. The main contributions of this paper are as follows:

(1) The traditional ASM algorithm is improved in three aspects, and the improved algorithm is applied to the facial feature point location, which solves the shortcomings of the traditional ASM algorithm, such as inaccurate location, low efficiency, and easy to fall into local minimum.

(2) With the help of the characteristics of deep learning unsupervised learning classification, the sample data are trained by deep belief network model to obtain a classifier based on four personality traits and facial features.

(3) Finally, through the comparative analysis of the classification results of psychological personality test, it verifies the feasibility and effectiveness of this method and provides a scientific basis for college teachers to make a reasonable psychological counseling plan.

The rest of the paper is organized as follows. In Section 2, a literature review is studied in detail, while Section 3 provides the detailed methodology. Section 4 provides detailed results and discussion. Finally, the paper is concluded in Section 5.

2. Literature Review

Firstly, face feature point location technology is the main content of this study. According to the number of feature points, facial feature point location methods can be divided into two categories: local feature point location method and global feature point location method.

The local feature point location method can only locate a small number of feature points at a time, and it is often used to locate local facial features. In [15], Akash et al. proposed a method to locate the center of pupil and mouth based on YCbCr color space distribution. The advantage of this method is that the location of human face can be determined quickly by determining the position of three points.

Different from the local feature point location method, the global feature point location method can locate many feature points at one time. The more popular methods include active shape model (ASM) and active appearance model (AAM). In [16], Wang et al. proposed to recognize human emotion through facial images. ASM method is used to reconstruct facial shape, extract facial feature points, and then use radial basis function network to recognize facial expression. In [16], Wang et al. used ASM in the driving fatigue detection system and introduced the average synthetic accurate filter algorithm to improve the positioning accuracy and attitude adaptability of the eye.

Secondly, for deep learning technology. The acquisition of learners' personality characteristics is an important prerequisite for the realization of student-centered accurate and personalized teaching, and learning behavior is an important basis for the analysis of learners' personality characteristics. Deep learning is a kind of machine learning. The main step of machine learning is to obtain the initial data through the sensor, and after the data is preprocessed, the data features are extracted and selected. After feature expression, the process of reasoning, prediction, and recognition can be carried out.

In [17], Zhao et al. used Rapid Miner data mining tools to explore the recognition effects of decision tree, naive Bayes, and support vector machine on five personality traits. The results show that the accuracy of the decision tree algorithm for the recognition of personality traits is higher than that of the other two algorithms, and the comprehensive recognition effect of the big five personality traits is the best. The sensitivity of recognition of different personality traits is different. The sensitivity of conscientious personality trait type is the highest, while that of neurotic personality trait is the lowest.

As a classical neural network learning method, deep belief network learning method can be regarded as a neural network composed of multiple restricted Boltzmann machine layers. Among them, the network is limited to two layers, namely, the data visual layer and the hidden layer. And only there is a connection between layers, and the nodes in the network exist independently. Among them, the data in the hidden layer can be trained to obtain the data in the data visual layer.

3. Methodology

3.1. Student Face Feature Point Localization Algorithm Based on Improved ASM. In the process of ASM modeling, it is very important to calibrate the feature points of the training sample image. The accuracy of feature point calibration of sample image is directly related to the representativeness of the shape model. Figure 1 shows an example of how this article has marked a set of pictures. Each image is manually calibrated with 32 feature points.

3.1.1. 2D Texture Model Based on Gabor Wavelet and Gradient Feature. The traditional ASM algorithm uses 1D

Training sample 1   Training sample 2   Tagged sample 1   Tagged sample 2

(a)   (b)   (c)   (d)

FIGURE 1: A set of tagged images. (a) Training sample 1. (b) Training sample 2. (c) Tagged sample 1. (d) Tagged sample 2.

gray information near the feature points as matching features and takes the distance between the gray value vector at the target candidate points and the local gray model at the mark points as the similarity measure between the candidate points and the mark points [18–21]. In the search process, only the low-frequency information of the face is used. The low-frequency components mainly describe the global features of the object, while the high-frequency components mainly describe the local details. Therefore, the calculation amount of this algorithm is small, but the positioning is not very accurate [22–24].

In order to solve the above problems caused by the 1D texture model, we introduce a new method to build the local model of feature points, that is, to build a 2D local texture model for each feature point, as shown in Figure 2.

As shown in Figure 2, the current feature point can reach its target position $P1$, so the positioning error will be reduced. For each marked feature point, select a rectangular region with side length of $m \times n$, and the current feature point is the center of the rectangular region. By extracting the Gabor feature or gradient feature of each point in the rectangular region, the corresponding Gabor texture model or gradient texture model of the feature point can be obtained [25]. Since there are N face images in the training set, for each feature point, N rectangular regions can be obtained; that is, N 2D texture models can be obtained. Each 2D texture model is arranged row by row into a long vector, and all the 2D texture models form a matrix with N rows and $m \times n$ columns.

Gabor wavelet transform can extract spatial frequencies and local features in multiple directions in the local region of the image [26]. Therefore, Gabor feature description of human face can achieve a good recognition effect. In image processing and computer vision, two-dimensional Gabor wavelet is mainly used. The two-dimensional Gabor wavelet is expressed as follows:

$$W(x, y; \theta, \lambda, \varphi, \sigma, \gamma) = \exp\left(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right)\cos\left(2\pi\frac{x'}{\lambda} + \varphi\right),$$

(1)

where (x, y) represents the pixel position in the time domain, $x' = x\cos\theta + y\sin\theta$, $y' = -x\sin\theta + y\cos\theta$. θ denotes the direction parameter, λ denotes the wavelength parameter of the triangle cosine function, φ denotes the phase offset parameter, σ denotes the standard deviation of
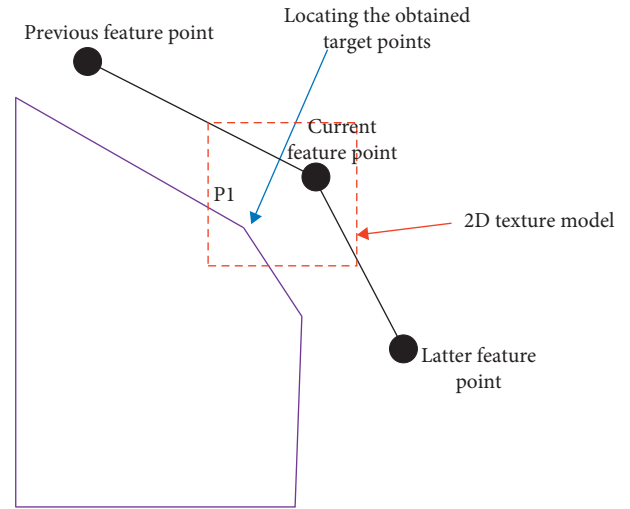
FIGURE 2: 2D local texture model.

the Gaussian function, and γ denotes the length-width ratio of the Gabor kernel. Two-dimensional Gabor wavelets with different shapes can be generated by selecting different parameters.

We use 40 Gabor wavelet functions combined with 5 frequencies and 8 directions to carry out the convolution operation on a point in the image, and 40 complex coefficients can be obtained, which are called the Gabor jet vector of the point. The 40 Gabor wavelet kernel functions adopted are shown in Figure 3.

Traditional ASM algorithm is to establish a local gray statistical feature texture model in a small area near the control point (sampling in the normal direction). Studies show that using gray scale as a local feature requires a lot of training sample images to obtain sufficient statistical information, and gray scale feature without gradient as a local feature has better robustness [27]. Since gray scale features are sensitive to illumination, in order to reduce the impact of illumination on positioning accuracy, and due to the limited number of training sample sets, a texture model based on 2D gradient features is established for each feature point of the highest level and sub-high-level pyramid subgraph of each image.

The 2D gradient texture model of each marker is generated by the following two steps: convolution and normalization.

FIGURE 3: Schematic diagram of 40 Gabor nuclei.

(1) Convolution. In this step, a 3 ∗ 3 convolution mask is used to produce a texture matrix for each pixel of the rectangular region (size $m \times n$) around the marker. Each element of the texture matrix corresponds to a pixel convolution value. The value of the element is a representation of some image feature of the pixel. This feature is determined by the convolution mask.

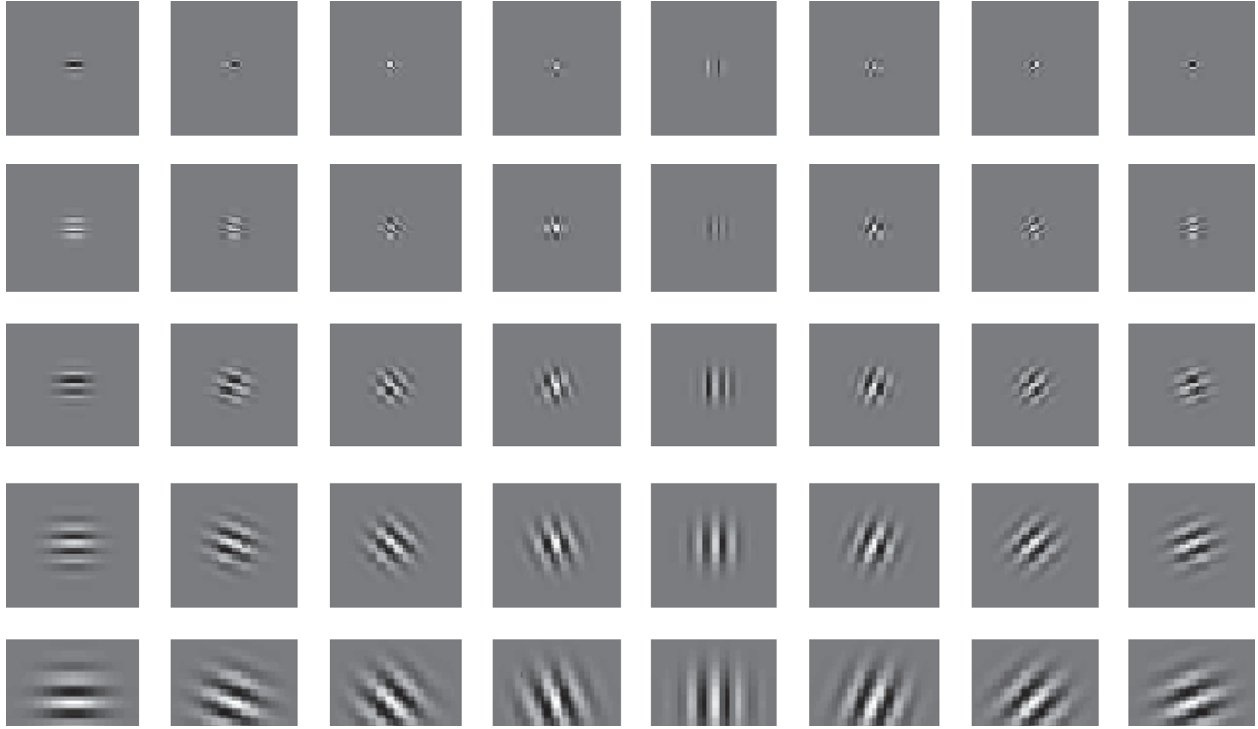Because the gradient features of feature points are to be extracted in this paper, the gradient mask is used, and the gradient mask matrix is

$$\mathbf{A} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 2 & -1 \\ 0 & -1 & 0 \end{bmatrix}. \qquad (2)$$

By convolution, the gradient texture matrix of the feature point can be obtained, which is 2D.

(2) Normalization Eliminates. In order to eliminate the influence of illumination and contrast, the gradient texture matrix also needs to be normalized. Each element of the gradient texture matrix is normalized by dividing the absolute value of all elements in the matrix.

By doing the same operation for the same marked point of each training sample image, N 2D gradient texture matrices can be obtained. Here, each 2D gradient texture matrix is still arranged into a long vector by row, and all the long vectors form a matrix, which has N rows and $m \times n$ columns. At this time, the mean and covariance matrices of 2D gradient texture matrix are obtained, and the mean and covariance matrices are used as the matching features of this point.

It is proved that using gradient information instead of gray information to build local feature texture model can effectively improve the matching accuracy and reduce the number of sample images to be trained.

3.1.2. A New Multiresolution Pyramid Decomposition Method.

In the training and searching phase of ASM under the framework of multiresolution analysis, the traditional ASM method first performs Gaussian filtering on the initial sample image and then samples the filtered image at intervals to get the pyramid image of the training sample. The resolution of these images decreases in turn, and the scale of the latter image is 1max 2 of the previous one. The initial sample image is Level 0 (layer 1). With the gradual decrease of resolution, the pyramid subimage obtained by this method contains less and less detail features, which is not conducive to the extraction of rich texture features, thus affecting the accuracy of feature point location.

In this paper, the size of the training picture and the test picture is 480 ∗ 640. The pyramid subimage is obtained by using db1 wavelet to decompose the sample image with 3-layer wavelet, as shown in Figure 4.

The sample image is decomposed by Gaussian filter of original ASM algorithm in three layers, and the pyramid subgraph obtained is shown in Figure 5.

It can be clearly seen from the experimental results of the above two graphs that, with the gradual reduction of resolution, the pyramid subgraph obtained by wavelet decomposition contains more detailed information than that obtained by original ASM. The rich detailed information is beneficial to the extraction of rich local texture features and

zero-order subgraph

(a)

first-order subgraph

(b)

secondary subgraph

(c)

tertiary subgraph

(d)

Figure 4: Pyramid subgraph obtained by wavelet decomposition. (a) Zero-order subgraph. (b) First-order subgraph. (c) Secondary subgraph. (d) Tertiary subgraph.
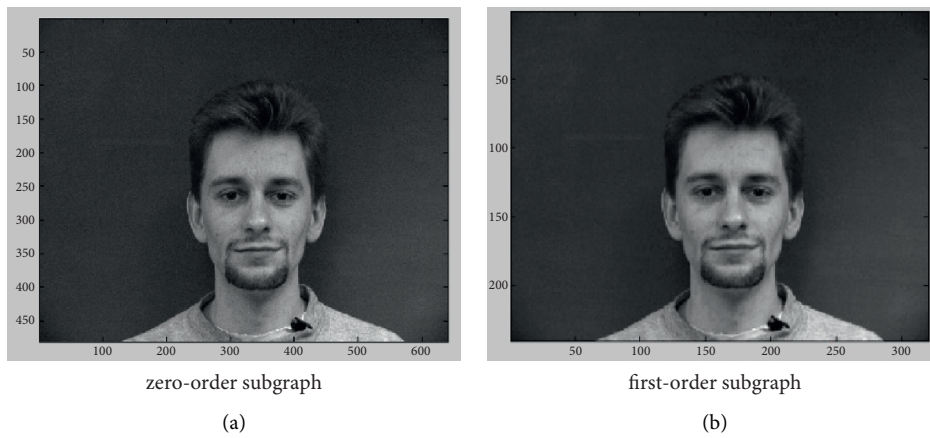

zero-order subgraph

(a)

first-order subgraph

(b)

Figure 5: Continued.

secondary subgraph

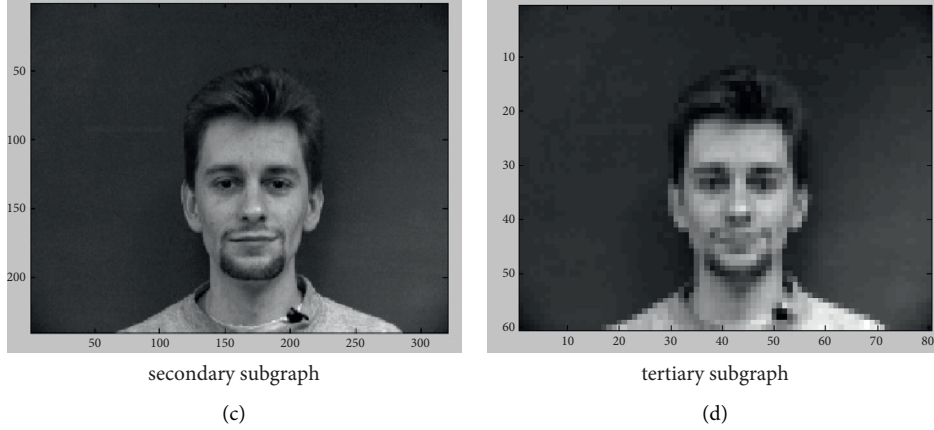(c)

tertiary subgraph

(d)

FIGURE 5: Pyramid subgraph obtained by Gaussian filter decomposition. (a) Zero-order subgraph. (b) First-order subgraph. (c) Secondary subgraph. (d) Tertiary subgraph.

the establishment of reliable texture model, so as to improve the accuracy of feature point localization.

3.1.3. Improved Multiresolution Pyramid Search Strategy.

In the traditional ASM algorithm, when multiresolution search is carried out, the search length of different layers is the same; that is, the same number of pixels are sampled for each feature point in each contour texture model of each layer. In this paper, the first layer (lowest layer) of the multiresolution pyramid is set as the original image, and the second layer (second layer) and the third layer (highest layer) are set as the pyramid subgraph obtained by wavelet decomposition.

This paper proposes to adaptively change the length of 2D texture search rectangle according to the different number of pyramid decomposition layers. Under low resolution, because some details of the image are ignored and the feature points are far away from the actual target position, a larger search range can be set, so as to obtain an ideal matching result for the overall contour. Correspondingly, at a higher resolution, because the feature points are close to the target position, the search range can be appropriately reduced, so as to obtain a relatively ideal local matching result. The edge length of 2D texture search rectangle adaptively decreases as the number of layers of multiresolution pyramid decreases. In this paper, we set the length of the search rectangle on the third layer to be L, the second layer to be 1/2, and the first layer to be $L/4$. The efficiency of this algorithm is improved because the computation is small at lower resolution.

3.2. Personality Trait Analysis Based on Deep Belief Network Model and Facial Features.

In order to explore the potential relationship between the four personality traits and facial features, we trained and learned the sample data through the deep belief network model, in order to obtain a trained

classifier based on the four personality traits and facial features of students.

3.2.1. Deep Belief Network Model.

In the field of computer vision, deep learning technology has shown its excellent performance. However, sometimes it is difficult to learn directly from the whole image using standard deep learning techniques, especially with complex data such as high-definition images. As an efficient deep learning algorithm, Deep Belief Networks (DBN) have gradually developed into a mainstream technology direction. Based on the principle of statistical mechanics, a random neural network Boltzmann machine model is generated, including a hidden layer and a visible layer, as shown in Figure 6.

On this basis, the principle architecture of multilayer restricted Boltzmann machine (RBM) is proposed, as shown in Figure 7, where $\mathbf{a} = (a_1, a_2, \ldots, a_{n_v})^T \in R^{n_v}$ is the bias vector of the visible layer, $\mathbf{b} = (b_1, b_2, \ldots, b_{n_h})^T \in R^{n_h}$ is the bias vector of the hidden layer, and $\mathbf{W} = (w_{i,j}) \in R^{n_h \times n_v}$ is the weight matrix between the hidden layer and the visible layer. Through the generative stacking technology, the depth belief network is finally produced by multiple restricted Boltzmann machines.

The constrained Boltzmann machine introduces a series of related probability distribution functions through the energy function. For the state vector $(\mathbf{v}, \mathbf{h})$ of a given set of neurons, its energy function can be expressed as

$$E(\mathbf{v}, \mathbf{h}|\boldsymbol{\theta}) = -\sum_{i=1}^{n_v} a_i v_i - \sum_{j=1}^{n_h} b_j h_j - \sum_{i=1}^{n_v} \sum_{j=1}^{n_h} h_j w_{i,j} v_i, \quad (3)$$

where $\mathbf{v}$ represents the state vector of neurons in the visible layer, and $\mathbf{h}$ represents the state vector of neurons in the hidden layer. n_v is the total number of all neurons in the visible layer, and n_h is the total number of all neurons in the hidden layer. $\theta = \{a_i, b_j, w_{i,j}\}$ is the regulator limiting the Boltzmann machine architecture.
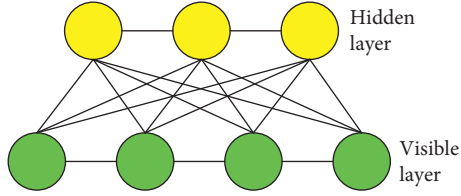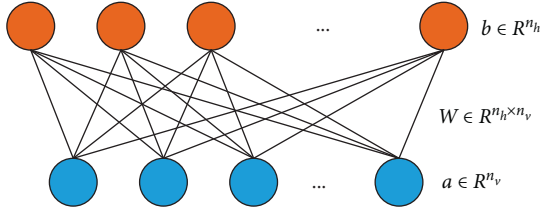
Figure 6: Boltzmann machine model.

$b \in R^{n_h}$

$W \in R^{n_h \times n_v}$

$a \in R^{n_v}$

Figure 7: RBM model.

Through the energy function defined in equation (3) above, the joint probability distribution of state $(\mathbf{v}, \mathbf{h})$ can be obtained as shown in

$$P(\mathbf{v}, \mathbf{h}|\boldsymbol{\theta}) = \frac{1}{Z(\boldsymbol{\theta})} e^{-E(\mathbf{v},\mathbf{h}|\boldsymbol{\theta})}, \tag{4}$$

$$Z(\boldsymbol{\theta}) = \sum_{v,h} e^{-E(\mathbf{v},\mathbf{h}|\boldsymbol{\theta})}. \tag{5}$$

In the formula, $Z(\theta)$ represents the normalized parameter. Let $p(\mathbf{v}|\theta)$ be the probability distribution of the visible layer vector $\mathbf{v}$; then $p(\mathbf{v}|\theta)$ can be calculated by the edge distribution of $P(\mathbf{v}, \mathbf{h}|\theta)$.

$$
\begin{aligned}
P(\mathbf{v}|\boldsymbol{\theta}) &= \sum_h P(\mathbf{v}, \mathbf{h}|\boldsymbol{\theta}) \\
&= \frac{1}{Z(\boldsymbol{\theta})} \sum_h e^{-E(\mathbf{v},\mathbf{h}|\boldsymbol{\theta})}.
\end{aligned} \tag{6}
$$

In the same way, we can get the probability distribution $p(\mathbf{h}|\theta)$ of the hidden layer vector $\mathbf{h}$.

$$
\begin{aligned}
P(\mathbf{h}|\boldsymbol{\theta}) &= \sum_v P(\mathbf{v}, \mathbf{h}|\boldsymbol{\theta}) \\
&= \frac{1}{Z(\boldsymbol{\theta})} \sum_v e^{-E(\mathbf{v},\mathbf{h}|\boldsymbol{\theta})}.
\end{aligned} \tag{7}
$$

By analyzing equations (6) and (7), it can be seen that in order to get $p(\mathbf{v}, |\theta)$ and $p(\mathbf{h}|\theta)$, the key step is to calculate the normalized parameter $Z(\theta)$. However, according to the definition of $Z(\theta)$, its computational complexity is relatively high. Due to the special principle of limiting Boltzmann machine model (the visible layer and the hidden layer are conditionally independent), when the state of all neurons in the visible layer is known, the probability of activation of a neural unit in the hidden layer can be calculated by

$$P\left(h_j = 1|\mathbf{v}, \boldsymbol{\theta}\right) = \sigma\left(b_j + \sum_i v_i w_{ij}\right), \tag{8}$$

where $\sigma(\cdot)$ represents the Sigmoid activation function.

Because there is no connection between all neural nodes in the same layer, the relationship between the values of all neural nodes in the same layer and the values of a single node is as follows:

$$P(\mathbf{h}|\mathbf{v}) = \prod_{j-1}^{n_h} P\left(h_j|v\right), \tag{9}$$

$$P(\mathbf{v}|\mathbf{h}) = \prod_{i-1}^{n_v} P\left(v_i|h\right). \tag{10}$$

The random gradient algorithm is usually used to find the maximum value in order to obtain the optimal adjustment factor in the network. The DBN model is shown in Figure 8.

3.2.2. Feature Classification Process. DBN training process is generally divided into two steps: (1) pretraining stage and (2) fine-tuning stage. The loss function required in the fine-tuning stage is shown in

$$L(x, y) = \|x - y\|_2^2, \tag{11}$$

where the symbol $\|\cdot\|_2$ represents the 2 norm of reconstruction error, x represents input data, and y represents reconstructed data.

The data comes from the face images of the tested students in the sample set, which includes a training sample set, a test sample set, and a text document of the training sample classification information. The feature classification process is shown in Figure 9.

Step 1: divide personality traits into three categories (e.g., low agreeableness, moderate agreeableness, and high agreeableness), input 32 feature information extracted based on active model, respectively, and form 32 feature vectors representing feature information.

Step 2: classify the data samples. The data in the data sample is divided into two folders, the training sample set and the test sample set, and all the sample data in the training sample set are labeled by categories to form a text document, which can be used in the later learning and training of image features.

Step 3: sample training. According to the deep belief network structure model, this paper adopts the deep belief network structure model containing five layers of RBM to train the training samples provided.

Step 4: test the sample. After training the training samples with the deep confidence network, the relevant weights and biases of each layer of DBN can be obtained. Then, the relevant classification results can be obtained by testing the test samples.
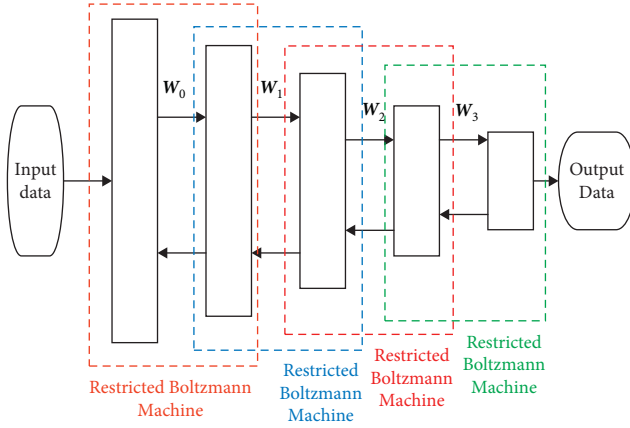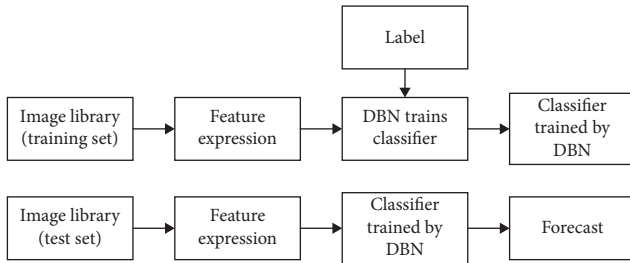
Figure 8: DBN model.

Figure 9: Feature classification process.

The classifiers of extroversion, openness, agreeableness, and conscientiousness can be established after the four personality traits are processed, respectively, according to the above steps.

4. Experimental Results and Analysis

4.1. Experimental Environment. The experiment is carried out on IMM face database [28–33], and the experimental simulation tool is MATLAB7.8. IMM face database consists of 240 images of six faces of 40 different people, all without glasses. The image size is $640 * 480$. 120 face images are selected from the library as the training set and 80 as the test images. In order to simplify the problem, the images used for training and testing were all neutral expressions of positive faces, and the illumination was relatively uniform. The feature points are marked on the external contour of the face and the edges of each facial organ. Each training image was calibrated with 58 feature points. After repeated tests, this paper sets the 2D texture rectangular search side lengths of the third layer, the second layer, and the first layer as 16, 8, and 4, respectively.

4.2. Performance Evaluation Indicators. In order to test the performance of the improved ASM algorithm, two functions are introduced to evaluate its performance. The first is the average error function:

$$E_{\text{ave}} = \frac{1}{N} \times \frac{1}{n} \sum_{i=1}^{N} \sum_{j=1}^{n} \text{abs}\,(x\,(i,\,j) - \text{pos}\,(i,\,j)), \qquad (12)$$

where N is the total number of test images and n is the number of feature points marked. $x\,(i,\,j)$ is the coordinate of the jth feature point marked by the ith subimage, and $\text{pos}\,(i,\,j)$ is the coordinate [34] of the target point located by ASM algorithm.

The other is the time function of the average cost:

$$T_{\text{ave}} = \frac{1}{N} \sum_{i=1}^{N} t_i, \qquad (13)$$

where t_i is the time it takes to process each picture [35, 36].

4.3. Comparison of Face Feature Points Location Results. The effect of improved ASM and traditional ASM for face feature point localization is shown in Figures 10 and 11, respectively.

Under the premise of 50 iterations, the average error between the improved ASM and the traditional ASM is compared, and the result is shown in Figure 12.

It can be seen from the experimental results that the improved ASM algorithm can more accurately locate the key face feature points and improve the positioning accuracy of face feature points. The performance comparison results between improved ASM and traditional ASM are shown in Table 1.

By comparing the results, it can be seen that the proposed method is more accurate than the original algorithm. The improved algorithm is less efficient than the original one because it builds a 2D texture model with a large amount of computation for each feature point and extracts Gabor features with a large amount of computation for the lowest pyramid subgraph of each feature point. However, it can be seen from the experimental results that the positioning accuracy of the proposed algorithm is more than doubled compared with the original algorithm, while the processing time is not doubled. Therefore, the shortcoming of the algorithm efficiency decline is acceptable.

4.4. Classifier Test. Through the training samples, we can get four classifiers after DBN training: extroversion, openness, agreeableness, and rigor.

The DBN model is compared with the convolutional neural network model, and the results of recognition accuracy and consumption time are shown in Table 2 and Figure 13.

As can be seen from Table 2 and Figure 13, the classifier trained by DBN is used to test four kinds of personality traits of the test sample. The accuracy of extroversion is 77.95%, the accuracy of openness is 81.16%, the accuracy of agreeableness is 90.63%, and the accuracy of rigor is 91.42%, totaling 85.29%, which is higher than that of the classifier trained by convolutional neural network model.

Thus, it can be seen that the classifier trained by DBN has a good effect on the prediction results of the four trait tests on the premise of providing facial features.

feature point location result 1

feature point location result 2

(a)　　　　　　　　　　　　　　　　(b)
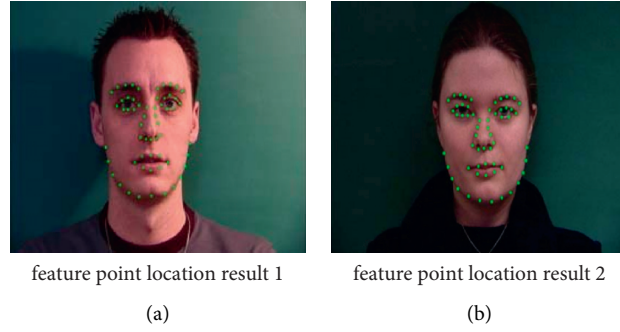
FIGURE 10: Effect of face feature point location using improved ASM. (a) Feature point location result 1. (b) Feature point location result 2.

feature point location result 1

feature point location result 2

(a)　　　　　　　　　　　　　　　　(b)
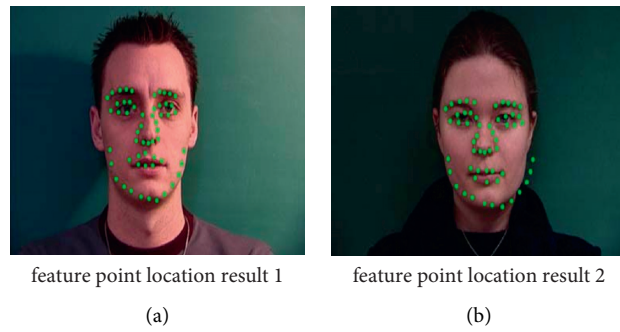
FIGURE 11: Effect of face feature point location using traditional ASM. (a) Feature point location result 1. (b) Feature point location result 2.

- - - traditional ASM
—— improved ASM
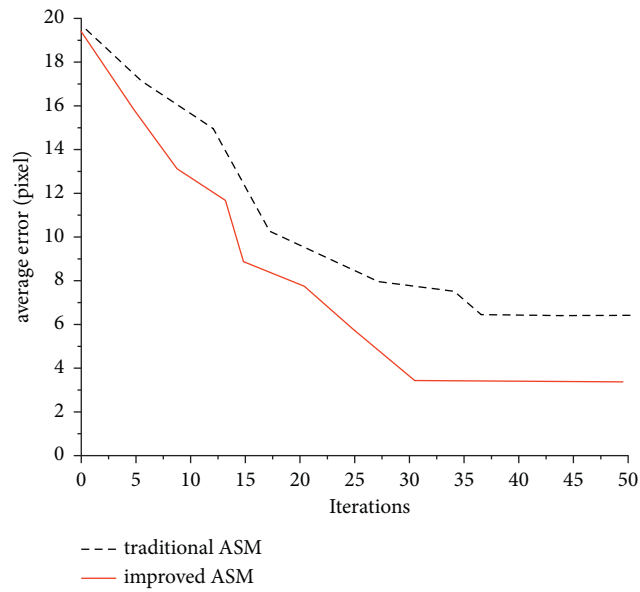
FIGURE 12: Comparison of average error between improved ASM and traditional ASM.

TABLE 1: Comparison of the performance of improved and traditional ASM.

Objective function	Improved ASM	Traditional ASM
E_{ave} (pixels)	3.16	6.34
T_{ave} (s)	4.67	3.09

TABLE 2: Comparison of the accuracy of two classifiers.

Machine learning approach	Classifier	Accuracy (%)
Convolution neural network model	Extroversion openness	73.88
		78.57
	Agreeableness	85.73
	Rigor	87.74
	Total	81.48
BP neural network model	Extroversion openness	69.05
		73.45
	Agreeableness	80.72
	Rigor	85.61
	Total	77.21
DBN model	Extroversion openness	77.95
		81.16
	Agreeableness	90.63
	Rigor	91.42
	Total	85.29
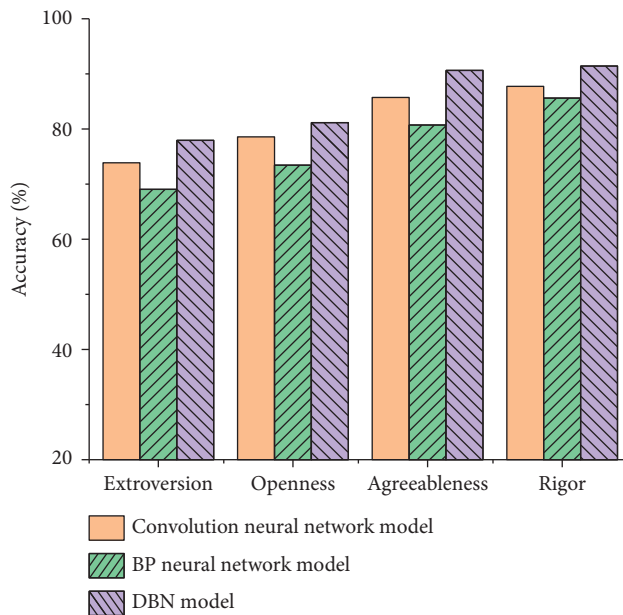
FIGURE 13: Comparison of the accuracy of two classifiers.

5. Conclusion

In this paper, a personality trait detection method based on ASM facial feature point location and DBN model is proposed. First of all, the traditional ASM algorithm is improved in three aspects, and the improved algorithm is applied to the facial feature point location, which solves the shortcomings of the traditional ASM algorithm, such as inaccurate location, low efficiency, easy to fall into local minimum, and so on. Secondly, with the help of the characteristics of deep learning unsupervised learning classification, the sample data are trained by the deep belief network model in order to get the classifier based on four personality traits and facial features. Finally, through the comparative analysis of the classification results of psychological personality test, it verifies the feasibility and effectiveness of this method and provides a scientific basis for college teachers to make a reasonable psychological counseling plan.

Data Availability

The data used in this paper are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] R. Zhang, L. Lin, R. Zhang, W. Zuo, and L. Zhang, "Bit-scalable deep hashing with regularized similarity learning for image retrieval and person Re-identification," *IEEE Transactions on Image Processing*, vol. 24, no. 12, pp. 4766–4779, 2015.

[2] T. M. Scholtens, F. Schreuder, S. T. Ligthart, J. F. Swennenhuis, J. Greve, and L. W. M. M. Terstappen, "Automated identification of circulating tumor cells by image cytometry," *Cytometry, Part A*, vol. 81A, no. 2, pp. 138–148, 2012.

[3] M. Wright, C.-J. Lin, E. O'Neill, D. Cosker, and P. Johnson, "3D gesture recognition: an evaluation of user and system performance, Lecture Notes in Computer Science," in *Proceedings of the International Conference on Pervasive Computing*, pp. 294–313, Springer, Berlin, Heidelberg, June 2011.

[4] H. Cheng, L. Yang, and Z. Liu, "Survey on 3D hand gesture recognition," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 26, no. 9, pp. 1659–1673, 2015.

[5] A. Mousavian, D. Anguelov, J. Flynn, and J. Kosecka, "3d bounding box estimation using deep learning and geometry," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7074–7082, IEEE, San Juan, PR, USA, June 2017.

[6] K. Sun, B. Xiao, D. Liu, and J. Wang, "Deep high-resolution representation learning for human pose estimation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5693–5703, IEEE, New York, NY, USA, June 2019.

[7] H. S. Fang, S. Xie, Y. W. Tai, and C. Lu, "Rmpe: regional multi-person pose estimation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2334–2343, IEEE, San Juan, PR, USA, June 2017.

[8] Z. Cao, G. Hidalgo, T. Simon, S. E. Wei, and Y. Sheikh, "Openpose: realtime multi-person 2d pose estimation using part affinity fields," 2018, http://arxiv.org/abs/1812.08008.

[9] B. Xiao, H. Wu, and Y. Wei, "Simple baselines for human pose estimation and tracking," in *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 466–481, Springer, Munich, Germany, September 2018.

[10] M. Andriluka, U. Iqbal, E. Insafutdinov et al., "Posetrack: a benchmark for human pose estimation and tracking," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5167–5176, IEEE, San Juan, PR, USA, June 2018.

[11] M. Munaro, S. Michieletto, and E. Menegatti, "An evaluation of 3d motion flow and 3d pose estimation for human action recognition," in *Proceedings of the RSS Workshops: RGB-D: Advanced Reasoning with Depth Cameras*, Berkeley, CA, USA, July 2013.

[12] S. Hadfield, K. Lebeda, and R. Bowden, "Natural action recognition using invariant 3D motion encoding," in *Proceedings of the European Conference on Computer Vision*, pp. 758–771, Springer, Glasgow, UK, September 2014.

[13] H. Rhodin, J. Spo¨rri, I. Katircioglu et al., "Learning monocular 3D human pose estimation from multi-view images," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 8437–8446, IEEE, San Juan, PR, USA, June 2018.

[14] L. Yang, S. Li, D. Lee, and A. Yao, "Aligning latent spaces for 3d hand pose estimation," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2335–2343, IEEE, Cambridge, MA, USA, June 2019.

[15] A. Akash, M. Akhand, M. A. H. Akhand, and N. Siddique, "Robust face detection integrating novel skin color matching under variant illumination conditions," *International Journal of Image, Graphics and Signal Processing*, vol. 13, no. 2, pp. 1–15, 2021.

[16] Y. Wang, Y. Zhao, Z. Guo, M. Qi, Y. Fan, and H. Meng, "Diffusion Tensor image segmentation based on multi-atlas active shape model," *Multimedia Tools and Applications*, vol. 78, no. 24, Article ID 34231, 2019.

[17] H. Zhao, Y. Liu, and S. Li, "Online learning behavior based personality recognition," *Open Education Research*, vol. 25, no. 5, pp. 110–120, 2019.

[18] W. Cai and Z. Wei, "PiiGAN: generative adversarial networks for pluralistic image inpainting," *IEEE Access*, vol. 8, Article ID 48451, 2020.

[19] X. Ning, P. Duan, W. Li, and S. Zhang, "Real-time 3D face Alignment using an encoder-Decoder network with an efficient Deconvolution Layerfficient deconvolution layer," *IEEE Signal Processing Letters*, vol. 27, pp. 1944–1948, 2020.

[20] W. Cai and Z. Wei, "Remote sensing image classification based on a cross-attention mechanism and graph convolution," *IEEE Geoscience and Remote Sensing Letters*, vol. 10, pp. 1–5, 2020.

[21] X. Ning, K. Gong, W. Li, and L. Zhang, "JWSAA: Joint weak saliency and attention aware for person re-identification," *Neurocomputing*, vol. 453, pp. 801–811, 2020.

[22] W. Cai, B. Liu, Z. Wei, M. Li, and J. Kan, "TARDB-Net: triple-attention guided residual dense and BiLSTM networks for hyperspectral image classification," *Multimedia Tools and Applications*, vol. 1, pp. 1–22, 2021.

[23] X. Yu, J. Yang, and Z. Xie, "Training SVMs on a bound vectors set based on Fisher projection," *Frontiers of Computer Science*, vol. 8, no. 5, pp. 793–806, 2014.

[24] S. E. Wei, V. Ramakrishna, T. Kanade, and Y. Sheikh, "Convolutional pose machines," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4724–4732, IEEE, Las Vegas, NV, USA, July 2016.

[25] Z. Cao, T. Simon, S. E. Wei, and Y. Sheikh, "Realtime multi-person 2d pose estimation using part affinity fields," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7291–7299, IEEE, Honolulu, HI, USA, July 2017.

[26] Y. Ding, J. Yang, J. Ponce, and H. Kong, "Homography-based minimal-case relative pose estimation with known gravity direction," in *Proceedings of the IEEE Transactions on Pattern Analysis and Machine Intelligence*, August 2020.

[27] X. Yu, F. Jiang, J. Du, and D. Gong, "A cross-domain collaborative filtering algorithm with expanding user and item features via the latent factor space of auxiliary domainsfiltering algorithm with expanding user and item features via the latent factor space of auxiliary domains," *Pattern Recognition*, vol. 94, pp. 96–109, 2019.

[28] Z. Wang, C. Zou, and W. Cai, "Small sample classification of hyperspectral remote sensing images based on sequential joint deeping learning modelfication of hyperspectral remote sensing images based on sequential joint deeping learning model," *IEEE Access*, vol. 8, Article ID 71353, 2020.

[29] W. Cai and Z. Wei, "Diversity-generated image inpainting with style extraction," 2019, http://arxiv.org/abs/1912.01834.

[30] Z. L. Yang, S. Y. Zhang, Y. T. Hu, Z. W. Hu, and Y. F. Huang, "Vae-Stega: linguistic steganography based on variational auto-encoder," *IEEE Transactions on Information Forensics and Security*, vol. 16, pp. 880–895, 2020.

[31] X. Ning, X. Wang, S. Xu et al., "A review of research on co-training. concurrency and computation: practice and experience," *Bio-Engineering Applications*, vol. 32, Article ID e6276, 2021.

[32] Y. Kim and D. Kim, "A CNN-Based 3D human pose estimation based on projection of depth and ridge data," *Pattern Recognition*, vol. 106, 2020.

[33] M. Dantone, J. Gall, C. Leistner, and L. Van Gool, "Human pose estimation using body parts dependent joint regressors," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3041–3048, IEEE, San Juan, PR, USA, June 2013.

[34] J. Su, A. X. Liu, Z. Sheng, and Y. Chen, "A partitioning approach to RFID identification," *IEEE/ACM Transactions on Networking*, vol. 28, no. 5, pp. 2160–2173, 2020.

[35] S. Johnson and M. Everingham, "Learning effective human pose estimation from inaccurate annotation," in *Proceedings of the 24th IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1465–1472, IEEE, Colorado Springs, CO, USA, June 2011.

[36] J. Su, R. Xu, S. Yu, B. Wang, and J. Wang, "Idle slots skipped mechanism based tag identification algorithm with enhanced collision detection," *KSII Transactions on Internet and Information Systems*, vol. 14, no. 5, pp. 2294–2309, 2020.

Research Article

SAR Target Recognition Using Improved Sparse Representation with Local Reconstruction

Li Ma [ID]

School of Information Engineering, China University of Geosciences, Beijing, Beijing 100083, China

Correspondence should be addressed to Li Ma; 1004185206@cugb.edu.cn

In order to handle the problem of synthetic aperture radar (SAR) target recognition, an improved sparse representation-based classification (SRC) is proposed. According to the sparse coefficient vector resulting from the global dictionary, the largest coefficient in each class is taken as the reference. Then, the surrounding neighborhoods of the sample with the largest coefficient are selected to construct the optimal local dictionary in each training class. Afterwards, the samples in the local dictionary are used to reconstruct the test sample to be identified. Finally, the decision is made according to the comparison of the reconstruction errors from different classes. In the experiments, the proposed method is verified based on the moving and stationary target acquisition and recognition (MSTAR) dataset. The results show that the proposed method has performance advantages over existing methods, which demonstrates its effectiveness and robustness.

1. Introduction

Synthetic aperture radar (SAR) is capable of measuring high-resolution images for effective ground observation and surveillance. Image interpretation technologies represented by SAR target recognition are widely used in military and civilian fields. SAR target recognition is a typical image pattern recognition problem, which aims to extract and classify the target of interest in SAR images [1, 2]. In order to improve the comprehensive performance of SAR target recognition, researchers extensively use advanced image feature extraction and classification algorithms for experimentation and verification. There are many types of features applied to SAR target recognition, including geometric shape features, projection transformation features, and electromagnetic scattering features. The geometric shape features describe the two-dimensional shape distribution of the target, such as area and contour [3–10]. The projection transformation features use mathematical projection or signal transformation algorithms to extract stable characteristics of the original images [11–16]. Electromagnetic scattering features describe the backscattering characteristics of the target in a specific radar frequency band, typically

including the scattering centers [17–20] and polarizations [21]. In the classification stage, a suitable classifier is selected to confirm the class of the extracted features. Early classification strategies were mainly based on the idea of the nearest neighbors, such as K-nearest neighbor (KNN) classifier [11]. With the development of pattern recognition technologies, new classifiers such as support vector machine (SVM) [22, 23], multilayer perceptron (MLP) [12], and adaptive boosting (AdaBoost) [24] emerged. In recent years, the deep learning technology has become a new favorite in the field of image interpretation and has also been widely used in SAR target recognition [25–30]. Sparse representation-based classification (SRC) derived from compressive sensing theory was also widely used in pattern recognition and image classification [31–37]. Researchers introduced SRC into SAR target recognition and verified its feasibility. Since then, more works have continued to improve the overall recognition performance by optimizing the solution algorithm and decision-making mechanism [32–36].

Compared with other classifiers, SRC uses a linear fitting idea to evaluate the similarity between the test sample and each training class, so no pretraining is required. In addition, it is not difficult to find from the results of the existing

literature that the sparse representation itself has a certain degree of robustness to the common extended operating conditions (EOCs) [38, 39] in SAR target recognition such as noise corruption and partial occlusion. Therefore, SRC has broad application prospects in SAR target recognition. In this paper, the traditional SRC is improved to enhance the performance in SAR target recognition. First, in the global dictionary, the test sample is reconstructed by SRC, and the sparse coefficient vector is obtained. Afterwards, the optimal local samples are selected in each classes according to a certain criterion. The criterion takes the sample with the largest coefficient as the reference and selects its surrounding neighborhoods to construct a local dictionary. Finally, the test sample is optimally reconstructed on the local dictionaries from different classes to obtain their corresponding reconstruction errors. The target class of the test sample is finally decided based on the principle of the minimum error. In the experiments, based on the moving and stationary target acquisition and recognition (MSTAR) dataset, the proposed method is tested under the standard operating condition (SOC) and three EOCs (configuration variance, depression angle difference, and noise corruption). Experimental results show that the proposed method can achieve superior performance over some existing methods in all four typical scenarios, verifying its effectiveness and robustness.

2. SRC

The sparse representation is based on the theory of compressive sensing and analyzes the characteristics of the sample by linearly representing the unknown sample on an overcomplete dictionary. Wright et al. first applied SRC in face recognition [31], that is, to determine the category of the test sample based on the reconstruction error of each class calculated based on the sparse representation coefficients. Specifically, a global dictionary $A = [A_1, A_2, \ldots, A_C] \in R^{d \times N}$ composed of multiple training classes is first constructed, where A_i represents the N_i atoms corresponding to the training samples in the ith class. For the test sample y to be identified, equation (1) is employed to perform the sparse linear represented:

$$\hat{x} = \arg \min_x \|x\|_0,$$
$$\text{s.t. } \|y - Ax\|_2^2 \leq \varepsilon, \tag{1}$$

where x is the sparse coefficient vector to be solved and ε is the set error threshold.

Since the direct solution of the optimization problem in equation (1) is very complicated, researchers tried to obtain high-confidence approximate solutions through the principle of equivalent approximation. For example, in [31], the ℓ_1 norm was used to replace the original ℓ_0 norm to convert the problem into a convex optimization one, which is easier to solve. In [32], an orthogonal matching pursuit algorithm (OMP) was developed based on a greedy mechanism to improve the overall solution efficiency. According to the solved sparse coefficient vector, the target class of the test sample can be judged according to its distributions in different classes. In [33], the decision was made according to the energy of nonzero coefficients in different classes. Among many principles, the criterion based on the minimum reconstruction error is the most widely used. The basic idea is to linearly reconstruct the test samples with samples of each class and then calculate the reconstruction error as follows:

$$r(i) = \|y - A_i x_i\|_2^2, \quad (i = 1, 2, \ldots, C), \tag{2}$$

where x_i is the coefficient vector distributed on the ith class and $r(i)$ is the reconstruction error of the ith class to the test sample. Finally, the target class of the test sample can be determined by comparing the errors from different classes.

Although the decision criteria in the traditional SRC have certain validity, their characterization ability for each class is not sufficiently exploited. In the minimum reconstruction error criterion, all samples in each class are used for reconstruction. In fact, due to the azimuthal sensitivity of SAR images, the training samples related to the test sample should share a similar azimuth angle. Therefore, in order to obtain a better reconstruction result, the test sample should be reconstructed and analyzed on the local dictionary.

3. Improvement of SRC for Target Recognition

3.1. Improved SRC with Local Reconstruction. This paper makes some improvement on traditional SRC for SAR target recognition. Before implementing the traditional SRC, the atoms of each class in the dictionary are arranged in ascending azimuth order to ensure that the azimuths between adjacent samples are the closest. Afterwards, the global sparse coefficients are solved according to equation (1). Then, the optimal local dictionary is selected and constructed in each class. Taking the ith class as an example, the atom with the maximum coefficient is used as the reference sample, and the surrounding neighborhood samples are chosen. Because SAR images are sensitive to azimuth, it is generally believed that they can maintain a high correlation within the interval of $\pm 5°$. Therefore, this paper refers to this criterion in selecting surrounding samples. When the azimuth error is lower than $5°$ with the reference sample, the candidate is incorporated into the local dictionary. Based on the local dictionary, this paper performs the optimal representation of the test sample class by class as follows:

$$\hat{\alpha} = \arg \min \left\{ \|y - A_L \alpha_L\|_2^2 + \lambda \|\alpha_L\|_2^2 \right\}, \tag{3}$$

where A_L represents the local dictionary selected on a certain class, α_L is the corresponding coefficient vector, and λ is the regularization coefficient. The above optimization has the following analytical solution:

$$\hat{\alpha} = \left(A_L^T A_L + \lambda * I \right)^{-1} A_L^T y, \tag{4}$$

where I represents the unit matrix.

Compared with the traditional SRC mechanism, the optimization problem in equation (3) emphasizes the reconstruction accuracy. In fact, in the local dictionary, the sparsity constraint is not established, and the optimal

reconstruction is more effective. Finally, according to the coefficient vector solved on each class, the corresponding reconstruction error for the test sample can be solved according to equation (2). Finally, the target class of the test sample is determined according to the principle of the minimum error.

3.2. Target Recognition Procedure. According to the above algorithms, the basic process of the recognition method proposed in this paper is shown in Figure 1. It can be further decomposed into the following steps:

> Step 1: the test sample is processed by SRC based on the global dictionary formed by all the training samples and the sparse coefficient vector is solved
>
> Step 2: the optimal local dictionary for each class is established according to Section 3.1
>
> Step 3: the optimal reconstruction of the test sample is performed on the local dictionary of each class and the corresponding reconstruction error is calculated
>
> Step 4: the target class of the test sample is determined according to the principle of the minimum error

In the specific implementation process, the principal component analysis (PCA) is used to extract feature vectors for all training and test samples so as to improve the overall classification efficiency. The OMP algorithm is used to solve the sparse coefficient vector resulting from the global dictionary.

4. Experiments

4.1. MSTAR Dataset. The performance of the proposed method is tested based on the MSTAR dataset. The dataset has been the authoritative benchmark for the testing and verification of SAR target recognition methods since its public release in the 1990s. Figure 2 shows the target in the dataset, including tanks, armored vehicles, and transport vehicles. The SAR image of each type of target contains omnidirectional azimuths and several depression angles, so that various conditions can be flexibly set to carry out experiments. According to existing literature, Table 1 gives a typical experimental setup, using SAR images of all the 10 types of targets at 17° and 15° depression angles as training and testing sets, respectively. In particular, the test samples of BMP2 and T72 targets have more configurations than their training samples (different configurations are marked by the notations in the parentheses). In general, under the experimental condition in Table 1, the differences between the training and test sets are relatively small, which can be approximated as SOC. In addition, according to the diversity of the image samples of the MSTAR dataset, several EOCs can be set or simulated, such as configuration variance, depression angle variance, and noise interference.

During the experiment, the proposed method is compared with several types of existing methods, focusing on comparison with traditional SRC-based methods, including the SRC1 method in [32], the SRC2 method in [33], and the SRC3 method in [34]. These three types of methods either

adopt different decision-making mechanisms or introduce new solution constraints. In addition, a new method based on CNN is also set up in the comparison method, i.e., ESENet, proposed in [29]. The follow-up experiments specifically include 4 types: 1 SOC and 3 EOCs. SOC focuses on evaluating the basic recognition performance of the method, and EOCs verify the robustness of the method, mainly the reliability in complex scenarios.

4.2. SOC. At first, the performance of the proposed method is tested under SOC using the training and test sets in Table 1. In this case, the similarities between the test sample and the training sample are relatively high, so the difficulty of the recognition problem is relatively low. However, since Table 1 involves 10 types of targets, the correct classification still faces certain challenges. Figure 3 shows the recognition results of the proposed method under SOC. In the confusion matrix, the diagonal element is the correct recognition rate of the corresponding target. Considering the correct recognition samples of 10 types of targets, the average recognition rate of the proposed method in this paper reaches 99.04%, which shows its effectiveness. Table 2 compares the recognition results of various methods under this condition. The recognition rate of the proposed method is significantly higher than those of the traditional SRC-based methods, indicating that proper local dictionary selection and optimal reconstruction can improve the performance of SAR target recognition. Compared with the ESENet method, the recognition rate of the proposed method is slightly higher. Due to the configuration differences that occurred in BMP2 and T72 in the test set, the adaptability of the trained networks has declined to a certain extent. In summary, the proposed method has certain performance advantages under SOC.

4.3. EOC-1: Configuration Variance. Table 3 sets the training and test samples under the condition of configuration variance, including 3 types of targets. Among them, the test samples and samples of BMP2 and T72 targets are from different configurations. Table 4 shows the average recognition rates of different methods under this situation. The comparison shows that the proposed method can maintain the highest performance under configuration variance, showing its superior robustness. Compared with the other three types of SRC-based methods, this paper optimizes the local dictionary in the global sparse coefficients and performs the optimal reconstruction class by class, which further improves the effectiveness of the proposed method. Under configuration variance, there are only small differences in the local structure of the targets between the test and training samples. The reconstruction in a single class rather than the global dictionary is helpful to discover such subtle differences, thereby improving the recognition accuracy. Compared with the results under SOC, the recognition performance of the ESENet method decreases the most significantly, mainly because the influence of the configuration variances is further aggravated at this time.
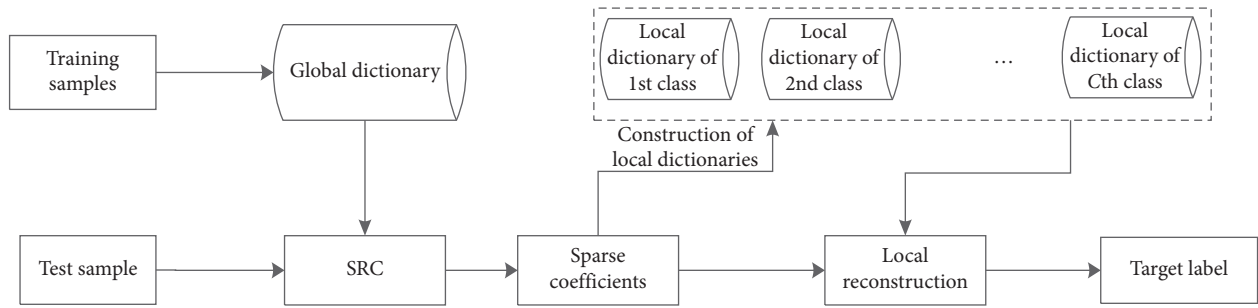
FIGURE 1: Procedure of SAR target recognition based on improved SRC.
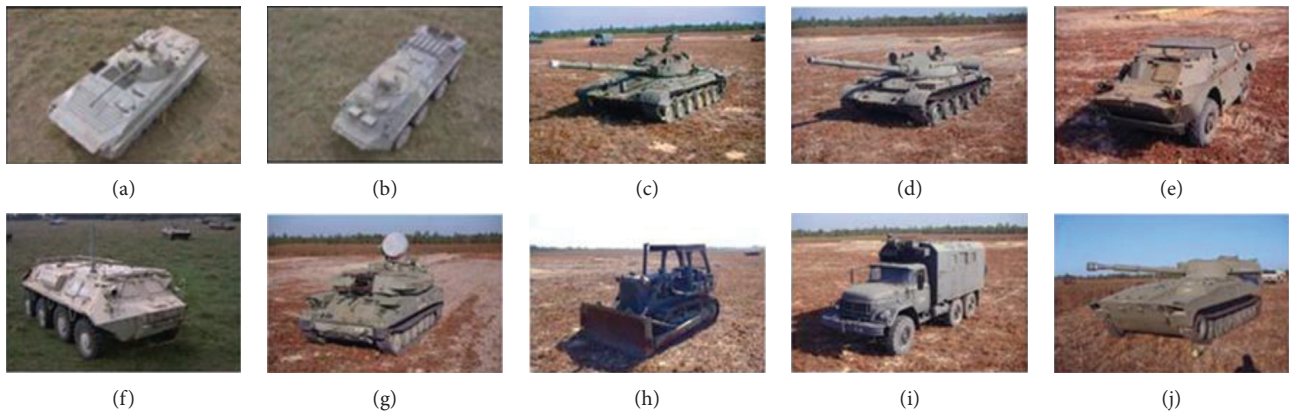
FIGURE 2: Optical images of targets in MSTAR dataset [35]. (a) BMP2, (b) BTR70, (c) T72, (d) T62, (e) BRDM2, (f) BTR60, (g) ZSU23/4, (h) D7, (i) ZIL131, and (j) 2S1.

TABLE 1: Training and test samples under SOC.

Target class	BMP2	BTR70	T72	T62	BRDM2	BTR60	ZSU23/4	D7	ZIL131	2S1
Training set	233 (9563)	233	232 (132)	299	298	256	299	299	299	299
Test set	587 (9563, 9566, C21)	233	582 (132, 812, S7)	273	274	195	274	274	274	274

FIGURE 3: Confusion matrix of the proposed method under SOC.

TABLE 2: Comparison of performance under SOC.

Method	Average recognition rate (%)
Proposed	99.04
SRC1	96.12
SRC2	96.63
SRC3	97.21
ESENet	98.78

4.4. EOC-2: Depression Angle Variance. The MSTAR dataset also includes SAR images of several types of targets at multiple different depression angles. As shown in Table 5, the training samples are images of 3 types of targets, i.e., 2S1, BRDM2, and ZSU23/4, from the depression angle of 17°; the test samples are from the depression angles of 30° and 45°, respectively. The large depression angle variance leads to a decrease in the similarity between the test and training samples, which brings about certain obstacles to correct recognition. Table 6 compares the average recognition rates

TABLE 3: Experimental condition under different configurations.

| Target class | Training set (17°) | | Test set (15°) | |
	Configuration	Sample amount	Configuration	Sample amount
BMP2	9563	233	9566 c21	196
				196
T72	132	232	812 s7	195
				191
BTR60	7532	256	7532	195
T62	A51	299	A51	273

TABLE 4: Comparison of performance under configuration variances.

Method	Average recognition rate (%)
Proposed	98.52
SRC1	94.76
SRC2	96.08
SRC3	97.34
ESENet	97.72

TABLE 5: Training and test samples under depression angle variance [37].

	Target class	2S1	BRDM2	ZSU23/4
Training set	Depression angle	17°	17°	17°
	Sample amount	299	298	299
Test set	Depression angle	30°	30°	30°
	Sample amount	288	287	288
	Depression angle	45°	45°	45°
	Sample amount	303	303	303

TABLE 6: Comparison of performance at different depression angles.

| Method | Average recognition rate (%) | |
	30°	45°
Proposed	95.45	73.28
SRC1	93.52	66.37
SRC2	94.03	67.78
SRC3	94.48	69.12
ESENet	95.03	67.26

of various methods at two test depression angles. The average recognition rate of the proposed method is 95.45% and 73.28% at the elevation angles of 30° and 45°, respectively, which has advantages compared with other methods. Especially at 45° with a notable depression difference, the advantages of the proposed method are more significant. Through proper local dictionary selection and optimal reconstruction, the characterization ability of different classes for the current test sample can be investigated to the greatest extent, so it can better adapt to the situation of depression angle variances.

4.5. EOC-3: Noise Corruption. In order to quantitatively test the noise robustness of the proposed method under different signal-to-noise ratios (SNR), the test samples in
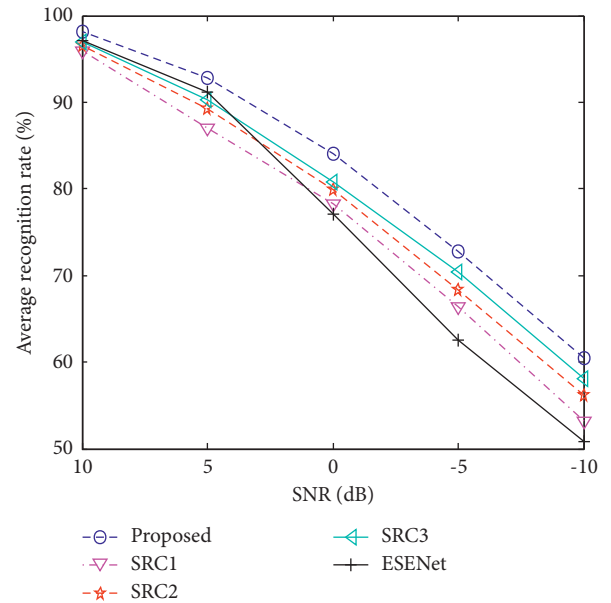
FIGURE 4: Comparison of performance under noise corruption.

Table 1 are used as the benchmark and different degrees of Gaussian white noise are added to them according to the ideas in [19]. Figure 4 plots the recognition rate curves of various methods with the change of the SNR. The proposed method can achieve the highest average recognition rate at each noise level, showing its better noise robustness. Compared with the ESENet method, several types of methods based on SRC are generally more robust, especially in the case of low SNR, verifying that sparse representation has a certain robustness to noise interference. The proposed method examines the representation ability of each class on a reliable local dictionary and can deal with noise corruption more effectively through the optimal reconstruction process.

5. Conclusion

The paper proposes an improved SRC for the SAR target recognition. On the basis of the global sparse coefficients obtained from traditional SRC, the local dictionary is constructed according to each training class. Considering the azimuthal sensitivity of SAR images, this paper takes the sample with the largest correlation coefficient in each class as the reference and selects some of its neighborhood samples with approaching azimuths to construct the local dictionary

with the strongest representation ability. Based on the local dictionary, the linear fitting is performed on the test sample according to the idea of optimal reconstruction, and finally the target class of the test sample is determined by comparing the reconstruction errors resulting from different classes. Based on the MSTAR dataset, the proposed method is tested under 4 typical conditions. The experimental results reflect that the proposed method can maintain superior performance under both SOC and EOCs, which proves its effectiveness and robustness for SAR target recognition.

Data Availability

The dataset used in this paper is publicly available.

Conflicts of Interest

The author declares that there are no conflicts of interest regarding the publication of this paper.

References

[1] K. El-Darymli, E. W. Gill, and P. McGuire, "Automatic target recognition in synthetic aperture radar imagery: a state-of-the-art review," *IEEE Access*, vol. 4, pp. 6014–6058, 2016.

[2] B. Ding and G. Wen, "Target reconstruction based on 3-D scattering center model for robust SAR ATR," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 7, pp. 3772–3785, 2018.

[3] B. Ding and G. Wen, "A region matching approach based on 3-D scattering center model with application to SAR target recognition," *IEEE Sensors Journal*, vol. 18, no. 11, pp. 4623–4632, 2018.

[4] M. Amoon and G. A. Rezai-rad, "Automatic target recognition of synthetic aperture radar (SAR) images based on optimal selection of Zernike moments features," *IET Computer Vision*, vol. 8, no. 2, pp. 77–85, 2014.

[5] S. Gishkori and B. Mulgrew, "Pseudo-Zernike moments based sparse representations for SAR image classification," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 55, no. 2, pp. 1037–1044, 2019.

[6] X. Zhang, Z. Liu, and S. Liu, "Sparse coding of 2D-slice Zernike moments for SAR ATR," *International Journal of Remote Sensing*, vol. 38, no. 2, pp. 412–431, 2017.

[7] P. Bolourchi, H. Demirel, and S. Uysal, "Target recognition in SAR images using radial Chebyshev moments," *Signal Image & Video Processing*, vol. 11, no. 6, pp. 1–8, 2017.

[8] C. Shan, B. Huang, and M. Li, "Binary morphological filtering of dominant scattering area residues for SAR target recognition," *Computational Intelligence and Neuroscience*, vol. 2018, Article ID 9680465, 3 pages, 2018.

[9] G. C. Anagnostopoulos, "SVM-based target recognition from synthetic aperture radar images using target region outline descriptors," *Nonlinear Analysis*, vol. 71, no. 2, pp. e2934–e2939, 2009.

[10] J. Tan, X. Fan, S. Wang et al., "Target recognition of SAR images by partially matching of target outlines," *Journal of Electromagnetic Waves and Applications*, vol. 33, no. 7, pp. 865–881, 2019.

[11] A. K. Mishra, "Validation of PCA and LDA for SAR ATR," in *Proceedings of the IEEE TENCON*, pp. 1–6, Hyderabad, India, November 2008.

[12] A. K. Mishra and T. Motaung, "Application of linear and nonlinear PCA to SAR ATR," in *Proceedings of the IEEE 25th International Conference Radioelektronika (RADIO-ELEKTRONIKA)*, pp. 349–354, Pardubice, Czech Republic, April 2015.

[13] Z. Cui, Z. Cao, and J. Yang, "Target recognition in synthetic aperture radar images via non-negative matrix factorisation," *IET Radar, Sonar & Navigation*, vol. 9, no. 9, pp. 1376–1385, 2015.

[14] G. Dong and G. Kuang, "Classification on the monogenic scale space: application to target recognition in SAR image," *IEEE Transactions on Image Processing*, vol. 24, no. 8, pp. 2527–2539, 2015.

[15] G. Dong, G. Kuang, and N. Wang, "SAR target recognition via joint sparse representation of monogenic signal," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 8, no. 7, pp. 3316–3328, 2015.

[16] X. Miao and Y. Liu, "Target recognition of SAR images based on complex bidimensional empirical mode decomposition," *Scientific Programming*, vol. 10, Article ID 6642316, 2021.

[17] H. C. China, R. L. Moses, and L. C. Potter, "Model-based classification of radar images," *IEEE Transactions on Information Theory*, vol. 46, no. 5, pp. 1842–1854, 2000.

[18] B. Ding, G. Wen, and J. Zhong, "Robust method for the matching of attributed scattering centers with application to synthetic aperture radar automatic target recognition," *Journal of Applied Remote Sensing*, vol. 10, no. 1, Article ID 16010, 2016.

[19] B. Ding, G. Wen, J. Zhong, and C. X. Yang, "A robust similarity measure for attributed scattering center sets with application to SAR ATR," *Neurocomputing*, vol. 219, pp. 130–143, 2017.

[20] B. Ding, G. Wen, and X. Huang, "Target recognition in synthetic aperture radar images via matching of attributed scattering centers," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 10, no. 7, pp. 3334–3347, 2017.

[21] W. Yang, J. Lu, and Z. Cao, "A new algorithm of target classification based on maximum and minimum polarizations," in *Proceedings of the CIE International Conference on Radar*, pp. 1–4, Beijing, China, October 2001.

[22] Q. Zhao and J. C. Principe, "Support vector machines for SAR automatic target recognition," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 37, no. 2, pp. 643–654, 2001.

[23] H. Liu and S. Li, "Decision fusion of sparse representation and support vector machine for SAR image target recognition," *Neurocomputing*, vol. 113, pp. 97–104, 2013.

[24] Y. Sun, Z. Liu, and S. Todorovic, "Adaptive boosting for SAR automatic target recognition," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 43, no. 1, pp. 112–125, 2007.

[25] M. Kang, K. Ji, X. Leng, X. Xing, and H. Zou, "Synthetic aperture radar target recognition with feature fusion based on a stacked autoencoder," *Sensors*, vol. 17, no. 1, p. 192, 2017.

[26] S. Chen, H. Wang, F. Xu, and Y. Q. Jin, "Target classification using the deep convolutional networks for SAR images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 6, pp. 1685–1697, 2016.

[27] J. Zhao, Z. Zhang, and W. Yu, "A cascade coupled convolutional neural network guided visual attention method for ship detection from SAR images," *IEEE Access*, vol. 6, pp. 50693–50708, 2018.

[28] R. Min, H. Lan, and Z. Cao, "A gradually distilled CNN for SAR target recognition," *IEEE Access*, vol. 7, pp. 42190–42200, 2019.

[29] L. Wang, X. Bai, and F. Zhou, "SAR ATR of ground vehicles based on ESENet," *Remote Sensing*, vol. 11, no. 11, p. 1316, 2019.

[30] P. Zhao, K. Liu, and H. Zou, "Multi-stream convolutional neural network for SAR automatic target recognition," *Remote Sensing*, vol. 10, no. 9, p. 1473, 2018.

[31] J. Wright, A. Y. Yang, and A. Ganesh, "Robust face recognition via sparse representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 2, pp. 210–227, 2009.

[32] J. J. Thiagaraianm, K. N. Ramamurthy, and P. Knee, "Sparse representations for automatic target classification in SAR images," in *Proceedings of the 4th International. Symposium. Communication., Control Signal Processing*, pp. 1–4, Limassol, Cyprus, March 2010.

[33] L. Zhang, Z. Tao, and B. Wang, "SAR image target recognition using kernel sparse representation based on reconstruction coefficient energy maximization rule," in *Proceedings of the IEEE ICASSP*, pp. 2369–2373, Shanghai, China, March 2016.

[34] X. Xing, K. Ji, H. Zou, and J. Sun, "Sparse representation based SAR vehicle recognition along with aspect angle," *Science World Journal*, vol. 2014, Article ID 834140, 10 pages, 2014.

[35] L. Yu, L. Wang, and Y. Xu, "Combination of joint representation and adaptive weighting for multiple features with application to SAR target recognition," *Scientific Programming*, vol. 2021, Article ID 9063419, 9 pages, 2021.

[36] C. Zhang, X. Jiang, and X. Liu, "Semi-supervised SAR ATR via multi-discriminator generative adversarial network," *IEEE Sensors Journal*, vol. 19, no. 17, pp. 7525–7533, 2019.

[37] B. Chen, X. Ma, and C. Liu, "SAR target recognition based on joint sparse representation of complementary features," in *Proceedings of the 2018 International Conference on Sensors, Signal and Image Processing*, pp. 29–34, Athens, Greece, October 2018.

[38] J. R. Diemunsch and J. Wissinger, "Moving and stationary target acquisition and recognition (MSTAR) model-based automatic target recognition: search technology for a robust ATR," in *Proceedings of the 5th SPIE Algorithms Synthetic Aperture Radar Image*, vol. 3370, pp. 481–492, Orlando, FL, USA, September 1998.

[39] E. Keydel, S. Lee, and J. Moore, "MSTAR extended operating conditions: a tutorial," in *Proceedings of the SPIE*, pp. 228–242, Cambridge, UK, December 1996.

Research Article

Abnormal Event Detection in Videos Based on Deep Neural Networks

Qinmin Ma ⓘ

School of Artificial Intelligence, Shenzhen Polytechnic, Shenzhen 518055, China

Correspondence should be addressed to Qinmin Ma; mqm@szpt.edu.cn

Abnormal event detection has attracted widespread attention due to its importance in video surveillance scenarios. The lack of abnormally labeled samples makes this problem more difficult to solve. A partially supervised learning method only using normal samples to train the detection model for video abnormal event detection and location is proposed. Assuming that the distribution of all normal samples complies to the Gaussian distribution, the abnormal sample will appear with a lower probability in this Gaussian distribution. The method is developed based on the variational autoencoder (VAE), through end-to-end deep learning technology, which constrains the hidden layer representation of the normal sample to a Gaussian distribution. Given the test sample, its hidden layer representation is obtained through the variational autoencoder, which represents the probability of belonging to the Gaussian distribution. It is judged abnormal or not according to the detection threshold. Based on two publicly available datasets, i.e., UCSD dataset and Avenue dataset, the experimental are conducted. The results show that the proposed method achieves 92.3% and 82.1% frame-level AUC at a speed of 571 frames per second on average, which demonstrate the effectiveness and efficiency of our framework compared with other state-of-the-art approaches.

1. Introduction

With the development of chip technology and cost reeducation of bandwidth and storage equipment cost, etc., network digital cameras have replaced traditional analog cameras and are widely deployed in museums, banks, airport, etc. In order to strengthen public safety protection and prevent crime, the video surveillance has entered the era of blowout. According to HIS Data Display [1], the new video surveillance cameras installed in 2016 worldwide will produce approximately 566 GB of data in one day. To 2023, the data amount is estimated to reach 3500 GB. The rapid growth of video data puts forward higher requirements for video understanding. Intelligent surveillance technology has replaced traditional video surveillance personnel to achieve real-time structured processing and analysis of massive video data. As one of the key technologies of intelligent monitoring technology, abnormal event detection is from real-time detection in massive surveillance video data, which

are a small number of abnormal events that are inconsistent with most normal events.

In recent years, abnormal event detection has gradually become a research hotspot in the field of computer vision and pattern recognition. The main difficulty is that the scenes of abnormal events are diverse. It is difficult to define an interface covering the boundaries of various possible abnormal events. A common solution is to define an abnormal event as a low probability event relative to a normal event, which enables statistical processing of abnormal events, deviated from expectations, and events that are inconsistent with normal samples are abnormal events. Same as the most popular ideas in the field of computer vision and pattern recognition, the existing methods for detecting abnormal events can be roughly divided into two steps [2–4]: event representation and anomaly detection model. Event representation is to extract appropriate features from the video to represent the event. Due to the ambiguity of event definition, the event can be characterized by object-level

features or pixel-level features. The former often uses object trajectory features [5] or object appearance characteristics [6] (such as sports history images and sports energy images) to indicate an event. However, object-level features rely on detecting and tracking objects, which is difficult to handle in a crowded scene, especially for moving objects that block each other. For pixel-level features, they are often extracted from two-dimensional image blocks or three-dimensional video cubes to represent, such as spatiotemporal gradient (STG) [7], optical histograms of optical flow (HOF) [8, 9], and mixture of dynamic textures (MDT) [4]. After obtaining the characteristics that represent the event, the next question is to build an anomaly detection model. The anomaly detection model is to establish rules or models for normal events. Then, the test event that violates the rules or does not conform to the model is treated as an exception. Common models are cluster-based detection models [10], detection model based on state inference [11], and detection model based on sparse reconstruction [8, 12]. Among them, the cluster-based detection model clusters similar normal events together. Therefore, samples far away from these cluster centers during the testing phase are regarded as abnormal events. The state inference model assumes that normal events will undergo a fixed change over time. And, the abnormal event does not conform to this change. For detection models based on sparse reconstruction, the main principle is that the reconstruction of normal events has a small error relative to the reconstruction of abnormal events.

Although the above methods have achieved certain results in previous studies, there is a problem because the event representation and anomaly detection models are designed separately. Such operations cause researchers to spend too much effort to design them separately, but these methods often fail; when the video scene changes, generalization ability is poor. Recently, deep learning has achieved excellent results in the fields of computer vision and pattern recognition and intelligent manufacturing, such as object recognition [6, 13], object detection [14], behavior recognition [15], and health diagnosis. The key to the success of deep learning methods is that the two steps of feature representation and pattern recognition are jointly optimized, which can maximize the performance of the joint collaboration between them. It can further improve the generalization ability of the method for different scenarios. Driven by the success of deep learning technology, researchers began to apply it to abnormal event detection [16–18]. In [16], a three-channel architecture was proposed which used autoencoder on each channel (Autoencoder) [17]. To learn features, a single-class support vector machine (SVM) is employed afterwards to predict the anomaly score of each channel. Finally, the abnormal scores of the three channels are merged as the final basis for judging abnormalities. Sabokrou et al. introduced a cascaded anomaly detection method, which detected abnormal events based on the reconstruction error of the autoencoder and the sparsity of the sparse autoencoder. Based on manual features and short video clips, Hasan et al. adopted the fully connected autoencoder and fully convolution autoencoder to learn the time regularity of normal events. Then, according to the

reconstruction error, the time regularity score of normal events was calculated to detect abnormalities. However, these methods are based on deep reconstruction treat samples that are different from normal samples as anomalies. It ignores the small probability of abnormal events. A large number of normal samples that did not appear are often misjudged as abnormal, leading to false alarms. Unlike these methods above, in this paper, we propose an end-to-end deep learning framework for abnormal event detection. The proposed method is based on variational autoencoder (VAE) [19–22], which can map high-dimensional raw input data to low-dimensional hidden layer representations through deep learning technology. And, it constrains the low-dimensional hidden layer representation to conform to a Gaussian distribution. Therefore, the hidden layer of the normal sample indicates that the probability value calculated for the Gaussian distribution is relatively large. The hidden layer of abnormal samples indicates that the probability value calculated for the Gaussian distribution will be relatively small. Actually, obtaining the hidden layer representation and constraining to a Gaussian distribution can, respectively, correspond to the two main steps of anomaly detection: event representation and anomaly detection model. In the proposed method, the two main steps are jointly optimized through an end-to-end deep learning framework, which can improve the generalization ability. Experimental results on two public datasets show that the proposed method has strong generalization ability and the detection performance reaches the level of current technology development.

2. VAE for Anomaly Detection

The overall process of the proposed method can be described as follows. During the training phase, the space-time cube of normal samples is densely sampled. The original pixels are directly used as the input of the VAE to learn the Gaussian distribution in the hidden layer representation of the input data. Then, for a test sample, the hidden layer representation of the test sample is obtained through the VAE, which calculates the probability that it belongs to the Gaussian distribution and uses it as an anomaly score. At last, the samples with abnormal scores below the threshold are judged to be abnormal. In this section, we first briefly introduce the principle of the autoencoder. Then, the proposed method of video abnormal event detection based on variational autoencoder is elaborated.

2.1. Principle of Autoencoder. Autoencoder [17] maps the input data to the hidden layer space to get its hidden layer representation. Through its hidden layer representation, the original input data can be reconstructed. Self-encoder by encoder $f_{\mathbf{w}_1}(\bullet)$ and decoder $g_{\mathbf{w}_2}(\bullet)$ composition can be expressed as

$$
\begin{aligned}
\mathbf{z} &= f_{\mathbf{w}_1}(\mathbf{x}), \\
\mathbf{x}' &= g_{\mathbf{w}_2}(\mathbf{z}),
\end{aligned}
\tag{1}
$$

where $\mathbf{x}$ and $\mathbf{x}\prime$ represent the input of the autoencoder and the input of the reconstruction, respectively, $\mathbf{z}$ is the hidden representation for $\mathbf{x}$, and $\mathbf{W}_1$ and $\mathbf{W}_2$ are the parameters of the neural network. In order to minimize the input $\mathbf{x}$ and reconstruct the input $\mathbf{x}\prime$, the reconstruction error between is obtained as follows:

$$\min_{\mathbf{W}_1, \mathbf{W}_2} \|\mathbf{x} - \mathbf{x}'\|_2^2. \tag{2}$$

The hidden layer representation of the autoencoder is often used as effective features, which directly enter into the subsequent pattern recognition model. In order to improve the expressive ability of hidden layer representation, the noise reduction autoencoder [23] and sparse autoencoder [17] were developed by introducing noise and increasing sparsity constraints. The hidden layer representation is robust and sparsity against partial damage of data.

Suing error-based reconstruction [18, 24, 25] or directly extracting the hidden layer representation as a feature [16], autoencoders have been successfully used to solve anomaly detection tasks. However, these methods ignore the probability model in which normal samples occur with high probability and abnormal samples occur with low probability. To solve this problem, we assume that the hidden layer representation of the normal sample conforms to the Gaussian distribution, and a video abnormal event detection method based on variational autoencoder is proposed.

2.2. Anomaly Detection Model Based on VAE. Given n normal training samples $\mathbf{x} = \{x_i \in \mathbb{R}^s\}_{i=1}^n$, where the dimension of the sample is s, then the VAE [19–22] learns that the hidden layer represents the Gaussian distribution in the space. In the hidden layer representation space, assuming that the training samples conform to the Gaussian distribution, which means that all training samples are clustered into one cluster center, the samples far from the cluster center are abnormal samples.

Specifically, the hidden layer representation $\mathbf{z}$ satisfies

$$\mathbf{z} \sim \mathcal{N}(0, \mathbf{I}), \tag{3}$$

where $\mathbf{I}$ is the identity matrix. Similar to the reconstruction process of the autoencoder, VAE makes the data generated by the model very similar to the input data. Similar to the architecture of traditional autoencoders, VAE also includes two neural networks:

(1) Inferred network: a probabilistic encoder $q_\phi(\mathbf{z}|\mathbf{x})$ w will enter $\mathbf{x}$ mapped to hidden representation $\mathbf{z}$ close to reality posterior distribution $p(\mathbf{z}|\mathbf{x})$

(2) Generate network: a generative decoder $p_\theta(\mathbf{x}|\mathbf{z})$, which expresses the hidden layer without relying on any specific input prior $\mathbf{z}$ reconstruction to original training data $\mathbf{x}$

Among them, ϕ and θ represent the parameters of the two networks, respectively. Denote the network as "Encoder," the training data $\mathbf{x}$ is mapped to hidden layer representation $\mathbf{z}$. The generative network can be seen as "decoder," and the hidden layer $\mathbf{z}$ refactors to training data $\mathbf{x}$.

According to the theory of VAE [20], the loss function can be expressed as

$$\mathscr{L}(\theta, \phi, \mathbf{x}) = E_{\mathbf{z} \sim q_\phi(\mathbf{z}|\mathbf{x})}[\log p_\theta(\mathbf{x}|\mathbf{z})] - D_{KL}(q_\phi(\mathbf{z}|\mathbf{x}) \| p(\mathbf{z})). \tag{4}$$

In (5), the first item $\mathbf{x}$ is the expected log likelihood of the training data, which facilitates the decoder $p_\theta(\mathbf{x}|\mathbf{z})$ to rebuild training data $\mathbf{x}$. It can be considered as reconstruction error. When the reconstruction effect is good, the value of this item is smaller. According to the principle of Monte Carlo sampling, for each sample in the training data $\mathbf{x} = \{x_i \in \mathbb{R}^s\}_{i=1}^n$, for $q_\phi(\mathbf{z}|\mathbf{x})$ collection n a z_i, $1 \leq i \leq n$, there is

$$E_{\mathbf{z} \sim q_\phi(\mathbf{z}|\mathbf{x})}[\log p_\theta(\mathbf{x}|\mathbf{z})] \approx \frac{1}{n} \sum_{i=1}^n \log p_\theta(x_i|z_i), \tag{5}$$

where z_i is the hidden layer representation for x_i.

The second item is Kullback–Leibler divergence between $q_\phi(\mathbf{z}|\mathbf{x})$ and $p(\mathbf{z})$ [9], which represent the distribution that the encoder wants to learn and the prior distribution represented by the hidden layer. Kullback–Leibler divergence can measure the difference between two probability distributions. For two similar probability distributions, the Kullback–Leibler divergence is very small. Based on the hypothesis, $q_\phi(\mathbf{z}|\mathbf{x})$ is the normal distribution $\mathbf{N}(\mu, \delta)$, and there is

$$q_\phi(\mathbf{z}|\mathbf{x}) = \frac{1}{\sqrt{2\pi}\boldsymbol{\sigma}} \cdot e^{-(\mathbf{z}-\mathbf{u})^2/2\sigma^2}. \tag{6}$$

According to (3), $p(\mathbf{z})$ can be further expressed as

$$p(\mathbf{z}) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\mathbf{z}^2/2}. \tag{7}$$

According to (6) and (7), the second term of (4) can be expressed as

$$-D_{KL}(q_\phi(\mathbf{z}|\mathbf{x}') \| p(\mathbf{z})) = -\int q_\phi(\mathbf{z}|\mathbf{x}') \log \frac{q_\phi(\mathbf{z}|\mathbf{x}')}{p(\mathbf{z})} d\mathbf{z} = -0.5(1 + \log \boldsymbol{\sigma}^2 - \boldsymbol{\mu}^2 - \boldsymbol{\sigma}^2). \tag{8}$$

Through the reparameterization method [24], the network parameters can be adjusted by (4), suing STD [26]. The VAE is essentially based on the autoencoder, which adds a

Kullback–Leibler divergence. The hidden layer representation obtained by the encoder not only can reconstruct the input samples but also conforms to a Gaussian distribution.

Therefore, it is possible to detect abnormal events through the learned VAE.

2.3. Prediction.
After learning the network weights of the VAE, for a test sample $\mathbf{y}$, the hidden layer representation $\mathbf{z}'$ from the inferred network $q_\phi(\mathbf{z}|\mathbf{x})$ can be obtained. According to (6), the probability of $\mathbf{z}'$ belonging to the Gaussian distribution is

$$p(\mathbf{z}') = \frac{1}{\sqrt{2\pi}\boldsymbol{\sigma}} \cdot e^{-(\mathbf{z}'-\mathbf{u})^2/2\boldsymbol{\sigma}^2}. \tag{9}$$

If the test sample $\mathbf{y}$ is a normal sample, it must appear in the high probability area of the Gaussian distribution. In contrast, the hidden layer of the abnormal sample indicates that the probability value calculated for the Gaussian distribution will be relatively small. Therefore, in order to infer whether the test sample is an abnormal sample, the threshold to make judgments can be set for $p(\mathbf{z}')$ as follows:

$$p(\mathbf{z}') \overset{\text{normal}}{\underset{\text{ab normal}}{\gtrless}} \delta, \tag{10}$$

where δ determines the threshold of the sensitivity of the detection method in this paper.

3. Experiment

In order to verify the effectiveness of the proposed method, experiments were conducted on two data sets, i.e., UCSD Ped1 dataset [8] and Avenue dataset [26]. And, the results are compared with several existing methods. Afterwards, we will introduce the experimental data, evaluation index, experimental details, and experimental results in detail.

3.1. Experimental Data and Evaluation Indicators.
UCSD Ped1 dataset: the dataset records scenes on the sidewalk through a fixed camera, and the lens angle is slightly tilted. It contains 34 normal and 36 anomaly samples with the size of 238×158. Each video clip contains 200 frames. Normal events are pedestrians on the sidewalk. The abnormal events mainly include bicycles, skate, small car, and pedestrians walking on the lawn.

Avenue dataset: the dataset uses a fixed camera to record the scene in front of the school corridor and the lens angle is slightly tilted. It contains 15 normal and 21 anomaly samples with the size of 360×240. The dataset has a total of 30,652 frames. Normal events include pedestrians walking parallel to the camera. And, abnormal events include people running, throwing objects, and loitering. In Figure 1, some examples of events in two datasets are given, in which the upper pictures from each figure are normal ones while those on the bottom are anomalies.

Frame-level evaluation index and pixel-level evaluation index [11] are used to evaluate the performance of the detection method. For frame-level evaluation indicators, if a frame in the test sample contains at least one abnormal pixel, it is determined that the frame is an abnormal frame. For pixel-level evaluation indicators, if the anomalous area overlaps with the real anomaly marked area by more than

40%, it is determined that the frame is an abnormal frame. Whether it is a frame-level evaluation index or a pixel-level evaluation index, the detection rate (True Positive Rate, TPR) and false alarm rate (False Positive Rate, FPR) are calculated at first. Then, by changing the threshold δ in (10), the area under the curve (AUC) can be plotted.

3.2. Experimental Setup.
For the two datasets, every frame is resized as 160×120. Each normal sample video clip is divided into the size of $10 \times 10 \times 5$ with nonoverlapping space-time cubes. Then, these space-time cubes are converted into vectors with the size of 500×1 and normalized as the network input to train the weight of the variational autoencoder. In the proposed network, there are four hidden layers with 500, 500, 2000, and 30 neurons respectively. It uses a completely symmetrical network structure. The optimizer chooses the Adam Optimizer [8], and the initial learning rate is set to be 0.001. And, after every 1000 iterations, the learning rate reduces to 1/10 and the process stops at 10,000 iterations. The parameters are set as $\rho_1 = 0.9$ and $\rho_2 = 0.999$ and the batch size is 100. In the testing phase, the test video is also divided into sizes of $10 \times 10 \times 1$ with nonoverlapping space-time cubes. They are input into the proposed network to obtain its hidden layer representation. Then, based on (10) whether the area is abnormal can be determined. The experimental hardware platform is NVIDIA GTX1070TI with video memory 8 GB. The software environment is Tensorflow and *Python*. In order to fully evaluate the performance of the proposed method, several comparison methods are drawn from current literatures, i.e., [7, 10, 17] and [22]. For simplicity, there are denoted as "Method 1," "Method 2," "Method 3," and "Method 4," respectively.

3.3. Results and Discussion.
Figure 2 gives the results on the UCSD Ped1 dataset, where Figures 2(a) and 2(b) show the frame-level and pixel-level ROC curves. Figure 2 also provides the ROC curves of the proposed method and comparison ones. In the first three methods, the two steps of event representation and the establishment of the anomaly detection model are carried out separately. Among them, Method 1 extracts mixed dynamic texture features and then establishes a statistical inference anomaly detection model. Method 2 extracts spatiotemporal gradient features and then adopts sparse reconstruction method for anomaly detection. Method 3 uses autoencoder to extract features and single-class support vector machine for anomaly detection. Method 4 is an end-to-end deep learning method. The results of these four methods are obtained from the corresponding papers, among them Method 4 does not provide ROC curve.

As can be seen from Figure 2, the proposed method achieved the best results on the frame-level evaluation criteria. On the pixel-level evaluation standard, the results of the proposed method are not much different from those of the other two methods, i.e., Method 2 and Method 3, but obviously better than that of Method 1. Table 1 shows the comparison results of different algorithms on the UCSD Ped1 dataset at the frame level and the pixel level. The proposed method achieves 92.3% frame-level AUC and
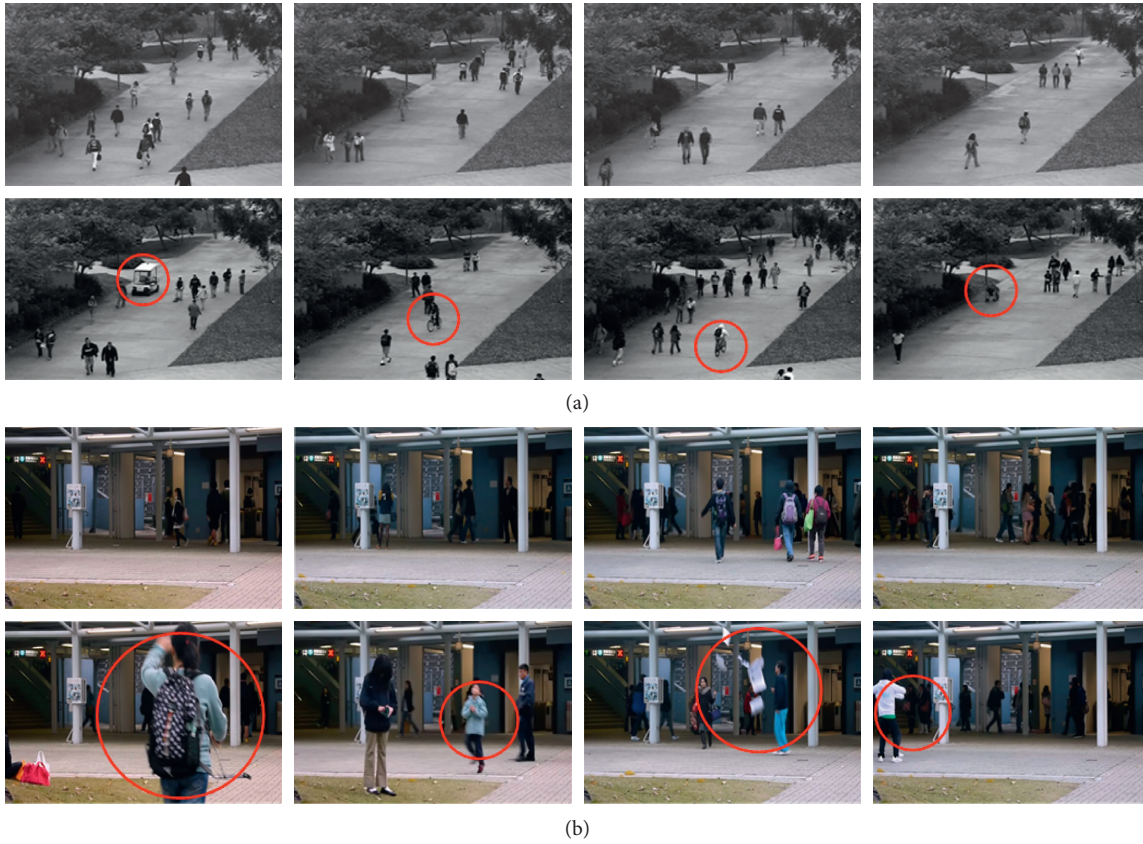
FIGURE 1: Examples of some events in the abnormal event detection dataset. (a)Examples from the UCSD Ped1 dataset. (b)Examples from the Avenue dataset.
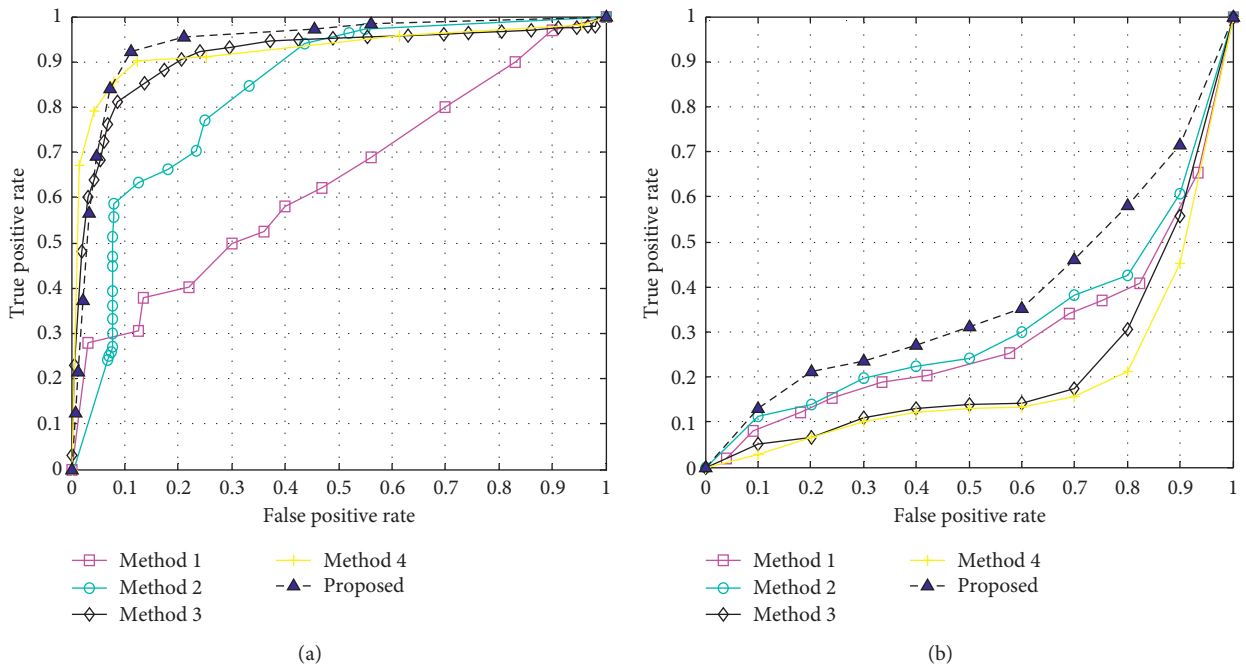
FIGURE 2: ROC curves for the UCSD Ped1 dataset. (a)Frame-level ROC. (b) Pixel-level ROC.

71.4% pixel-level AUC, which are better than all other comparison methods. It is worth noting that learning temporal regularity is also an end-to-end deep learning method. However, the experimental results are clearly lower than the proposed method. This is because the method uses each frame of the video as the input of the neural network.

TABLE 1: Comparison with the existing methods in terms of AUC% for the UCSD Ped1 dataset.

Method type	Frame-level AUC (%)	Pixel-level AUC (%)
Proposed	93.1	66.4
Method 1	82.3	45.1
Method 2	92.2	64.1
Method 3	91.9	65.2
Method 4	82.3	63.7

TABLE 2: Comparison with the existing methods in terms of frame-level AUC% for the Avenue dataset.

Method type	Frame-level AUC (%)
Ours	82.5
Method 2	81.1
Method 4	78.6

FIGURE 3: Examples of the detection results.

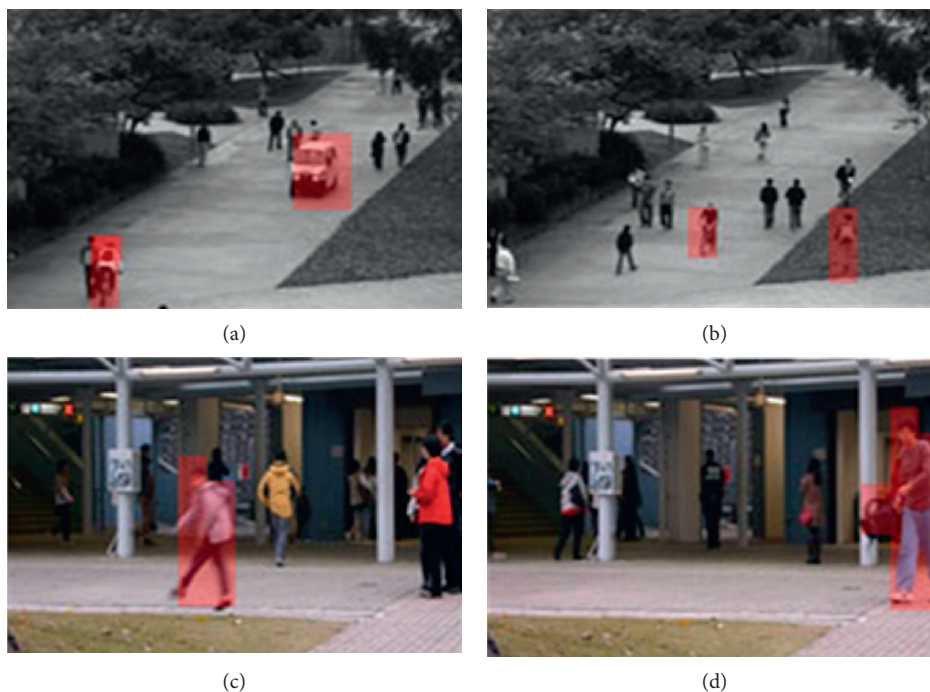Table 2 shows the frame-level detection results on the Avenue dataset. On the Avenue dataset, only Method 2 and Method 4 are tested. And, Method 4 does not give the corresponding ROC curve. Compared with the other two methods, the proposed method achieves 82.1% results in frame-level AUC, which is higher than the other two methods by 1.3% and 3.8%, respectively. The results prove that the proposed method achieves high detection accuracy and good generalization on the Avenue dataset.

Figure 3 shows examples of partially correct detection results on two datasets. Among them, (a) and (b) are the test results of the USCD Ped1 dataset and (c) and (d) are the test results from the Avenue dataset. It can be observed from Figure 3 that the proposed method can detect different types of abnormal events, including bicycle, trolley, skateboard, and trolley. So, its performance for anomaly detection can be further validated.

Table 3 shows the comparison of detection speed between the proposed method and other one on the UCSD ped1 dataset. The results of the comparison methods come from their corresponding articles. The hardware environment of the whole experiment process is Intel Core i7-8700 k 3.7 GHz CPU, NVIDIA GeForce GTX 1070Ti (8 GB video memory) GPU and 16 GB RAM memory. The computing platform is *Python* 3.7 and Tensorflow 1.7. As can be seen from Table 3, the detection speed of the proposed method is 571 fps, which obviously surpasses the detection speed of other comparison methods.

TABLE 3: Running time comparison on the UCSD Ped1 dataset.

Method type	Computing platform	CPU (GHz)	GPU	RAM (GB)	Detection speed (fps)
Proposed	*Python* 3.7+Tensorflow1.7	3.7	NVIDIA GTX 1070Ti	16	571
Method 1	—	3.0	—	2.0	0.04
Method 2	MATLAB 2012	3.4	—	8.0	143.5
Method 3	MATLAB 2015	3.5	—	16	120
Method 4	MATLAB 2015	2.1	Nvidia quadro K4000	32	0.11

4. Conclusion

In this paper, a method of video anomaly detection and location based on VAE is proposed using an end-to-end deep learning framework. The method assumes that all normal samples conform to a Gaussian distribution. The probability value of the abnormal sample in the Gaussian distribution is relatively small. In the proposed method, the two steps of event representation and establishment of anomaly detection model are, respectively, converted into the hidden layer representation and Gaussian distribution constraint in the VAE. In addition, the two steps are jointly optimized to improve the accuracy and generalization ability of the method. The quantitative results in the two public datasets show that the proposed method has reached the current technological development level. The next step of the research will consider the realization of the proposed method on more complex datasets.

Data Availability

The datasets used in this paper are publicly available.

Conflicts of Interest

The authors author declares that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

References

[1] H. Song, C. Sun, X. Wu, M. Chen, and Y. Jia, "Learning normal patterns via adversarial attention-based autoencoder for abnormal event detection in videos," *IEEE Transactions on Multimedia*, vol. 22, no. 8, pp. 2138–2148, 2020.

[2] S. Lee, H. G. Kim, and R. M. Ro, "BMAN: bidirectional multi-scale Aggregation networks for abnormal event detection," *IEEE Transactions on Image Processing*, vol. 29, pp. 2395–2408, 2019.

[3] O. Popppla and K. Wang, "Video-based abnormal human behavior recognition -a review," *IEEE Transactions on Systems, Man, and Cybernetics - Part C: Applications and Reviews*, vol. 42, no. 6, pp. 865–878, 2012.

[4] A. Bera, S. Kim, and D. Mancoch, "Real-time anomaly detection using trajectory-level crowd behavior learning," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 1289–1296, IEEE, Las Vegas, NV, USA, June 2016.

[5] O. Ye, J. Deng, Z. Yu, T. Liu, and L. Dong, "Abnormal event detection via feature expectation subgraph calibrating classification in video surveillance scenes," *IEEE Access*, vol. 8, pp. 97564–97575, 2020.

[6] C. Lu, J. Shi, and J. Jia, "Abnormal event detection at 150 FPS in MATLAB," in *Proceedings of IEEE International Conference on Computer Vision*, pp. 2720–2727, ICCV), December 2013, Sydney, Australia.

[7] Y. Cong, J. Yuan, and J. Liu, "Abnormal event detection in crowded scenes using sparse representation," *Pattern Recognition*, vol. 46, no. 7, pp. 1851–1864, 2013.

[8] Y. Yuan, J. Fang, and Q. Wang, "Online anomaly detection in crowd scenes via structure analysis," *IEEE Transactions on Cybernetics*, vol. 45, no. 3, pp. 562–575, 2015.

[9] W. Li, V. Mahadevan, and N. Vasconcelos, "Anomaly detection and localization in crowded scenes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 1, pp. 18–32, 2014.

[10] J. Kwon and K. M. Lee, "A unified framework for event summarization and rare event detection from multiple views," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 9, pp. 1737–1750, 2015.

[11] H. Ren, W. Liu, S. I. Olsen, S. Escalera, and T. B. Moeslund, "Unsupervised behavior-specific dictionary learning for abnormal event detection," in *Proceedings of the British Machine Vision Conference*, Swansea, Wales, September 2015.

[12] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778, IEEE, Las Vegas, NV, USA, June 2016.

[13] Z. Cai, N. Vasconcelos, and R.-C. N. N. Cascade, "Delving into high quality object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2348–2357, IEEE, Salt Lake City, UT, USA, June 2018.

[14] Y. Zhao, Y. Xiong, and D. Lin, "Recognize actions by disentangling components of dynamics," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2348–2357, IEEE, Salt Lake City, UT, USA, June 2018.

[15] D. Xu, Y. Yan, E. Ricci, and N. Sebe, "Detecting anomalous events in videos by learning deep representations of appearance and motion," *Computer Vision and Image Understanding*, vol. 156, pp. 117–127, 2017.

[16] Q. Zhang, L. T. Yang, Z. Chen, and P. Li, "A survey on deep learning for big data," *Information Fusion*, vol. 42, pp. 146–157, 2018.

[17] M. Sabokrou, M. Fathy M, and M. Hoseini, "Video anomaly detection and localization based on the sparsity and reconstruction error of autoencoder," *IET Electronic Letter*, vol. 52, no. 13, pp. 1122–1124, 2016.

[18] M. Hasan, J. Choi, J. Neumanny, A. K. Roy-Chowdhury, and L. S. Davis, "Learning temporal regularity in video sequence," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778, IEEE, Las Vegas, NV, USA, June 2016.

[19] R. Xie, N. M. Jan, K. Hao, L. Chen, and B. Huang, "Supervised variational autoencoders for soft sensor modeling with missing data," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 4, pp. 2820–2828, 2020.

[20] Y. Zerrouki, F. Harrou, N. Zerrouki, A. Dairi, and Y. Sun, "Desertification detection using an improved variational autoencoder-based approach through ETM-landsat satellite data," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 202–213, 2021.

[21] Y. Wang, B. Dai, G. Hua, J. Aston, and D. Wipf, "Recurrent variational autoencoders for learning nonlinear generative models in the presence of outliers," *IEEE Journal of Selected Topics in Signal Processing*, vol. 12, no. 6, pp. 1615–1627, 2018.

[22] T. Zhao, F. Li, and P. Tian, "A deep-learning method for device activity detection in mMTC under imperfect CSI based on variational-autoencoder," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 7, pp. 7981–7986, 2020.

[23] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," in *Proceedings of the International Conference on Learning Representations*, Banff, CA, USA, April 2014.

[24] P. Vincent, H. Larcochele, Y. Bengio, and P.-A. Manzagol, "Extracting and composing robust features with denoising autoencoders," in *Proceedings of the 25th International Conference on Machine Learning*, pp. 1096–1103, Helsinki, Finland, June 2008.

[25] M. Song, A. Montanari, and P.-M. Nguyen, "A mean field view of the landscape of two-layer neural networks," *Proceedings of the National Academy of Sciences*, vol. 115, no. 33, pp. E7665–E7671, 2018.

[26] T. Wang and H. Snoussi, "Detection of abnormal visual events via global optical flow orientation histogram," *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 6, pp. 988–998, 2014.

Research Article

Learning Deep Attention Network from Incremental and Decremental Features for Evolving Features

Chuxin Wang [1] and Haoran Mo[2]

[1]*Hebei University of Environmental Engineering, Qin Huang Dao 066000, China*
[2]*Innopolis University, Innopolis, Russia*

Correspondence should be addressed to Chuxin Wang; wangchuxin@hebuee.edu.cn

In many real-world machine learning problems, the features are changing along the time, with some old features vanishing and some other new features augmented, while the remaining features survived. In this paper, we propose the cross-feature attention network to handle the incremental and decremental features. This network is composed of multiple cross-feature attention encoding-decoding layers. In each layer, the data samples are firstly encoded by the combination of other samples with vanished/ augmented features and weighted by the attention weights calculated by the survived features. Then, the samples are encoded by the combination of samples with the survived features weighted by the attention weights calculated from the encoded vanished/ augmented feature data. The encoded vanished/augmented/survived features are then decoded and fed to the next cross-feature attention layer. In this way, the incremental and decremental features are bridged by paying attention to each other, and the gap between data samples with a different set of features are filled by the attention mechanism. The outputs of the cross-feature attention network are further concatenated and fed to the class-specific attention and global attention network for the purpose of classification. We evaluate the proposed network with benchmark data sets of computer vision, IoT, and bio-informatics, with incremental and decremental features. Encouraging experimental results show the effectiveness of our algorithm.

1. Introduction

1.1. Background. In the machine learning problems, a basic assumption is the data samples have consistent and stable features. These features are usually generated by a set of sensors and used by the machine learning models as inputs. However, in many real-world applications, this assumption does not hold, and the features are changing with some old features vanishing and some new features added. For example, in the application of environmental monitoring, different sensors are deployed, including gravimetric, optical, and electrochemical sensors [1–4]. These sensors have different life cycle lengths and different working conditions. Thus, some sensors expired sooner than the others; thus, the corresponding features vanished sooner. Meanwhile, some other sensors can be used for a long time to continue to generate features. Their features will be surviving along the

data collection process. Moreover, with the development of sensors, some new sensors are produced and deployed and begin to generate newly augmented features. As a result, the working sensors are evolving over time and the features are changing accordingly. Some old features are vanishing and some new features are augmented, while the remaining features survive. This scenario makes the feature not stable and challenges the stable feature assumption of most popular machine learning settings [5–8]. This problem is called the incremental and decremental feature (IDF) problem. Given the importance of the IDF problem, surprisingly, only very few works have been done to solve it directly [6], and the performance is not satisfying.

Meanwhile, the deep attention network has been a popular method for the machine learning area. Attention mechanism represents a data instance not only by itself but also by paying attention to the other instances weighted by

the attention weights. The attention weights are usually calculated according to the instance features and then normalized by the softmax function [9–12]. There are two types of attention network: self-attention [13–16] and cross-attention networks [17–20]. The self-attention mechanism calculates the attention weight of each instance from itself, while the cross-attention mechanism usually calculates attention weights according to the similarity between itself and the other instances. However, most existing attention network only pays attention to instances and assuming the features are stable. Thus, the attention mechanism of existing methods cannot be applied to the IDF problem.

In this paper, we propose a novel solution for the IDF problem with a cross-feature attention network. Our solution pays attention to the vanished, survived, and augmented features to bridge the gaps among the features of evolving sensors. This is the first work of attention mechanism across features, and it fits the nature of the IDF problem.

1.2. Related Works. In this section, we review the related works of IDF; even there are only very few such existing works.

  (i) Hou and Zhou [6] developed an algorithm to handle the incremental and decremental features and the streaming data instances. This algorithm has two stages. The first stage is to compress the vanished features by learning a classifier in the vanished feature space so that the important information of the vanished features is embedded in the trained classifier. The second stage is an expanding stage. It will not only include the augmented features in this stage but also try to balance the vanished features, survived features, and augmented features. The balancing is conducted by imposing the classification responses with/without the augmented features. Moreover, the learning strategy is one-pass learning, which takes only one training sample to update the model in each iteration.

  (ii) Ma et al. [7] designed a transfer learning method for the domains, where only a part of the features are shared, while the other features are different. This problem setting is similar to the IDF, given the partially shared features space. To be specific, the target domain not only has the source domain's feature but also has some newly augmented features. To solve this problem, this method also imposes the target domain's classification responses of data with/without augmented features to be consistent with each other. Moreover, the features shared across the source and target domains are also jointly regularized to be consistently sparse, i.e., the importance of the same feature should be consistent across two domains.

  (iii) Wu et al. [21] proposed a feature selection algorithm to handle the streaming features. In this scenario, the features are not known from the very beginning of the learning process, but come in a one-by-one

way, while the number of training samples remains the same. The algorithm is designed to select the most important features from the streaming feature set. The selected features should be not only relevant but also nonredundant. The feature selection is performed in an iterative algorithm. When the algorithm receives a new feature, the algorithm first determines if it is relevant to the class. If not relevant and is also redundant, it will be dropped. Otherwise, this feature is selected. The problem of streaming features is a special case of the IDF, where it only handles the streaming incremental features but ignores the decremental features.

Among these existing methods, the IDF problem is solved by imposing consistency of classification responses with/without augmented features, in both works of [6, 7]. The intersection of vanished/augmented/survived features is not explored directly. Thus, the cross-feature information is not utilized effectively to boost the learning performance.

1.3. Our Contribution. To fill this gap, in this paper, we propose the first attention network to pay attention from one feature set to another one. The motivation to do so is that we believe even the feature changes in the sample batches collected from a different time, and they have an inner relationship and they are complimentary for the purpose of classification of the samples. The sensor evolving changes the observed features, but actually, the features should be complete and consistent in an ideal situation where all the sensors do not expire and are all deployed at the very beginning. Thus, we would like to recover the vanished features for the new batch of data, and also recover the augmented features for the old batch of data. For this purpose, we encode each sample by paying attention to the vanished/augmented features. To explore the feature relationship, we calculate the attention weight by the survived features. In this way, we have a vanished/augmented feature-attention code vector for each sample, and even it has no vanished/augmented features, by bridging itself to the samples with vanished/augmented features with help of survived-feature attention. With the vanished/augmented feature-attention code vectors, we pay attention back to the survived features by encoding each sample as the combination of other samples' survived features. The attention weights are again calculated by the codes of the last cross-attention layers. Decoders are also applied to recover the original features from the code vectors, and the recovered features are inputs of the next cross-attention layers. In this way, we design a deep cross-feature attention network to represent the samples with IDF. The encoded vectors of the network are further represented by a set of class-specific attention networks and a global attention network for the purpose of classification.

Our contribution is threefold:

  (1) We design a novel deep neural network with new cross-attention layers for the purpose of learning from evolving

(2) We propose a novel learning algorithm to optimize the parameters of the network in a supervised way

(3) We evaluate the proposed algorithm experimentally regarding parameter sensitivity, running time, and comparison to other algorithms

1.4. Paper Organization. This paper is organized as follows. In Section 2, we describe the new network of cross-feature attention. In Section 3, we evaluate the proposed method experimentally. In Section 4, the conclusion is given.

2. Cross-Feature Attention Network

2.1. Problem Setting. In this section, we discuss a learning problem with changing features. Suppose we have training set of n data samples, and these samples belong to two batches. One batch has no data samples generated from a set of historical sensors, and the other batch has $n_c = n - n_o$ data samples generated from the current set of sensors. Compare to the current sensor set, some old sensors have vanished, some new senors have been added to the current sensor set, and the remaining sensors remain the same. The old data batch is denoted as $X^o = \{(x_i, y_i, \psi_i)\}_{i=1}^{n_o}$, where $x_i \in R^{d1}$ is the ith sample's feature vector of the d_1 vanished sensors, $y_i \in R^{d2}$ is its feature vector of the d_2 survived sensors, and $\psi_i \in \{1, \ldots, C\}$ is its class label. The current data batch is denoted as $Xc = \{y_i, z_i, \psi_i\}_{i=n_o+1}^{n}$, where $y_i \in R^{d2}$ is the ith sample's feature vector of the d_2 survived sensors, while $z_i \in R^{d3}$ is its feature vector of the d_3 newly added sensors, and ψ_i is its class label. The overall training data set is $X = Xo \cup Xc$, and the learning problem is to learn a model to predict the class label of a test data sample with features of the current sensors.

2.2. Network Architecture. To represent each data sample of both current and old batches, we propose a deep cross-feature attention representation network and a discriminative network to separate samples of different classes.

2.2.1. Cross-Feature Attention Layers. Given the ith data sample, to represent it, we propose to pay attention from itself to three feature spaces, which are the vanished sensor space, survived sensor space, and newly added sensor space.

(1) Attention to Vanished Sensor Space. We firstly pay attention from the ith sample to the old batch Xo, even the ith sample is from the current batch. To this end, we calculate its similarity to the jth sample of Xo in the feature space of the survived sensors shared by both batches. The similarity between the ith and jth sample is calculated as

$$s(y_i, y_j; A) = y_i^\top A y_j, \quad j \in Xo, \tag{1}$$

where $A \in R^{d2 \times d2}$ is the parameter matrix of the similarity function. The attention score from the ith sample to the jth

sample regarding the survived sensors is obtained by applying a softmax function to the similarity scores:

$$\alpha_{ij} = \frac{\exp(s(y_i, y_j; A))}{\sum_{j\prime \in Xo} \exp(s(y_i, y_{j\prime}; A))}. \tag{2}$$

With the attention scores calculated from survived sensor features, we represent the ith sample by combining the transformed features of the vanished sensor features weighted by these attention scores:

$$f_i = \sum_{j: j \in Xo} \alpha_{ij} \Theta^\top x_j, \tag{3}$$

where Θ is the transforming matrix. Please note, in this attention-based representation of the ith sample, the attention scores are calculated in the survived sensor space, while the attention base vectors are in the vanished sensor features space.

(2) Attention to Newly Added Sensor Space. We also pay attention to the current batch of training samples in the space of newly added sensors. To this end, we calculate the attention weights in the space of the survived sensors and use them to weigh the samples of the current batch in the new sensor space. We firstly calculate the similarity between the ith sample ($i \in X$) and the jth sample of the current batch:

$$s(y_i, y_j; B) = y_i^\top B y_j, \quad j \in Xc, \tag{4}$$

where B is the similarity function parameter matrix. Accordingly, we apply a softmax function to the similarities to calculate the attention weights:

$$\beta_{ij} = \frac{\exp(s(yi, yj; B))}{\sum_{j\prime \in Xc} \exp(s(yi, yj\prime; B))}. \tag{5}$$

The new representations of the ith sample by the attention to the current training batch in the space of newly added sensor space is the combination of the transformed samples with the above weights:

$$h_i = \sum_{j: j \in Xc} \beta_{ij} \Phi^\top z_j, \tag{6}$$

where Φ is the transforming parameter matrix. The cross-feature mapping is performed from newly added features with weights of the survived features.

(3) Attention to Survived Sensor Space. Paying attention to the samples of the entire data set is weighted by the representations of the above two layers. Given the ith sample, we firstly concatenate the two vectors of the last two attention layers, f_i and h_i, to a longer vector, $t_i = \begin{bmatrix} f_i \\ h_i \end{bmatrix} \in R^{d1+d3}$. With this vector, we calculate the similarity between two samples, the ith and jth samples:

$$s(ti, tj; E) = t_i^\top E t_j, \quad j \in X, \tag{7}$$

where E is the similarity function parameter matrix. The attention weights are calculated by softmax:

$$\gamma_{ij} = \frac{\exp\left(s\left(t_i, t_j; E\right)\right)}{\sum_{j_\prime \in X} \exp\left(s\left(t_i, t_{j_\prime}; E\right)\right)}. \tag{8}$$

The attention layer output vector of the ith sample is the combination of the features of survived sensors weighted by attention weights in (8):

$$g_i = \sum_{j:\ j \in X} \gamma_{ij} \Psi^\top y_j, \tag{9}$$

where Ψ is the transforming matrix.

(4) Decoding Layer. With the above three layers of cross-attentions, we have three representation vectors f_i, g_i, and h_i. We can further decode the sensor features from these vectors for the next layers' inputs in a deep network architecture. The decoding layers are dense layers with activation layers:

$$\begin{aligned} x_i^{\text{new}} &= \varphi\left(W^\top f_i\right), \quad \forall i \in Xo, \\ y_i^{\text{new}} &= \varphi\left(V^\top g_i\right), \quad \forall i \in X, \\ z_i^{\text{new}} &= \varphi\left(R^\top h_i\right), \quad \forall i \in Xc, \end{aligned} \tag{10}$$

where W, V, and R are the dense layer parameter matrices and $\varphi(\cdot)$ is the activation function.

Given the above base layers, we build a multiple layer cross-feature attention network by feeding the outputs of the decoding layer to the next layers of old and newly added sensor attention layers. In the lth layer, the output of the $l-1$th layer is x_i^{l-1}, y_i^{l-1}, and z_i^{l-1} for the lth sample, and itwill be used to estimate f_i^l and h_i^l of this layer according to (3) and (6):

$$\begin{aligned} f_i^l &= \sum_{j:\ j \in Xo} \alpha_{ij}^l \Theta_l^\top x_j^{l-1}, \\ h_i^l &= \sum_{j:\ j \in Xc} \beta_{ij}^l \Phi_l^\top z_j^{l-1}. \end{aligned} \tag{11}$$

Then, f_l iand f_i^l will be used to recalculate the weights of (8), γ_{ij}^l, and finally estimate g_i^l as

$$g_i^l = \sum_{j:\ j \in X} \gamma_{ij}^l \Psi_l^\top y_j^{l-1}. \tag{12}$$

The decoding layer is applied to generate the outputs of this layer:

$$\begin{aligned} x_i^l &= \varphi\left(W_l^\top f_i^l\right), \quad \forall i \in Xo, \\ y_i^l &= \varphi\left(V_l^\top g_i^l\right), \quad \forall i \in X, \\ z_i^l &= \varphi\left(R_l^\top h_i^l\right), \quad \forall i \in Xc. \end{aligned} \tag{13}$$

The cross-feature attention layer is shown in Figure 1. We can see that, in this layer, the input data has three sets of features, and the attention is paid from one feature to another. To be more specific, the new presentation of one feature is the combination of samples in this feature space, but the weights of attention are estimated from another feature space.

Suppose we have L layers of cross-feature attention and the outputs of the last attention layers are f_i^L, g_i^L, and h_i^L. In our implementation, we set the layer number L to 12. They
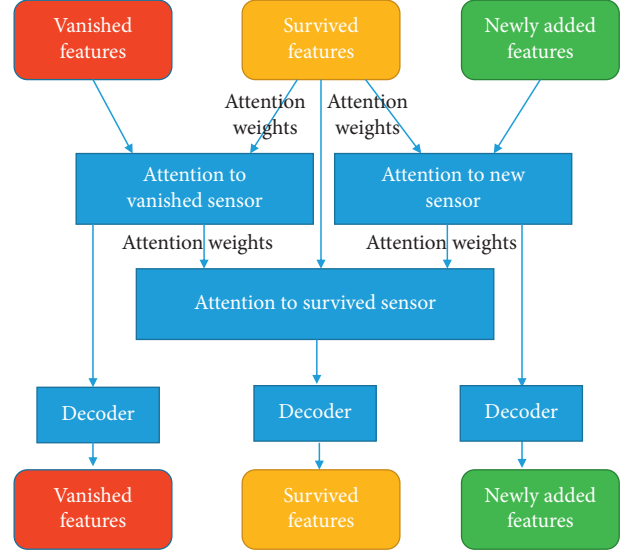

FIGURE 1: Cross-feature attention layer.

are further concatenated as a long vector to represent the ith sample as follows:

$$u_i = \begin{bmatrix} f_i^L \\ g_i^L \\ h_i^L \end{bmatrix}. \tag{14}$$

This vector will be the input of the next class-specific attention network for the purpose of classification.

2.2.2. Class-Specific Attention Layers. Given the ith samples cross-feature attention representation, u_i, and its class label, ψ_i, we have used two class-specific attention layers to map it to the space of its won class and the entire data set of all classes.

(1) Class-Specific Attention Layer. To represent the ith sample, we pay attention to the samples of the same class, j: $\psi_i = \psi_j$. The attention weight is again calculated according to the similarity between u_i and u_j:

$$s\left(u_i, u_j; \iota_\psi\right) = \varphi\left(\iota_\psi^\top \begin{bmatrix} u_i \\ u_j \end{bmatrix}\right), \quad \forall j:\ \psi_i = \psi_j, \tag{15}$$

where the similarity function is based on the concatenation of u_i and u_j, a dense layer parameterized by ι_ψ, and an activation layer $\varphi(\cdot)$. The attention weights from the ith sample to class ψ are calculated by softmax normalization over the samples of the class ψ:

$$\delta_{ij}^\psi = \frac{\exp\left(s\left(u_i, u_j; \iota_\psi\right)\right)}{\sum_{j_\prime:\ \psi_{j_\prime} = \psi} \exp\left(s\left(u_i, u_j; \iota_\psi\right)\right)}, \quad \forall j:\ \psi_j = \psi. \tag{16}$$

The class-specific attention representation of the ith sample regarding to class ψ is the combination of the weighted samples of class ψ:

$$P_i^\psi = \sum_{j: \ \psi j = \psi} \delta_{ij}^\psi \Omega_\psi^\top u j, \tag{17}$$

where Ω_ψ the projection matrix.

(2) Global Attention Layer. Beside the class-specific attention layers, we also build a global attention layer to represent the ith sample to all the data samples of the entire data set. The attention weights are calculated from the ith sample to all samples, $j \in X$. The similarity between the ith and jth samples is also based on a concatenation, a dense, and a activation layer:

$$s(u_i, u_j; \pi) = \phi\left(\pi^\top \begin{bmatrix} u_i \\ u_j \end{bmatrix}\right), \quad \forall j: \ j \in X, \tag{18}$$

where π is the dense layer parameter. Accordingly, the weights of attention are normalized by a softmax:

$$\varpi_{ij} = \frac{\exp\left(s(u_i, u_j; \pi)\right)}{\sum_{j': \ j' \in X} \exp\left(s\left(u_i, u_{j'}; \pi\right)\right)}, \quad \forall j \in X. \tag{19}$$

The global attention representation of the ith sample is

$$q_i = \sum_{j: \ j \in E} \varpi_{ij} \Upsilon^\top u_j, \tag{20}$$

where Υ the the projection matrix.

With these layers, for each data sample, we have two representations, which are class-specific attention vector, $p_i^{\psi i}$, and a global attention vector, q.

2.3. Network Training. There are many parameters of the proposed network. To learn these parameters, we firstly model a minimization problem with the training data set and then develop an iterative optimization algorithm to solve it.

2.3.1. Objective Function. To train the parameters of the cross-feature attention network and class-specific/global attention network, we consider the following two problems:

(1) Minimization of Within-Class Scattering. For each class ψ, we hope that its samples' representations of this class are not scattered so that they can be gathered as close as possible. The samples' class-specific representations are $p_i^\psi|_{i: \ \psi_i = \psi}$. To

measure the within-class scattering, we first calculate the mean vector of this class as

$$\mu^\psi = \frac{1}{n_\psi} \sum_{i: \ \psi_i = \psi} p_i^\psi, \tag{21}$$

where n_ψ is the number of samples of the class ψ. The within-class scattering measure of class ψ is calculated as

$$S_\psi^W = \sum_{i: \ \psi_i = \psi} \text{Tr}\left((p_i^\psi - \mu_\psi)(p_i^\psi - \mu_\psi)\top\right), \tag{22}$$

where $\text{Tr}(\cdot)$ is the trace of a matrix. The following minimization problem is modeled to optimize the parameters,

$$\min \sum_{\psi=1}^{C} S_\psi^W, \tag{23}$$

so that, for all the classes, the within-class scattering is minimized jointly.

(2) Maximization of Interclass Scattering. We also propose to maximize the scattering of different classes. For this purpose, we firstly calculate an overall mean vector over the entire data set, using the global attention representations:

$$\mu = \frac{1}{n} \sum_{i: \ i \in X} q_i. \tag{24}$$

Meanwhile, in the global attention representation space, we also calculate the mean vectors for each class, ψ:

$$\rho^\psi = \frac{1}{n\psi} \sum_{i: \ \psi i = \psi} q_i. \tag{25}$$

The interclass scattering is measured as

$$S_B = \sum_{\psi=1}^{C} n_\psi \text{Tr}\left((\rho^\psi - \mu)(\rho^\psi - \mu)^\top\right). \tag{26}$$

To separate different classes, we propose to maximize it:

$$\max S_B. \tag{27}$$

The overall objective of this problem is the combination of (23) and (27).

A minimization problem is proposed as follows to learn the parameters of the network:

$$\text{Min}_\Pi \left\{ o(\Pi) = \sum_{\psi=1}^{C} S_\psi^W - S_B + C\|\Pi\|_F^2 = \sum_{\psi=1}^{C} \sum_{i: \ \psi i = \psi} \text{Tr}\left((p_i^\psi - \mu_\psi)(p_i^\psi - \mu_\psi)^\top\right) - \sum_{\psi=1}^{C} \text{Tr}\left((p_i^\psi - \mu)\psi^\top\right) + C\|\Pi\|_F^2 \right\}, \tag{28}$$

where $\Pi = \left\{(A_l, B_l, E_l, \Theta_l, \Phi_l, \Psi_l, W_l, V_l, R_l)|_{1=1}^{L}, \ (\iota_\psi, \Omega_\psi) |_{\psi=1}^{C}, \pi, \Upsilon\right\}$, $\|\Pi\|_F^2$ is the squared ℓ_2 norms of the parameters to

prevent the overfitting problem, and C is the weight of the squared ℓ_2 norm term.

2.3.2. Objective Optimization. To solve the problem of (28), we employ the algorithm of Adam [22]. This algorithm is based on gradient descent updating of the parameters, and its optimization is for the stochastic objective. The lower-order moments is updated adaptively.

2.4. Model Inference. With the trained parameters of the network, we can inference the class label of a test sample. Its survived feature vector is y, and its newly added feature vector is z. We firstly represent it by the trained cross-feature network as u, according to (14). Then, for each class, we calculate its class-specific representation p_ψ and a global representation q. Then, we calculate the distance between the specific representation of the test sample and the class mean regarding to the class ψ:

$$s_w(\psi) = \mathrm{Tr}\Big(\big(p_\psi - \mu_\psi\big)\big(p_\psi - \mu_\psi\big)^\top\Big). \tag{29}$$

Moreover, we also update the mean vector of this class in the global representation space by

$$\rho_\psi = \frac{1}{n\psi + 1}\left(q + \sum_{i:\ \psi i = \psi} q_i\right). \tag{30}$$

With the updated class mean, we recalculate its distance to the overall mean vector μ as follows:

$$s_b(\psi) = \mathrm{Tr}\Big(\big(\rho_\psi - \mu\big)\big(\rho_\psi - \mu\big)^\top\Big). \tag{31}$$

The overall score of assigning the test sample the class ψ is the difference of $s_b(\psi)$ and $s_w(\psi)$:

$$\eta(\psi) = s_b(\psi) - s_w(\psi). \tag{32}$$

It measures how close the test sample is to the class ψ and how it makes the class ψ far away from the other classes. The test sample is assigned to the class which gives the largest score:

$$\psi^* = \arg\max_{\psi=1}^{C} \eta(\psi). \tag{33}$$

3. Experiments

In this section, we evaluate the proposed method cross-feature attention network (CFAN) experimentally. We firstly introduce the data sets and the experimental setting, then give the experimental results, and summarize the observations from the results.

3.1. Data Sets and Experimental Protocol

3.1.1. Data Sets. In the experiment, we use four data sets as benchmarks, including two computer vision data sets, an Internet of things (IoT) data set, and a bio-informatics data set. The statistics of the data sets is given in Table 1. The details of the data sets are as follows.

 (i) Satimage is an image data set. It has 6,431 images, and the problem is to categorize each image into one of the 7 categories, including red soil, cotton crop, grey soil, damp grey soil, soil with vegetation stubble, mixture class (all types present), and very damp grey soil. Each image is represented by 36 features, which are the 9 pixels in the neighborhood of 4 spectral bands [23].

 (ii) MINIST is a hand-written digit data set. It has a training set of 60,000 images and a test set of 10,000 images. Each image has 28×28 pixels and a class label of ten digits; thus, it is a 10-class classification problem [24].

 (iii) SensIT vehicle is an acoustic data set. It has 78,823 training samples and 19,705 test samples. For each sample, it has 50 features. The learning problem of this data set is a 3-class classification problem [25].

 (iv) Protein data set is a bio-informatics data set. It has 32,661 training data samples, 6,621 test data samples, and 2,871 evaluation samples. Each sample has 357 features. The problem for this data set is a 3-class classification problem [26].

3.1.2. Protocol. To perform the experiments, we split the features set of each data set into three subsets, so that each set has an equal size. The three feature sets are the vanished features, survived features, and newly added features, respectively. To create the training and test data set, we use the 10-fold cross-validation protocol. A data set is split into 10 folds of equal sizes, and each fold is used as a test set, while the other 9 folds are combined to form a training set. We firstly train the model over the training set and then test it over the test set. To measure the accuracy of the classification, classification accuracy is used. It is calculated as the rate of correctly classified samples from the overall test samples.

3.2. Experimental Results. In this section, given the data and protocol, we perform the experiments and report the results. We evaluate the proposed method from three different aspects, including the sensitivity to the tradeoff parameter, the running time, and the performance compared to the state of the art.

3.2.1. Parameter Sensitivity. In the objective of our model in (28), there is a tradeoff parameter C to balance the regularization term and other terms. It controls the weight of the model complexity. To evaluate how it affects the performance of the model, we plot the curves of the accuracy against the changing values of C in Figure 2. From the figure, we observe that, with the increase of the weight of the regularization term, the accuracy is improved slightly. However, generally speaking, the performance keeps stable regarding the change of the value of C. A simpler model with a larger value of C can improve the quality of the model, but the improvement is not significant.

3.2.2. Running Time Analysis. The running time of the training process of the model is also studied in the

Table 1: Statistics of data sets.

Data set	# of data sample	# of class	# of feature
Satimage	6,431	7	36
MINIST	70,000	10	784
SensIT vehicle	98,528	3	50
Protein	42,153	3	357

Figure 2: Sensitivity curve of tradeoff parameter C.

Figure 3: Running time of training and test process.
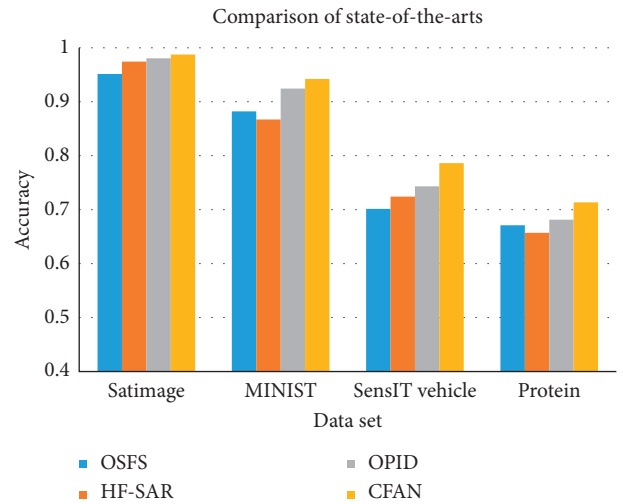
Figure 4: Comparison of accuracy of state of the arts.
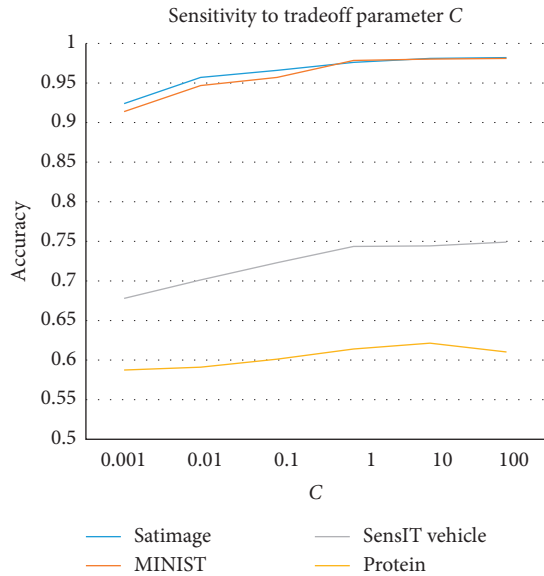
experiment. Moreover, the running time of the classification of the test samples is also reported. The running time over four benchmark data sets is given in Figure 3.

(1) From the figure, we can see that the training time is longer than the test time for each data set. This is natural because the training algorithm scans each training sample for many iterations, and the number of training samples is also larger than the test sample. Meanwhile, in the test process, each test sample is only scanned once.

(2) SensIT Vehicle and MINIST have a longer running time than that of Satimage and Protein since the data set sizes are larger.

(3) With ten thousands of samples, the running time of training and test is only hundreds of seconds. This indicates the efficiency of the algorithm.

3.2.3. Comparison to State of the Arts. We compared our algorithm CFAN against the other three state-of-the-art algorithms, including

(i) One-pass incremental and decremental learning approach (OPID) [6]

(ii) Heterogenous feature-based structural adaptive regression (HF-SAR) [7]

(iii) Online streaming feature selection (OSFS) [21]

The accuracy of these methods is reported in Figure 4. We have the following observations from this figure:
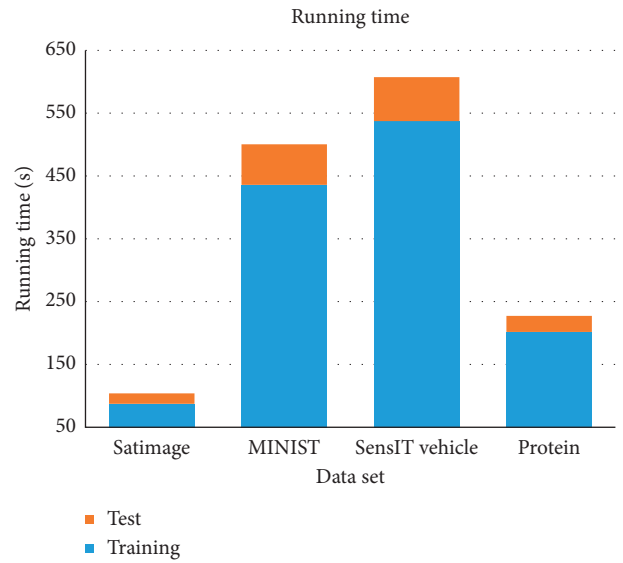
(1) In all the cases, the CFAN algorithm keeps outperforming the compared methods, especially in the most challenging data sets, SensIT Vehicle and Protein. This is a strong indicator of the effectiveness and advantage of CFAN over the compared methods. The main reason is the power of the cross-feature attention layers which takes advantage of the essential nature of the changing features due to the evolving sensors.

(2) The second best algorithm is OPID, which is also specially designed for the IDF problem. However, because it used as a linear model to model the evolving features, it fails to capture the complex pattern of the features. In contrast, CFAN used the deep attention layers for this purpose, thus giving much better results.

(3) HF-SAR and OSFS give the worst results. They can handle some special cases of IDF but are not perfect solutions for this problem.

4. Conclusion

In this paper, we proposed a novel solution for the machine learning problem with evolving features. In the process of learning, old features vanished, while new features are argumented, and the remaining features survived. To handle the evolving features, we use a deep network structure with a newly designed cross-feature attention layer. This layer fills the gap between the survived features and vanished/argumented features, by paying attention from/to different feature sets. In this way, data samples with a different sets of features are mapped to a common feature space. To learn the attention network parameters, we proposed to construct the class-specific attention layer to minimize the within-class scattering and the global attention layer for the maximization of interclass scattering. Experimental results show the stability of the algorithm and the outperforming of the proposed algorithm against the other methods.

Data Availability

The dataset used in this paper are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] J. Hwang, C. Shin, and H. Yoe, "Study on an agricultural environment monitoring server system using wireless sensor networks," *Sensors*, vol. 10, no. 12, Article ID 11189, 2010.

[2] P. Jiang, H. Xia, Z. He, and Z. Wang, "Design of a water environment monitoring system based on wireless sensor networks," *Sensors*, vol. 9, no. 8, pp. 6411–6434, 2009.

[3] M. F. Othman and K. Shazali, "Wireless sensor network applications: a study in environment monitoring system," *Procedia Engineering*, vol. 41, pp. 1204–1210, 2012.

[4] G. Xu, W. Shen, and X. Wang, "Applications of wireless sensor networks in marine environment monitoring: a survey," *Sensors*, vol. 14, no. 9, Article ID 16932, 2014.

[5] J. Dong, Y. Cong, G. Sun, T. Zhang, and X Xu, "Evolving metric learning for incremental and decremental features," 2020, https://arxiv.org/abs/2006.15334.

[6] C. Hou and Z. H. Zhou, "One-pass learning with incremental and decremental features," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 11, pp. 2776–2792, 2017.

[7] Z. Ma, Y. Yang, N. Sebe, and A. G. Hauptmann, "Knowledge adaptation with PartiallyShared features for event DetectionUsing few exemplars," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 9, pp. 1789–1802, 2014.

[8] H. V. Ribeiro, S. Mukherjee, and X. H. Zeng, "The advantage of playing home in nba: microscopic, team-specific and evolving features," *PLoS One*, vol. 11, no. 3, Article ID e0152440, 2016.

[9] A. Vaswani, N. Shazeer, N. Parmar et al., "Attention is all you need," 2017, https://arxiv.org/abs/1706.03762.

[10] X. Wang, X. He, Y. Cao, M. Liu, and T. S. Chua, "KGAT: knowledge graph attention network for recommendation," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 950–958, Anchorage, AK, USA, August 2019.

[11] Y. Yan, J. Qin, B. Ni et al., "Learning multi-attention context graph for group-based re-identification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 4, 2020.

[12] Y. Zhang, Z. Guo, and W. Lu, "Attention guided graph convolutional networks for relation extraction," 2019, https://arxiv.org/abs/1906.07510.

[13] J. Lee, I. Lee, and J. Kang, "Self-attention graph pooling," in *Proceedings of the International Conference on Machine Learning*, pp. 3734–3743, PMLR, Long Beach, CA, USA, June 2019.

[14] P. Shaw, J. Uszkoreit, and A. Vaswani, "Self-attention with relative position representations," 2018, https://arxiv.org/abs/1803.02155.

[15] H. Zhang, I. Goodfellow, D. Metaxas, and A. Odena, "Self-attention generative adversarial networks," in *Proceedings of International Conference on Machine Learning*, pp. 7354–7363, PMLR, Long Beach, CA, USA, June-2019.

[16] H. Zhao, J. Jia, and V. Koltun, "Exploring self-attention for image recognition," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 10076–10085, Seattle, WA, USA, April 2020.

[17] Y. Hao, Y. Zhang, K. Liu et al., "An end-to-end model for question answering over knowledge base with cross-attention combining global knowledge," in *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*, vol. 1, pp. 221–231, Vancouver, Canada, July 2017.

[18] R. Hou, H. Chang, B. Ma, S. Shan, and X. Chen, "Cross attention network for few-shot classification," 2019, https://arxiv.org/abs/1910.07677.

[19] Z. Huang, X. Wang, L. Huang, C. Huang, Y. Wei, and W. Liu, "Ccnet: criss-cross attention for semantic segmentation," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 603–612, Seoul, South Korea, November 2019.

[20] K. H. Lee, X. Chen, G. Hua, H. Hu, and X. He, "Stacked cross attention for image-text matching," in *Proceedings of the European Conference on Computer Vision*, pp. 201–216, ECCV), Munich, Germany, September 2018.

[21] X. Wu, K. Yu, W. Ding, H. Wang, and X. Zhu, "Online feature selection with streaming features," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 5, pp. 1178–1192, 2012.

[22] D. P. Kingma and J. Ba, "Adam: a method for stochastic optimization," 2014, https://arxiv.org/abs/1412.6980.

[23] Satimage, 2014,https://datahub.io/machine-learning/satimage.

[24] Mnist Digits Classification Dataset, 2014,https://keras.io/api/datasets/mnist/.

[25] M. F. Duarte and H. Y. Hen, "Vehicle classification in distributed sensor networks," *Journal of Parallel and Distributed Computing*, vol. 64, no. 7, pp. 826–838, 2004.

[26] J. Y. Wang, "Application of support vector machines in bioinformatics," M.Sc. thesis, Department of Computer Science and Information Engineering, National Taiwan University, Taipei, Taiwan, 2002.

Research Article

Combination of Joint Representation and Adaptive Weighting for Multiple Features with Application to SAR Target Recognition

Liqun Yu [ID], Lu Wang, and Yongxing Xu

School of Economics and Management, Harbin University of Science and Technology, Harbin 150080, China

Correspondence should be addressed to Liqun Yu; yuliqun@hrbust.edu.cn

For the synthetic aperture radar (SAR) target recognition problem, a method combining multifeature joint classification and adaptive weighting is proposed with innovations in fusion strategies. Zernike moments, nonnegative matrix factorization (NMF), and monogenic signal are employed as the feature extraction algorithms to describe the characteristics of original SAR images with three corresponding feature vectors. Based on the joint sparse representation model, the three types of features are jointly represented. For the reconstruction error vectors from different features, an adaptive weighting algorithm is used for decision fusion. That is, the weights are adaptively obtained under the framework of linear fusion to achieve a good fusion result. Finally, the target label is determined according to the fused error vector. Experiments are conducted on the moving and stationary target acquisition and recognition (MSTAR) dataset under the standard operating condition (SOC) and four extended operating conditions (EOC), i.e., configuration variants, depression angle variances, noise interference, and partial occlusion. The results verify the effectiveness and robustness of the proposed method.

1. Introduction

Synthetic aperture radar (SAR) target recognition has been researched for decades since 1990s [1]. According to the comprehensive review of current literature, the existing SAR target recognition methods can be divided into different aspects. From the aspect of target descriptions, the methods can be categorized as template-based and model-based ones. In the former, the references for the test sample are described by SAR images from different conditions, e.g., azimuths, depression angles, backgrounds, called training samples [2–4]. In the latter, the target characteristics are generated by models including CAD and scattering center models [5–10]. From the aspect of the decision engine, these methods are distinguished as feature-based and classifier-based ones. The former employs or designs specific features for SAR images so the discrimination can be exploited. The latter adopts or develops suitable classifiers for SAR target recognition so the overall performance can be improved. According to previous works, the features used in SAR target recognition cover geometric ones, transformation ones, and electromagnetic

ones. The geometric shape features describe the target area and contour distributions [11–21], such as the Zernike moments, outline descriptors. The transformation features can be further divided into two sub-categories as projection and decomposition ones. The former aims to find the optimal projection directions through the learning of training samples, so the high dimension of the original images can be reduced efficiently. Typical algorithms for projection features include principal component analysis (PCA) [22], nonnegative matrix factorization (NFM) [23], etc. The latter decomposes the original image through a series of signal bases to obtain different layers of descriptors. The representation algorithms for decomposition features include wavelet decomposition [24], monogenic signal [25, 26], bidimensional empirical mode decomposition (BEMD) [27], etc. The electromagnetic features focus on radar backscattering characteristics of targets, e.g., the attributed scattering center [28–32]. Classifiers are usually applied after feature extraction to make the final decisions. The classifiers in previous SAR target recognition methods were mainly inherited from the traditional pattern recognition field, such as K nearest

neighbor (KNN) [22], support vector machine (SVM) [33, 34], sparse representation-based classification (SRC) [34–36], and joint sparse representation [37–39]. In recent years, with the development of deep learning theory, the relevant models represented by convolutional neural networks (CNN) [40–43] have also been continuously applied to SAR target recognition with high effectiveness.

Considering the properties of different types of features, the multifeature SAR target recognition methods were developed to combine their strengths. These methods can be generally divided into parallel fusion, hierarchal fusion, and joint fusion. The parallel fusion classifies different features independently and further fuses their decisions [44, 45]. The hierarchal fusion classifies different features sequentially and a reliable decision in the former stage can avoid the remaining works [43, 46, 47]. The joint fusion mainly makes use of the multitask learning algorithms, which classify different features in the same framework, such as joint sparse representation [39]. Based on the previous works, this paper proposes a SAR target recognition method via a combination of joint representation of multiple features and adaptive weighting. Three types of features, i.e., Zernike moments, NMF, and monogenic signal, are used to describe the target characteristics in SAR images, which reflect the target shape, pixel distribution, time-frequency properties, respectively. In this sense, the three features have good complementarity. The joint sparse representation model [48, 49] is used to represent the three features, which employs their inner correlation to improve the representation accuracy. In the traditional decision-making mechanism based on joint sparse representation, the reconstruction errors of different tasks are directly added, and then the decision is made according to the minimum error. Actually, different tasks have different weights because they have different discrimination capabilities, so the idea of equal weights has certain shortcomings. As a remedy, this paper uses the adaptive weighting algorithm proposed in [50] to obtain the optimal weights for different features. For the reconstructed error vectors of different types of features, the adaptive weights are solved and used for linear fusion. Finally, the target label of the test sample is decided based on the fused error vector. In the experiments, tests and verifications are carried out on the moving and stationary target acquisition and recognition (MSTAR) dataset. The results of typical experimental setups show the effectiveness and robustness of the proposed method.

2. Extraction of Multiple Features

2.1. Zernike Moments.
The moment features are useful features to describe the shape and outline distribution of a region. The famous Hu moments could maintain good effectiveness for image with the relatively low noise level. However, for SAR images with strong noises and rotations and translations, the Hu moments may lose their adaptability. The Zernike moments can maintain high rotation invariance and noise robustness, which are more suitable for describing the regional features of SAR images [11–13].

With the form of $f(r, \theta)$ in polar coordinates, the Zernike moments of the input image are calculated as follows:

$$Z_{nl} = \frac{n+1}{\pi} \int_0^{2\pi} \int_0^1 [V_{nl}(r, \theta)] * f(r, \theta) r \mathrm{d}r \mathrm{d}\theta, \quad (1)$$

where $n = 0, 1, \ldots, \infty$; $l = 0, \pm 1, \ldots$; $n - |l|$ is an even and $|l| \le n$ 。

The Zernike polynomials $V_{nl}(r, \theta) = R(r)e^{il\theta}$ are a set of orthogonal complete complex-valued functions on the unit circle $x^2 + y^2 \le 1$, which complies with the following constraints:

$$\int_0^{2\pi} \int_0^1 [V_{nl}(r, \theta)] * V_{mk} r \mathrm{d}r \mathrm{d}\theta = \frac{\pi}{n+1} \delta_{mn} \delta_{kl}. \quad (2)$$

Based on Zernike moments, the rotation invariants can be generated as follows:

$$Z_{nl} = Z_{nl} \times Z_{n,-l} (l = 0, 1, 2, \ldots). \quad (3)$$

Before calculating the Zernike moment of an image, it is necessary to place the center of the image at the origin of the coordinates and map the pixels to the inside of the unit circle. Based on the principle of Zernike moments, the moments of any order can be obtained. In comparison, the higher-order moments contain more detailed information about the objects in the image. With reference to [11], this paper selects the Zernike moments at the 6th, 7th, 8th, 9th, 10th, and 11th orders (i.e., [n, m] = {[0, 0, 1, 1, 1, 1, 6, 7, 8, 9, 10, 11]}), to construct a feature vector, which describes the target area in the SAR image.

2.2. NMF.
NMF provided a way to efficiently reduce data dimension. Different from traditional PCA, NMF brings in the nonnegativity constraint and the resulting projection matrix could better maintain the valid information as validated in previous works [23].

For an input matrix $D \in \mathrm{R}^{n \times m}$, it is decomposed by NMF as follows:

$$D \approx WH, \quad D_{ij}, W_{i,u}, H_{u,j} \ge 0, \quad (4)$$

where $0 \le i \le n - 1, 0 \le j \le m - 1$, and $0 \le u \le r - 1, W \in \mathrm{R}^{n \times r}$ and $H \in \mathrm{R}^{r \times m}$. The reconstruction error is employed to evaluate the decomposition precision, which is defined as the square Euclidean distance as follows:

$$\underset{W,H}{\arg \min} \|D - WH\|^2 = \sum_{0 \le i < n, 0 \le j < m} \left(D_{i,j} - (WH)_{i,j} \right)^2. \quad (5)$$

The above objective function can be iteratively updated to find the solutions as follows:

$$H_{au} \leftarrow H_{au} \frac{\left(W^T D\right)_{au}}{\left(W^T WH\right)_{au}},$$

$$W_{ia} \leftarrow W_{ia} \frac{\left(DH^T\right)_{ia}}{\left(WHH^T\right)_{ia}}, \quad (6)$$

where $0 \le a < r, 0 \le u < m$, and $0 \le i < n$.

With the solution of the matrix W, its transpose W^{-1} is used as the projection matrix for feature extraction. With reference to [23], this paper employs NMF to obtain an 80-dimension feature vector for an input SAR image.

2.3. Monogenic Signal.

As a 2D extension of the traditional analytic signal, the monogenic signal has been successfully applied to feature extraction of SAR images [25, 26]. Denote the input image as to where $z = (x, y)^T$ represent the pixel locations. The monogenic component is calculated as follows:

$$f_M(z) = f(z) - (i, j)f_R(z), \tag{7}$$

where i and j are imagery units along with different directions.

Based on $f_M(z)$, three monogenic features can be generated to describe the local amplitude, local phase, and local orientation:

$$\text{amplitude:} \ A(z) = \sqrt{f(z)^2 + |f_R(z)|^2},$$

$$\text{phase:} \ \varphi(z) = a\tan 2(|f_R(z)|, f(z)) \in (-\pi, \pi],$$

$$\text{orientation:} \ \theta(z) = a\tan 2\left(\frac{f_y(z)}{f_x(z)}\right) \in \left(-\frac{\pi}{2}, \frac{\pi}{2}\right], \tag{8}$$

where $f_x(z)$ and $f_y(z)$ are the i-imaginary and j-imaginary components of the monogenic component, respectively.

As reported in previous works, the monogenic features could reveal the time-frequency properties of the original SAR image, including the intensity distribution, structural, and geometric information. With reference to [25], this paper reorganizes the three features in a vector, called monogenic feature vector.

3. Joint Classification with Adaptive Weights

3.1. Joint Sparse Representation.

The joint sparse representation is an extended version of traditional sparse representation, which handles several related problems simultaneously [48, 49]. As the inner correlations of different sparse representations are exploited, the overall reconstruction precision can be improved. For the multiple features from the same SAR image, they are related and suitable to be represented by joint sparse representation. In the following, the basic process of jointly representing multiple features is described. Assume there are M different features from the sample y, denoted as $[y^{(1)}, \cdots, y^{(M)}]$, a general form of joint sparse representation is as follows:

$$\min_A \left\{ g(A) = \sum_{l=1}^M \left\| y^{(l)} - D^{(l)} * a^{(l)} \right\| \right\}, \tag{9}$$

where $D^{(l)}$ is the global dictionary corresponding to the lth feature; $A = [a^{(1)}, \ldots, a^{(M)}]$ is a matrix established by the coefficient vectors by different features.

It can be analyzed that the objective function in equation (9) is equal to the solutions of the sparse representation problems of different features separately. In this sense, it can hardly make use of the inner correlations of different features. As a remedy, the joint sparse representation model in previous works imposed ℓ_0/ℓ_1 norm on the coefficient matrix A with a new objective function as follows:

$$\min_A g(A) + \eta \|A\|_{0,1}, \tag{10}$$

where η is the regularization factor.

During the solution of equation (10), the coefficient vectors of different features tend to share a similar pattern because of the constraint of ℓ_0/ℓ_1 norm. Therefore, the inner correlations among different features can be employed. It is reported and validated that simultaneous orthogonal matching pursuit (SOMP) [48] and multitask compressive sensing [49] are suitable for solving the problem. With the solution of the coefficient matrix $\hat{A} = [\hat{a}^{(1)}, \ldots, \hat{a}^{(M)}]$, the reconstruction errors of different training classes can be calculated to further determine the target label as follows:

$$\text{identity}(y) = \min_{i=1, \cdots, C} \sum_{l=1}^M \left\| y^{(l)} - D_i^l \hat{a}_i^{(l)} \right\|_2, \tag{11}$$

where D_i^l is the local dictionary of the lth feature with regard to the ith class; $\hat{a}_i^{(l)}$ is the corresponding coefficient vector.

3.2. Decision Fusion with Adaptive Weights.

Equation (11) gives the basic decision-making mechanism of the traditional joint sparse representation model applied to classification. In essence, this is a linear weighted fusion algorithm with the same weight. That is, it is considered that the contributions of different features to the final recognition are consistent. However, in the actual process, the effectiveness of different features for recognition is often different, so special consideration is required. Fusion by linear weighting is an effective method for processing multisource information, and its core element is to scientifically determine the weights of different components. To this end, this paper adopts the adaptive weight determination algorithm proposed in [50]. To simplify the description, take the reconstruction error vectors of the two types of features (denoted as d_i^1 and d_i^2) as examples to describe their fusion process.

Step 1. Define $\beta_r = h - \min(d_1^r, \ldots, d_C^r)/h$, in which $h = \sum_{i=1}^C d_i^r$, $r = 1, 2$

Step 2. Normalize d_i^1 and d_i^2 using $d_i^1 = (d_{\max}^1 - d_i^1)/(d_{\max}^1 - d_{\min}^1)$ and $d_i^2 = (d_{\max}^2 - d_i^2)/(d_{\max}^2 - d_{\min}^2)$, in which $d_{\max}^1$ and $d_{\max}^1$ are the maximum and minimum of d_i^1; $d_{\max}^2$ and $d_{\min}^2$ are the maximum and minimum of d_i^2;

Step 3. Reorganize d_i^1 and d_i^2 in an ascending manner to obtain the new sequences as $e_1^1 \le e_2^1 \le \cdots \le e_C^1$ and $e_1^2 \le e_2^2 \le \cdots \le e_C^2$ to further define $w = (e_2^1 - e_1^1) + (e_2^2 - e_1^2)$ $w_1 = (e_2^1 - e_1^1)/w$, and $w_2 = (e_2^2 - e_1^2)/w$.

Step 4. $f_i = \beta_1 w_1 d_i^1 + \beta_2 w_2 d_i^2 \, (i = 1, 2, \ldots, C)$ is achieved as the fused reconstruction error. The target label is decided as the class with the minimum error $k = \arg\min_i f_i$.

The above algorithm analyzes the distributions of single reconstruction error vectors while comparing their individual characteristics. So, the result weights could better reflect the importance of different components than the traditional experiential weights such as the equal ones. For the reconstruction error fusion of the three types of features in this paper, the same idea is adopted, and the specific algorithm can be found in [50]. Figure 1 shows the basic process of the proposed method. The three types of features produce the reconstruction error vectors corresponding to each training class under the joint sparse representation model, respectively. The final reconstruction error vector is obtained using the adaptive weighted fusion algorithm. Finally, the target label of the test sample is determined according to the minimum error.

4. Experiments and Analysis

4.1. Experimental Setup. The MSTAR dataset is used to test and analyze the proposed method. Figure 2 shows the 10 types of typical vehicle targets included in the dataset, e.g., tanks, armored vehicles, and trucks. For each target, the MSTAR dataset collects samples in a relatively complete azimuth range with several depression angles. Table 1 shows the basic training and test sets used in subsequent experiments, which are from two depression angles of 17° and 15°. Accordingly, the test and the training sets have only a small depression angle difference, and their overall similarity is relatively high. Such a situation is generally considered as a standard operating condition (SOC). On this basis, some simulation algorithms, including noise addition and occlusion generation, are developed to obtain test samples under extended operating conditions (EOC). Furthermore, samples from different target configurations and depression angles can be employed to set up EOCs like configuration variants and depression angle variances. Therefore, based on the above conditions, the performance of the proposed method can be investigated and verified in a comprehensive way.

Some reference methods selected from published works are used for comparison, including ones using single features and ones using multiple features. The former three use single features, i.e., Zernike moments [11], NMF [23], and monogenic signal [25], which are consistent with the proposed method. The latter three decision fusion strategies including parallel fusion [45], hierarchical fusion [46], and joint classification [39], in which the three classified features are the same as the proposed method. Especially, the joint classification in the reference methods only performs joint sparse representation and directly adds the reconstruction results of different features with no adaptive weighting.

4.2. SOC. With reference to the basic training and test sets in Table 1, the proposed method is tested under SOC. Figure 3 presents the recognition results of the proposed method in the form of a confusion matrix. According to the corresponding relationship between the horizontal and vertical coordinates, the diagonal elements mark the correct recognition rates of different categories. Define the average recognition rate P_{av} as the proportion of the test samples correctly classified in the entire test set. The P_{av} of the proposed method is calculated to be 99.38%. Table 2 shows the P_{av}s of all the methods which are achieved according to the same process on the same platforms. The effectiveness of the proposed method can be intuitively validated with its highest P_{av}. Compared with the three types of methods using single features, the multifeature methods achieve obvious advantages, reflecting the complementarity between different features. Among the four multifeature methods, the idea of joint classification has some predominance over the parallel fusion and hierarchal fusion mainly because the inner correlations of different features are exploited. In comparison with the joint classification method, the proposed one further enhances the overall recognition performance by introducing adaptive weights, verifying the effectiveness of the proposed strategy.

4.3. EOC1-Configuration Variants. For different military applications, the same target may have different models. When the test sample and the training sample come from different models, the difficulty of the target recognition problem will increase. Table 3 shows the training and test sets under the condition of model difference, where the test samples and training samples of BMP2 and T72 are from different models. Table 4 shows the identification results of the proposed method for different models. It can be seen from the recognition rate that there are differences in the similarity between different test models and the reference models in the training set. Table 5 compares the P_{av}s of different methods under current conditions. The performance advantage of the multifeature method compared with the single-feature method is still obvious. In the framework of joint classification, the proposed method makes full use of the classification advantages of different features for model differences by introducing adaptive weights, thereby improving the final recognition performance.

4.4. EOC2-Depression Angle Variances. The test sample to be classified may come from a different depression angle from the training samples. Considering the sensitivity of SAR images to view angles, it is difficult to correctly classify test samples from different depression angles. Table 6 sets up the training and test samples with different depression angles, in which the test set including samples from 30° to 45° depression angles. Figure 4 shows the P_{av}s of different methods at two depression angles. It shows that the large depression angle difference (45°) causes significant performance degradations of all the methods. Comparing the results at the two depression angles, the P_{av}s of the proposed method are the highest, verifying its robustness. Based on multifeature joint representation, the proposed method adaptively obtains the weights of different features, so their effectiveness for depression angle variances can be better utilized.
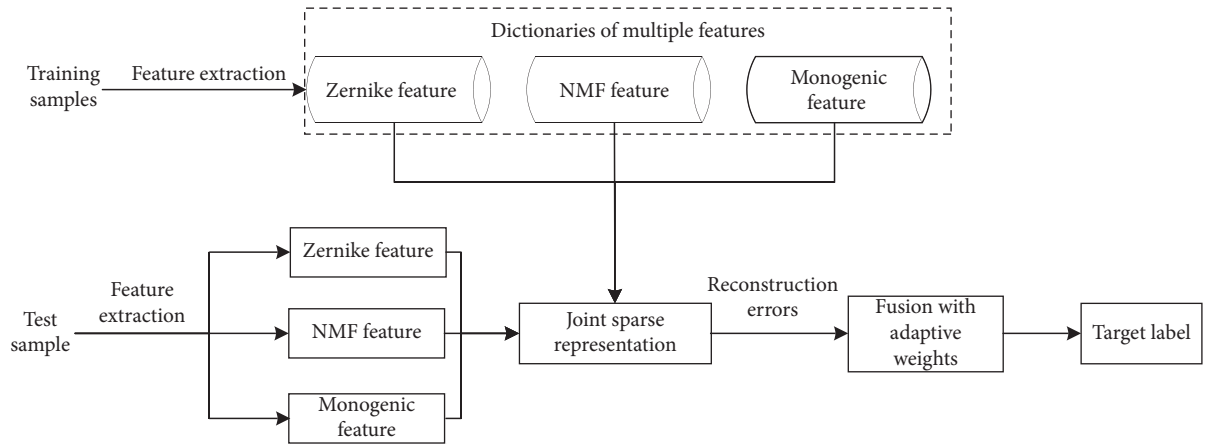
FIGURE 1: Procedure of SAR target recognition based on joint classification of multiple features with adaptive weights.
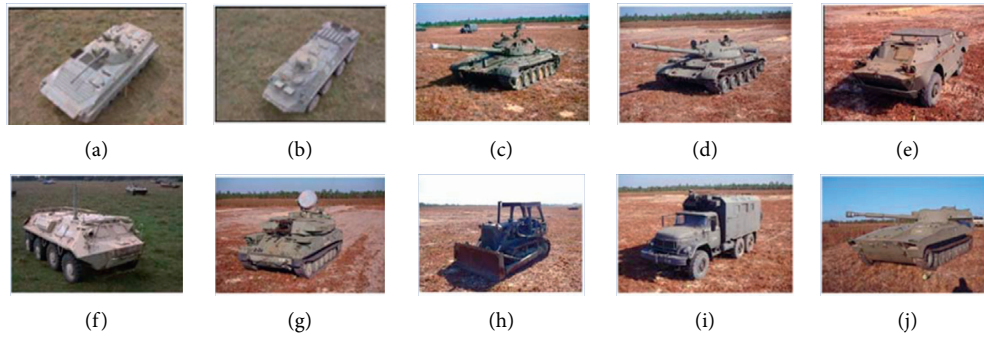

FIGURE 2: Optical images of ten classes of targets in MSTAR dataset.

TABLE 1: Basic training and test sets used in the experiments.

Type	Training set		Test set	
	Depression angle	Number of samples	Depression angle	Number of samples
BMP2		233		195
BTR70		233		196
T72		232		196
T62		299		273
BRDM2	17°	298	15°	274
BTR60		256		195
ZSU23/4		299		274
D7		299		274
ZIL131		299		274
2S1		299		274

4.5. EOC3-Noise Corruption. In the actual process, the obtained test samples of noncooperative targets are contaminated by varying degrees of noise. With the difference of the signal-to-noise ratio (SNR) between the test and the training samples increasing, their correlation decreases simultaneously. Based on the basic training and test sets in Table 1, this paper uses noise simulation to construct test sets with different SNRs using the original test samples [31], including -10 dB, -5 dB, 0 dB, 5 dB, and 10 dB. The proposed and reference methods are tested under different noise levels, and the statistical results are

shown in Figure 5. Intuitively, noise corruption has a significant impact on recognition performance. In contrast, the proposed method maintains the highest P_{av} at each SNR, verifying its robustness. Similar to the results under SOC, the performance predominance of the multifeature methods is still obvious over the single-feature ones. Among the three types of single-feature methods, the ones using Zernike moments and monogenic features are more robust than the ones using NMF features, which also reflects the different sensitivities of different features to noise interference. By effectively fusing different

Figure 3: Confusion matrix of the proposed method under SOC.

Table 2: Recognition performance of different methods under SOC.

Method	Proposed	Zernike	NMF	Mono	Parallel fusion	Hierarchical fusion	Joint classification
P_{av} (%)	99.38	98.42	98.56	98.74	99.08	99.13	99.18

Table 3: Training and test samples under configuration variants.

Type	Training set			Test set		
	Depression angle	Configuration	Number of samples	Depression angle	Configuration	Number of samples
BMP2		9563	233		9566	428
					C21	429
BRDM2		—	298		—	0
BTR70		—	233		—	0
	17°			15°, 17°	812	426
					A04	573
T72		132	232		A05	573
					A07	573
					A10	567

Table 4: Recognition results of the proposed method under configuration variants.

Type	Configuration	Decided class				P_{av} (%)
		BMP2	BRDM2	BTR70	T72	
BMP2	9566	422	1	2	3	98.60
	C21	425	2	0	2	99.07
	812	2	0	3	421	98.83
	A04	2	3	2	566	98.78
T72	A05	2	2	6	563	98.25
	A07	2	1	3	567	98.95
	A10	3	1	4	559	98.59
Overall						98.76

TABLE 5: Recognition performance of different methods under configuration variants.

Method	Proposed	Zernike	NMF	Mono	Parallel fusion	Hierarchical fusion	Joint classification
P_{av} (%)	98.76	97.67	97.12	97.36	97.82	98.08	98.12

TABLE 6: Training and test samples under depression angle variances.

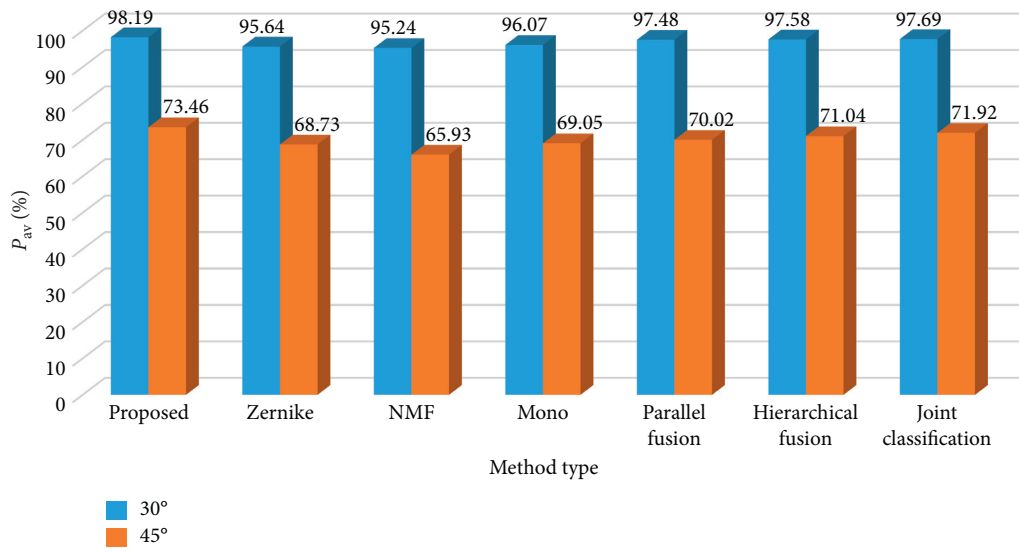| Type	Training set		Test set	
	Depression angle	Number of samples	Depression angle	Number of samples
2S1		299	30°	288
			45°	303
BRDM2	17°	298	30°	287
			45°	303
ZSU23/4		299	30°	288
			45°	303

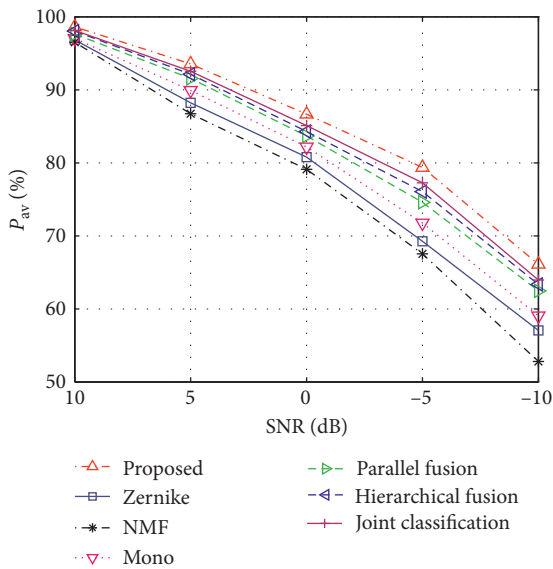FIGURE 4: Recognition performance of different methods at 30° and 45° depression angles.

FIGURE 5: Recognition performance of different methods under noise corruption.

features with joint classification and adaptive weighting, they can be integrated to improve the reliability of the final decision in the proposed method.

4.6. EOC4-Partial Occlusion. Occlusion situations are also very common in practical applications. As a result, part of the ground target cannot be illuminated by the radar waves with no echoes. Using a similar idea of noise simulation, this paper constructs the test sets of different occlusion levels based on the test samples in Table 1 according to the target occlusion model in [31]. Afterwards, the recognition results of the proposed and reference methods are obtained as shown in Figure 6. The comparison shows that the proposed method maintains the highest P_{av}s at different occlusion levels, reflecting its robustness. The proposed method comprehensively exploits multiple types of features and uses adaptive weights to further employ the advantages of the specified features that are more effective for occlusion situations. Therefore, the final recognition results are improved.
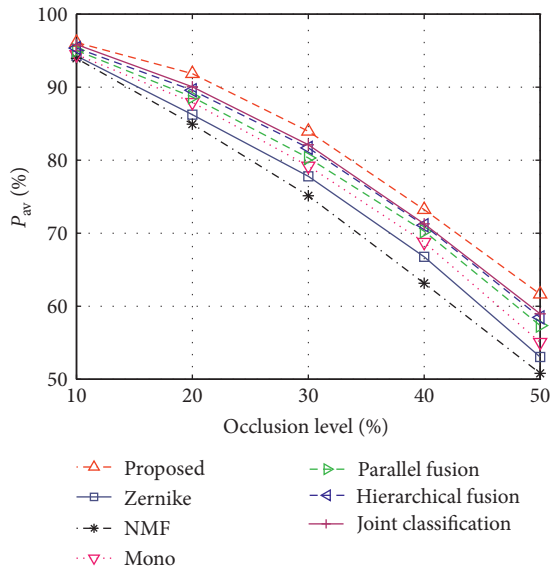
FIGURE 6: Recognition performance of different methods under partial occlusion.

5. Conclusion

This paper applies joint sparse representation and adaptive weighting to SAR target recognition. For the reconstruction error vectors of the three types of features resulting from the joint sparse representation, their corresponding weights are determined adaptively, reflecting different contributions of different features to the final classification result. Based on the MSTAR data set, experiments were carried out to test the recognition performance under a typical SOC and four representative EOCs. The proposed method achieves a P_{av} of 99.38% for 10-class targets under SOC and superior performance over the reference ones under different EOCs. The experimental results show that the high effectiveness and robustness of the proposed method, which has certain advantages and potentials in practical uses.

Data Availability

The dataset used in this paper is publicly available.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

References

[1] K. El-Darymli, E. W. Gill, P. McGuire, D. Power, and C. Moloney, "Automatic target recognition in synthetic aperture radar imagery: a state-of-the-art review," *IEEE Access*, vol. 4, pp. 6014–6058, 2016.

[2] L. M. Novak, G. J. Owirka, W. S. Brower, and A. L. Weaver, "The automatic target-recognition system in SAIP," *Lincoln Laboratory Journal*, vol. 10, no. 2, pp. 187–202, 1997.

[3] L. M. Novak, G. J. Owirka, and W. S. Brower, "Performance of 10- and 20-target MSE classifiers," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 36, no. 4, pp. 1279–1289, 2000.

[4] S. M. Verbout, W. W. Iring, and A. S. Hanes, "Improving a template-based classifier in a SAR automatic target recognition system by using 3-D target information," *Lincoln Laboratory Journal*, vol. 6, no. 1, pp. 53–76, 1993.

[5] J. R. Diemunsch and J. Wissinger, "Moving and stationary target acquisition and recognition (MSTAR) model-based automatic target recognition: search technology for a robust ATR," *SPIE Algorithms Synthetic Aperture Radar Imagery*, vol. 3370, pp. 481–492, 1998.

[6] T. D. Ross, J. J. Bradley, L. J. Hudson, and M. P. O'Connor, "So what's the problem? —an MSTAR perspective," *SPIE Algorithms Synthetic Aperture Radar Imagery*, vol. 3721, pp. 662–672, 1999.

[7] E. Keydel, S. Lee, and J. Moore, "MSTAR extended operating conditions: a tutorial," *SPIE*, vol. 2757, pp. 228–242, 1996.

[8] J. Zhou, Z. Shi, X. Cheng et al., "Automatic target recognition of SAR images based on global scattering center model," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 10, pp. 3713–3729, 2011.

[9] B. Ding and G. Wen, "Target reconstruction based on 3-D scattering center model for robust SAR ATR," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 7, pp. 3772–3785, 2018.

[10] B. Ding and G. Wen, "A region matching approach based on 3-D scattering center model with application to SAR target recognition," *IEEE Sensors Journal*, vol. 18, no. 11, pp. 4623–4632, 2018.

[11] M. Amoon and G. a. Rezai-rad, "Automatic target recognition of synthetic aperture radar (SAR) images based on optimal selection of Zernike moments features," *IET Computer Vision*, vol. 8, no. 2, pp. 77–85, 2014.

[12] S. Gishkori and B. Mulgrew, "Pseudo-Zernike moments based sparse representations for SAR image classification," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 55, no. 2, pp. 1037–1044, 2019.

[13] X. Zhang, Z. Liu, S. Liu, D. Li, Y. Jia, and P. Huang, "Sparse coding of 2D-slice Zernike moments for SAR ATR," *International Journal of Remote Sensing*, vol. 38, no. 2, pp. 412–431, 2017.

[14] P. Bolourchi, H. Demirel, and S. Uysal, "Target recognition in SAR images using radial Chebyshev moments," *Signal Image & Video Processing*, vol. 11, no. 6, pp. 1–8, 2017.

[15] C. Clemente, L. Pallotta, D. Gaglione, A. De Maio, and J. J. Soraghan, "Automatic target recognition of military vehicles with Krawtchouk moments," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 53, no. 1, pp. 493–500, 2017.

[16] B. Ding, G. Wen, C. Ma et al., "Target recognition in synthetic aperture radar images using binary morphological operations," *Journal of Applied Remote Sensing*, vol. 10, no. 4, Article ID 046006, 2016.

[17] S. Cui, F. Miao, Z. Jin et al., "Target recognition of synthetic aperture radar images based on matching and similarity evaluation between binary regions," *IEEE Access*, vol. 7, pp. 154398–154413, 2019.

[18] C. Shan, B. Huang, and M. Li, "Binary morphological filtering of dominant scattering area residues for SAR target recognition," *Computational Intelligence and Neuroscience*, vol. 2018, Article ID 9680465, 15 pages, 2018.

[19] G. C. Anagnostopoulos, "SVM-based target recognition from synthetic aperture radar images using target region outline descriptors," *Nonlinear Anal.*vol. 71, no. 2, pp. e2934–e2939, 2009.

[20] J. Tan, X. Fan, S. Wang et al., "Target recognition of SAR images by partially matching of target outlines," *Journal of Electromagnetic Waves and Applications*, vol. 33, no. 7, pp. 865–881, 2019.

[21] X. Zhu, Z. Huang, and Z. Zhang, "Automatic target recognition of synthetic aperture radar images via Gaussian mixture modeling of target outlines," *Optik*, vol. 194, Article ID 162922, 2019.

[22] A. K. Mishra and T. Motaung, "Application of linear and nonlinear PCA to SAR ATR," in *Proceedings of the IEEE 25th International Conference Radioelektronika (RADIOELEKTRONIKA)*, pp. 349–354, Pardubice, Czech Republic, April 2015.

[23] Z. Cui, Z. Cao, J. Yang, J. Feng, and H. Ren, "Target recognition in synthetic aperture radar images via nonnegative matrix factorisation," *IET Radar, Sonar & Navigation*, vol. 9, no. 9, pp. 1376–1385, 2015.

[24] W. Xiong, L. Cao, and Z. Hao, "Combining wavelet invariant moments and relevance vector machine for SAR target recognition," in *Proceedings of the IET International Radar Conference*, pp. 1–4, Xi'an, China, April 2009.

[25] G. Dong and G. Kuang, "Classification on the monogenic scale space: application to target recognition in SAR image," *IEEE Transactions on Image Processing*, vol. 24, no. 8, pp. 2527–2539, 2015.

[26] G. Dong, G. Kuang, N. Wang, L. Zhao, and J. Lu, "SAR target recognition via joint sparse representation of monogenic signal," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 8, no. 7, pp. 3316–3328, 2015.

[27] M. Chang, X. You, and Z. Cao, "Bidimensional empirical mode decomposition for SAR image feature extraction with application to target recognition," *IEEE Access*, vol. 7, pp. 135720–135731, 2019.

[28] H. C. China, R. L. Moses, and L. C. Potter, "Model-based classification of radar images," *IEEE Transactions on Information Theory*, vol. 46, no. 5, pp. 1842–1854, 2000.

[29] B. Ding, G. Wen, J. Zhong et al., "Robust method for the matching of attributed scattering centers with application to synthetic aperture radar automatic target recognition," *Journal of Applied Remote Sensing*, vol. 10, no. 1, Article ID 016010, 2016.

[30] B. Ding, G. Wen, J. Zhong, C. Ma, and X. Yang, "A robust similarity measure for attributed scattering center sets with application to SAR ATR," *Neurocomputing*, vol. 219, pp. 130–143, 2017.

[31] B. Ding, G. Wen, X. Huang, C. Ma, and X. Yang, "Target recognition in synthetic aperture radar images via matching of attributed scattering centers," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 10, no. 7, pp. 3334–3347, 2017.

[32] X. Zhang, "Noise-robust target recognition of SAR images based on attribute scattering center matching," *Remote Sensing Letters*, vol. 10, no. 2, pp. 186–194, 2019.

[33] Q. Zhao and J. C. Principe, "Support vector machines for SAR automatic target recognition," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 37, no. 2, pp. 643–654, 2001.

[34] H. Liu and S. Li, "Decision fusion of sparse representation and support vector machine for SAR image target recognition," *Neurocomputing*, vol. 113, pp. 97–104, 2013.

[35] J. J. Thiagaraianm, K. N. Ramamurthy, P. Knee et al., "Sparse representations for automatic target classification in SAR images," in *Proceedings of the 4th International Symposium on Communications, Control and Signal Processing*, pp. 1–4, Monticello, IL, USA, October 2010.

[36] H. Song, K. Ji, Y. Zhang, X. Xing, and H. Zou, "Sparse representation-based SAR image target classification on the 10-class MSTAR data set," *Applied Sciences*, vol. 6, no. 1, p. 26, 2016.

[37] H. Zhang, N. M. Nasrabadi, Y. Zhang, and T. S. Huang, "Multi-view automatic target recognition using joint sparse representation," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 48, no. 3, pp. 2481–2497, 2012.

[38] B. Ding and G. Wen, "Exploiting multi-view SAR images for robust target recognition," *Remote Sensing*, vol. 9, no. 11, p. 1150, 2017.

[39] S. Liu and J. Yang, "Target recognition in synthetic aperture radar images via joint multifeature decision fusion," *Journal of Applied Remote Sensing*, vol. 12, no. 1, Article ID 016012, 2018.

[40] D. E. Morgan, "Deep convolutional neural networks for ATR from SAR imagery," in *Proceedings of the SPIE*, pp. 1–13, Bellingham, WA, USA, April 2015.

[41] S. Chen, H. Wang, F. Xu et al., "Target classification using the deep convolutional networks for SAR images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 6, pp. 1685–1697, 2016.

[42] H Furukawa, "Deep learning for target classification from SAR imagery data augmentation and translation invariance," *IEICE Technical Report*, vol. 109, pp. 13–17, 2017.

[43] C. Jiang and Y. Zhou, "Hierarchical fusion of convolutional neural networks and attributed scattering centers with application to robust SAR ATR," *Remote Sensing*, vol. 10, no. 6, p. 819, 2018.

[44] U. Srinivas and V. Monga, "Meta-classifiers for exploiting feature dependence in automatic target recognition," in *Proceedings of the IEEE Radar Conference*, pp. 147–151, Florence, Italy, October 2011.

[45] R. Huan and Y. Pan, "Decision fusion strategies for SAR image target recognition," *IET Radar, Sonar & Navigation*, vol. 5, no. 7, pp. 747–755, 2011.

[46] Z. Cui, Z. Cao, J. Yang et al., "A hierarchical propelled fusion strategy for SAR automatic target recognition," *EURASIP Journal on Wireless Communications and Networking*, vol. 39, pp. 1–8, 2013.

[47] B. Feng, W. Tang, and D. Feng, "Target recognition of SAR images via hierarchical fusion of complementary features," *OPTIK*, vol. 217, Article ID 164695, 2020.

[48] J. A. Tropp, A. C. Gilbert, and M. J. Strauss, "Algorithms for simultaneous sparse approximation. Part II: convex relaxation," *Signal Processing*, vol. 86, no. 3, pp. 589–602, 2006.

[49] S. Ji, D. Dunson, and L. Carin, "Multitask compressive sensing," *IEEE Transactions on Signal Processing*, vol. 57, no. 1, pp. 92–106, 2009.

[50] Y. Xu and Y. Lu, "Adaptive weighted fusion: a novel fusion approach for image classification," *Neurocomputing*, vol. 168, pp. 566–574, 2015.