Wireless Communications and Mobile Computing

Semantic Mobile Computing within the Internet of Things and Web of Things 2021

Lead Guest Editor: Chi-Hua Chen
Guest Editors: Lingjuan Lyu, Ting Bi, and Feng-Jang Hwang

Semantic Mobile Computing within the Internet of Things and Web of Things 2021

Wireless Communications and Mobile Computing

Semantic Mobile Computing within the Internet of Things and Web of Things 2021

Lead Guest Editor: Chi-Hua Chen
Guest Editors: Lingjuan Lyu, Ting Bi, and Feng-Jang Hwang

Contents

Retracted: Age Label Distribution Learning Based on Unsupervised Comparisons of Faces
Wireless Communications and Mobile Computing
Retraction (1 page), Article ID 9825958, Volume 2023 (2023)

Retracted: Research on Sports Class Load Monitoring System Based on Threshold Classification Algorithm
Wireless Communications and Mobile Computing
Retraction (1 page), Article ID 9794634, Volume 2023 (2023)

Retracted: Effects of Attentional Withdrawal on Chinese Disabled Biathletes' Shooting Performance
Wireless Communications and Mobile Computing
Retraction (1 page), Article ID 9782397, Volume 2023 (2023)

Retracted: Detection of Power Data Outliers Using Density Peaks Clustering Algorithm Based on *K*-Nearest Neighbors
Wireless Communications and Mobile Computing
Retraction (1 page), Article ID 9890424, Volume 2023 (2023)

Online Multiplayer Tracking by Extracting Temporal Contexts with Transformer
Xiao Han, Yongbin Wang, Shouxun Liu, and Cong Jin
Research Article (10 pages), Article ID 6177973, Volume 2022 (2022)

[Retracted] A Mobile Intelligent Mine Platform with a Hybrid Fuzzy NN and ATT-CNN Prewarning Model
Yuliang Wu, Chao Wu, Jun Wang, Xiangnan Zhang, and Wei Chen
Research Article (14 pages), Article ID 4545936, Volume 2022 (2022)

[Retracted] Detection of Power Data Outliers Using Density Peaks Clustering Algorithm Based on K-Nearest Neighbors
Qingpeng Li, Lei Chen, and Yuhan Wang
Research Article (7 pages), Article ID 2203137, Volume 2022 (2022)

[Retracted] Power IoT System Architecture Integrating Trusted Computing and Blockchain
Yong Yan, Jinqian Chen, Ying Yao, Shaoyong Guo, Ao Xiong, and Wang Zhang
Research Article (13 pages), Article ID 2891917, Volume 2022 (2022)

A QoS-Based Fairness-Aware BBR Congestion Control Algorithm Using QUIC
Yi Han, Mengjie Zuo, Huijun Yuan, Yi Zhong, Zhenhui Yuan, and Ting Bi
Research Article (16 pages), Article ID 7222030, Volume 2022 (2022)

MISD-SLAM: Multimodal Semantic SLAM for Dynamic Environments
Yingxuan You, Peng Wei, Jialun Cai, Weibo Huang, Risheng Kang, and Hong Liu
Research Article (13 pages), Article ID 7600669, Volume 2022 (2022)

Contents

Contents

Hindawi
Wireless Communications and Mobile Computing
Volume 2023, Article ID 9873492, 1 page
https://doi.org/10.1155/2023/9873492

Retraction

Retracted: Featureless Blood Pressure Estimation Based on Photoplethysmography Signal Using CNN and BiLSTM for IoT Devices

Wireless Communications and Mobile Computing

Received 10 October 2023; Accepted 10 October 2023; Published 11 October 2023

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. Li, L. N. Harfiya, and C. Chang, "Featureless Blood Pressure Estimation Based on Photoplethysmography Signal Using CNN and BiLSTM for IoT Devices," *Wireless Communications and Mobile Computing*, vol. 2021, Article ID 9085100, 10 pages, 2021.

Hindawi
Wireless Communications and Mobile Computing
Volume 2023, Article ID 9851453, 1 page
https://doi.org/10.1155/2023/9851453

Retraction

Retracted: Power IoT System Architecture Integrating Trusted Computing and Blockchain

Wireless Communications and Mobile Computing

Received 10 October 2023; Accepted 10 October 2023; Published 11 October 2023

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. Yan, J. Chen, Y. Yao, S. Guo, A. Xiong, and W. Zhang, "Power IoT System Architecture Integrating Trusted Computing and Blockchain," *Wireless Communications and Mobile Computing*, vol. 2022, Article ID 2891917, 13 pages, 2022.

Hindawi
Wireless Communications and Mobile Computing
Volume 2023, Article ID 9850836, 1 page
https://doi.org/10.1155/2023/9850836

Retraction

Retracted: A System for Trusted Recovery of Data Based on Blockchain and Coding Techniques

Wireless Communications and Mobile Computing

Received 10 October 2023; Accepted 10 October 2023; Published 11 October 2023

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. Chen, Y. Yan, S. Guo, Y. Ren, and F. Qi, "A System for Trusted Recovery of Data Based on Blockchain and Coding Techniques," *Wireless Communications and Mobile Computing*, vol. 2022, Article ID 8390241, 12 pages, 2022.

Hindawi
Wireless Communications and Mobile Computing
Volume 2023, Article ID 9808517, 1 page
https://doi.org/10.1155/2023/9808517

Retraction

Retracted: Antialiasing Attention Spatial Convolution Model for Skin Lesion Segmentation with Applications in the Medical IoT

Wireless Communications and Mobile Computing

Received 10 October 2023; Accepted 10 October 2023; Published 11 October 2023

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] P. T. Le, C. Chang, Y. Li, Y. Hsu, and J. Wang, "Antialiasing Attention Spatial Convolution Model for Skin Lesion Segmentation with Applications in the Medical IoT," *Wireless Communications and Mobile Computing*, vol. 2022, Article ID 1278515, 15 pages, 2022.

Hindawi
Wireless Communications and Mobile Computing
Volume 2023, Article ID 9785461, 1 page
https://doi.org/10.1155/2023/9785461

Retraction

Retracted: Remote Judgment Method of Painting Image Style Plagiarism Based on Wireless Network Multitask Learning

Wireless Communications and Mobile Computing

Received 10 October 2023; Accepted 10 October 2023; Published 11 October 2023

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. Wang, "Remote Judgment Method of Painting Image Style Plagiarism Based on Wireless Network Multitask Learning," *Wireless Communications and Mobile Computing*, vol. 2021, Article ID 1345974, 8 pages, 2021.

Hindawi
Wireless Communications and Mobile Computing
Volume 2023, Article ID 9764052, 1 page
https://doi.org/10.1155/2023/9764052

Retraction

Retracted: Research on Acquisition and Tracking Algorithm of Global Satellite Positioning Receiver Based on UWB

Wireless Communications and Mobile Computing

Received 10 October 2023; Accepted 10 October 2023; Published 11 October 2023

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. Yang, D. Sun, and H. Ruan, "Research on Acquisition and Tracking Algorithm of Global Satellite Positioning Receiver Based on UWB," *Wireless Communications and Mobile Computing*, vol. 2021, Article ID 9025086, 9 pages, 2021.

Hindawi
Wireless Communications and Mobile Computing
Volume 2023, Article ID 9761846, 1 page
https://doi.org/10.1155/2023/9761846

Retraction

Retracted: Finite-Time Attitude Cooperative Control of Multiple Unmanned Aerial Vehicles via Fast Nonsingular Terminal Sliding Mode Control

Wireless Communications and Mobile Computing

Received 10 October 2023; Accepted 10 October 2023; Published 11 October 2023

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. Han, X. Tuo, Y. Tang, and P. He, "Finite-Time Attitude Cooperative Control of Multiple Unmanned Aerial Vehicles via Fast Nonsingular Terminal Sliding Mode Control," *Wireless Communications and Mobile Computing*, vol. 2022, Article ID 4324626, 11 pages, 2022.

Hindawi
Wireless Communications and Mobile Computing
Volume 2023, Article ID 9894320, 1 page
https://doi.org/10.1155/2023/9894320

Retraction

Retracted: A Mobile Intelligent Mine Platform with a Hybrid Fuzzy NN and ATT-CNN Prewarning Model

Wireless Communications and Mobile Computing

Received 10 October 2023; Accepted 10 October 2023; Published 11 October 2023

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. Wu, C. Wu, J. Wang, X. Zhang, and W. Chen, "A Mobile Intelligent Mine Platform with a Hybrid Fuzzy NN and ATT-CNN Prewarning Model," *Wireless Communications and Mobile Computing*, vol. 2022, Article ID 4545936, 14 pages, 2022.

Hindawi
Wireless Communications and Mobile Computing
Volume 2023, Article ID 9838651, 1 page
https://doi.org/10.1155/2023/9838651

Retraction

Retracted: Analysis of Feasibility Modeling of Multimode Physical Fitness Training for Long-Distance Runners

Wireless Communications and Mobile Computing

Received 11 July 2023; Accepted 11 July 2023; Published 12 July 2023

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] X. Xia, "Analysis of Feasibility Modeling of Multimode Physical Fitness Training for Long-Distance Runners," *Wireless Communications and Mobile Computing*, vol. 2021, Article ID 5817396, 8 pages, 2021.

Hindawi
Wireless Communications and Mobile Computing
Volume 2023, Article ID 9836176, 1 page
https://doi.org/10.1155/2023/9836176

Retraction

Retracted: Study on Regional Control of Tourism Flow Based on Fuzzy Theory

Wireless Communications and Mobile Computing

Received 11 July 2023; Accepted 11 July 2023; Published 12 July 2023

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] W. Shang and G. Chuangle, "Study on Regional Control of Tourism Flow Based on Fuzzy Theory," *Wireless Communications and Mobile Computing*, vol. 2021, Article ID 9648879, 7 pages, 2021.

Hindawi
Wireless Communications and Mobile Computing
Volume 2023, Article ID 9825958, 1 page
https://doi.org/10.1155/2023/9825958

Retraction

Retracted: Age Label Distribution Learning Based on Unsupervised Comparisons of Faces

Wireless Communications and Mobile Computing

Received 11 July 2023; Accepted 11 July 2023; Published 12 July 2023

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. Li, Z. Deng, W. Xu, Z. Li, and H. Liu, "Age Label Distribution Learning Based on Unsupervised Comparisons of Faces," *Wireless Communications and Mobile Computing*, vol. 2021, Article ID 1996803, 7 pages, 2021.

Hindawi
Wireless Communications and Mobile Computing
Volume 2023, Article ID 9794634, 1 page
https://doi.org/10.1155/2023/9794634

Retraction

Retracted: Research on Sports Class Load Monitoring System Based on Threshold Classification Algorithm

Wireless Communications and Mobile Computing

Received 11 July 2023; Accepted 11 July 2023; Published 12 July 2023

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. Zhao, "Research on Sports Class Load Monitoring System Based on Threshold Classification Algorithm," *Wireless Communications and Mobile Computing*, vol. 2021, Article ID 3891453, 10 pages, 2021.

Hindawi
Wireless Communications and Mobile Computing
Volume 2023, Article ID 9782397, 1 page
https://doi.org/10.1155/2023/9782397

Retraction

Retracted: Effects of Attentional Withdrawal on Chinese Disabled Biathletes' Shooting Performance

Wireless Communications and Mobile Computing

Received 11 July 2023; Accepted 11 July 2023; Published 12 July 2023

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.

In addition, our investigation has also shown that one or more of the following human-subject reporting requirements has not been met in this article: ethical approval by an Institutional Review Board (IRB) committee or equivalent, patient/participant consent to participate, and/or agreement to publish patient/ participant details (where relevant).

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. Li and S. Li, "Effects of Attentional Withdrawal on Chinese Disabled Biathletes' Shooting Performance," *Wireless Communications and Mobile Computing*, vol. 2021, Article ID 7264264, 9 pages, 2021.

Hindawi
Wireless Communications and Mobile Computing
Volume 2023, Article ID 9890424, 1 page
https://doi.org/10.1155/2023/9890424

Retraction

Retracted: Detection of Power Data Outliers Using Density Peaks Clustering Algorithm Based on *K*-Nearest Neighbors

Wireless Communications and Mobile Computing

Received 11 July 2023; Accepted 11 July 2023; Published 12 July 2023

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. Li, L. Chen, and Y. Wang, "Detection of Power Data Outliers Using Density Peaks Clustering Algorithm Based on *K*-Nearest Neighbors," *Wireless Communications and Mobile Computing*, vol. 2022, Article ID 2203137, 7 pages, 2022.

Hindawi
Wireless Communications and Mobile Computing
Volume 2022, Article ID 6177973, 10 pages
https://doi.org/10.1155/2022/6177973

Research Article

Online Multiplayer Tracking by Extracting Temporal Contexts with Transformer

Xiao Han,[1,2] **Yongbin Wang**,[2,3] **Shouxun Liu**,[1] **and Cong Jin**[1]

[1]*School of Information and Communication Engineering, Communication University of China, Beijing, China*
[2]*State Key Laboratory of Media Convergence and Communication, Communication University of China, Beijing, China*
[3]*Collaborative Innovation Center, Communication University of China, Beijing, China*

Correspondence should be addressed to Yongbin Wang; ybwang@cuc.edu.cn

Received 31 October 2021; Revised 4 August 2022; Accepted 13 September 2022; Published 11 October 2022

Academic Editor: Ting Bi

Sports competition is one of the most popular programs for many audiences. Tracking the players in sports game videos from broadcasts is a nontrivial challenge for computer vision researchers. In sports videos, the direction of an athlete's movement changes quickly and unpredictably. Mutual occlusion between athletes is also more frequent in team competitions. However, the rich temporal contexts among the adjacent frames have been excluded from consideration. To address this dilemma, we propose an online transformer-based learnable framework in an end-to-end fashion. We use a transformer architecture to extract the temporal contexts between the successive frames and add them to the network training, which is robust to occlusion and complex direction changes in multiplayer tracking. We demonstrate the effectiveness of our method on three sports video datasets by comparing them with recently advanced multiplayer trackers.

1. Introduction

With the increasing number of spectators who prefer sports games, video analysis of sports games has received more and more increasing attention in the field of computer vision. Multiple player tracking has become an urgent demand for sports video analysis, which has substantial benefits for intelligently editing game videos. For example, the identity information and movement information of athletes obtained by automatic tracking can be visualized and quantified. The audience can quickly focus on interested athletes and obtain more interactivity from the sports game. If tracking information can be obtained in real time, broadcasters can use them to assist in broadcasting, editing the highlight clips, and commenting on the game. The scope and amount of movement for each player can provide objective clues for assessing the player's abilities and formulating specific strategies. It is essential to correctly track multiple players even under various challenging conditions.

Several efforts have been made to address this issue. The traditional method uses Bayesian inference to solve the association problem. For instance, associating the identities of the isolated tracks by exploiting the graph constraints and similarity measures [1]. They formulate it as a Bayesian network inference problem. Reference [2] proposed a dual-mode two-way Bayesian inference approach that dynamically switches between an offline general model and an online dedicated model to address single isolated object tracking and multiple occluded objects tracking integrally by forward filtering and backwards smoothing. With the development of CNNs, researchers have begun to use deep learning to extract features. Reference [3] first utilized Faster R-CNN [4] to generate an initial detection, and the associating step is modeled as a minimum-cost network flow problem. An adaptive multiple scale sampling scheme based on spatially proximate foreground regions [5] is very helpful for preserving the underlying states of tracked objects even with severe occlusions. Despite some success, multiplayer tracking has a few problems. For example, most of the previous methods are offline, which cannot meet real-time requirements due to the use of information after the current frame. In addition, they all follow the standards of tracking by detection. Having the two models separately executed may lead to efficiency problems. In

the real scene, there are still some difficulties in sports game video broadcasts. In the case of live broadcasts, athletes need to be tracked in real time so that player analysis and highlight editing can be quickly presented to the audience.

Multiobject tracking generally refers to pedestrians or vehicles. Compared with multiobject tracking where the movement direction is relatively stable, multiplayer tracking has additional difficulties. Multiplayer tracking has rapid and frequent changes in the direction of a player's movement, and there are frequent occlusion and disappearance-reproduction problems in team sports. To effectively deal with these problems, most of the previous multiplayer tracking methods are offline and not end-to-end. In this paper, we propose an end-to-end online multiplayer tracking model using a transformer structure to extract the temporal domain information between adjacent frames and add it to the model training, which better solves the occlusion and the problem of sudden changes in direction.

Our contribution in this work can be summarized as follows:

(a) *Temporal contextual information among the successive frames is optimized by the transformer structure*

(b) *The previous frames' feature map is reused as an input of the current frame encoder, which can quickly and effectively associate the tracking box and reduce the missed tracking*

(c) *Our method is online and end-to-end, which is more concise and robust compared with previous multiplayer tracking models*

We use the MOT challenge evaluation metrics to perform comparative experiments on three sports datasets. Our model performs well on some indicators.

2. Related Work

2.1. Multiple Object Tracking. In the generalized MOT, the tracked objects include pedestrians, vehicles, animals, athletes, cells, and some rigid objects. MPT is a more specific task of research objects in the field of MOT. We first review the related work on player tracking, including both multiobject tracking and multiplayer tracking. Then, we briefly discuss the differences between multiple object tracking (MOT) and multiple player tracking (MPT). Finally, we also discuss the application of transformer architecture in the computer vision community.

Object tracking is a vital and basic task in computer vision. It has been applied in various real-world areas, such as security monitoring systems, autonomous driving, and video understanding. Single-object tracking (SOT) mainly provides the position information of the object to be tracked in the first frame and locates the object in the following consecutive video frames. However, MOT has no prior knowledge. The position coordinates of all objects in each frame and the corresponding identity ID to each object need to be labeled in the MOT task to distinguish the objects in the inner classification.

The current mainstream MOT algorithms are divided into two categories. The tracking-by-detection method, such as SORT [6] and Deep SORT [7]. First, a series of bounding boxes are extracted through conventional object detection methods, and then, based on the relationship between the previous and subsequent frames, the bounding boxes containing the same object are assigned the same ID. Recently, there have also been many studies on jointly learning the detector and data association, including JDE [8] and FairMOT [9]. Our transformer-based method allows object detection and appearance embedding to be learned in an end-to-end model. Thanks to the attention architecture in the transformers [10], the method we propose can learn efficiently and obtain the bounding box and identity ID simultaneously.

The MOT algorithm can also be divided into offline and online methods. When trying to determine the object location and ID information in a certain frame, the offline tracking algorithm [11–13] can use the information after the current frame. Because of the availability of more global information, offline algorithm results are often more accurate, but they generally consume more time. In contrast, online tracking algorithms [14–17] can only use current and past information to predict the current frame. The online tracking method is very suitable for automatic driving, navigation, program live broadcasting, and other tasks that require high real-time performance. Compared with offline methods, the performance of online methods tends to be less accurate because the methods cannot use future information to repair previous errors. To some extent, the multiplayer tracking studied in this paper belongs to MOT. However, compared with ordinary pedestrian and vehicle tracking, multiple player tracking has difficulties, such as a more similar appearance within the class, unstable movement direction and speed, and more occlusions and collisions. By making use of the attention's unique structure, which correlates the entire input sequence, we carefully modify the classic transformer structure to obtain an online multiplayer tracking model. Additionally, our model is in an end-to-end form.

2.2. Multiple Player Tracking. Some contributions have been proposed in previous studies on soccer player tracking [18, 19]. However, as mentioned above, there are three main differences between player tracking and pedestrian tracking.

(1) The appearance of the athletes is more similar. Most pedestrian clothes are diverse in color and style. In contrast, athletes, especially those from the same team, wear the same uniform. However, the jersey number can be used as a clue feature to distinguish the players

(2) Pedestrians generally move in a uniform speed and in a straight line, while the athlete's movement direction and speed are unpredictable due to their frequent drastic swerves and sudden speed changes, which will increase the difficulty of tracking due to severe deformation

(3) In addition, compared to pedestrians, athletes will crash into each other and collide more frequently

To resolve these difficult problems, previous researchers have made some developments. Pallavi et al. [18] is a tracking-by-detection method that solves the problems of player detection, labeling, and tracking in broadcast soccer videos. To accommodate the frequent disappearance and reappearance of targets, [20] views the date association in multiplayer tracking as a Markov Chain Monte Carlo (MCMC) problem. A dual-mode two-way Bayesian inference approach was proposed in [2] that dynamically switches between an offline general model and an online dedicated model to deal with single isolated object tracking and multiple occluded players tracking integrally by forward filtering and backwards smoothing. Liu et al. [21] extracted a set of "Game Context Features" from the video with off-field interference removed to represent the current state of the field. Then, a random decision forest combining the current trajectory and the context features selects the best affinity model for a certain athlete at a certain moment.

2.3. Transformer. The transformer structure was first proposed in the machine translation task. Transformer architectures are based on a self-attention mechanism that learns the relationships between elements of a sequence. In [10], the transformer abandons the traditional CNN and RNN, and the entire network uses the attention mechanism. More precisely, the transformer only consists of self-attention and a feedforward neural network. In contrast from the sequential structure of the RNN, the parallel computing system of the transformer structure has better parallelism and conforms to the GPU framework. Since then, the transformer model has gained increasing popularity in NLP tasks, such as text classification, machine translation, and question answering. Breakthroughs from transformer networks in Natural Language Processing (NLP) domain have sparked great interest in the computer vision community to adapt these models for vision tasks. Transformers are gradually being used in many vision tasks, such as image recognition, image enhancement, target detection, and image segmentation. To bridge the gap between the fields, many studies have made some modifications when introducing the transformers and their variant transfer learning into visual tasks. For example, [22] focused on completely migrating the transformer to the image classification task and completely abandoned the CNN. The input image is divided into patches, and then, each patch is flattened. The subsequent operation is similar to the BERT [23] in machine translation. Based on the CNN and the transformer, [24] completely removed the postprocessing steps of the previous detection algorithms that rely on artificial a priori for NMS, anchor generators, and constructs a completely end-to-end target detection framework. The aforementioned methods merely use the transformer structure in image-level vision tasks. After that, [25] also introduced transformers to the visual tracking community for the first time. By virtue of the key-query mechanism in the attention architecture, they can track new targets in the joint detection and tracking pipeline. Inspired by [25], we carefully adjusted the structure of the transformer to adapt to the task of multiplayer tracking, and we obtained a competitive result.

3. Approach

As mentioned above, MPT can be regarded as a MOT problem where the tracking objects are athletes, and the purpose is to obtain the position coordinates and identity information of all athletes in consecutive frames. We offer a mathematical formulation of MPT. Given an image frames sequence as the input, devoted by $I = \{I_1, I_2, \cdots, I_t, \cdots\}$, where I_t is the tth frame. We employ $S_t = \{S_t^1, S_t^2, \cdots, S_t^i, \cdots, S_t^m\}$ to devote the state of the tth frame, where m represents the total number of athletes in the tth frame, and S_t^m is the state of the ith athlete in the tth frame. Athlete's state include position, size, speed, direction, and appearance. The trained model is given a sequence of frames and outputs the trajectory T, identity d, and position and size information (x, y, w, h) of all athlete targets in each frame. We employ $p_t^i = (T, d, x, y, w, h)$ to denote the output result of the ith athlete in the tth frame.

To solve the serious deformation and similar appearance of athletes in the MPT task, it is necessary to obtain more accurate identity information and position coordinates of the athletes. In this section, we introduce our proposed model in detail. After that, the settings in the training and inference process are discussed.

3.1. Framework. There are four core components in our model, which are the backbone network to extract the feature map, the encoder component, the decoder tracking component, and the matching component, as shown in Figure 1.

Thanks to the transformer architecture [10], we can exploit the rich temporal contexts among the adjacent frames in the video flow via an encoder-decoder structure. An overview of our transformer-based architecture is illustrated in Figure 2. The encoder part is simple and the same as the classic transformer encoder structure. The encoder takes the feature maps of two consecutive frames as a pair of inputs. The feature map calculated in the previous frame is retained and reused as part of the input of the current encoder, which can reduce computational consumption. Specifically, in the self-attention structures, we use the classic dot-product to calculate the correlation. The add and norm represent residual connection and instance normalization [26], respectively. In the decoder module, we use the same decoder structure to complete object detection and tracking propagation. The difference between the two decoder operations is the input query. The detection branch is the same as DETR [24], which is the learned query, while the query of the tracking propagation branch is the feature of the target provided by the tracker in the previous frame. In the decoding process, the cross-attention block bridges features from the previous frame and the current frame to propagate temporal contexts.

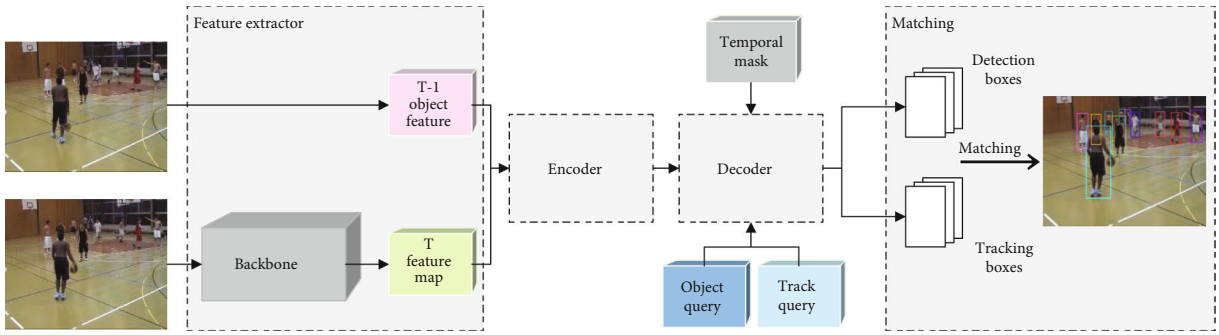


Figure 1: The transformer-based framework of the proposed. The CNN module is used to extract the features of the input frame. The global feature maps of the previous frame and the current frame are fed to the encoder, and then, the combined global feature map of two consecutive frames is input into the decoder as a common key. The temporal mask is beneficial for suppressing the background changes transformed from the previous frame temporally and concentrates on the target player. The object detection features of the current frame and the tracked object features of the previous frame are input into two decoders with a shared structure, and then, we obtain detection boxes and tracking boxes. Finally, IoU matching is employed to associate detection boxes with tracking boxes.

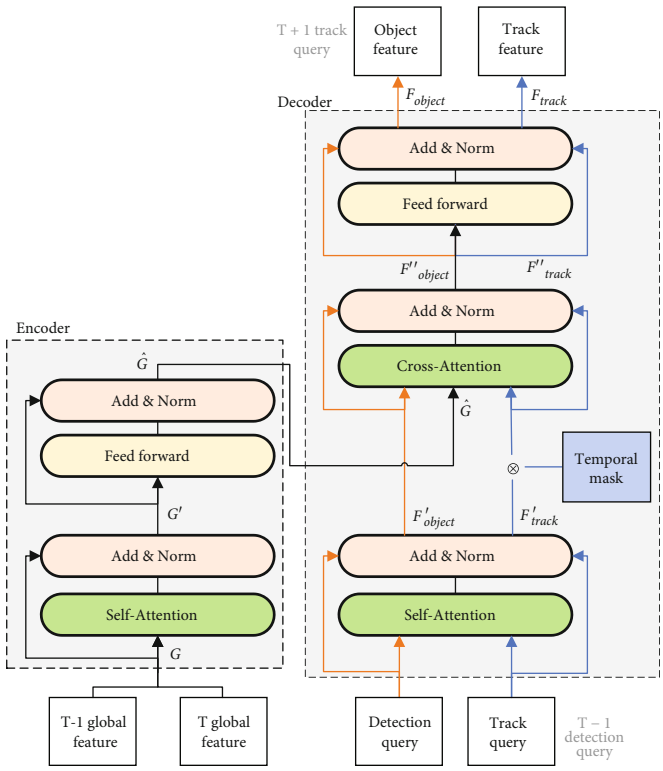


Figure 2: Transformer architecture in our method.

(1) *Encoder*

In the encoder part, the global feature map g_{t-1} of the previous frame is retained. We combine g_{t-1} and the global feature map of the current frame g_t in the series, denoted by G. The similarity matrix $A_{G\longrightarrow G}$ is calculated by the self-attention block as follows:

$$A_{G\longrightarrow G} = \text{Attention}(G, G). \tag{1}$$

Attention(G, G) represents the self-attention operation, as shown in the green box in Figure 2. Then, the assembled global feature G is transformed through $A_{G\longrightarrow G}G$. As shown in Figure 2, the next operation is residual connection and instance normalization as follows:

$$G' = \text{Ins.Norm}(A_{G\longrightarrow G}G + G). \tag{2}$$

We use instance normalization, denoted by Ins.Norm in our encoder-decoder structure. In the following experiments, we compare several mainstream normalization methods. Experimental results show that instance normalization is the best performer among them. G' is the output of the hidden layer after residual connection and instance

normalization. According to [24], the encoder and decoder include a fully connected feed forward network after the attention block. The feed forward network consists of two linear transformations with a ReLU activation in between as follows:

$$\mathrm{FFN}\left(G'\right) = \mathrm{ReLU}\left(G' W_1 + b_1\right) W_2 + b_2. \tag{3}$$

Obtained by Equations (2) and (3), the output of the encoder denoted by $\widehat{G}$ is as follows:

$$\widehat{G} = \mathrm{Ins.Norm}\left(\mathrm{FFN}\left(G'\right) + G'\right). \tag{4}$$

By virtue of the self-attention structure, the global feature map of two consecutive frames can be aggregated to generate $\widehat{G}$. $\widehat{G}$ will be input as the common key into the next two shared decoders.

(2) *Decoder*

In the decoder, two decoders that share the network structure—detection decoder and track decoder, as shown in the orange connection and blue connection, are used to generate the player track boxes and the player detection boxes of the current frame, respectively. The orange line in Figure 2 represents the object decoder. The learnable detected player feature in the current frame is used as its input. We concatenate the representation of all the player's patches into an object query F_{object} in the current frame. As shown by the orange line in Figure 2, the first self-attention block, including residual connection and normalization, outputs the middle layer feature F'_{object}, which is expressed as follows:

$$F'_{\mathrm{object}} = \mathrm{Ins.Norm}\left(A_{F_{\mathrm{object}} \longrightarrow F_{\mathrm{object}}} F_{\mathrm{object}} + F_{\mathrm{object}}\right). \tag{5}$$

Based on the common feature $\widehat{G}$ in Equation (4) and the middle layer feature F'_{object} in Equation (5), we can compute the similarity matrix as follows:

$$A_{\widehat{G} \longrightarrow F'_{\mathrm{object}}} = \mathrm{Attention}\left(\widehat{G}, F'_{\mathrm{object}}\right). \tag{6}$$

Then, the cross-attention matrix $A_{\widehat{G} \longrightarrow F_{\mathrm{object}}}$ is fed to the residual connection and normalization layer as follows:

$$F'_{\mathrm{object}}{}' = \mathrm{Ins.Norm}\left(A_{\widehat{G} \longrightarrow F'_{\mathrm{object}}} \widehat{G} + F'_{\mathrm{object}}\right), \tag{7}$$

where $F'_{\mathrm{object}}{}'$ is the middle feature exported by the cross-attention block, including the residual connection and normalization. Furthermore, the feed forward network is added to the end of the decoder. In the detection encoder, we finally calculate the object feature, which is detected from the aggregated global feature map $\widehat{G}$ by the object query. This object feature $\widehat{F_{\mathrm{object}}}$ is the next frame's track query as follows:

$$\widehat{F_{\mathrm{object}}} = \mathrm{Ins.Norm}\left(\mathrm{FFN}\left(F'_{\mathrm{object}}{}'\right) + F'_{\mathrm{object}}{}'\right). \tag{8}$$

The blue connection represents the track decoder, as seen in Figure 2. The detected object feature map of the previous frame is fed to the track decoder as the track query F_{track} of the current frame. Similar to the first self-attention block in the aforementioned detection decoder, F'_{track} is the feature of the middle layer as follows:

$$F'_{\mathrm{track}} = \mathrm{Ins.Norm}\left(A_{F_{\mathrm{track}} \longrightarrow F_{\mathrm{track}}} F_{\mathrm{track}} + F_{\mathrm{track}}\right). \tag{9}$$

To leverage the temporal context information between the two consecutive frames and transform the temporal motion prior [27], we construct a Gaussian Radial Basis Function—temporal feature for the track query as follows:

$$m(y) = \exp\left(-\frac{\|y - c\|^2}{2\sigma^2}\right), \tag{10}$$

where c is the ground truth of the object position. Temporal mask matrix M_{temp} is the temporal feature ensemble. Similar to the aforementioned detection decoder, after the cross-attention block, we obtain the middle layer feature as follows:

$$F'_{\mathrm{track}}{}' = \mathrm{Ins.Norm}\left(A_{\widehat{G} \longrightarrow F_{\mathrm{track}}} \widehat{G} + F'_{\mathrm{track}} \otimes M_{\mathrm{temp}}\right), \tag{11}$$

where $\otimes$ is the elementwise multiplication. Finally, $F'_{\mathrm{track}}{}'$ is fed to the feed forward network block, and the final output of the track decoder is $\widehat{F_{\mathrm{track}}}$ as follows:

$$\widehat{F_{\mathrm{track}}} = \mathrm{Ins.Norm}\left(\mathrm{FFN}\left(F'_{\mathrm{track}}{}'\right) + F'_{\mathrm{track}}{}'\right). \tag{12}$$

3.2. Training. As shown in Figure 1, a pair of adjacent frames are fed to our model during training, which comes from the same clip video. If the input is a single picture, translation conversion and other operations must be performed on the original picture to generate an input that simulates adjacent frames. The output of the backbone is the feature map of the current frame. The encoder module takes the feature map of the previous frame and the feature map of the current frame as input. These features come from the output of the feature extraction module. The feature map of the previous frame is retained and reused for current tracking. Cross-attention is the information exchange between the encoder and the decoder. Both decoder operations are detection tasks. On the one hand, the loss of the detection branch is the same as that of DETR, and the set prediction loss is used for the distribution of the ground truth during training. Then, the regression and classification loss are calculated. On the other hand, for the branch of the track, there is no need to allocate ground truth again when there are track characteristics, and the classification and regression loss are directly calculated.

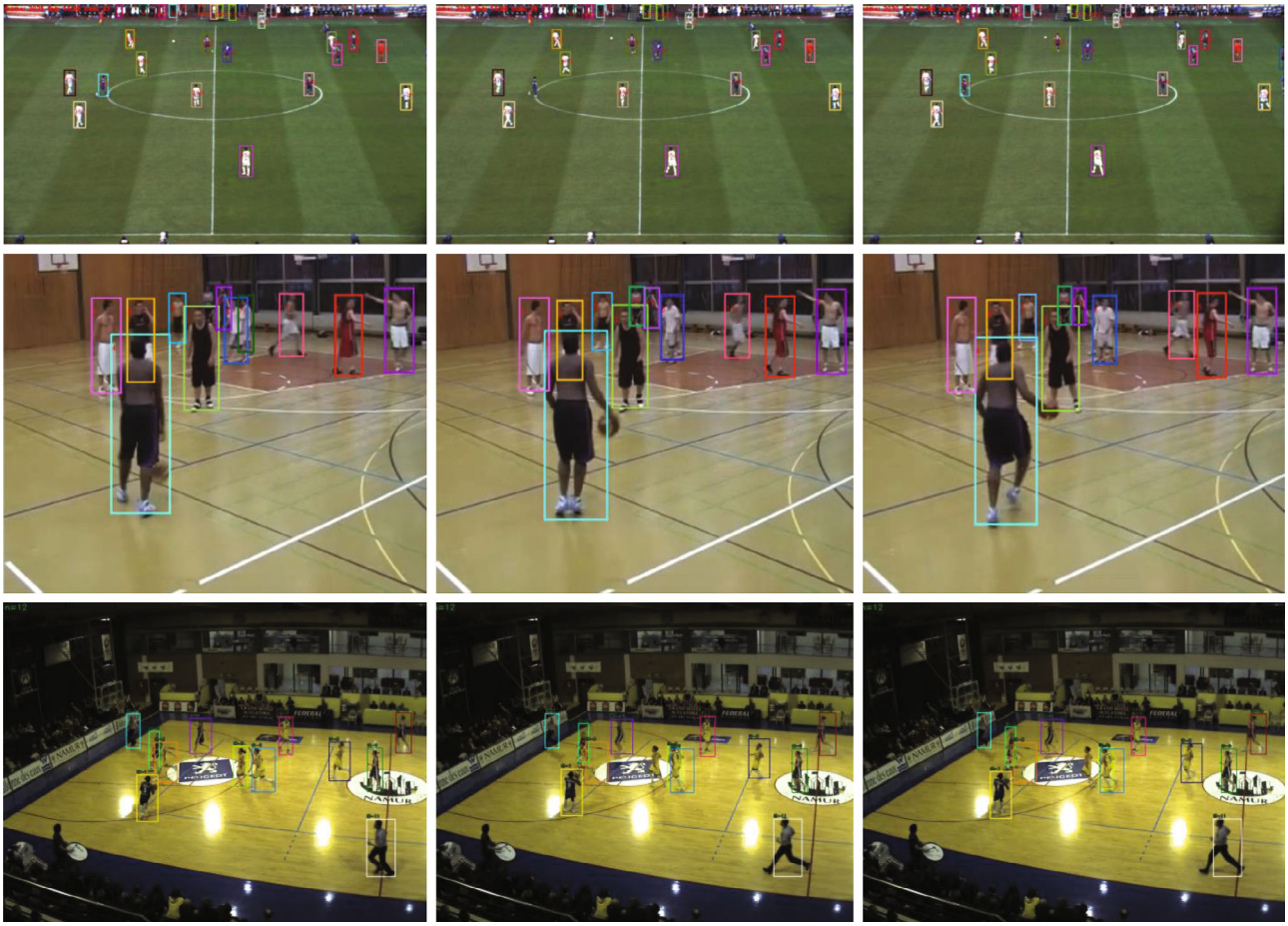


Figure 3: Sample tracking results from the APIDIS, EPFL-basketball, and ISSIA CNR-soccer datasets.

The final matching loss is defined as follows:

$$L_{\text{track}} = \lambda_{\text{cls}} \cdot L_{\text{cls}} + \lambda_{L_1} \cdot L_1 + \lambda_{\text{iou}} \cdot L_{\text{iou}}, \tag{13}$$

where L_{cls} is the focal loss of the predicted classifications and ground truth category labels. L_1 is the L_1 loss, and L_{iou} is the IoU loss between the normalized center coordinates and the height and width of the predicted boxes and the ground truth box. Then, λ_{cls}, λ_{L_1}, and λ_{iou} are the coefficients of each module.

During training, the output of the decoder will minimize the cross-entropy with the ground truth, but there is no correct answer in the inference process. To solve this problem and enable the model to detect new targets in time, we add exposure bias during the training process.

3.3. Inference. In the inference process, our method first obtains the global feature map of this frame through the CNN feature extractor and detects the players in it. Then, the two global feature maps of the first frame are combined and input into the model. After that, our method performs the object transform and box association for the next frames until the entire video sequence is completed. We use track rebirth tips [8], which are often used for tracking tasks to enhance the robustness where athletes' occlusions and movement directions change suddenly.

4. Experiments

To evaluate the proposed method, we present the experimental setting and performance compared with several MOT algorithms. There is no uniform evaluation standard for multiple tracking tasks. For fairness, we use the metrics in the MOT challenge. To verify that the encoder-decoder structure in our method makes good use of the information between the adjacent frames and the rebroadcasts of the tracking information of the previous frame to the current frame, we also perform ablation experiments.

4.1. Datasets

4.1.1. ISSIA-soccer. This dataset [28] is collected for a football match broadcast. The resolution is 1920*1080, and the frame rate is 25 fps. There are 6 videos, each of which is 2 minutes. At the beginning of every video, the first 300 frames without annotated information are used to remove the background initialization. Although the resolution of this dataset is high, the camera is far away from the target of tracking players, and the characteristics of the athletes are blurred.

4.1.2. APIDIS. The videos in this dataset [29] come from 7 fixed cameras, 5 of which are ordinary wide-angle cameras and 2 of which are fish-eye panoramic cameras taken from

Table 1: Comparison of tracking performance on the expanded APIDIS basketball dataset.

Methods	*Mode*	*MOTA*↑	*MOTP*↑	*IDF1*↑	*MT*↑	*ML*↓	*FP*↓	*FN*↓	*FPS*↓
CEM [36]	Offline	64.0	76.9	46.7	45.1	23.6	1598	3190	1.1
MHT [14]	Offline	73.2	78.6	50.3	54.7	22.9	890	2787	0.8
ELP [37]	Offline	74.7	80.4	55.9	55.3	21.4	801	2561	3.6
SORT [6]	Online	74.8	80.3	52.0	56.3	22.7	763	2757	17.2
Ours	Online	75.3	80.6	56.1	57.0	20.0	767	2538	15.1

Table 2: Comparison of tracking performance on the ISSIA-soccer dataset.

Methods	*Mode*	*MOTA*↑	*MOTP*↑	*IDF1*↑	*MT*↑	*ML*↓	*FP*↓	*FN*↓	*FPS*↓
CEM [36]	Offline	62.8	66.8	38.3	35.9	25.6	1062	2168	1.7
MHT [14]	Offline	63.5	68.2	39.8	36.2	22.1	306	2280	1.3
ELP [37]	Offline	67.5	65.1	43.6	38.9	24.3	334	2017	3.6
SORT [6]	Online	69.1	71.5	46.7	49.1	22.5	329	2033	17.2
Ours	Online	72.4	75.7	50.5	53.1	19.5	366	2480	23.2

Table 3: Comparison of tracking performance on the EPFL-basketball dataset.

Methods	*Mode*	*MOTA*↑	*MOTP*↑	*IDF1*↑	*MT*↑	*ML*↓	*FP*↓	*FN*↓	*FPS*↓
CEM [36]	Offline	59.0	63.3	37.9	48.0	41.1	2834	5863	5.4
MHT [14]	Offline	61.2	61.0	40.1	47.3	36.4	3037	5974	3.8
ELP [37]	Offline	66.4	65.9	45.3	45.6	39.0	2785	5359	6.0
SORT [6]	Online	67.1	72.9	52.9	53.8	38.8	2944	4990	21.5
Ours	Online	68.7	68.5	50.6	54.5	31.7	2767	4550	19.0

the top of the venue. The size is 1600×1200, and the frame rate is almost 22 fps on average. There are a total of 1500 frames in the dataset with tracking information, including 2 referees and 10 basketball players. However, this dataset has slightly fewer labeled frames. To enable the dataset to be trained in the deep network, we expand the annotation information. We choose a clip with a length of 30 s from 5 ordinary wide-angle camera original videos. Then, we supplement their tracking information. Finally, 4800 frames can be used for tracking training.

4.1.3. EPFL-basketball. The EPFL-basketball dataset [30] was taken in the school basketball hall, with 4 fixed cameras standing on the ground, shooting from the four directions of the basketball court. There are 4 segments, each of which is a 6-minute video taken from four views. The resolution is 360*288, and the frame rate is 25 fps. This dataset has a low resolution, which poses challenges for detection and tracking.

4.2. Implementation Details. In the feature extraction module, we use ResNet-34 [31] as the backbone. We use the parameters after pretrained on the COCO dataset [32]. For simplicity, we fix the weight of the pretrained ResNet-34 and only fine-tune the fully connect connected layers. We train our model with the Adam optimizer for 30 epochs with a starting learning rate of e^{-4}, and the learning rate decays to e^{-5} at 20 epochs. The batch size is set to 12. We use standard data augmentation techniques, including rotation, scaling, and color jittering.

4.3. Evaluation Metrics. There is no established standard of evaluation for multi-player tracking. For the sake of fairness, it is feasible to utilize the MOT metrics [33, 34] to measure the multiplayer tracking methods. Multiobject tracking accuracy (MOTA) mainly considers all object matching errors in tracking, including the ratio of misses in the sequence, false positives, and of mismatches. Multiobject tracking precision (MOTP) represents the accuracy of the target position. The closer MOTP is to 1, the higher the positioning accuracy of the tracker. Mostly tracked targets (MT) mean the ratio of groundtruth trajectories that are covered by a track hypothesis for at least 80% of their respective lifespan. Mostly lost targets (ML) represent the ratio of groundtruth trajectories that are covered by a track hypothesis for at most 20% of their respective lifespan [35]. False positive (FP) is the total number of false positive, and false negative (FN) expresses the total number of false negatives. The IDF1 score is the ratio of correctly identified detections to the average number of groundtruth and computed detections. Frame per second (FPS) indicates the speed of tracking processing. IDS is the number of ID switches, that is, the tracking object ID is different from its historical ID, which often occurs when multiple objects block each other.

Table 4: Ablation study on the temporal mask.

Temporal mask	*IDF1*↑	*MOTA*↑	*MOTP*↑	*MT*↑	*ML*↓
With	55.3	59.7	77.3	53.1	25.7
Without	56.1	75.3	80.6	57.0	20.0

Table 5: Temporal contexts between consecutive frames can improve the tracking effect on the expanded APIDIS basketball dataset.

Adding feature	*IDF1*↑	*MOTA*↑	*MOTP*↑	*MT*↑	*ML*↓
Current & Current	53.5	67.9	69.0	51.9	25.9
Translated & Current	50.1	66.7	74.1	55.8	23.4
Previous & Current	56.1	75.3	80.6	57.0	20.0

Table 6: Ablation experiment of normalization.

Normalization	*IDF1*↑	*MOTA*↑	*MOTP*↑
BatchNorm [38]	61.3	59.7	77.3
LayerNorm [39]	60.7	55.3	79.5
GroupNorm [40]	62.5	66.9	80.4
PowerNorm [41]	62.0	63.4	79.2
InstanceNorm [26]	56.1	75.3	80.6

4.4. Result Analysis. We first compare our method with other state-of-the-art multiple object tracking and multiple player tracking pipelines. An example of our multiplayer tracking result of the proposed method on three sports datasets is shown in Figure 3. Then, we verified the effect of the proposed method on multiplayer tracking for each component and evaluated our extracted temporal context tracking by comparing it to changing the input of the decoder.

4.4.1. Compared with Other Trackers. We compare the proposed method with other commonly classical multiobject tracking methods CEM [36], MHT [14], ELP [37], and SORT [6] on three datasets APIDIS, ISSIA-soccer, and EPFL-basketball. As shown in Tables 1–3, by virtue of the rich temporal context information between two consecutive frames, our method suppresses the comparative methods on MOTA and IDF1 on three datasets. Video is streaming media, and the temporal context information in it is crucial for continuous tracking.

However, most current trackers [14, 36, 37] tackle the task by frame-by-frame object tracking, where the temporal relationship between consecutive frames is largely ignored. Among them, SORT [6] records the object state from the previous frame in the object state management in the tracking stage, and this idea is widely adopted in the multiobject tracking model. Although historical frames are considered in some of the above methods, video frames are still considered independent and do not contribute to each other. We directly use the combined features of the previous frame and the current frame to generate tracking track boxes and use the interframe context to predict the position of the target of the previous frame in the current frame. The experimental results show that our method can better manage the problem of multiobject tracking in sports, such as for the players with similar appearances and complex movement states. However, our method does not perform as well as the previous research results on the FPS indicator because we introduce a more complex transformer structure in the pipeline, which increases its calculation time.

4.4.2. Temporal Mask. Due to the uniqueness of the attention mechanism, the transformer pays more attention to the drastically changing parts of the image, including the background that should not be highlighted. We leverage the temporal mask mechanism to suppress the background changes transformed from the previous frame temporally and to concentrate on the target player. As we can see from Table 4, the performance of our approach with the mechanism outperforms that without it, which proves that the effectiveness of the temporal mask mechanism can suppress the effect of changing the background on the tracking results.

4.4.3. Temporal Contexts. We also conduct comparative experiments on temporal contexts, as shown in Table 5.

Experimental results show that adding the object feature map of the previous frame has a better effect and reduces the probability of missing tracking. To verify that the transformer structure in our method is conducive to extracting the temporal contexts between consecutive frames, we test the use of different feature maps as the input of the encoder part, and the current feature map is combined into a composite. "Current & Current" means that the two current frame feature maps are used as a pair of input encoders. "Translated & Current" means that the current frame is subjected to random scaling and translating operations to the feature map and the original frame's feature map as input. "Previous & Current" is the combination of the global feature map of the previous frame and the current frame and the input encoder described above. As seen in Table 5, if the object feature map of the previous frame is not used, the MOTA result will be reduced by 7.4-8.6%.

4.4.4. Feature Normalization. If the value range of each dimension of the input matrix has a large difference, it will cause a large difference in the slope of the loss function in each direction, and training will become difficult. To address this dilemma, previous researchers have proposed a variety of effective feature normalization methods. To make the model training more efficient. We conduct comparative experiments on different normalization methods, and the results are shown in Table 6. Instance normalization is pixelwise to calculate the mean and the standard deviation, which corresponds to the pixel-to-pixel correspondence between the two frames in our method of cross-attention. The results show that instance normalization is more suitable for our player tracking task.

5. Conclusion

In this paper, we propose a novel model for multiplayer tracking in broadcast sports game videos. We take advantage of the transformer-based structure and make full use of the temporal contexts between consecutive frames. Extensive experiments are conducted to demonstrate that after adding the temporal context information, our model improves the results in the sports videos. Deep neural networks and transformer networks have achieved tremendous success in many vision applications. We will utilize the tracker for better sports video analysis.

Data Availability

We used the public datasets ISSIA-soccer, APIDIS, and EPFL-basketball, which are marked in the paper.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The result of this paper is supported by the National Key R&D Program of China (Grant No. 2019YFB1406201).

References

[1] P. Nillius, J. Sullivan, and S. Carlsson, "Multi-target tracking - linking identities using Bayesian network inference," in *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, vol. 2, pp. 2187–2194, New York, NY, USA, 2006.

[2] J. Xing, H. Ai, L. Liu, and S. Lao, "Multiple player tracking in sports video: a dual-mode two-way Bayesian inference approach with progressive observation modeling," *IEEE Transactions on Image Processing*, vol. 20, no. 6, pp. 1652–1667, 2011.

[3] L. Kong, D. Huang, and Y. Wang, "Long-term action dependence-based hierarchical deep association for multi-athlete tracking in sports videos," *IEEE Transactions on Image Processing*, vol. 29, pp. 7957–7969, 2020.

[4] 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.

[5] W. Kim, S.-W. Moon, J. Lee, D.-W. Nam, and C. Jung, "Multiple player tracking in soccer videos: an adaptive multiscale sampling approach," *Multimedia Systems*, vol. 24, no. 6, pp. 611–623, 2018.

[6] A. Bewley, Z. Ge, L. Ott, F. Ramos, and B. Upcroft, "Simple online and realtime tracking," in *2016 IEEE International Conference on Image Processing (ICIP)*, pp. 3464–3468, Phoenix, AZ, USA, 2016.

[7] N. Wojke, A. Bewley, and D. Paulus, "Simple online and real-time tracking with a deep association metric," in *2017 IEEE International Conference on Image Processing (ICIP)*, pp. 3645–3649, Beijing, China, 2017.

[8] P. Bergmann, T. Meinhardt, and L. Leal-Taixe, "Tracking without bells and whistles," in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 941–951, Seoul, Korea (South), 2019.

[9] Y. Zhang, C. Wang, X. Wang, W. Zeng, and W. Liu, "FairMOT: on the fairness of detection and re-identification in multiple object tracking," *International Journal of Computer Vision*, vol. 129, no. 11, pp. 3069–3087, 2021.

[10] 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*, Long Beach, CA, USA, 2017.

[11] L. Chen, X. Peng, and M. Ren, "Recurrent metric networks and batch multiple hypothesis for multi-object tracking," *IEEE Access*, vol. 7, pp. 3093–3105, 2019.

[12] L. Leal-Taixé, C. Canton-Ferrer, and K. Schindler, "Learning by tracking: Siamese CNN for robust target association," in *2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 418–425, Las Vegas, NV, USA, 2016.

[13] J. Son, M. Baek, M. Cho, and B. Han, "Multi-object tracking with quadruplet convolutional neural networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3786–3795, Honolulu, HI, USA, 2017.

[14] C. Kim, F. Li, A. Ciptadi, and J. M. Rehg, "Multiple hypothesis tracking revisited," in *2015 IEEE International Conference on Computer Vision (ICCV)*, pp. 4696–4704, Santiago, Chile, 2015.

[15] S. Anton Milan, H. Rezatofighi, A. Dick, I. Reid, and K. Schindler, "*Online multi-target tracking using recurrent neural networks*," in *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence (AAAI'17)*, pp. 4225–4232, San Francisco, CAL, USA, 2017.

[16] J. Yin, W. Wang, Q. Meng, R. Yang, and J. Shen, "A unified object motion and affinity model for online multi-object tracking," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 2020, pp. 6768–6777, Virtual, 2020.

[17] H. Zhou, W. Ouyang, J. Cheng, X. Wang, and H. Li, "Deep continuous conditional random fields with asymmetric inter-object constraints for online multi-object tracking," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 29, no. 4, pp. 1011–1022, 2019.

[18] V. Pallavi, J. Mukherjee, A. K. Majumdar, and Shamik Sural, "Graph-based multiplayer detection and tracking in broadcast soccer videos," *IEEE Transactions on Multimedia*, vol. 10, no. 5, pp. 794–805, 2008.

[19] M. Manafifard, H. Ebadi, and H. Abrishami Moghaddam, "A survey on player tracking in soccer videos," *Computer Vision and Image Understanding*, vol. 159, pp. 19–46, 2017.

[20] J. Liu, X. Tong, W. Li, T. Wang, Y. Zhang, and H. Wang, "Automatic player detection, labeling and tracking in broadcast soccer video," *Pattern Recognition Letters, Video-based Object and Event Analysis*, vol. 30, no. 2, pp. 103–113, 2009.

[21] J. Liu, P. Carr, R. T. Collins, and Y. Liu, "Tracking sports players with context-conditioned motion models," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1830–1837, Portland, OR, USA, 2013.

[22] A. Dosovitskiy, L. Beyer, A. Kolesnikov et al., "An image is worth 16x16 words: transformers for image recognition at scale," 2020, https://arxiv.org/abs/2010.11929.

[23] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "*BERT: pre-training of deep bidirectional transformers for language understanding*," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pp. 4171–4186, Minneapolis, Minnesota, 2019.

[24] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, and S. Zagoruyko, "End-to-end object detection with transformers," in *16th European Conference on Computer Vision*, pp. 213–229, Cham, 2020.

[25] P. Sun, J. Cao, Y. Jiang et al., "TransTrack: multiple object tracking with transformer," 2020, https://arxiv.org/abs/2012.15460.

[26] D. Ulyanov, A. Vedaldi, and V. Lempitsky, "*Instance normalization: the missing ingredient for fast stylization*," 2017, https://arxiv.org/abs/ArXiv:1607.08022.

[27] N. Wang, W. Zhou, J. Wang, and H. Li, "Transformer meets tracker: exploiting temporal context for robust visual tracking," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1571–1580, Nashville, TN, USA, 2021.

[28] T. D'Orazio, M. Leo, N. Mosca, P. Spagnolo, and P. L. Mazzeo, "A semi-automatic system for ground truth generation of soccer video sequences," in *2009 Sixth IEEE International Conference on Advanced Video and Signal Based Surveillance*, pp. 559–564, Genova, Italy, 2009.

[29] C. De Vleeschouwer and D. Delannay, *Basket Ball Dataset from the European Project APIDIS*, 2009.

[30] F. Fleuret, J. Berclaz, R. Lengagne, and P. Fua, "Multicamera people tracking with a probabilistic occupancy map," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 2, pp. 267–282, 2008.

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

[32] T.-Y. Lin, M. Maire, S. Belongie et al., "Microsoft COCO: common objects in context," in *Computer Vision – ECCV 2014*, Lecture Notes in Computer Science, pp. 740–755, 2014.

[33] K. Bernardin and R. Stiefelhagen, "Evaluating multiple object tracking performance: the CLEAR MOT metrics," *EURASIP Journal on Image and Video Processing*, vol. 2008, Article ID 246309, 10 pages, 2008.

[34] E. Ristani, F. Solera, R. Zou, R. Cucchiara, and C. Tomasi, "Performance measures and a data set for multi-target, multi-camera tracking," in *Computer Vision – ECCV 2016 Workshops*, Lecture Notes in Computer Science, pp. 17–35, 2016.

[35] N. Ran, L. Kong, Y. Wang, and Q. Liu, "A robust multi-athlete tracking algorithm by exploiting discriminant features and long-term dependencies," in *MultiMedia Modeling*, Lecture Notes in Computer Science, I. Kompatsiaris, B. Huet, V. Mezaris, C. Gurrin, W.-H. Cheng, and S. Vrochidis, Eds., pp. 411–423, Springer International Publishing, 2019.

[36] A. Milan, S. Roth, and K. Schindler, "Continuous energy minimization for multitarget tracking," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 1, pp. 58–72, 2014.

[37] N. McLaughlin, J. M. Del Rincon, and P. Miller, "Enhancing linear programming with motion modeling for multi-target tracking," in *2015 IEEE Winter Conference on Applications of Computer Vision*, pp. 71–77, Santiago, Chile, 2015.

[38] S. Ioffe and C. Szegedy, "Batch normalization: accelerating deep network training by reducing internal covariate shift," in *Proceedings of the 32nd International Conference on Machine Learning*, pp. 448–456, Lille, France, 2015.

[39] J. L. Ba, J. R. Kiros, and G. E. Hinton, "Layer normalization," 2016, https://arxiv.org/abs/1607.06450.

[40] Y. Wu and K. He, "Group normalization," in *Proceedings of the European conference on computer vision (ECCV)*, pp. 3–19, Munich, Germany, 2018.

[41] S. Shen, Z. Yao, A. Gholami, M. W. Mahoney, and K. Keutzer, "PowerNorm: rethinking batch normalization in transformers," in *Thirty-seventh International Conference on Machine Learning*, Vienna, Austria, 2020.

Hindawi
Wireless Communications and Mobile Computing
Volume 2023, Article ID 9894320, 1 page
https://doi.org/10.1155/2023/9894320

Retraction

Retracted: A Mobile Intelligent Mine Platform with a Hybrid Fuzzy NN and ATT-CNN Prewarning Model

Wireless Communications and Mobile Computing

Received 10 October 2023; Accepted 10 October 2023; Published 11 October 2023

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. Wu, C. Wu, J. Wang, X. Zhang, and W. Chen, "A Mobile Intelligent Mine Platform with a Hybrid Fuzzy NN and ATT-CNN Prewarning Model," *Wireless Communications and Mobile Computing*, vol. 2022, Article ID 4545936, 14 pages, 2022.

Hindawi
Wireless Communications and Mobile Computing
Volume 2022, Article ID 4545936, 14 pages
https://doi.org/10.1155/2022/4545936

Research Article

A Mobile Intelligent Mine Platform with a Hybrid Fuzzy NN and ATT-CNN Prewarning Model

Yuliang Wu,[1] **Chao Wu**,[2] **Jun Wang**,[3] **Xiangnan Zhang**,[4] **and Wei Chen**[5]

[1]*Blockchain Research Center of China, School of Economic Information Engineering, Southwestern University of Finance and Economics, Emergency Management Institute, Chengdu University, Chengdu, China*
[2]*Shandong University of Science and Technology, College of Intelligent Equipment, Tai'an, China*
[3]*Fintech Innovation Center, Financial Intelligence & Financial Engineering Key Laboratory of Sichuan Province, School of Economic Information Engineering, Southwestern University of Finance and Economics, Chengdu, China*
[4]*School of Economic Information Engineering, Southwestern University of Finance and Economics, Chengdu, China*
[5]*School of Mechanical Electronic & Information Engineering, China University of Mining and Technology (Beijing), Beijing 100083, China*

Correspondence should be addressed to Chao Wu; sjh2100@163.com and Wei Chen; chenwdavior@163.com

Received 10 November 2021; Revised 8 July 2022; Accepted 25 July 2022; Published 18 August 2022

Academic Editor: Chi-Hua Chen

Given the existence of coal production risk, effective prewarning is important to the reliability and safety of coal mine. So, the development of a risk prewarning system has become an important safety management tool. To improve the prediction ability and the supervision level of safety production, and handle different multidimensional (temporal and spatial) information for risk prewarning, we built a new platform based on the Internet, cloud platform, mobile communication, GIS, and artificial intelligence technology, i.e., a mobile intelligent mine platform. The terminal of the platform provides real-time queries and procedures of coal mine production and risk prewarning and provides data support and technical means for daily supervision, remote networking analysis, law enforcement inspection, and emergency rescue. The prewarning model of safety risk is an essential means to realize prewarning. The complexity of production of coal mine leads to the dynamic characteristics, fuzziness, and randomness of coal mine accidents. The complex nonlinear relationship between index and risk level leads to low accuracy of the traditional back propagation (BP) neural network prewarning method. A novel model based on a compensation fuzzy neural network (NN) and an attention mechanism-convolutional neural network (ATT-CNN) are a critical part of the new design. First, to full use of the convolutional network to get a larger receptive field, one-dimensional time series is transformed into two-dimensional matrix as the input of the CNN network by mapping. The neural network is utilized to extract the advanced features of the input signal. The results are finally output through a fully connected classifier. The model fuses multisource data at the feature level, employs the temporal and spatial relationships of monitoring data, and dynamically evaluates the risk. The experiment shows that the proposed model achieves impressive performance in both quantitative and qualitative evaluations and has improved the model generalization ability. The combination of data integration, remote examinations, and approval from existing information systems enables this platform to provide dynamic reminders of approval information, various risk prewarning, and management process automation through a mobile network.

1. Introduction

As coal demand and total mining volume rise, most mines have entered the stage of deep coal mining [1]. Deep mining excavation is more difficult and has more distinctive problems and potential safety hazards than shallow mining. The risk factors of the environment always affect the safe production of coal mines. Therefore, real-time monitoring and effective prediction in the process of coal mining have become major issues to be resolved for the healthy development of the coal industry [2, 3].

The risk estimation is normally completed through the quantitative and qualitative assessment of on-site environment, equipment and personnel operation factors by technical experts with data analysis, checklist designing and risk matrix forming, and so on. Nowadays, many literature has put forward new methods for risk assessment. Some experts and scholars use machine learning (ML) methods to predict and evaluate the risk of coal mines. However, because data processing and feature definition and selection are cumbersome, the cost of ML increases, which prevents its widespread application. In recent years, the rapid development of deep learning (DL) theory and technology has avoided the problem of relying on artificial feature engineering that afflicts traditional ML algorithms and thus has attracted extensive attention. DL can fit very complex nonlinear functions through a deep network structure and can automatically extract features.

Another harsh reality is that underground staff cannot get timely information regarding the safety of the surrounding environment, thus preventing them from predicting potential danger and evaluating their safety in real time. Every mineworker must be kept informed of any possible risk in an adequate and timely manner so that they can precisely evaluate their safety.

The powerful mobility and capability of intelligent mobile terminals ensure staff can master the overall state at the same time, issue strategies in time, and solve problems quickly. The combination of mobile terminals and cloud service technology can cover the risk prewarning service to every staff member, which is meaningful to the coal mine intelligence.

To solve the bottleneck of risk prediction and evaluation in coal mines and provide risk safety assessment information for underground personnel, this paper constructs an intelligent mobile mine platform and establishes a coal mine risk prewarning model based on a deep convolutional model.

Therefore, developing an information early warning system that can accurately evaluate the risk level of coal mine safety hazards in time and feedback the corresponding risk level countermeasures to supervisors in time is a key problem to be resolved urgently.

The rest of this paper is organized as follows. Section 2.1 gives an overview of research status at home and abroad. Section 2.2 then describes the problems. Section 2.3 introduces the points of contribution. Section 3 presents the design of an intelligent mobile mine system infrastructure. Section 4 provides a coal mine safety prewarning combination model. Section 5 illustrates the main function of the platform. Next, Section 6 summarizes key technologies and features. Conclusion is finally drawn in Section 7.

2. Theoretical Background and Motivation

2.1. Research Status at Home and Abroad. In an effort to improve mineral resource development, preserve the ecological environment, and avoid geological disasters, many scholars have proposed a series of solutions. Most of these solutions for the prewarning of typical dynamic disasters in coal mines involve high-level technologies such as ML.

Deng et al. [4] proposed a prediction model of spontaneous coal combustion in coal mine goafs based on a random forest algorithm. Qiu [5] used a deep belief network to extract the features of gas concentration time series and established a particle swarm optimization-support vector regression (PSO-SVR) model for gas concentration temporal and spatial modeling and trend prediction. To accurately predict rockburst disasters, Tian et al. and Gong et al. [6, 7] proposed a domain-aware deep neural network (DA-DNN) model based on improved adaptive moment estimation (Adam) and dropout. To construct the gas concentration distribution map, Cheng et al. [8] combined long short-term memory (LSTM) and a fully connected (FC) neural network (NN) to construct an LSTM-FC model to predict gas concentrations at different locations.

Tong and Cui [9] constructed a convolutional neural network (CNN) to classify and identify the underground environment, equipment status, and personnel behavior in video and audio to reduce the occurrence of coal mine risk disasters and accidents. Geng [10] proposed a hybrid model based on a dilated causal temporal convolutional network (DCT-CNN) and a long short-term memory recurrent neural network combined with a convolution network (CNN-LSTM) to predict the risk of local points on the roof.

Su et al. [11] built a safety risk assessment model for county coal mining areas with analytic network process (ANP) and probabilistic neural network. Built on Internet, cloud platform, smart mine, Internet of Things, geographic information system (GIS), and other technologies, Li [12] designed a platform for risk grading control and accident hidden danger investigation and control in coal mine. To improve the prediction accuracy of chaotic time series, Huang et al. [13] proposed a prediction model based on hybrid neural network and attention mechanism. Wang et al. [14] studied an intelligent hierarchical management and control and information early warning system for coal mine safety risks based on improved particle swarm optimization (PSO) and CNN. Yang et al. [15] proposed a new method for fault diagnosis of rolling bearings based on CNN deep learning based on attention mechanism. Wang [16] studied the classification and management strategy of coal mine safety risk and uses PSO to improve the error back propagation of CNN by utilizing the training parameters and error functions of CNN as PSO particles and fitness functions, respectively.

Liu [17] built a risk precontrol management system for mine safety that uses hazard identification and risk assessment as its basis, risk precontrol as its core, and unsafe behavior control as its focus. Li et al. [18] showed that the combination of the fuzzy analytic hierarchy process (AHP) and Bayesian network is feasible and applicable and can be used as a decision-making tool to prevent coal mine gas explosions and provide decision-makers with a technical guide for managing coal mine gas explosion risks. Li et al. calculated the weights from experts using a fuzzy AHP method based on subjective and objective expert weights [19]. Analysts and decision-makers can use the proposed batch normalization (BN) model in a coal mine as a decision support tool. Wu et al. built a mine platform with a risk

control function for coal mines [20]. Some researchers [21–23] utilized ML to study prewarning scenarios of coal mine pressure and gas and verified the effectiveness of ML. Mathatho et al. proposed a hierarchical approach consisting of principal component analysis (PCA) and an artificial neural network (ANN) model to improve the prediction accuracy of methane levels [24].

The above literature review shows that ML can mine the effective information contained in monitoring data and has made some progress in the mining field. These works provide ideas for research on risk prewarning.

Another key aspect of mining engineering for underground staff is transmitting environmental information with location and assessment data in real time. Various types of monitoring information are sent to a security model, the output of which is risk prediction. These outputs with geographic information are sent to specific areas for broadcasting to wireless nodes. Zheng et al. [25] built a client/server-(C/S-) based visual danger source risk precontrol mobile phone management information system. Zhai et al. [26] described the network structure, software architecture, software design, system deployment ,and system testing of Lu'an group's mobile monitoring information platform. Tan and Li [27] introduced the architecture and functional module design of a visual mobile office platform for coal mines. The mobile office employed a mobile internet and the Internet of Things (IoT). Chen [28] proposed the idea and scheme for the construction of a mobile application system platform for coal mines.

The above literature shows that research on ML is of great significance to improve risk prediction capabilities.

2.2. Problems. The above studies have achieved a certain amount of prediction progress from different perspectives, which have a great push to this problem. Although these technologies such as BP neural networks can achieve satisfactory accuracy, they still rely heavily on expert knowledge to extract features. The BP neural network algorithm with defects of local minimization and slow convergence speed makes it difficult to accurately and efficiently evaluate the safety risk of a coal mine. However, due to the complexity of the risk mechanism and the diversity of various factors, the following deficiencies still exist in practice:

(1) Most methods determine the weight of each index. The determination of the weight by the staff is subjective; the weight is sometimes set by the experience of staff

(2) Risk prewarning is complex and nonlinear. The risk is the result of multifactor. Some of these analytic features are definite and quantitative, while others are qualitative and fuzzy. The quantitative model of risk assessment is difficult to describe accurately by mathematical formulas and theories, which are multivariable, coupled, and highly complex

(3) The relationship between temporal and spatial multivariable analyses has not been studied enough

(4) The existing systems have a strong focus on the functions of the central node and neglect network terminal personnel. Although much comprehensive information is already available on the central nodes, information sharing and risk precontrol tools are seldom available to mining staff

All of these factors place higher requirements on the predictive model, communication, and terminals, and it is still necessary to explore new prediction methods. Therefore, we design a mobile intelligent mine platform and propose a new model with artificial intelligence for risk prewarning.

2.3. Points of Contribution. At present, integrated automation systems (IASs) have been built in most coal mines. Making use of IASs has enabled the monitoring and control of the important parts of underground equipment and systems [29].

An intelligent analysis of data is lacking in traditional coal mine automation and information systems. A dispatcher can only "transmit messages;" however, these messages are difficult to transmit to all mineworkers in real time.

To resolve the aforementioned issues and meet the practical requirements of underground security and the mining system, a mobile smart mine is constructed based on the comprehensive automation technology of coal mines, online data detection, computers, mobile networks, and mining-specific technologies. We propose a new prewarning model and integrate mobile security monitoring systems, geographic information systems (GIS), risk prewarning, and mobile communication into mobile smart mines and embed their functions into mobile apps. These designs are designed to play essential roles in miner safety assurance, production scheduling, and disaster relief.

The research innovation points are the following:

(1) A novel mobile intelligent mine platform is proposed for risk prewarning with a new combination model. Mobile internet technology concerns positioning, security monitoring, and risk evaluation. With smart terminals, underground staff can get the latest risk assessment information

(2) A novel model is proposed to construct a multivariable prewarning model. The NN prewarning model is always running and works in conjunction with the CNN model at a specific time for mutual confirmation

(3) ATT is incorporated into CNN to improve prediction accuracy

(4) A hazard prewarning model based on an attention mechanism–convolutional neural network (ATT-CNN) is proposed to deeply mine the behaviors or characteristics of the mining environment. It not only handles the information within all sequential data but also preserves the sequential spatiality. The model solves the problem of multisource information fusion and employs the temporal and spatial relationships of each monitoring variable

In short, by integrating IoT technology, mobile internet technology, and DL theories, we construct a comprehensive intelligent mobile mine platform that aims to provide risk prewarning that significantly improve the risk control efficiency of coal enterprises.

3. Materials and Methods in the Design of a Mobile Intelligent Mine System Infrastructure

This paper proposes an intelligent mine information system that senses every person, all equipment, and each aspect of the surroundings. The ultimate goal of this system is to allow staff members to utilize up-to-date technology and to run a coal mine with scientific and comprehensive management. Figure 1 shows the logic block diagram of the system. In this system, data related to all equipment and environmental information are sent to the intelligent mobile platform. After information fusion and risk assessment, the assessment results and relevant details are transferred to the frontline staff and supervisor through the mobile communication network as soon as possible in text, curves, and even multimedia.

A background application service center creates a seamless connection between the existing monitoring system and the intelligent mobile terminal in a coal mine. We call this center the network management platform of the system. The platform completes the instant messaging between the coal terminal and the mobile terminal simultaneously. The platform receives real-time data from the monitoring system and stores them in a database of background. The intelligent mobile terminal accesses the internet through a cellular network or IEEE 802.11 wireless networking technology (Wi-Fi). Then, the platform database is available to mobile applications. The firewall in the system is used to realize security isolation between the industrial control network and the existing information network.

The mobile system has five major functions:

(1) System integration: all coal mine subsystem data integration, functional integration, and interconnection services

(2) Multilevel networking: remote, real-time, and multilevel networking of coal mine subsystems and intuitive reflection of mine real-time status

(3) Safety information management: basic mine information, safety production regulations, safety facilities and equipment, mine plans, and staff safety information [30]

(4) Security monitoring mobile office: incident management, regulatory bulletin and notification, safety production statements, data statistics, and query functions

(5) Wireless remote monitoring of the mine: mine managers, technicians, and crew access and control comprehensive mine information using a fourth generation (4G) mobile network

Under the special conditions of mobile application scenarios in coal mines, the mobile information platform is composed mainly of 3 modules, as shown in Figure 2.

4. Coal Mine Safety Prewarning Combination Model

4.1. Coal Mine Risk Prewarning Management. In our design, the coal mine risk precontrol management system includes hazard identification, hazard risk assessment, formulation and implementation of management standards and measures, and feedback of inspection. The task analysis method works for hazard identification, and the risk matrix method works for risk assessment. Depending on the assessment results, the system classifies the risks into different levels. Each identified hazard is ranked by a corresponding risk level, management standard measures, and supervision measures for product safety. The whole process is shown in Figure 3.

4.2. Construction of the Coal Mine Safety Prewarning Combination Model. The basic idea of the model and application is shown in Figure 4. The prewarning period T of the network model is shorter than the period L of the ATT-CNN model. This guarantees that the technical staff has sufficient time to determine the accurate weights and membership functions for the overall validity and reliability of the model. An NN dominates the prewarning process, and the ATT-CNN provides additional verification for the timely and effective prewarning of the coal mine safety status. According to the equipment data acquisition period and quantity, the prewarning period T is generally set as 5 minutes, 10 minutes, or 15 minutes by the staff.

The NN prewarning model works with the ATT-CNN during a specific time for mutual confirmation. If the two kinds of evaluation results are the same, the network model is valid and can continue to be applied. Otherwise, the weights of the output results of the two models are set so that the prewarning result value is adjusted, the original sample is expanded by adding the adjusted data, and the network training is carried out again to obtain a new network model structure. In a changing environment, the new network structure learns "new knowledge," which improves the reliability of the prewarning results in subsequent stages. The retrained NN is used as a model in the follow-up cycle until it is compared with the ATT-CNN of the new stage again and adjusted.

The compensation fuzzy NN model is used for daily prewarning, and according to its calculated results, the corresponding prewarning is issued. A combination model composed of compensation, a fuzzy NN, and the ATT-CNN is applied at a certain time. The final warning result is determined according to whether the two warning results are consistent. If the two results are the same, the prewarning is issued according to the result; otherwise, the prewarning is adjusted according to their respective weights. At the same time, the training samples are extended and fed to the NN again for training to obtain a new network model for the following warning cycle.

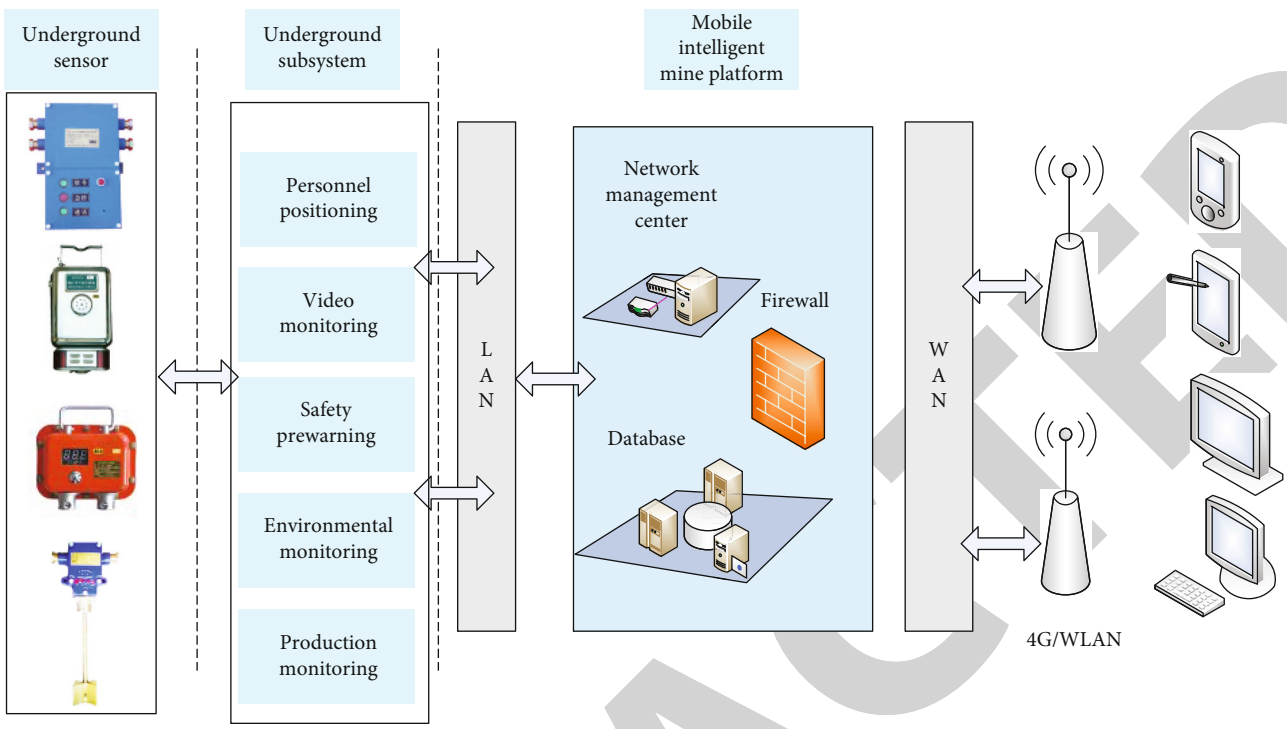


Figure 1: Mobile intelligent mine system infrastructure.

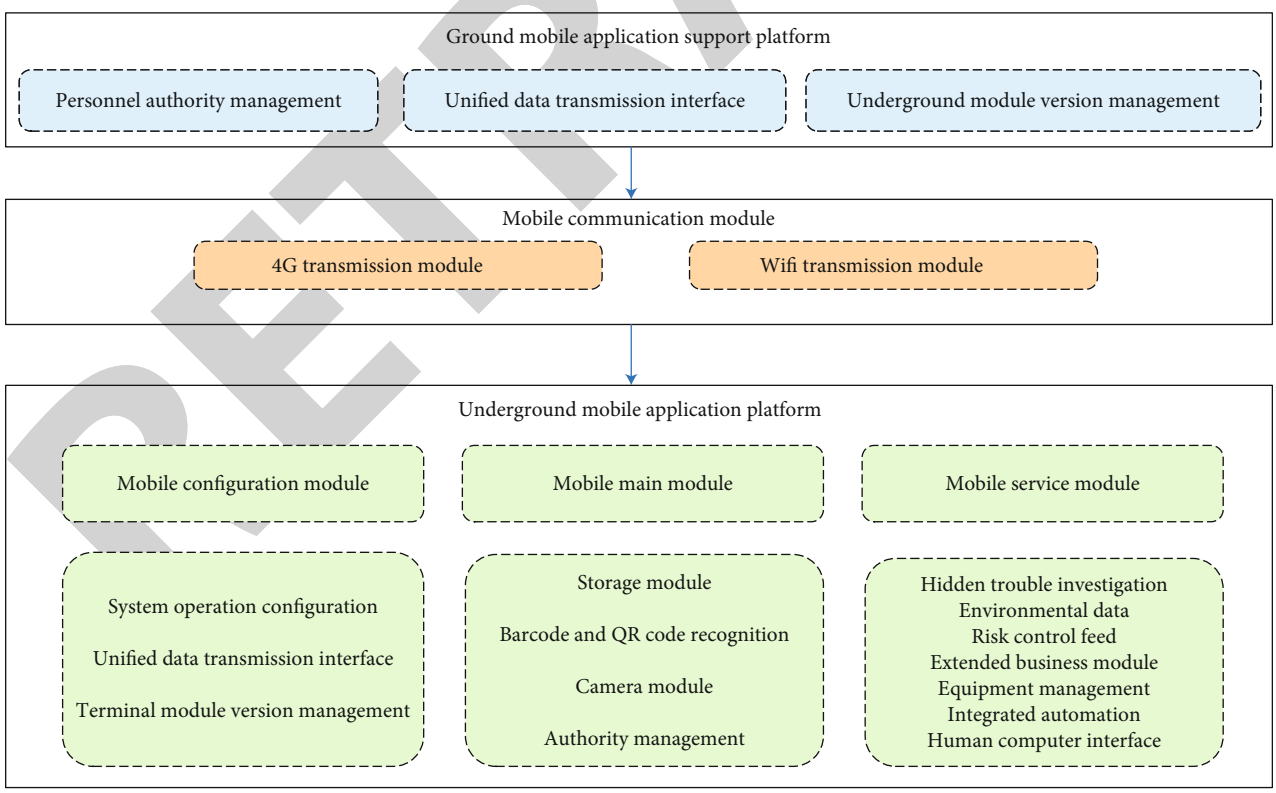


Figure 2: Mobile application system platform for risk prewarning. (1) Ground mobile application support platform: data support and configuration management for the regular operation of underground mobile applications are achieved through a support platform. (2) Mobile communication module: this is mainly for the underground wireless signals. (3) Underground mobile application platform: it includes a mobile configuration module, a main mobile module, and a mobile service module.

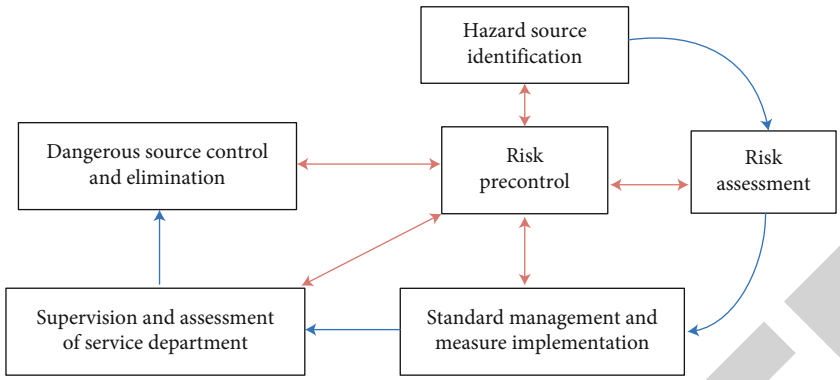


Figure 3: Diagram of risk prewarning management.

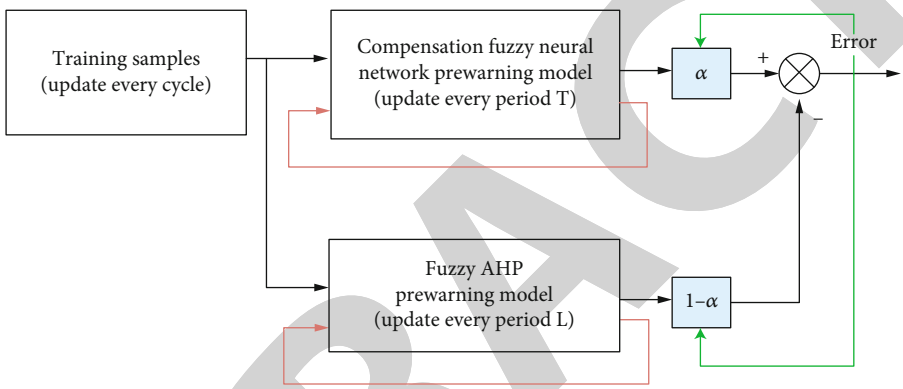


Figure 4: Basic idea of the model and application.

After the prewarning combination model is determined, the prewarning period of the NN model can be selected as days or weeks, and the period of the ATT-CNN model can be selected as weeks, months, or longer to ensure that sufficient time is available for accurate weights and membership functions to be determined, thereby ensuring the overall effectiveness and reliability of the method. The combination model is more valuable in the process of mutual confirmation between the two models for prewarning.

If the values of the monitoring indicators are $(x_1\, x_2 \cdots x_n)$, the prewarning values y_1 and y_2 are obtained when the monitoring indicators are substituted into the network model and the ATT-CNN model. The prewarning criteria already determine the corresponding alarm level. If the prewarning values y_1 and y_2 are the same, the network model is applied and regarded as effective. If the values are different, the calculation formula of the prewarning value is adjusted. Assuming that the weight of the network model result is α, the weight of the ATT-CNN model is 1-α. The adjusted prewarning value is $y* = \alpha y_1 + (1-\alpha)y_2$, where the value of α is [0, 1]. Assuming that the network model used in the above calculation was trained on a particular number of samples m, the m samples are expanded to $m+p$ samples, where $(x_{m+p1}\, x_{m+p2} \cdots x_{m+pn})$ denotes the n monitoring indicators, y_{m+p} is the adjusted prewarning value of $y*$, and the expanded $m+p$ samples are recorded as follows:

$$\begin{bmatrix} x_{11} & x_{11} & \cdots & x_{11} & y_1 \\ x_{21} & x_{22} & \cdots & x_{2n} & y_2 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} & y_m \\ x_{m+p1} & x_{m+p2} & \cdots & x_{m+pn} & y_{m+p} \end{bmatrix}. \quad (1)$$

A new NN structure is built after network retraining is carried out on the $m+p$ samples. The new structure is not adjusted until it is compared with the ATT-CNN model at a later stage. Through the above processing, the two combined models provide a solution for the variable weight problem. Different model parameters are adopted at various stages to adapt to the dynamic characteristics of the system. The two models confirm each other simultaneously to ensure the accuracy and effectiveness of the model and improve the reliability and credibility of the warning results.

A review of the risk evaluation of the factors impacting training samples shows that the physical meanings are different, and the numerical values fluctuate considerably. Normalization should be performed before predicting risks. To eliminate the order-of-magnitude differences and avoid

excessive training errors, the method used in this study is the normalization of the data by the maximum and minimum as follows:

$$x_k' = \frac{x_k - x_{\min}}{x_{\max} - x_{\min}}, \tag{2}$$

where $x_{\min}$ is the minimum number in the data series, and $x_{\max}$ is the maximum number in the series.

The compensated fuzzy NN is a multiple-input single-output (MISO) system that consists of a single-valued fuzzy generator, a Gaussian membership function, a product inference rule, a negative-positive compensation operation, and an improved center-of-gravity antifuzzifier. Let A_i^k be the fuzzy set in the domain U, B^k be the fuzzy set in domain V, and x_i and y be linguistic variables, with $i = 1, 2 \cdots, n$; $k = 1, 2... m$. The fuzzy membership function is as follows:

$$\begin{aligned} u_{A_i^k}(x) &= \exp\left[-\left(\frac{x_i - a_i}{\alpha_i^k}\right)^2\right], \\ u_{B^k}(y) &= \exp\left[-\left(\frac{y - b}{\beta^k}\right)^2\right], \end{aligned} \tag{3}$$

where a and α are the center and width of the input membership function, respectively, b and β are the center and width of the output membership function, respectively, $x_1, x_2, \cdots, x_n$ is defined as the input X, and $U_1 \times U_2 \times \cdots \times U_n$ is defined as the domain U. For a fuzzy subset A' in domain U, an output fuzzy subset B' is generated in the output domain V according to the k-th fuzzy rule. The maximum algebraic product synthesis operation is used for the fuzzy reasoning, and then the fuzzy set B' on V is derived from the fuzzy inference rules as follows:

$$\mu_{B'k}(y) = \sup_{X \in U}\left(\mu_{A_1^k \times A_2^k \times \cdots \times A_n^k \longrightarrow B^k(X,y)} \cdot \mu_{A'}(X)\right). \tag{4}$$

The product operation is adopted for the fuzzy implication as follows:

$$\mu_{A \longrightarrow B} = \mu_A(x)\mu_B(x). \tag{5}$$

The negative operation is as follows:

$$u^k = \prod_{i=1}^{n} \mu A_i^k(x_i). \tag{6}$$

The active operation is as follows:

$$v^k = \left[\prod_{i=1}^{n} \mu A_i^k(x_i)\right]^{1/n}. \tag{7}$$

The compensation operation is as follows:

$$\mu_{A_1^k \times \cdots \times A_n^k}(X) = \left(u^k\right)^{1-\gamma}\left(v^k\right)^{\gamma} = \left[\prod_{i=1}^{n} \mu A_i^k(x_i)\right]^{1-\gamma+\gamma/n} \tag{8}$$

where γ is the compensating degree, with $\gamma \in [0, 1]$; hence, the following is obtained:

$$\mu_{B'k}(y) = \sup_{X \in U}\left\{\mu_{B^k}(y)\mu_{A'}(X)\left[\prod_{i=1}^{n} \mu A_i^k(x_i)\right]^{1-\gamma+\gamma/n}\right\}. \tag{9}$$

Single-value fuzzification is adopted, with $\mu_{A'}(X) = 1$, $\mu_{B^k}(b^k) = 1$; then, the following is obtained:

$$\mu_{B'k}\left(b^k\right) = \left[\prod_{i=1}^{n} \mu A_i^k(x_i)\right]^{1-\gamma+\gamma/n} \tag{10}$$

where $f(X)$ is defined as the ant-gelatinization function as follows:

$$f(X) = \frac{\sum_{k=1}^{m} b^k \delta^k \left[\prod_{i=1}^{n} \mu A_i^k(x_i)\right]^{1-\gamma+\gamma/n}}{\sum_{k=1}^{m} \delta^k \left[\prod_{i=1}^{n} \mu A_i^k(x_i)\right]^{1-\gamma+\gamma/n}}. \tag{11}$$

The objective function is as follows:

$$E^p = \frac{1}{2}\left[f(x^p) - y^p\right]^2. \tag{12}$$

The b center of the output membership function is trained according to the gradient descent method, as are β, a, and α. The training formulas are as follows:

$$\begin{aligned} b^k(t+1) &= b^k(t) - \eta \left.\frac{\partial E^p}{\partial b^k}\right|_t, \\ \beta^k(t+1) &= \beta^k(t) - \eta \left.\frac{\partial E^p}{\partial \beta^k}\right|_t, \\ a^k(t+1) &= a^k(t) - \eta \left.\frac{\partial E^p}{\partial a^k}\right|_t, \\ \alpha^k(t+1) &= \alpha^k(t) - \eta \left.\frac{\partial E^p}{\partial \alpha^k}\right|_t. \end{aligned} \tag{13}$$

Let $r = c^2/c^2 + d^2, r \in [0, 1]$; then, the following is

obtained:

$$c(t+1) = c(t) - \eta\left\{\frac{2c(t)d^2(t)}{c^2(t)+d^2(t)}\right\}\frac{\partial E^p}{\partial r}\bigg|_t,$$
$$d(t+1) = d(t) - \eta\left\{\frac{2d(t)c^2(t)}{c^2(t)+d^2(t)}\right\}\frac{\partial E^p}{\partial r}\bigg|_t, \quad (14)$$
$$r(t+1) = \frac{c^2(t+1)}{c^2(t+1)+d^2(t+1)},$$

where η is the learning rate and $t = 0, 1, 2, \cdots$.

4.3. ATT-CNN Algorithm. The simulation results of some researchers [31, 32] show that the ATT-CNN model is better in prediction accuracy and performance when tested on several chaotic time series. Inspired by that work, we propose an ATT-CNN algorithm for risk prediction.

Considering the massive amount of data involved, the computational complexity of traditional FC NNs is unfeasible. The ATT-CNN allows its input layer to input any dimension for adaptive topology. Its local connection and weight-sharing characteristics reduce the network-free parameters and speed up the calculations. We can thoroughly explore the risk status information contained in the original signal using the ATT-CNN for end-to-end automatic learning, extract features without expert experience, and obtain a better nonlinear relationship between monitoring data and risk prewarning.

The translation and scaling of the input signal are highly invariant in time and space in the ATT-CNN, which is also not sensitive to noise. These features are theoretically suitable for data processing under the complex conditions of coal mines.

Furthermore, in the ATT, the weight of each feature is calculated, and then each component is weighted and summed. The larger the weight is, the more significant the contribution of the element to the current recognition. The essence of the ATT can be expressed as a mapping of a query (Query) to a series of key-value pairs (Key-Value). The idea of the ATT is shown in Figure 5.

In the ATT, the constituent elements in the source are regarded as a series of <Key, Value> data pairs. Given an element Query in the target, the weight coefficient of Value is obtained for each *Key* by calculating the similarity or correlation between Query and each Key. Then, Value is weighted and summed to obtain the final attention value.

Therefore, the ATT is used to perform a weight sum of *Value* in Source, and Query and Key are used to calculate the corresponding weight coefficients of Value. The ATT is calculated as follows:

$$\text{Attention}(\text{Query}, \text{Source}) = \sum_{i=1}^{Lx} \text{Similarity}(\text{Query}, \text{Key}_i) * \text{Value}_i, \quad (15)$$

where $Lx = \|\text{Source}\|$ is the length of the source; the meaning of the formula is as described above.

The process focuses on the calculation of the weight coefficient. The larger the weight is, the greater the focus on the corresponding Value; i.e., the weight represents the importance of information, and Value is the corresponding information.

Based on the CNN algorithm, the ATT-CNN is generated by adding the ATT to enhance the invariance of the CNN in time and space and to further extract features. That is, the attention layer is added between the input layer and the convolutional layer in the CNN. The ATT calculates the attention weight a_j of the input feature Z_j to weight Z_j. The weighted feature Z_j' replaces the original Z_j as the input of the next layer. After progressing through the pooling layer, the FC layer, and the output layer, the final classification model is obtained.

The algorithm structure diagram based on the ATT-CNN is shown in Figure 6. The ATT is introduced in the convolutional layer to learn a weight distribution that is used to weight the original output of the convolutional layer.

By performing feature aggregation on the feature maps generated by the convolutional layer, each feature map is turned into a feature vector that can be considered a global receptive field. The weight vector a_j learned by the ATT is matched and multiplied by the feature map of the convolutional layer. Finally, the new feature Z_j' learned under the ATT network is obtained. The BN layer reduces the impact of changes in the middle layer data distribution. The activation layer enhances the nonlinearity of the model. The rectified linear unit (ReLU) accelerates the convergence of the CNN as an activation function. When the parameters are adjusted by the backpropagation learning method, the ReLU makes the weights of the shallow layer easier to train. Max pooling has the advantage of obtaining features that are unrelated to location information. The FC layer is usually used with softmax regression to classify the convolutional layer features and the pooling layer.

These operations together constitute a feature extraction layer. By stacking multiple feature extraction layers, feature extraction is realized layer by layer, and classification is realized in the last layer of the FC layer and softmax layer.

The crossentropy function is often used in combination with softmax in ML. We take the crossentropy function as the objective function. After using the error backpropagation algorithm to calculate the derivative of the objective function relative to the NN parameters, the optimization process employs the Adam algorithm.

5. Platform System Main Function

5.1. Production Scheduling and Information Management System. The production scheduling and information management system includes the system management module, the production scheduling information module, the query and report module, and other functional modules.

System management module is as follows: this module enables management of the department, account and role,

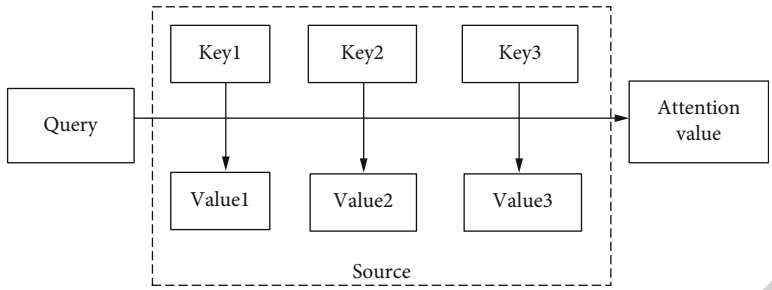


Figure 5: Idea of the attention mechanism.

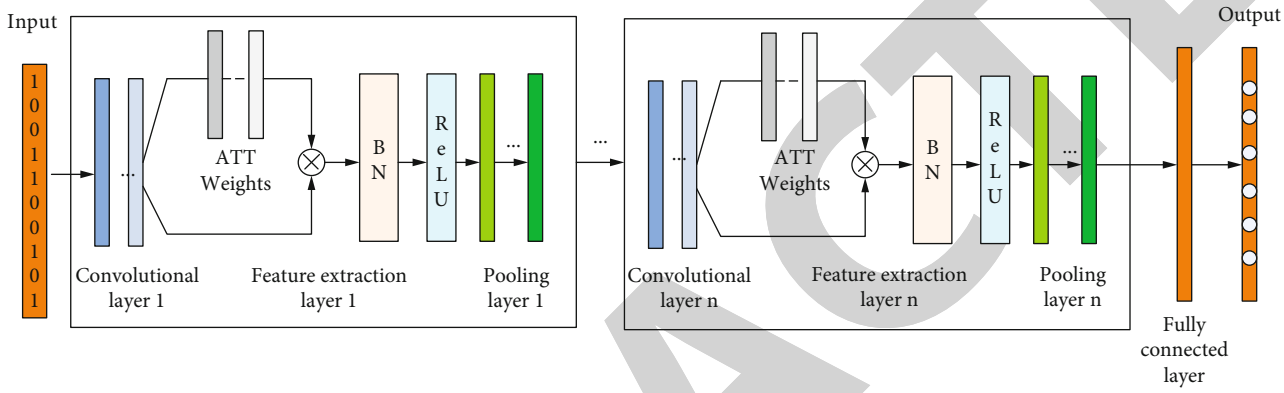


Figure 6: Algorithm structure based on the ATT-CNN.

password maintenance, landing inquiry, and management of the common tools and menu.

(1) Production scheduling and information module: this module provides the completion of much information management, such as the monthly manufacturing program, daily drifting footage, daily drilling footage, daily water inflow, daily production information, gas monitoring, a list of leaders on duty, the weekly production schedule report, the address book, the driving moraine frontline, drill hole location information, and all information maintenance

(2) Report view: this view provides daily, weekly, and monthly production reports, monthly manufacturing program data, tunneling data, drilling data, water inflow data, and the report from the leader on duty

(3) Report upload: mine flood data, work accident data, and monthly reports are uploaded

5.2. Data Acquisition System. The functions of real-time data collection, storage, statistics, and analysis are available for every subsystem, including risk monitoring, underground substations, domestic water supply, online monitoring of fans, gas drainage, the water supply of the air shaft, personnel positioning, and safety information management [33].

5.3. Mine Mobile Information Platform. The mobile information platform supports multiple functions, e.g., real-time parameter monitoring, alarm recording, and trend curve query functions. It can be used in 12 systems, including mine risk monitoring, personnel positioning, safety information management, production reports, address books, belt transportation, ventilation monitoring, wind well water supply, domestic water supply, gas drainage, and central substations.

6. Key Technologies and Features

6.1. Open Crossplatform Web Services. As a software component, web services are also a popular new technology that serves other applications by open standards for web protocols and data formats (such as HyperText Transfer Protocol (HTTP), Extensible Markup Language (XML), and Simple Object Access Protocol (SOAP)). It can be used as a service element to construct a distributed architecture system with advantages such as decentralized architecture dynamic integration, burden balance, and cell upgrades [34].

Open crossplatform web services can readily integrate different applications developed from different programming languages and run on different operating systems by adopting open standards.

In our design, by depending on web services and integrating the mobile 4G network and PC network, we can schedule work management in optional modes, which can significantly reduce an application's geographical and temporal constraints.

6.2. Visual Analysis Method. The system adopts advanced graphical analysis tools and strategies [35]. With the use of animation, the system changes the traditional original lengthy, complicated summary, and statistical data into

intuitive, clear pictures with animation. Under the visual data conditions, we can apprehend the statistical data by only a glance owing to the sharp contrast.

6.3. Advanced Concept of Ajax Technology. The system has been implemented using asynchronous JavaScript and XML (Ajax). Ajax was chosen because it offers a more definitive and better experience. Ajax is not a single technology but a group of technologies. HyperText Markup Language (HTML) and Cascading Style Sheets (CSS) can be used in combination to mark up and style information.

Ajax is a web development technique that uses many web technologies on the client side to create asynchronous web applications. With Ajax, web applications can communicate with a server asynchronously (in the background) without interfering with the display and behavior of the existing page. By decoupling the data interchange layer from the presentation layer, Ajax allows extension web applications to change content dynamically without reloading the entire page. The following technologies are incorporated into Ajax:

(1) HTML (or Extensible HyperText Markup Language (XHTML)) and CSS for presentation

(2) The document object model (DOM) for the dynamic display of and interaction with data

(3) JavaScript object notation (JSON) or XML for data interchange and Extensible Stylesheet Language Transformations (XSLT) for data manipulation

(4) The XML Http Request object for asynchronous communication

(5) JavaScript to bring these technologies together

XML is no longer required for data interchange; therefore, XSLT is no longer required to manipulate data. JSON is often used as an alternative format for data interchange, along with other formats such as preformatted HTML or plain text.

Asynchronous HTML and HTTP (AHAH) involves using the XML Http Request to retrieve (X) HTML fragments, which are then inserted directly into the web page.

6.4. Examples of Hazard Prediction. We trained our models on one machine with a RX 5500 XT 8GB graphics processing unit (GPU). This model uses Keras 2.3, a DL framework based on Python 3.7.3 (with TensorFlow 2.1.0 as the back end), to quickly build a deep hybrid NN. The ATT-CNN parameters are shown in Table 1.

The strip and padding of the convolution layer are set to 1, the learning rate is 1*e*-3, and the batch size is 256. The size of the convolution kernel is 3×3, and the maximum number of network iterations is 1000. The dropout layer is not employed in the experiments since the data came from sensors.

Coal and gas outbursts (CGOs) are the result of the common effects of ground stress, coal seam gas, and the physical and mechanical properties of coal structures. The main factors affecting gas outbursts are uncertain. The prediction of CGO risk can be regarded as the prediction of an ambiguous event determined by multiple factors. In the following, this risk prediction process is explained.

Combining the analysis of the risk factors for CGOs and the ranking of the importance of outburst influence by the triangular fuzzy comprehensive evaluation method, the input variables of the NN prediction model are the following:

(1) Initial gas release velocity

(2) Index of coal destruction

(3) Gas pressure

(4) Index of the geological structure

(5) Index of the coal firmness coefficient

(6) Permeability coefficient

(7) Gas content

(8) Attenuation coefficient of the drilling gas emission volume

(9) Depth of mining

The output vector is the characteristic of the CGO. The coal seams or areas are divided into 4 categories according to the degree of danger by combining relevant theories and knowledge of CGOs, detailed rules for the prevention and control of CGOs, and an evaluation index system for the risk of CGOs. The output vector contains the four types of coal seam outburst hazards, represented by four nodes, where 1000 means serious danger, 0100 means greater danger, 0010 means average risk, and 0001 means small risk.

The number of input layers n in the NN is 9 according to the triangular fuzzy method, the output layer l is 4 by the classification of the prominent danger, and the number of nodes of the hidden layer neurons is set to 17 from the empirical formula. Therefore, (9, 17, 4) denotes the structure of the fuzzy NN. The sample data are shown in Table 1.

The data are obtained from the database copy of monitoring system of Xuyong Coal Mine in Sichuan Province, where the gas content increases with the coal depth. The relation between gas content and coal seam depth is not a simple linear relation such that the gradient of content is large in shallow areas and small in deep areas. The relative gas emission is $20.1\,m^3/t$. The variation gradient of the gas content is generally between 3 and $8\,m^3/t{\cdot}daf/100\,m$. The permeability coefficients of coal seams are different in spatial distribution due to the other occurrence conditions. There are abnormal features in local areas such that the permeability coefficient shows apparent characteristics of the partition and subband.

For colliery gas, the gas concentration is related to not only its own historical data but also the influence of other factors on the coal seam, such as the influence of the underground water volume, the tunneling speed in the process of underground coal mining, and the grade of the mine

Table 1: The detailed settings of the ATT-CNN parameters.

No.	Item	Kernel size	Amount	Output size	Activation function
1	Input layer				
2	CNN layer 1	3×3	48	$8 \times 8 \times 48$	ReLU
3	Pooling layer 1	3×3	1	$6 \times 6 \times 48$	
4	ATT layer 1				
5	CNN layer 2	3×3	128	$6 \times 6 \times 128$	ReLU
6	Pooling layer 2	3×3	1	$4 \times 4 \times 128$	
7	ATT layer 2				
8	CNN layer 3	3×3	384	$4 \times 4 \times 384$	ReLU
9	CNN layer 4	3×3	256	$4 \times 4 \times 256$	ReLU
10	Pooling layer 4	3×3	64	$2 \times 2 \times 256$	
11	FC	1×1024	1	4×1	ReLU
12	Output layer			4×1	Softmax

Table 2: Data of the coal seam gas outburst risk sample.

No.	b1	b2	b3	b4	b5	b6	b7	b8	b9	b10	b11
a1	c19	18.24	0.5	0.65	0.5	0.33	0.73	9.814	0.657	139	0100
a2	c19	18.24	0.5	1.75	0.8	0.33	0.73	13.909	0.657	135	0100
a3	c19	10.17	0.5	1.2	0.8	0.54	0.54	9.591	1.09	182	0010
a4	c19	10.17	0.5	0.85	0.5	0.54	0.54	9.943	1.09	160	0010
a5	c19	20.34	0.6	1.2	0.9	0.24	0.24	13.305	1.535	341	1000
a6	c19	20.34	0.6	1.55	0.5	0.24	0.24	14.982	1.535	212	1000
a7	c19	20.34	0.6	0.84	0.9	0.24	0.24	11.278	1.535	232	1000
a8	c19	20.34	0.6	1.2	0.9	0.19	0.19	13.303	1.535	342	1000
a9	c19	24.01	0.5	0.48	0.9	0.21	0.21	8.299	0.193	200	1000
a10	c24	12.54	0.5	0.85	0.5	0.55	0.55	10.121	0.193	290	0100
a11	c24	12.54	0.5	0.7	0.9	0.55	0.55	9.059	0.588	292	0001
a12	c24	10.17	0.5	1.6	0.5	0.34	0.34	16.326	0.588	316	0001
a13	c24	10.17	0.5	1.65	0.9	0.34	0.34	16.54	0.588	288	0001

Table 3: Results of the CGO risk prediction based on the neural network model network mode.

No.	v1	v2	v3	v4	Output	Predicted level of danger
a1	0.0038	0.9764	0.0012	-0.0026	0100	Medium
a2	0.0034	1.0084	0.0107	-0.0231	0100	Medium
a3	-0.018	-0.026	0.9935	0.0985	0010	General
a4	0.0186	0.0429	0.9788	-0.075	0010	General
a5	1.0213	-0.085	0.0386	-0.0703	1000	Serious
a6	0.9901	-0.006	-0.005	0.017	1000	Serious
a7	1.0058	0.0254	0.011	-0.0002	1000	Serious
a8	0.9841	0.0646	-0.0438	0.0498	1000	Serious
a9	-0.003	1.0042	-0.0083	0.0258	0100	Medium
a10	-0.013	0.9877	0.0115	0.0238	0100	Medium
a11	0.0075	0.0035	0.0103	0.9341	0001	Smaller
a12	0.0029	0.0011	0.004	1.0321	0001	Smaller
a13	-4*E*-04	0.0036	-0.0033	0.9884	0001	Smaller

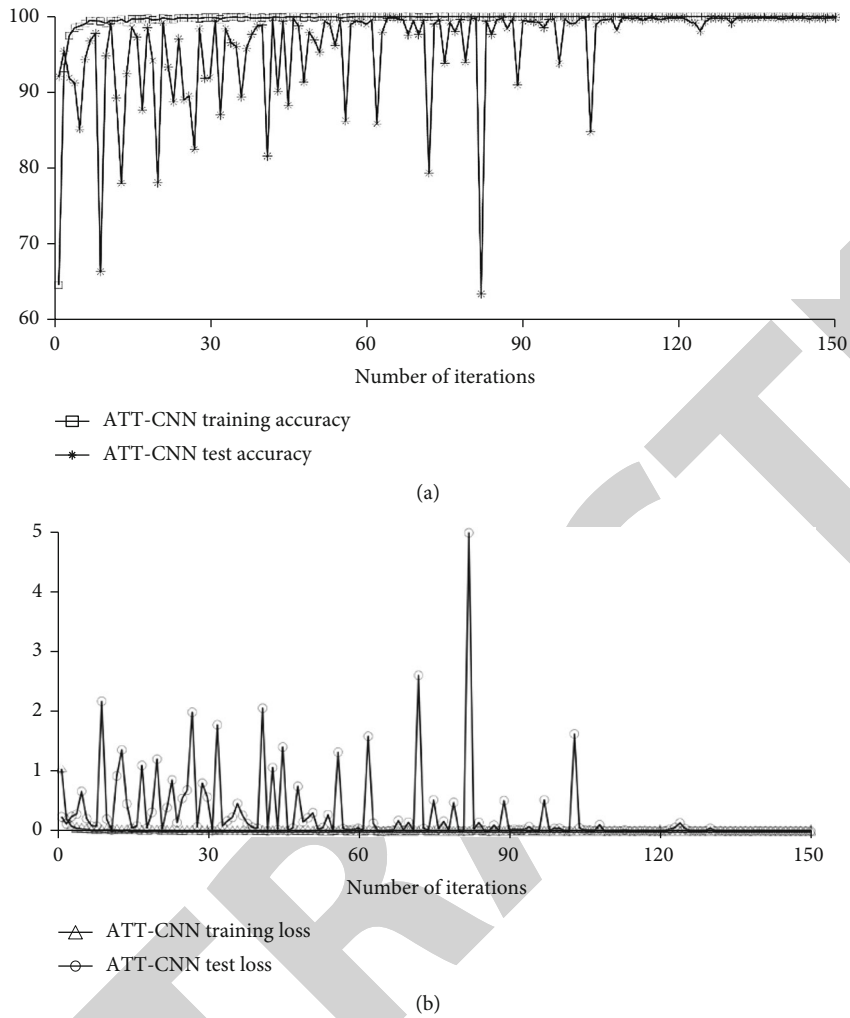


Figure 7: (a) Model accuracy diagram of the epoch process of the ATT-CNN model. (b) Loss value curve diagram of the epoch process of the ATT-CNN model.

surrounding rock itself. The input of the gas concentration prediction model is 10.

The gas data used in this paper comprise a total of 17280 groups of gas concentration samples collected every 5 minutes for 60 consecutive days. For training, the data are divided into a training set and a test set at a ratio of 7 : 3, in which the samples in the first 42 days are used as the training set and the remaining 18 days are used as the verification set. In this paper, there is no manual feature handle except for the standardization and one hot coding of the data.

In Table 2, b1 is the coal seam location; b2 denotes the initial speed of the gas release (mmHg); b3 represents the coal destruction type, and b4 stands for the gas pressure (MPa); b5 denotes the geological structure type; b6 represents the coal firmness coefficient; b7 stands for the coal seam permeability coefficient (m/MPa 2 d); b8 denotes the gas content (m^3/t); b9 represents the drilling gas attenuation coefficient (d-1), b10 stands for the depth of mining (m), and b11 denotes the output expectations.

The number of neurons in the network's hidden layer is 17, and the number of neurons in the output layer is 4. The first layer transfer function is a logarithmic sigmoid function, and the second layer transfer function is a linear function of purelin.

The predicted CGO risk results based on the NN model after training are shown in Table 3.

The predicted values of the model are in good agreement with the output data. The ATT-CNN simulation result is shown in Figure 7. By continuously updating the loss curve, the parameters learned by the model gradually increase, and the fitting ability is steadily improved. After 120 iterations, the test accuracy is stable. After 150 iterations, the model terminates training.

When the accuracy of the model is relatively high, the corresponding error loss value is relatively low. When the test accuracy rate oscillates, the test loss value curve also oscillates. Therefore, the accuracy and loss are consistent in reflecting the changes in the model's parameters and the state of data fitting in the iterative process. The performance of signal classification can be verified by combining these two indicators.

As a vital part of the model, the ATT is the key to obtaining the feature sequence, affecting the classification accuracy. Additionally, combining the opinions of on-site experts and

comprehensive factors, we believe that the ATT-CNN model is suitable for describing the objective phenomenon, where 0.7 is set as the weight of the ATT-CNN and 0.3 is set as the weight of the NN.

The objective state of the system at a certain moment is theoretically unique. The prediction of the fuzzy NN is not always consistent with the ATT-CNN in three aspects: prewarning results, environmental factors, and equipment factors. The cause lies in the insufficient quantity of training data of the model. If inconsistent results between the ATT-CNN and the NN occur, the final risk levels are regulated by the weight. Periodically compare performance of the ATT-CNN model to results of the fuzzy NN and switch to your predictions when they become more accurate.

In order to train prediction methods that aim to work well for a different kind of risk, it is sufficient to provide training data from others sensor for model.

7. Conclusions

Risk prewarning is critical in the coal mine safety management. In this paper, a mobile intelligent mine platform with a combination model is proposed to achieve the risk prewarning and information transmission to personnel. Any staff member with a mobile terminal working in the platform can obtain data for management and risk prewarning at any time.

The proposed risk prewarning model is formed by a compensation fuzzy NN with a novel ATT-CNN hybrid prewarning model focuses on the hide features and evaluates these features under learned weights and aggregates not only the information within all sequential data but also the sequential spatiality.

CNN processes high-dimensional data without pressure and can automatically extracts features. The sliding window of CNN has no sequential relationship, and there is no mutual influence between different convolution kernels. Therefore, it has a very high degree of parallel freedom which is a very good advantage.

ATT can be regarded as an automatic weighting. It learns a weight distribution and then employs on the feature. That is ATT can assign higher weights to the features with greater impacts on prewarning to improve the prediction accuracy. ATT has a great advantage that it can visualize the attention matrix to show what parts the neural network pays more attention.

The ATT-CNN takes full advantage of the powerful nonlinear learning ability and avoids the traditional reliance on experts for knowledge extraction. The model is free from artificial determination of the index weights, which are difficult to allocate reasonably, and is a completely data-driven approach.

The experimental results show that our proposed hybrid prewarning model is very effective for risk prewarning in terms of both quantitative and qualitative evaluations.

The hybrid model in this paper can effectively solve the problem of large capacity, multidimensional, and unbalanced time series classification and can become more detailed by increasing the number of output nodes.

Although the hybrid model achieves effective performance on test data, the model has a large number of parameters and calculations, which is not suitable for running on mobile terminals and embedded devices. The design of a lighter and faster risk prewarning model suitable for the wireless terminals of personnel will be our future research topic.

Data Availability

The raw data required to reproduce these findings cannot be shared at this time as the data also forms part of an ongoing study.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

This research was supported by the National Natural Science Foundation of China (Grant Nos. 51874300 and U1510115), the 2019 Key Research and Development Plan Project of Shandong Province (Grant No. 2019, GGX101011), and the Science and Technology Development Plan Project of Shandong Province Tai'an City (Grant No. 2020GX041). We are grateful to Fang Miao for her valuable help.

References

[1] J. Zhang, Q. Li, and Y. Zhang, "Definition of deep coal mining and mining response analysis," *Journal of Coal Society*, vol. 44, no. 5, pp. 1314–1325, 2019.

[2] M. He, X. Ma, J. Wang, J. Zhang, and Y. Liu, "Analysis of mine appearance characteristics of automatic roadway forming face with roof cutting and pressure relief of composite roof in medium thick coal seam," *Journal of Rock Mechanics and Engineering*, vol. 37, no. 11, pp. 2425–2434, 2018.

[3] G. Xu, Z. Huang, Z. Fan et al., "Roof disaster type, monitoring and prevention technology system of working face," *Coal Science and Technology*, vol. 49, no. 2, pp. 1–11, 2021.

[4] J. Deng, C. Lei, K. Cao, L. Ma, C. Wang, and X. Zhai, "Random forest method for predicting coal spontaneous combustion in gob," *Journal of China Coal Society*, vol. 43, no. 10, pp. 2800–2808, 2018.

[5] C. Qiu, *Study on gas concentration prediction methods based on PSO-SVR, working paper*, China University of Mining and Technology, Beijing, 2017.

[6] R. Tian, H. Meng, and S. Cheng, "Prediction of intensity classification rock burst based on depth neural network," *Journal of China Coal Society*, vol. 45, no. S1, pp. 191–201, 2020.

[7] S. Gong, Y. Ta, and W. Wang, "Prediction and evaluation of coal mine coal bump based on improved deep neural network," *Geofluids*, vol. 2021, Article ID 7794753, 11 pages, 2021.

[8] Z. Cheng, L. Ma, and Y. Zhang, "Prediction of spatial distribution of gas concentration based on LSTM-FC model," *Computer Engineering and Applications*, vol. 56, no. 16, pp. 258–264, 2020.

Hindawi
Wireless Communications and Mobile Computing
Volume 2023, Article ID 9890424, 1 page
https://doi.org/10.1155/2023/9890424

Retraction

Retracted: Detection of Power Data Outliers Using Density Peaks Clustering Algorithm Based on *K*-Nearest Neighbors

Wireless Communications and Mobile Computing

Received 11 July 2023; Accepted 11 July 2023; Published 12 July 2023

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. Li, L. Chen, and Y. Wang, "Detection of Power Data Outliers Using Density Peaks Clustering Algorithm Based on *K*-Nearest Neighbors," *Wireless Communications and Mobile Computing*, vol. 2022, Article ID 2203137, 7 pages, 2022.

Hindawi
Wireless Communications and Mobile Computing
Volume 2022, Article ID 2203137, 7 pages
https://doi.org/10.1155/2022/2203137

Research Article

Detection of Power Data Outliers Using Density Peaks Clustering Algorithm Based on K-Nearest Neighbors

Qingpeng Li,[1] Lei Chen ,[2] and Yuhan Wang[1]

[1]*Nanchang Power Supply Company, State Grid Jiangxi Electric Power Company, Nanchang 330200, China*
[2]*School of Information Engineering, Nanchang Institute of Technology, Nanchang 330099, China*

Correspondence should be addressed to Lei Chen; 15161875401@163.com

Received 28 October 2021; Accepted 9 July 2022; Published 27 July 2022

Academic Editor: Chi-Hua Chen

As an important research branch in data mining, outlier detection has been widely used in equipment operation monitoring and system operation control. Power data outlier detection is playing an increasingly vital role in power systems. Density peak clustering (DPC) is a simple and efficient density-based clustering algorithm with a good application prospect. Nevertheless, the clustering results by the DPC algorithm can be greatly influenced by the cutoff distance, indicating that the results are highly sensitive to this parameter. To address the shortcomings of the DPC algorithm and take the characteristics of power data into consideration, we propose a DPC algorithm based on K-nearest neighbors for the detection of power data outliers. The proposed DPC algorithm introduces the idea of K-nearest neighbors and uses a unified definition of local density. In the DPC algorithm, only one parameter (K) needs to be determined, thus eliminating the influence of cutoff distance on the clustering result of the algorithm. The experimental results showed that the proposed algorithm can achieve accurate detection of power data outliers and has broad application prospects.

1. Introduction

With the construction and development of smart grids, power enterprises have accumulated a large amount of power data in various forms, from different sources, and with complex structures. With the rapid progression of artificial intelligence [1], effective power data mining can be achieved. Not only has it promoted the transformation of the power grid from the traditional physical model-based business model to a data-based one [2, 3] but also it has empowered the power grid enterprises to embrace the new digital economy.

Anomalies in the power industry mainly include loss of primary attributes of power data, inconsistent statistical caliber of power data, abnormal power consumption behaviors of customers, and power equipment failure. Since anomaly detection can detect abnormal power consumption behaviors in power operations, it is widely used in the power system [4]. In the early days, the method used to detect power consumption anomalies was relatively simple. In most cases, technicians should go onsite to diagnose the problem. For this method, technicians should have a great amount of experience. It would lead to the waste of human and material resources and the detected power fault time lagging behind the actual fault time so that the operation status of the power grid cannot be maintained in real time. Additionally, the results obtained using this method were not highly accurate and were highly correlated with the experience of technicians. When data-driven methods are used to detect power data outliers, the reliance on human resources can be reduced and the power grid operation status can be monitored in real time. Besides, when the power grid is in an alert or abnormal state, alarms can be issued or the power supply can be cut off to prevent the spread of faults and reduce the economic loss of the power grid. Given this, the data-driven methods will be the new trend and promising direction of power anomaly detection [5, 6].

Currently, algorithms for the detection of power data outliers include algorithms based on the probability statistical model, classification-based algorithms, distance-based algorithms, and clustering-based algorithms.

In outlier detection algorithms based on probability statistical models, it is assumed that the detected data fit a statistical model, such as a parametric model like the Gaussian mixture model or a nonparametric model like kernel density estimation. The outliers are detected by fitting the detected data to the statistical model and then comparing the deviation of the detected data with the model to determine whether an outlier is present in the data or not. Methods for the detection of data outliers based on parametric statistical models [7] and nonparametric statistical models [8] have been introduced. The experimental results showed that these methods could obtain good detection results when the statistical laws of the data were simple and the data size was small. Nevertheless, power data shows a typical temporal coupling pattern and is high dimensional. For such datasets, these methods often fail to achieve good detection results. Although these methods have a good theoretical basis, the practical application of these methods is limited by the fact that the specific statistical model fitting the detected data cannot be known in advance, resulting in blind detection and uncertain results.

The classification-based outlier detection method is a semi-supervised learning method [9]. This type of method operates in a two-phase fashion, that is, the training phase and the testing phase. The training phase requires sufficient labeled samples to train the classifier, and then the classifier determines whether the detected data are normal or outliers. The drawback of this method is that it requires enough labeled samples to train the classifier, and the performance of the classifier directly affects the detection accuracy. The neural network is a classification-based outlier detection algorithm with good self-learning capability, but its derivation process lacks interpretability [10].

For distance-based outlier detection, it is assumed that a data object is an outlier if it is far away from other points. The distance-based outlier detection method is simpler than the statistical model-based method because it is easier to define a distance-based metric for a dataset than to determine the distribution of the dataset. Fan et al. [11] proposed the outlier detection algorithm with personalized K-nearest neighbors (PKNN). In this algorithm, the number of K-nearest neighbors of each sample is determined automatically by the algorithm without any human intervention, so that samples in dense areas have more nearest neighbors and samples in sparse areas have fewer, which is more consistent with the actual distribution of the dataset. The idea of this type of method is simple, but the time complexity is high because of the required calculation of the distance between two data points. Additionally, the method is sensitive to the parameter K, thereby obtaining highly variable detection results for different values of K.

The clustering-based algorithm for outlier detection is an unsupervised learning method. This type of algorithm assumes that normal data belong to one or more clusters, and samples that do not belong to any cluster are considered outliers. Outlier detection is the process of identifying outliers in a dataset through cluster analysis. Clustering algorithms that have been frequently used for outlier detection include DBSCAN, BIRCH, CLARANS, STING, CLIQUE, and KNN [12, 13]. The clustering-based algorithms for outlier detection can obtain good detection results, but their time complexity is generally high and the clustering results are greatly influenced by the parameters.

The density peak clustering (DPC) algorithm is a novel density-based clustering method proposed by Rodriguez and Laio [14] in 2014. The main idea of this algorithm lies in the portrayal of cluster centers. The authors considered that the cluster centers were composed of many samples with a higher density and larger relative distance. The DPC algorithm can automatically determine the number of clusters and achieve any status quo clustering, which is a hot research topic in cluster analysis. However, the DPC algorithm has some problems. First, the clustering results by this algorithm are dependent on the cutoff distance, which can be hardly determined. Second, the definition of the local density of the algorithm does not take the data size and its distribution into account, resulting in unideal clustering accuracy [15].

Based on the characteristics of power data, we proposed a DPC algorithm based on K-nearest neighbors for the detection of power data outliers. Currently, the DPC algorithm has been seldomly applied for outlier detection. In this paper, we redefined the local density of the DPC algorithm by integrating the K-nearest neighbors. By considering the local characteristics of data and using a unified definition of local density, the original algorithm was improved. Meanwhile, only one parameter (K) needs to be determined, which eliminates the influence of cutoff distance on clustering performance, and its value can be easily determined. The experimental results showed that the proposed algorithm can achieve accurate detection of power data outliers.

2. DPC Algorithm

The basic idea of the DPC algorithm is as follows: (1) density peaks have a high local density and are surrounded by neighbors with lower density; (2) the density peaks are relatively far from each other. In the DPC algorithm, two variables are introduced to characterize the density and distance of Sample i, namely, the local density ρ_i and the relative distance δ_i to the nearest sample with a higher local density. The local density ρ_i can be calculated by:

$$\rho_i = \sum_j x\left(d_{ij} - d_c\right), \qquad \chi(x) = \begin{cases} 1 & x < 0 \\ 0 & x \geq 0 \end{cases}, \tag{1}$$

where d_{ij} is the Euclidean distance between Samples i and j. When the size of the dataset is small, the local density is calculated by the Gaussian kernel function as follows.

$$\rho_i = \sum_j \exp\left(-\frac{d_{ij}^2}{d_c^2}\right). \tag{2}$$

The relative distance of Sample i to the nearest sample with higher local density is denoted by δ_i, which can be calculated by:

$$\delta_i = \begin{cases} \min\limits_{j:\rho_j>\rho_i} \left(d\left(x_i, x_j\right)\right), & \text{if } \exists \text{j s.t.} \rho_j > \rho_i \\ \max\limits_{j} \left(d\left(x_i, x_j\right)\right), & \text{otherwise} \end{cases}. \quad (3)$$

If Sample i has the maximum local density, the corresponding relative distance will also be the largest.

The DPC algorithm takes the samples with the larger ρ_i and δ_i as density peaks. To find the density peaks, the DPC algorithm draws a decision diagram with ρ_i on the horizontal axis and δ_i on the vertical axis to select the density peaks. To better represent the density peaks, the DPC algorithm defines a decision value γ_i.

$$\gamma_i = \rho_i \times \delta_i. \quad (4)$$

The DPC algorithm considers the samples with a high local density and large distance as the density peaks. It means that the points with high γ_i shall be selected as the density peaks. After the density peaks are found, the remaining samples shall be allocated to the cluster that the nearest samples with a higher density than them belong.

3. Detection of Power Data Outliers Using DPC Based on K-Nearest Neighbors

3.1. DPC Algorithm Based on K-Nearest Neighbors. The local density of the DPC algorithm using either the Gaussian kernel or the cutoff kernel is related to the cutoff distance, and the optimal cutoff distance varies greatly among different datasets [16]. Additionally, the local density of the DPC algorithm is mainly determined by the samples within the range of cutoff distance, and the samples beyond the range contribute little to the local density. Owing to this, density peaks are more likely to appear in the region with high local density. For data with uneven density distribution, the cluster centers are concentrated in dense regions and there are no cluster centers in sparse regions. After analysis, it can be known that the relative density of a sample and its K-nearest neighbors can truly reflect whether the sample is a density peak or not.

Definition 1. Local density based on K-nearest neighbors. The local density of new samples can be defined by:

$$\rho_i = \exp\left(-\frac{\sum_{j=knn(i)} d_{ij}^{\ 2}}{\sum_{j=knn(i)} \sum_{v=knn(j)} d_{vj}^{\ 2}}\right), \quad (5)$$

where d_{ij} is the Euclidean distance between Samples i and j, $knn(i)$ is the set of K-nearest neighbors of Sample i, $\sum_{j=knn(i)} d_{ij}^{\ 2}$ is the sum of the Euclidean distance between Sample i and its K-nearest neighbors, indicating the degree of outlierness of Point i. The larger the value of $\sum_{j=knn(i)} d_{ij}^{\ 2}$, the greater the degree of outlierness, the more locally sparse Sample i. $\sum_{j=knn(i)} \sum_{v=knn(j)} d_{vj}^{\ 2}$ indicates the sum of the degree of outlierness of K-nearest neighbors of Sample i. The larger the value, the greater the local density of the point.

When the local density is defined in this way, the local density of the sample is only related to its K-nearest neighbors, thereby eliminating the interference of the distant irrelevant points. Also, the calculated local density is the relative density of the point and its K-nearest neighbors. It means that the local density of the sample in clusters with varying density distribution can be adjusted so that the local density of the sample in sparse clusters increases and the local density of the sample in dense clusters decreases. In doing so, the density peak in sparse clusters can be found more easily and thus improving the clustering performance for datasets with large differences in density.

3.2. Principle of Outlier Detection. Upon determination of (ρ_i, δ_i) of all sample points in the dataset, the decision diagram of ρ_i and δ_i shall be drawn. The points with both large ρ and δ are identified from the decision diagram, and these points are used as the cluster centers of the dataset. From the perspective of outlier detection, the points with smaller ρ and larger δ can also be visually seen in the decision diagram, and these points are identified as outliers.

Considering the characteristics of power data, we assumed that outliers should satisfy the following conditions: (1) the local density is less than the threshold of local density, that is, $\rho_i < \rho_f$; (2) the relative distance is greater than the threshold of relative distance, that is, $\delta_i > \delta_f$. Based on this, the outliers of the power data can be determined. The threshold of local density ρ_f can be calculated by:

$$\rho_f = \frac{1}{N} \sum_{i=1}^{N} \rho_i - \varepsilon_\rho. \quad (6)$$

The threshold of relative distance δ_f can be calculated by:

$$\delta_f = \frac{1}{N} \sum_{i=1}^{N} \delta_i - \varepsilon_\delta, \quad (7)$$

where N denotes the total number of samples in the power dataset, and ε_ρ and ε_δ denote the empirical parameters.

3.3. Procedures of Outlier Detection. Input: power load data X, number of neighbors K (the experimental result is optimal when K is 25).

Output: outliers.

Step 1: Data preprocessing. Preprocess the load data, such as replacing the missing values with the mean substitution method

Step 2: Calculate the Euclidean distance between samples and construct the distance matrix of samples

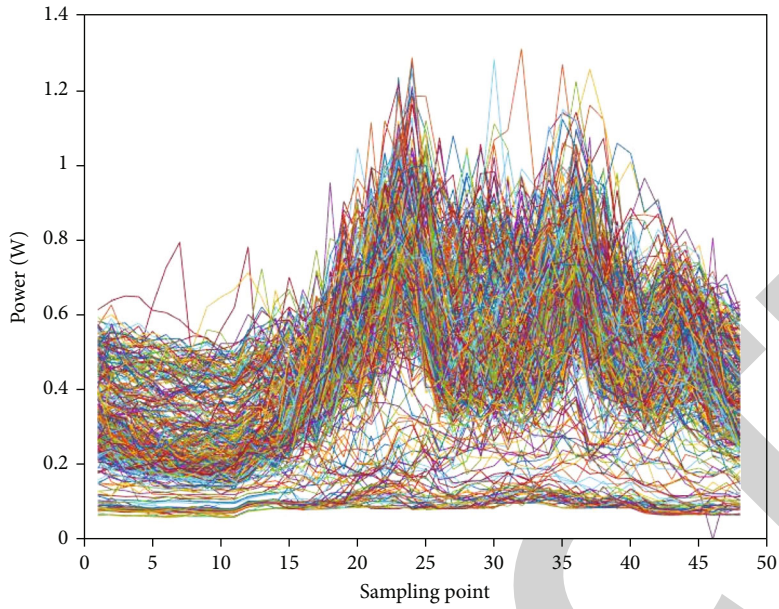


Figure 1: Daily load profile of single transformer.

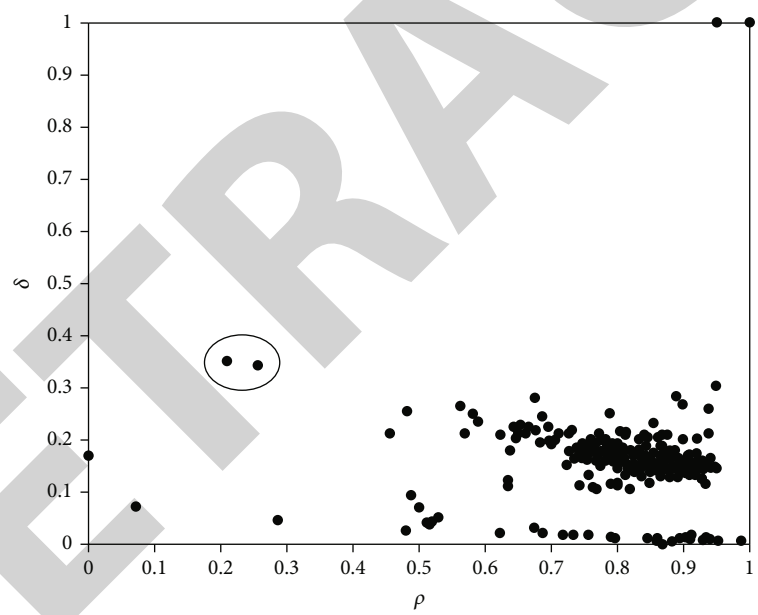


Figure 2: Outlier detection decision diagram of single transformer.

Step 3: Calculate the local density ρ_i and relative distance δ_i of samples according to Eqs. (5) and (3), respectively

Step 4: Set the threshold of local density and relative distance based on the domain expert's experience according to Eqs. (6) and (7)

Step 5: Identify Sample i with $\rho_i < \rho_f$ and $\delta_i > \delta_f$ as outliers and output outliers

4. Results and Analysis

4.1. Data Source. To verify the effectiveness of the DPC algorithm based on K-nearest neighbors for detection of power data outliers, we used the load data of AC 10 kV distribution transformers in a region for 366 days from January 1, 2020, to December 31, 2020. The power load data belongs to the storage sector, and its collection frequency is 0.5 h. The daily load profile has 48 data points. Case 1 is the data of a single transformer, and Case 2 is the daily load data of 10 transformers during 6 days from September 22, 2020, to September 27, 2020. In this paper, the mean substitution method was used, i.e., the mean of all the values except the missing point was used to replace the missing value of the attribute.

4.2. Outlier Detection of Load Profiles of Single Transformer. The load data profile of a single AC 10KV distribution

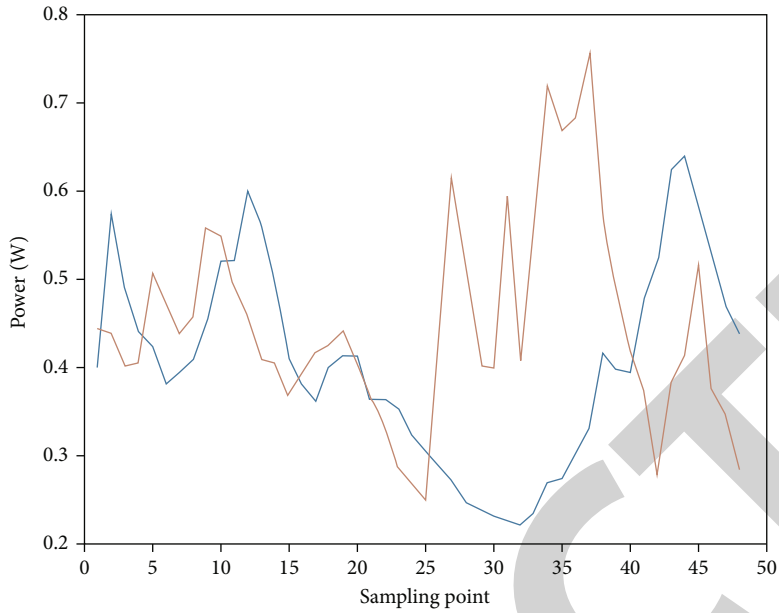


Figure 3: Outliers of single transformer.

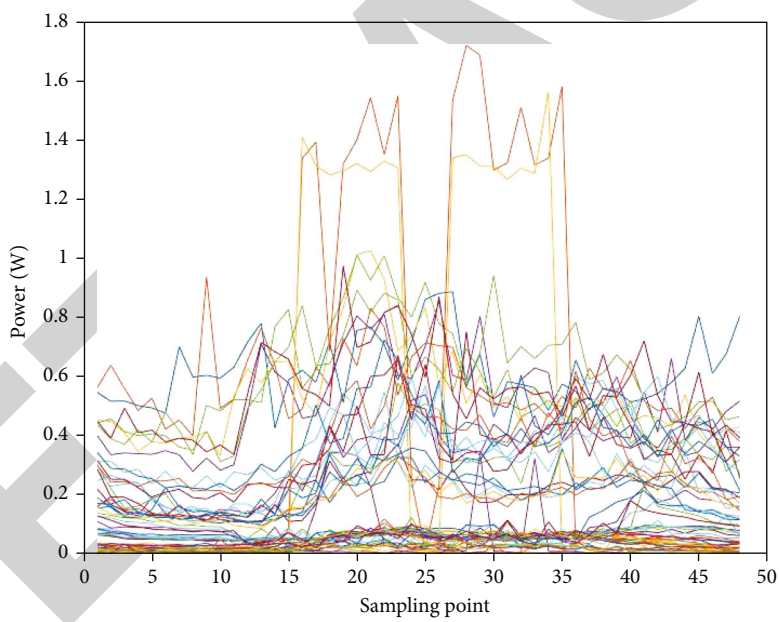


Figure 4: Daily load profile of 10 transformers during 6 days.

transformer during 366 days from January 1, 2020, to December 31, 2020, is shown in Figure 1.

As shown in Figure 1, the daily load trend of this transformer was more or less the same, but few profiles deviated from the normal operation pattern to a large extent. According to the steps of the DPC algorithm based on K-nearest neighbors for detection of power data outliers, the outlier detection decision diagram of this transformer was drawn, as shown in Figure 2.

As shown in Figure 2, the local density and relative distance of most of the samples fell in the region where the local density was higher than 0.4 and the relative distance was less than 0.3, and only very few samples had local density and relative distance fall beyond the above region. According to the principle that outliers should have low local density and large relative distance, the distribution of outliers was identified. The outliers were marked with hollow circles in Figure 2.

The empirical parameters were set as $\varepsilon_\rho = 0.4$ and $\varepsilon_\delta = 0.14$. Then, the outliers in the power data were found according to Eqs. (6) and (7), which were the points circled in Figure 2. The outliers of the load profile of a single transformer are shown in Figure 3.

As shown in Figures 1 and 3, the DPC algorithm based on K-nearest neighbors for detection of power data outliers

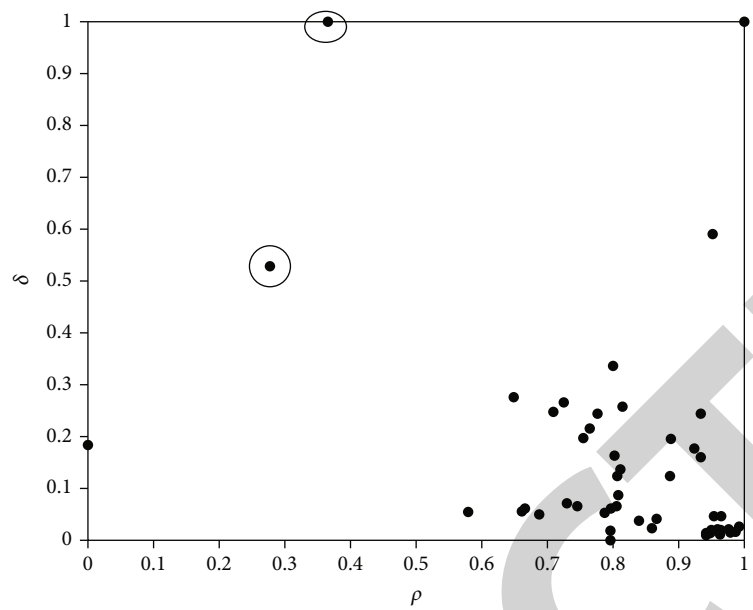


Figure 5: Outlier detection decision diagram of 10 transformers.

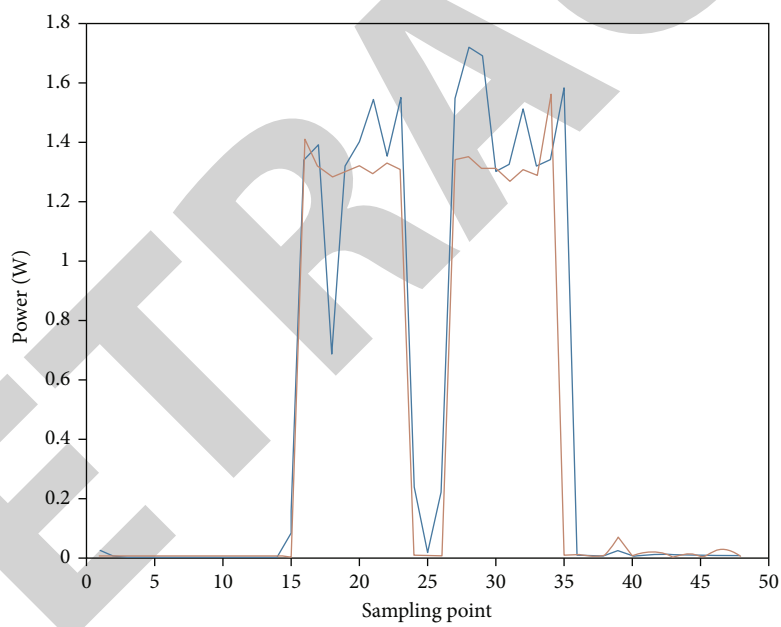


Figure 6: Outliers of 10 transformers.

can detect the profiles that are different from the conventional electricity consumption pattern from the load data. Among the total 366 daily load data, two power data outliers were detected, one on September 26, 2020, and the other on September 27, 2020. The blue and red profiles in Figure 3 represented the daily load profiles on September 26, 2020, and September 27, 2020, respectively. As shown in Figures 1 and 3, an electricity consumption peak should have appeared at sampling point 25 under normal conditions. However, sampling point 25 on September 27 reached the minimum electricity consumption in the day, and the electricity consumption at sampling point 25 on September 26 was also very low, and no peak of electricity consumption appeared between sampling points 25 and 35, which was inconsistent with the normal electricity consumption pattern of this industry. Therefore, this point was identified as an outlier.

4.3. Outlier Detection of Load Profiles of Multiple Transformers. The outlier detection was conducted simultaneously for the daily load data of 10 transformers, and the steps are the same as in Case 1. The daily load profiles of 10 transformers during 6 days from September 22, 2020, to September 27, 2020, are shown in Figure 4. As observed, the outliers were found in the data of few days.

Hindawi
Wireless Communications and Mobile Computing
Volume 2023, Article ID 9851453, 1 page
https://doi.org/10.1155/2023/9851453

Retraction

Retracted: Power IoT System Architecture Integrating Trusted Computing and Blockchain

Wireless Communications and Mobile Computing

Received 10 October 2023; Accepted 10 October 2023; Published 11 October 2023

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. Yan, J. Chen, Y. Yao, S. Guo, A. Xiong, and W. Zhang, "Power IoT System Architecture Integrating Trusted Computing and Blockchain," *Wireless Communications and Mobile Computing*, vol. 2022, Article ID 2891917, 13 pages, 2022.

Hindawi
Wireless Communications and Mobile Computing
Volume 2022, Article ID 2891917, 13 pages
https://doi.org/10.1155/2022/2891917

Research Article

Power IoT System Architecture Integrating Trusted Computing and Blockchain

Yong Yan,[1] **Jinqian Chen**,[2] **Ying Yao**,[1] **Shaoyong Guo**,[2] **Ao Xiong**,[2] **and Wang Zhang**[2]

[1]*State Grid Zhejiang Electric Power Research Institute, Hangzhou, China 310014*
[2]*State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing, China 100876*

Correspondence should be addressed to Jinqian Chen; chen_jinqian@bupt.edu.cn

Received 14 July 2021; Accepted 2 April 2022; Published 9 May 2022

Academic Editor: Chi-Hua Chen

With the construction of power Internet of Things, the network scale is constantly expanding, the network structure is increasingly complex, and the number of terminals is increasing rapidly, resulting in huge pressure on network security protection, and it is urgent to build a credible operation environment to ensure that the Internet of Things runs safely and reliably. This paper proposes a power Internet of Things architecture combining trusted computing and blockchain. The architecture ensures the credibility and security of the terminal based on remote integrity verification and access control policy verification. A method for malicious node detection based on blockchain (MMNDB) is placed in acquisition layer. It is used to monitor the status of each acquisition devices in time, detect the malicious terminal in time, isolate the suspicious terminal with low credibility from the power Internet of Things, and ensure credibility of the acquisition terminal and the data collected. Under premise of ensuring the communication efficiency of the power communication network, this architecture can greatly ensure the stable working of the power Internet of Things.

1. Introduction

The definition of IoT was first proposed by Kevin Ashton [1], the founder of the Automatic Identification Center of the Massachusetts Institute of Technology. Currently, IOS/IEC defines IoTs as an infrastructure that connects people to things and information systems, on top of which are intelligent services capable of processing and responding to information. The Internet of Things is an information carrier based on the Internet, traditional telecommunications networks, etc. It enables all ordinary physical objects that can be independently addressed to form an interconnected network. At present, scholars have put forward the concept of ubiquitous power IoT. The ubiquitous Internet of Things revolves around the power system, making full use of information technologies to realize the power system's interconnectedness of everything and everyone, forming a comprehensive state observation, efficient modern information processing, and application of intelligent services system. It is convenient and flexible [2].

In trusted computing, many organizations have different understandings of trust. ISO/IEC stresses the need to be resistant to a certain degree of virus and physical attack [3]. The IEEE definition emphasizes that it must be provable [4]. The Trusted Computing Group (TCG) defines trusted willfulness as follows: An entity is trusted if it always behaves in the expected way and toward the expected goal [5]. TCG's trusted computing technology approach is to improve the security of computer systems by introducing a trusted platform module (TPM) on the hardware platform. This technology approach has been generally recognized by the industry. At the same time, the blockchain is a technology that allows a fully decentralized ledger to be created for transactions. All nodes in the network have access to its ledger. It is being used in an increasing number of fields [6–8].

Nowadays, the scale of network continues to expand, the network structure is increasingly complex, and the number of terminals has increased sharply. This has caused great pressure on network security protection. It is urgent to build a credible operating environment to ensure the network quality of the power Internet of Things and ensure the Internet of Things Safe and reliable operation of business. However, traditional trusted computing is used to solve the security problems of the architecture. It aims to establish a trust chain in computing system and extend trust relationship from the underlying hardware to the upper application to enhance the security of the computing system. The issue of trust in the data space also requires urgent attention. Decentralization is a feature of blockchain, and it can be used to solve the data space's trust problem through a consensus method. As a result, it has become a development trend to integrate blockchain and trusted computing to build a high-credibility environment of physical networks and data logic spaces to support the construction of a trusted power Internet of Things.

The main contributions of this article are as follows:

(i) This paper proposes a power Internet of Things architecture that integrates blockchain and trusted computing. This architecture enables the device to have strong identity certification by deploying trusted chips in Internet of Things devices to ensure that the power Internet of Things has high-confidence characteristics. At the same time, the architecture achieves reliable data collection, secure transmission, reliable storage, and efficient use based on blockchain technology

(ii) This paper proposes a method for detecting malicious nodes based on reputation. This method evaluates the credibility of power Internet of Things nodes from multiple dimensions and uses smart contracts to automatically execute malicious node detection algorithms, isolate malicious nodes in time, and ensure the stable and safe operation of power Internet of Things

(iii) This paper designs a network quality evaluation model based on multidimensional evaluation indicators. The model uses a combined weight optimization model based on the least squares method to obtain the optimal weight value to assign the weight of the evaluation index, thereby improving the accuracy of the evaluation result

(iv) According to the results of simulation experiments, the proposed strategy for detecting malicious nodes in this paper is more accurate, lower false alarm rate, and shorter detection time. At the same time, the power Internet of Things architecture proposed in this article can isolate malicious nodes in a timely manner after being attacked, maintain good network quality, and ensure the continuity of smart grid business

2. Related Work

Different international organizations have provided different IOTS reference architectures, and their development has been widely recognized and used [9]. Traditional Internet of Things architecture employs a tiered system, which divides the Internet of Things into three layers: perception, network, and application. Comprehensive perception, dependable transmission, and intelligent processing are the features of the corresponding layer. Sensors, RFIDs, QR codes, and other information-gathering objects are found at the perception layer, as are smart devices and actuators for control; the network layer takes data from the perception layer and transfers it to the application layer, receiving a response from the application layer. The information will be routed and managed at the network layer, and the instructions will be passed to the perception layer; then there is another layer: The application layer contains middleware, application infrastructure, and numerous IoT services, with the purpose of providing interfaces for Internet of Things applications [10]. However, this traditional three-layer architecture does not build a credible operating environment and cannot well meet the requirements of power Internet of Things security protection.

Blockchain can be integrated with the Internet of Things to achieve decentralization and high accuracy of network resource management and access control, while ensuring data and identity privacy [11, 12], ensuring the stable operation of the Internet of Things system. The use of blockchain in the energy sector would increase energy security and efficiency by offering a distributed platform for systems, hence boosting the Internet of Things' efficacy. People can immediately obtain energy information without the intervention of a third party, and real and high-quality data can be freely transmitted across devices. Furthermore, a blockchain-based IoT system aids the long-time running of smart grid devices [13]. There are still several issues with applying blockchain technology directly to the Internet of Things system due to a lack of sufficient computational resources and network bandwidth. At the same time, the security of blockchain nodes is also facing severe challenges.

For these reasons, some researchers have proposed to improve the overall security of related application systems by integrating blockchain with trusted computing.

Hybridchain is a solution presented by Wang et al. [14], which integrates blockchain with a TEE. Hybridchain can provide excellent performance while maintaining confidentiality. This technique, however, has yet to be successfully implemented in IoT applications.

In view of the problems of data privacy breach and difficulty in confirming data rights in the existing data transaction solutions, Zhang et al. [15] proposed a data transaction solution that combines blockchain and trusted computing technology with business requirements to build a new data transaction system. Compared with the traditional data transaction model, this solution uses the decentralized characteristics of the blockchain to store data index information, transaction process, and data usage records, which realizes the credibility and confirmation of data and

prevents data privacy. However, data transaction is too cumbersome and requires high performance on the blockchain and trusted computing platform; otherwise, it cannot meet the real-time demand for massive data transactions.

Gao et al. [16] integrated trusted computing and blockchain technology to propose a secure medical IoT data analysis framework, which adds the hardware-based software protection extension technology (SGX) in trusted computing to edge computing nodes. This technology can provide an encrypted trusted execution space in the memory for processing and analyzing medical data and storing analysis results. This part of the area is invisible to edge computing nodes, thus avoiding data leakage. The blockchain network of this framework is used to store data analysis results and result access control strategies to realize the credible use of data analysis results. However, the SGX technology used in this framework is based on Intel chips, which has compatibility and performance limitations. At the same time, it has high memory performance requirements for edge computing nodes.

Liu et al. [17] proposed a credible network architecture based on blockchain-B-TNC. B-TNC fully integrates the security features of blockchain decentralization, tamper-proof, and traceability and realizes a stronger network trust model. However, this method focuses on cross-domain network connection and is not suitable for interaction with IoT terminals.

In summary, preliminary work has been carried out on the integration of trusted computing and blockchain to build a high-credibility cyberspace and application system [18, 19]. However, it is still necessary to integrate trusted computing and blockchain to build a highly trusted power Internet of Things.

3. High-Credibility Power Internet of Things Architecture Integrating Trusted Computing and Blockchain

The high-credibility power Internet of Things architecture that integrates trusted computing and blockchain mainly include three parts: a trusted protection layer, a trusted power Internet of things layer, and a trusted data management layer. To solve potential data untrustworthiness of power IoT, this architecture adds a trusted chip to the collection terminal so that each collection terminal has a strong identity certificate. The trusted power Internet of Things layer works on top of the protection layer and ensures the credibility of data transmission and business applications in the Internet of Things through trusted terminals and trusted edge servers. The trust data management layer uses the blockchain to store the trust relationship and energy data of the collection terminal and uses the traceable and immutable characteristics of the blockchain to ensure that the trusted control data in the TPCM and the energy data in the Internet of Things are not illegally tampered with. Blockchain provides credible supervision of credible energy data and improves the sharing efficiency of credible energy data (see Figure 1).

The credible protection layer is the basis for ensuring the credibility of the entire power communication network architecture. The collection terminals and edge servers of the traditional power Internet of Things work in an open environment, so they are susceptible to various attacks, resulting in untrustworthy or leakage of energy data. In this architecture, the trusted platform control module (TPCM) is implanted in the collection terminal, edge server, and server of the blockchain node. It can achieve the credibility of the operating environment. TPCM builds a chain of trust when the terminal is started, tracks the log information recorded in the process of trust chain construction, and encapsulates the log information into transactions and uploads it to the blockchain as trusted control data. Each verifier can quickly pass the trusted control data. The correctness of the trust data is verified, the trust information of the terminal is effectively supervised, and the data is prevented from being maliciously tampered with.

The credible power Internet of Things network aims to extend the trust relationship from the collection terminal to the entire network. This architecture collects energy data through collection terminals with trusted chips, uses edge agents to filter invalid information, manages trusted energy data through the IoT management center and cloud system, and publishes energy data to the blockchain network. The trusted Internet of Things realizes the flexible recording of various trusted collection terminals and the interconnection of trusted energy data. At the same time, the data center can provide real-time data analysis and provide data sharing interfaces for other data applications, which greatly reduces the construction cost and technical complexity of the Internet of Things, and improve safety.

Data credibility management is mainly completed through the blockchain, which provides the entire architecture with trusted sharing of energy data and trusts computing security management. The blockchain stores the trusted control data of the TPCM. When a terminal is abnormal, or the control data is tampered with by an illegal attack, the blockchain can quickly find out the specific data through traceability. Moreover, deal with suspicious terminals to ensure the credibility of the credible power Internet of Things working environment. The efficiency of mass data sharing of traditional power Internet of Things is extremely low, and this architecture manages the energy data collected by the terminal through the blockchain and uses the decentralized characteristics of the blockchain to prevent the problem of major failure caused by the high concurrency of massive data. The use of the traceability feature of the blockchain to monitor trusted energy data prevents the problem of illegal data tampering that may exist in the traditional energy data center. At the same time, when a user application makes a data request, the data in the blockchain can be directly accessed. The need for cumbersome iterative requests and the coordination of data sources between different departments greatly simplify the processing process and improve the efficiency of the power Internet of Things.

3.1. Trusted Protection Layer. TPCM is the core component to ensure the credibility of network architecture calculations.

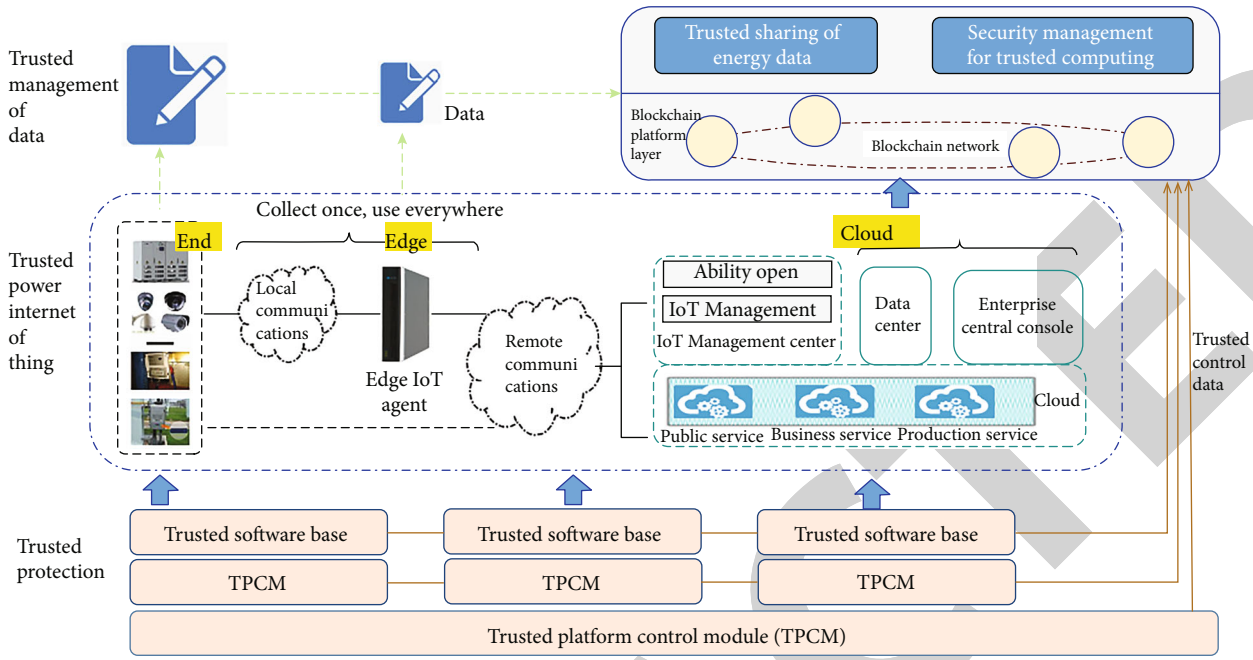


Figure 1: High-credibility power IoT architecture integrating trusted computing and blockchain.

It is responsible for credibility measurement and protection of nodes. It can build a trusted system, prevent malicious access, and effectively avoid real-time attacks in the process of server running. As the root of trust of the system, TPCM supplements root of trust control function based on TPM to realize active control and measurement of the system.

When the system is initialized, TPCM will actively verify the BIOS before the CPU starts. After the verification is passed, the CPU can run, ensuring active control of the entire platform. TPCM will first evaluate the integrity of the entire system before all applications run and generate integrity values. When the program starts, the integrity measurement process checks the status of each application and module and saves the resultant integrity measurement value in PCR, which cannot be tampered with. The PCR log is kept by TPCM. TPCM uploads the data fingerprint obtained by calculating the PCR log data and other related information to the blockchain. The appropriate steps are as follows (see Figure 2):

(1) Extract the data fingerprint of the previous PCR log information, and perform a hash operation together with this log information, user signature information, and user public key [3]

(2) Transfer the data fingerprint of this log information obtained by the hash operation to the blockchain

The system administrator can use logs to ensure the validity of the integrity authentication process.

3.2. Data Trust Management. This section mainly introduces the high-credibility power information communication platform that integrates blockchain and trusted computing. After the platform is installed and deployed, it provides administrator control, asset management, and security policy configuration. The unified management of the host computer facilitates the operation of the administrator. At the same time, the security log of the host in the trusted computing environment is also collected, which is convenient for the administrator to view and manage.

As shown in Figure 3, the data trusted management platform can implement user management, log management, node management and other functions. User management controls administrator users, controlling user login, information modification, and user permissions. We can also use security options to strengthen the system's own security from the system administrator identity, administrator use frequency, and password security. Node management is the unified management of hosts in a trusted computing environment, including node registration approval, whitelist scanning, allocation strategies, group management, modification of risk levels and groups, resetting node strategies, and deleting nodes. The platform can also collect security logs. The management center log performs auditing and log query functions of the administrator's system operation behavior. The content of the log includes the administrator's login, logout and all write operations. The node cluster log supports the auditing of event types and has a built-in default risk level.

4. Malicious Node Detection of High-Credibility Power Internet of Things

The architecture proposed a method of malicious node detection in the power IoT based on blockchain, which is called MMNDB. First of all, we introduce the data structure of the blockchain. The smart contract of the blockchain is designed in Sec. 4.2. This section presents the formal

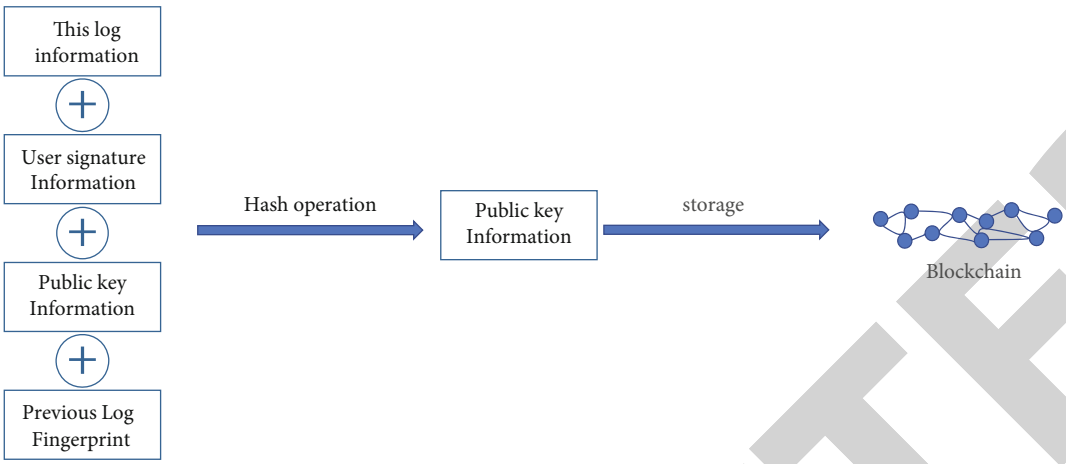


Figure 2: TPCM Llog Ffingerprint on Tthe bBlockchain.

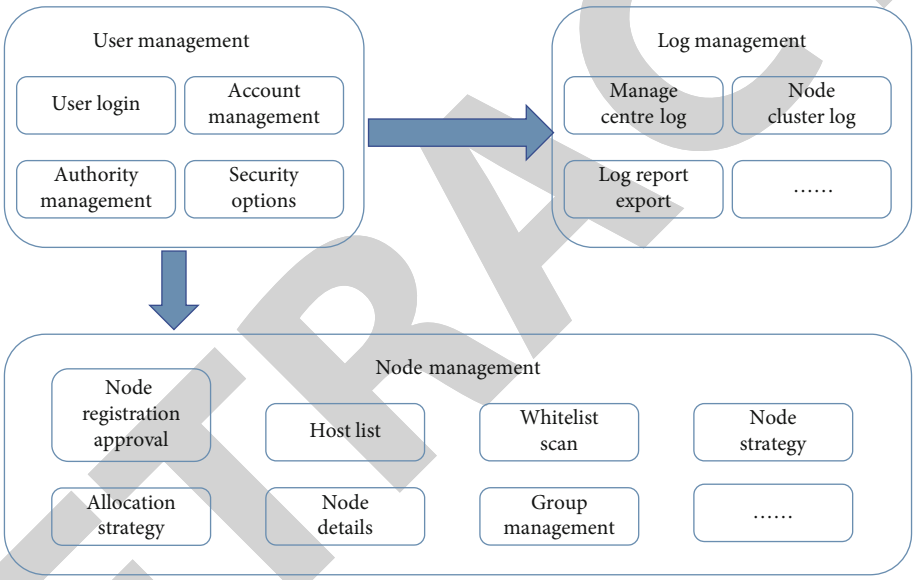


Figure 3: Data trust management platform architecture.

expression of the smart contract and discusses indicators of evaluating malicious nodes. Finally, the detailed process of malicious node detection is also expressed.

4.1. The Blockchain Data Structure for Detection of Malicious Nodes. In order to describe the blockchain for the detection of malicious nodes, this paper determines a block data structure based on blockchain for the detection of malicious nodes ($B_{bdmn} - BDS$). $B_{bdmn} - BDS$ records all data of communication.

Block header and block body are the two fundamental portions of a blockchain data structure, as shown in Figure 4. The block body contains the primary information of the power IoT gathering equipment, such as location, HMD, ID, PLR, MDR, CDR, and HBRT (see Sec. 4.2)

In a malicious node detection technology, the reputation of the collection terminal is evaluated in real-time from multiple angles. Check the terminals with low reputation value in time.

Here, $T_1 - T_n$ represents each piece of equipment, Hash1 denotes the hash pointer of T1 data, and Hash12 denotes the hash pointer of Hash1 + Hash2, resulting in a unique Merkle-root.

B_{bdmn} employs a "block + chain" chain data structure, and it uses a Merkle Tree formed by a hash pointer to store the information acquired by each block. Once the data of any block is adjusted, such a data structure allows the hash pointer of the block to be changed. Furthermore, the data is recorded by various types of devices utilizing this data structure. When the entire network is freed, the risk of hostile manipulation is reduced, safety and fairness are ensured, and the detection process is made easier.

4.2. Smart Contract for Detection of Malicious Nodes. An expression for the smart contract is as follows:

$$B_{bdmn} - SC = (LS, ES, Sen, \theta, P, \sigma, LOC, TIM, HCD). \quad (1)$$

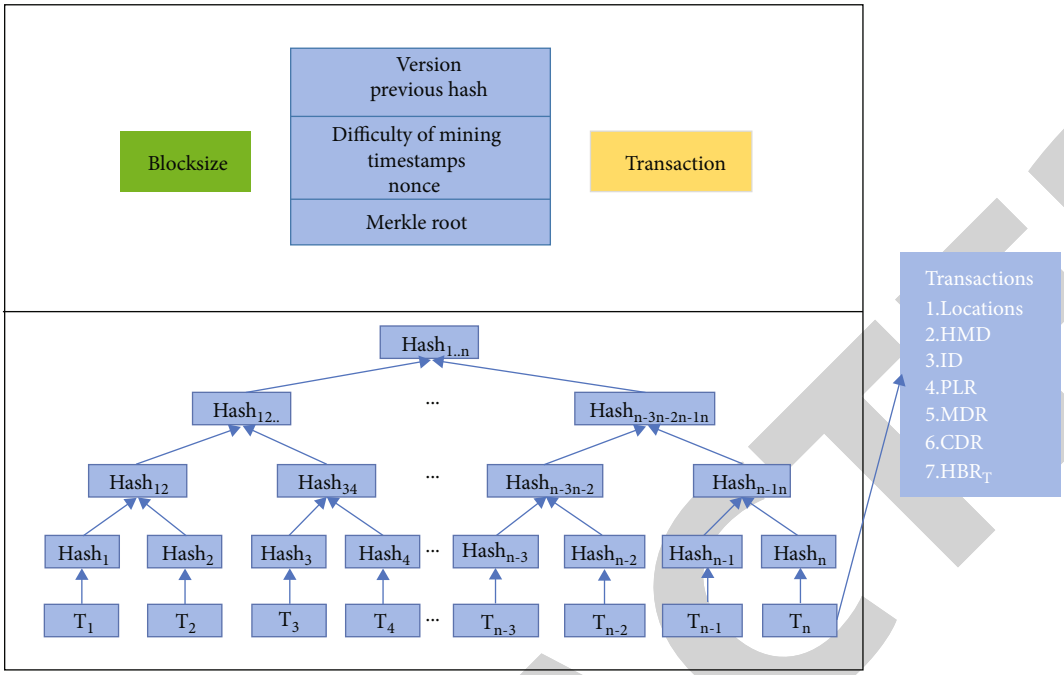


Figure 4: Block data structure.

Among them, LS is generally trusted to deploy smart contract, ES is the server to execute contract, Sen is the collection device to be detected, θ is the assessment criteria, P is reputation points, σ is threshold of a node that is judged to be malicious, and LOC is the location of the device to be detected. TIM is timestamp of the message forwarded by the node. HCD is hash of control data of the node.

As shown in Figure 5, the steps are as follows:

Step 1. The LS deploys the smart contract $B_{bdmn} - SC$

Step 2. The LS authorizes the ES to execute the $B_{bdmn} - SC$

Step 3. In a trusted computing environment, (2)-(6) is used to calculate the communication information of collection equipment to obtain the corresponding PLR, MDR, CDR, and HBR_T. These evaluation indicators will be stored in the blockchain periodically

Step 4. The ES use the Judge() function of smart contract to check hash of control data (HCD), the timestamp, and location information of the terminal to be detected. If the control data is changed or the timestamp and location information is not as expected, the node can be judged to be malicious directly. Finally, the id of malicious nodes will be added to Set_{id} of malicious nodes

Step 5. The NodeCommunicationScore() uses the value of PLR, MDR, CDR, and HBR_T stored in the blockchain to calculate the CS value by (6)

Step 6. Compute the amount of P using Formula (7)

Step 7. The ES uses the Detect() to cast the ID of malicious nodes as fellows: The system administrator sets threshold σ. When $P \leq \sigma$, the collection device is recorded to be malicious device. The id of this node will be added to Set_{id} of malicious nodes.

Step 8. According to the Set_{id} of malicious nodes, LS publishes malicious node information to the whole network by using consensus and isolates them from the credible network.

It is determined preliminarily that the current network is normal and the credible score value of the current network communication is calculated. The standards are expressed as follows:

PLR: Malicious devices and insecure links will increase the loss tolerance of the network, so the packet loss rate can be used as a measure of malicious nodes.

$$PLR = 1 - \left(\frac{n_{\text{recv}}}{\text{n}_{\text{send}}}\right), \tag{2}$$

where n_{recv} presents number of packets successfully received. n_{send} is the amount of data packets sent in the time period.

MDR: Bad devices may purposefully lengthen the time it takes to process data. To assure real-time data, the server calculates the message delay rate by obtaining the time to forwarding data and setting a time interval T_d:

$$MDR = \begin{cases} (T_{sensor}/T_d) & T_d \\ 1 & T_{sensor} \geq T_d \end{cases}. \tag{3}$$

CDR: The response time required by malicious nodes is usually longer. The server notes the total time from receiving

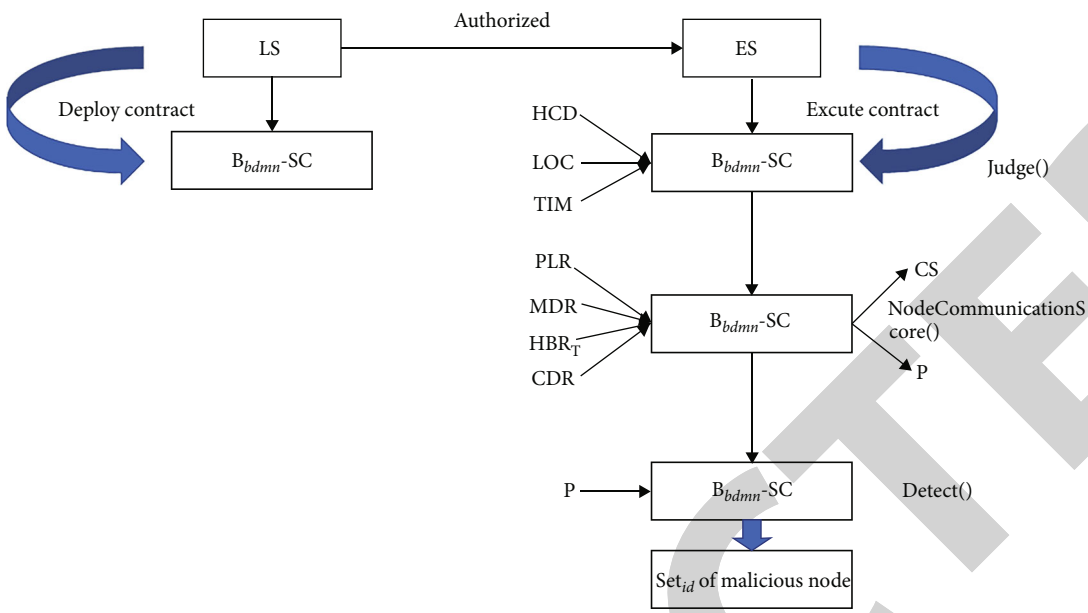


Figure 5: Steps of malicious node detection.

the request to the requester receiving the response from the node to be detected within a certain time interval.

$$CDR = \frac{(T_{hd} + T_{td})}{T_e}, \quad T_{td} = Ms/Nb + \frac{pd}{ts}. \tag{4}$$

Among them, T_{hd} is the time to process data. T_{td} is the time for message propagation, and Ms is the message size. Nb is network bandwidth. ts is the propagation speed, and T_e is the predetermined time period.

Historical behavior: The server calculates the proportion of malicious communication of the node to be detected in a certain time interval and uses this proportion as a reference standard for evaluating reputation:

$$HBR_T = Num_{MB}/Num_{NB} + Num_{MB}, \tag{5}$$

where Num_{MB} is the number of malicious network communications. Num_{NB} represents the number of normal network communications.

CS of the device under test is as follows:

$$\mathrm{CS} = \lambda_1 * PLR + \lambda_2 * MDR + \lambda_3 * CDR + \lambda_4 HBR_T. \tag{6}$$

Depending on the application context, each evaluation factor (λ) can be changed, and the sum of the evaluation factors needs to be 1.

The entire communication reputation P of the device under test is calculated by the overall amount of malicious network communication (N_{MB}) and the overall amount of normal communication N_{NB}:

$$P = \frac{e^{N_{NB}}}{e^{N_{NB}} + e^{N_{MB}} + 1}. \tag{7}$$

5. Simulation

To assess the method's effectiveness of malicious node detection and the network architecture, the corresponding experiments are designed based on NS2. This paper used Ethereum to build B_{bdmn}. The experiment evaluated the method's effectiveness from several aspects, as shown in Table 1.

5.1. Detection Rate. When the number of total nodes is 80, the influence of the different numbers of malicious nodes on the detection rate is tested. There are 10 rounds of testing:

$$\text{detection rate} = \sum N_{tmnd}/(N_{tmn} * R). \tag{8}$$

N_{tmnd} is the number of times that maliscious nodes are detected in each round. N_{tmn} presents the number of total malicious nodes. R is the number of rounds.

In terms of the detection rate of malicious nodes, MMNDB and the approach proposed in [17] are compared in Figure 6. In the initial stage, the detection rates of the two scenarios are nearly identical, as seen below. As the simulation progressed, the number of malicious nodes in the network grew, and the figure shows that MMNDB has a greater detection rate than the technique in [17]. This is because the strategy described in this research uses more trustworthy data to detect malicious nodes. Furthermore, this article uses a combination of attributes to determine

Table 1: Simulation parameter.

Parameter	Value
Simulation area	1000 m ∗ 1000 m
MAC protocol	802.11DCF
Initial battery power	10 J
Transfer protocol	UDP
Packet size	1040byte

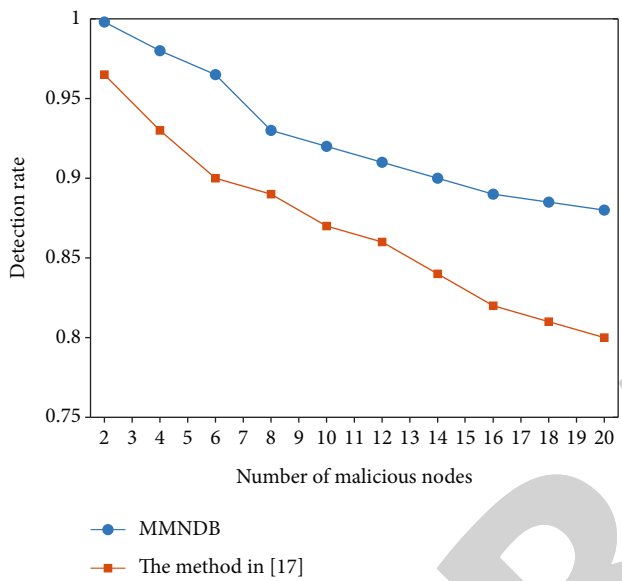


Figure 6: The detection rate of malicious nodes.

whether or not a node is malevolent. When it comes to detecting malicious nodes, MMNDB has a higher sensitivity.

5.2. False-Positive Rate. When the number of malicious nodes is 20, the influence of the different number of total nodes on the false-positive rate is tested. The false-positive rate of the two methods is compared in Figure 7. There are ten rounds of testing:

$$\text{false positive rate} = \sum N_{fp}/(N_{tmnd} * R). \tag{9}$$

N_{fp} is the number of false positives. N_{tmnd} is the number of times malicious nodes are detected in each round. R is the number of rounds.

The chart demonstrates that at the initial simulation stage, the false-positive rate is about the same. When the total number of nodes reaches 60, the false-positive rate of this technique is much lower than the algorithm in [17]. The false-positive rate of the algorithm in [17] increases dramatically as the number of total nodes grows. Because the more nodes there are in the network, the more difficult it is for the management system to monitor the behavior of nodes. Then, the attacker can tamper with communication data or detection algorithms easily. The normal nodes can be defiled as malicious nodes by the attacker to destabilize the network. However, MMNDB can ensure integrity of

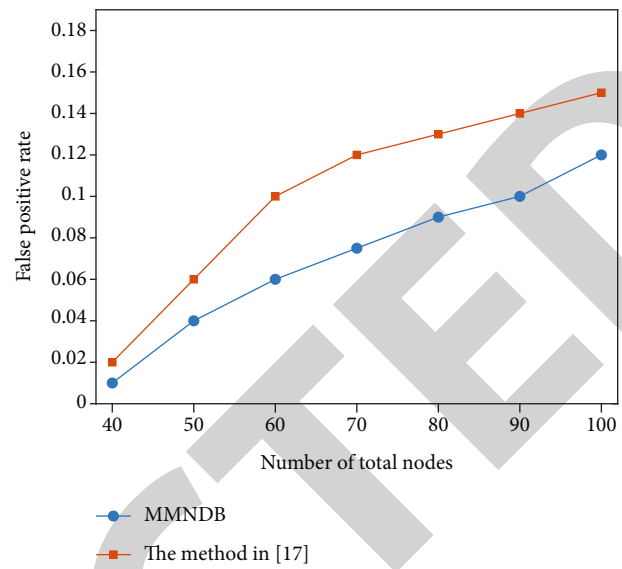


Figure 7: False-positive rate.

communication evaluation data and credibility of detection process based on blockchain. Therefore, this approach has a lower false-positive rate than the comparison algorithm. Moreover, it can ensure that the security and stability of the network operation are better.

5.3. Detection Time. When the number of malicious nodes is 20, the number of total nodes is 80. The detection time of the two methods is compared in Figure 8. As shown in the figure, the number of malicious nodes in the power Internet of Things decreases gradually with the increase of detection time. However, the proposed method can detect more malicious nodes simultaneously than [17]. Because the scheme determines the node reputation from more dimensions, the scheme has a higher detection accuracy. At the same time, the scheme uses the smart contract method to automatically execute the malicious node detection algorithm, which has a higher execution speed and can detect and isolate the malicious node in time.

5.4. Mean Throughput Speed. Throughput can be defined as the data successfully transmitted over some time. By testing the throughput of the normal network (NN) and the network with malicious nodes (NWMN), the network uses MMNDB to detect malicious nodes, and the effectiveness of the method proposed by this paper can be proved:

$$\text{Mean Throughput speed} = \frac{\sum PR}{tsp - tst}. \tag{10}$$

where $\sum$PR = size of the total receiving packet, tst = the start time, and tsp = end time.

The throughput of different types of networks was tested under the condition that malicious nodes were randomly distributed with a probability of 30%, as shown in Figure 9. As the number of nodes increases, the throughput increases. However, the throughput of a network with malicious nodes

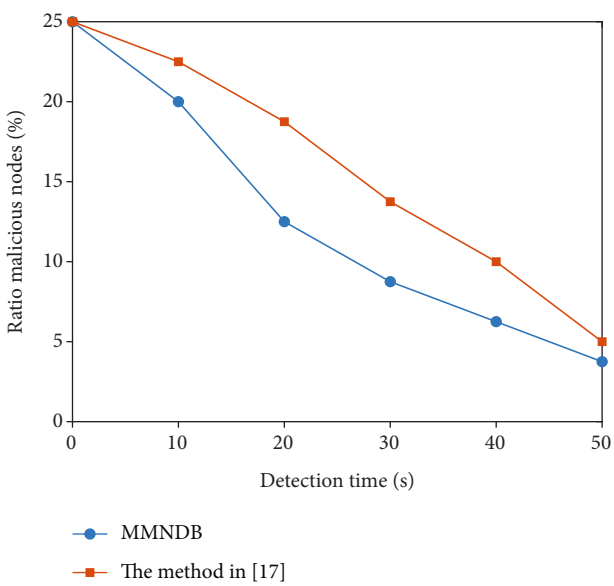


Figure 8: Detection time.

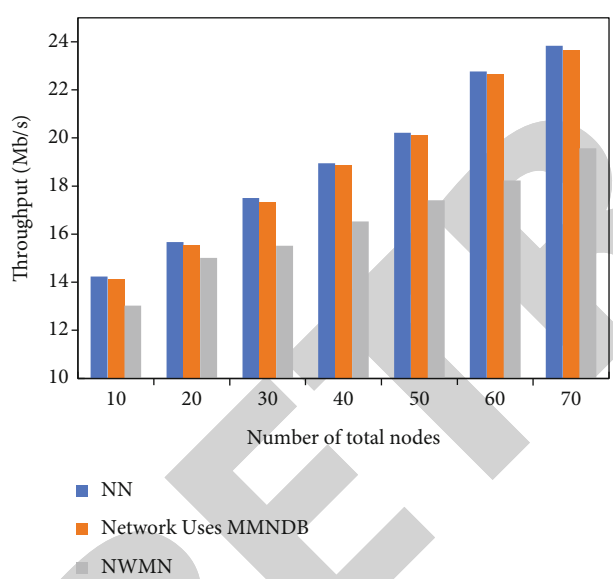


Figure 9: Throughput.

is the lowest. The throughput of network uses MMNDB that is affected by malicious nodes barely. Therefore, it can resist the attacks of malicious nodes on the network and ensure the efficient circulation of the network effectively.

5.5. End-to-End Latency. The overall time it takes for data to travel from the source to the destination is referred to as end-to-end latency, and it encompasses all of the numerous latencies experienced by the package from the sender to the receiver.

The average end-to-end latency (AEL) is as follows:

$$AEL = \frac{\sum(T_{pr} - T_{ps})}{N_{pr}} \tag{11}$$

where T_{pr} = time of receiving the package, T_{ps} = time of sending the package, and N_{pr} = number of received packages.

The end-to-end latency of different types of networks was tested under the condition that malicious nodes were randomly distributed with a probability of 30%, as shown in Figure 10. As the number of nodes increases, the latency of networks with malicious nodes increases sharply. However, the latency of the network uses MMNDB that is influenced by malicious nodes slightly. Therefore, it can reduce the negative effect of malicious nodes on the whole network effectively.

5.6. Package Delivery Rate. The package delivery rate (PDR) represents the success rate of packet transmission within a given time interval. It is as follows:

$$PDR = \frac{\sum N_r}{\sum N_g} * 100, \tag{12}$$

where N_r = number of package received and N_g = number of package generated.

The package delivery rate of different types of networks was tested under the condition that malicious nodes were randomly distributed with a probability of 30%, as shown in Figure 11. As the number of nodes increases, the package delivery rate decreases. However, the PDR of networks with malicious nodes goes down dramatically. Meanwhile, the network uses MMNDB to accurately isolate the malicious nodes to ensure that the packet loss rate is reduced efficiently.

5.7. Network Quality Evaluation. The evaluation index is the state parameter that reflects the quality of network service. Establishing evaluation indexes and building a network quality evaluation model can quantify the evaluation process, and the objective evaluation of network service quality can be achieved. In order to make the network service quality reference, this paper chooses the following evaluation indicators for research:

(1) Stability refers to the degree to which the measured values of various indicators in the network deviate from the normal values after being disturbed and affected. In this experiment, the stability of each service throughput needs to be calculated. The less the deviation, the better the stability of the network; otherwise, the worse the stability of the network. The specific calculation formula is as follows:

$$S = |X_i - \bar{X}|. \tag{13}$$

S denotes the system's stability, X_i is the measured value of each indicator, and $\bar{X}$ is the average measurement value of each indicator.

(2) Response time represents the time from the start of the request message from the IoT terminal to the end of service execution. It includes data transmission time, requests queuing time, system processing

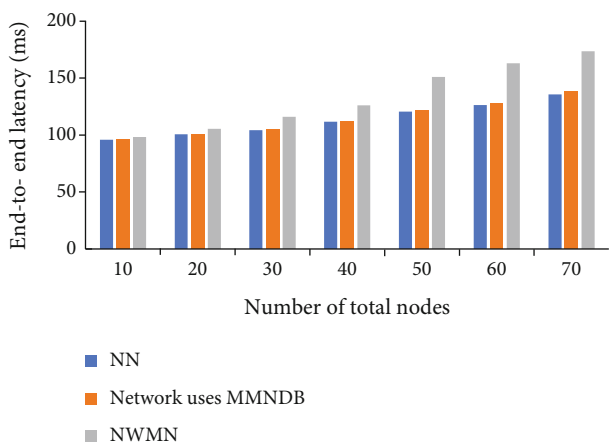


Figure 10: End-to-end latency.

time, data read, and write time. Response time is the indicator of the timeliness of the system to provide services. The shorter the response time, the better the system performance and the higher the service quality

(3) Utilization refers to the hardware resource overhead of an IoT system when executing a process. This paper uses server CPU utilization and link utilization as evaluation indicators. Excessive utilization may lead to equipment consumption. If the utilization rate is too low, it is not easy to utilize resources fully. Therefore, for the whole Internet of Things system, the average utilization rate is taken to reflect the overall situation of the system. Figure 12 shows the evaluation index of this experiment

According to the above indexes, the original evaluation index matrix V in the network quality evaluation problem is established as

$$\boldsymbol{V} = \begin{bmatrix} v_{11} & v_{12} & \cdots & v_{1n} \\ v_{21} & v_{22} & \cdots & v_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ v_{m1} & v_{m2} & \cdots & v_{mn} \end{bmatrix}. \tag{14}$$

As the value range of each attribute value in the original evaluation index matrix is not uniform, data preprocessing of matrix V is required. In this paper, the range method is used to get the standard evaluation matrix Z. In the evaluation process, it is necessary to assign weights to relevant evaluation indicators. At the moment, there are two types of weighting methods: subjective weighting and objective weighting. Subjective weighting reflects the preference of decision-makers, but subjective factors influence the results. Objective weighting is based on mathematical theory, but it ignores the preference of decision-makers. In order to effectively avoid the limitation of single weight, this paper adopts the combined weight optimization model based on the least square method to obtain the optimal weight value. We define grouping weights as $\omega = (\omega_1, \omega_2, \cdots, \omega_n)$. The specific operations are as follows:

(1) Subjective weight $\boldsymbol{U} = (u_1, u_2, \cdots, u_n)$ can be determined by modified step analysis

(2) Objective weight $\boldsymbol{V} = (v_1, v_2, \cdots, v_n)$ can be determined using the improved G1 method

(3) Lagrangian functions are constructed using least squares models:

$$\begin{aligned} L(\omega, \lambda) &= g(\omega) + \lambda\left(\sum_{j=1}^{m} \omega_j - 1\right) \\ &= \sum_{i=1}^{m}\sum_{j=1}^{n}\left\{\left[(u_j - \omega_j) z_{ij}\right]^2 + \left[(v_j - \omega_j) z_{ij}\right]^2\right\} \\ &\quad + 4\lambda\left(\sum_{j=1}^{m} \omega_j - 1\right). \end{aligned} \tag{15}$$

(4) The optimal comprehensive weight was obtained by Lagrange function:

$$\boldsymbol{\omega} = \boldsymbol{A}^{-1}\left[\boldsymbol{B} + \frac{1 - \boldsymbol{E}^T \boldsymbol{A}^{-1} \boldsymbol{B}}{\boldsymbol{E}^T \boldsymbol{A}^{-1} \boldsymbol{E}} E\right]. \tag{16}$$

Among them:

$$\begin{cases} \boldsymbol{A} = \operatorname{diag}\left[\sum_{i=1}^{m} {z_{i1}}^2, \sum_{i=1}^{m} {z_{i2}}^2, \cdots, \sum_{i=1}^{m} {z_{\text{in}}}^2\right] \\ \boldsymbol{E} = [1, 1, \cdots, 1]^T \\ \boldsymbol{B} = \left[\sum_{i=1}^{m} \frac{(u_1 + v_1){z_{i1}}^2}{2}, \cdots, \sum_{i=1}^{m} \frac{(u_n + v_n){z_{\text{in}}}^2}{2}\right]^T \\ \boldsymbol{\omega} = [\omega_1, \omega_2, \cdots, \omega_n] \end{cases}. \tag{17}$$

The weighted standardized evaluation matrix is obtained by multiplying the optimal comprehensive weight and standardized evaluation index matrix:

$$Y = \begin{bmatrix} \omega_1 z_{11} & \omega_2 z_{12} & \cdots & \omega_n z_{1n} \\ \omega_1 z_{21} & \omega_2 z_{22} & \cdots & \omega_n z_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ \omega_1 z_{m1} & \omega_2 z_{m2} & \cdots & \omega_n z_{mn} \end{bmatrix}. \tag{18}$$

Based on the weighted normalization matrix, this paper establishes a multi-attribute network quality evaluation model. The combined weight was calculated by MATlab software. The combined weights are shown in Table 2.

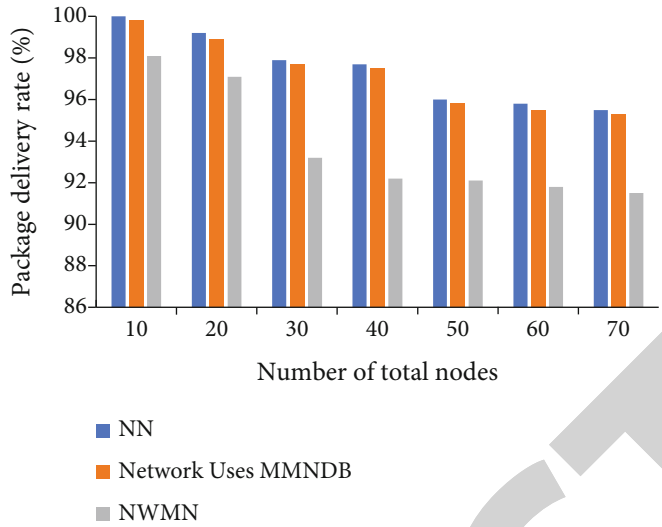


Figure 11: Package delivery rate.

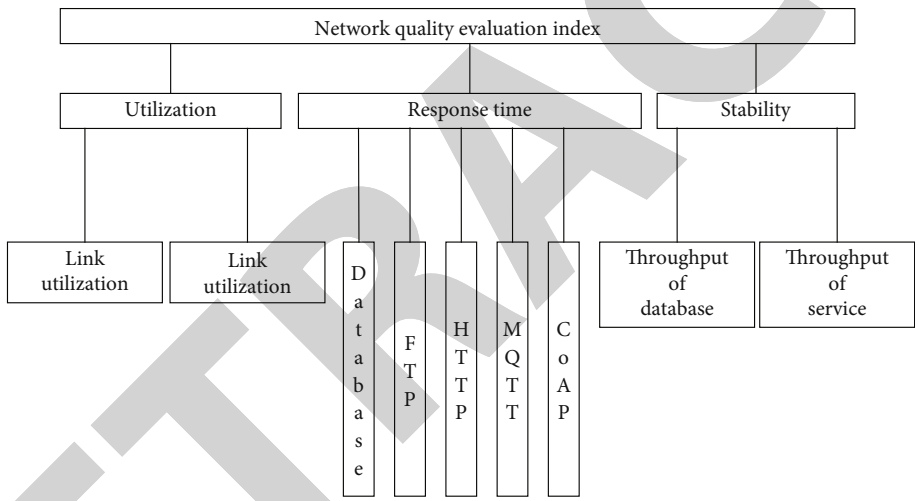


Figure 12: Network quality evaluation index.

Table 2: Combined weights.

Weight	Value
Link utilization	0.0982
CPU utilization	0.0815
Response time of database	0.1054
Response time of FTP	0.0316
Response time of HTTP	0.0316
Response time of MQTT	0.0376
Response time of CoAP	0.0883
Stability of database throughput	0.2431
Stability of database throughput	0.2826

The experimental process of network quality assessment is as follows:

(1) In the initial stage of the experiment, NS2 software was used to build the simulation network. The total number of IoT nodes in the IoT system was 100, including 30 malicious nodes and two servers. MQTT/CoAP protocol was used for communication between IoT nodes and servers. The FTP/HTTP protocol is used to transfer data between the Internet of Things client and server

(2) In the experimental stage, two malicious nodes were added to the Internet of Things system every 5 minutes, and data was collected according to the evaluation indicators given in Figure 12. The network quality assessment model designed in this paper is used to evaluate the network quality between the network architecture that integrates blockchain and trusted computing technology and the traditional IoT architecture every 2 minutes. The results are shown in Figure 13

As shown in Figure 13, with the increase of experiment time, the network quality of the two architectures increases

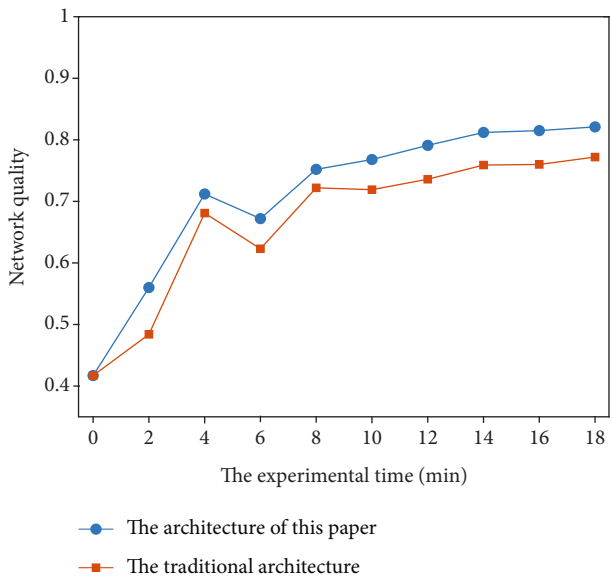


Figure 13: Network quality evaluation experiment.

gradually and eventually tends to be stable, but the network architecture proposed in this paper has better network quality.

5.8. Result Analysis. Based on the above simulation analysis results, under the same conditions, this paper proposed a malicious node detection method with higher accuracy compared with the literature [17], which can effectively guarantee the normal communication of the power Internet of Things. Meanwhile, this method has a lower false-positive rate and detection time. According to a network quality evaluation model, the quality of the proposed network architecture is tested. The test results show that the proposed network architecture can identify malicious nodes quickly and accurately, ensuring the stable operation of the power Internet of Things. The method proposed in this paper uses blockchain technology to store data of communication quality. The corresponding data must be obtained from the blockchain system during detection, so there are certain defects in detection efficiency. However, the method uses the trusted computing technology, which is complicated in function implementation and has disadvantages of poor compatibility.

6. Conclusion

The combination of trusted computing and blockchain technology can achieve trust transfer in trusted computing. However, the current blockchain has the disadvantages of low performance and high hardware requirements. Trusted computing also has the disadvantages of poor compatibility and slower computing speed. Therefore, no one has combined the applicable blockchain technology and trusted computing technology in the power Internet of Things. The high-credibility power Internet of Things architecture that integrates trusted computing and blockchain combines the benefits of trusted computing with the benefits of blockchain. It solves the problem of the centralization of former power IoTs and ensures the safety of every node based on trusted computing. The architecture is based on access control policy verification to ensure the security of terminal access. The existing malicious node detection methods [20–22] are mainly to realize the judgment by analyzing the node's log, behavior, configuration, and other data. However, these data can easily be tampered with by attackers to conceal malicious behavior. As a result, the current method has the disadvantages of low detection rate, high false alarm rate, and long detection time. As a result, at the collection layer, a malicious node detection technique based on blockchain and reputation is designed. The low suspicious terminal is isolated from the high-credibility power information communication network, ensuring the credibility of the collection terminal. In conclusion, the high-credibility power Internet of Things architecture, which combines trusted computing and blockchain, dramatically improves system security while also maintaining the efficiency of power Internet of Things communication.

Although the system architecture proposed in this paper has many advantages above, it still has some defects. Considering the application scenarios comprehensively, the production of trusted computing chips usually needs to be adapted according to the different power equipment, which has poor compatibility and is difficult to be modified according to the application scenarios. At the same time, this architecture uses blockchain to store relevant data, which is insufficient in writing and reading. Therefore, in future work, the focus will be on improving the compatibility of the trusted computing platform and the efficiency of blockchain access.

Data Availability

The result data used to support the findings of this study are included within the article.

Conflicts of Interest

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

Acknowledgments

This work is supported by the Key Science and Technology Project of State Grid Corporation of China, No.5700202019374A0000.

References

[1] W. H. Hassan, "Current research on Internet of Things (IoT) security: a survey," *Computer Networks*, vol. 148, pp. 283–294, 2019.

[2] R. Hou, G. Ren, C. Zhou, H. Yue, H. Liu, and J. Liu, "Analysis and research on network security and privacy security in ubiquitous electricity Internet of Things," *Computer Communications*, vol. 158, pp. 64–72, 2020.

[3] Common Criteria Project Sponsoring Organisation, *Common Criteria for Information Technology Security Evaluation. ISO/*

Hindawi
Wireless Communications and Mobile Computing
Volume 2022, Article ID 7222030, 16 pages
https://doi.org/10.1155/2022/7222030

Research Article

A QoS-Based Fairness-Aware BBR Congestion Control Algorithm Using QUIC

Yi Han,[1] Mengjie Zuo,[1] Huijun Yuan,[1] Yi Zhong,[1] Zhenhui Yuan,[2] and Ting Bi[3]

[1]*School of Information Engineering, Wuhan University of Technology, Wuhan 430070, China*
[2]*Department of Computer and Information Science, Northumbria University, Newcastle Upon Tyne, UK*
[3]*Department of Computer Science, Maynooth University, Co. Kildare, Ireland*

Correspondence should be addressed to Yi Zhong; zhongyi@whut.edu.cn

Received 29 October 2021; Accepted 30 March 2022; Published 29 April 2022

Academic Editor: Basem M. Elhalawany

Congestion control is a fundamental technology to balance the traffic load and the network. The Internet Engineering Task Force (IETF) Quick UDP Internet Connection (QUIC) protocol has flexible congestion control and at the same time possesses the advantages of high efficiency, low latency, and easy deployment at the application layer. Bottleneck bandwidth and round-trip propagation time (BBR) is an optional congestion control algorithm adopted by QUIC. BBR can significantly increase throughput and reduce latency, in particular over long-haul paths. However, BBR results in high packet loss in low bandwidth and low fairness in multi-stream scenarios. In this article, we propose the enhanced BBR congestion control (eBCC) algorithm, which improves the BBR algorithm in two aspects: (1) 10.87% higher throughput and 74.58% lower packet loss rate in the low-bandwidth scenario and (2) 8.39% higher fairness in the multi-stream scenario. This improvement makes eBCC very suitable for IoT communications to provide better QoS services.

1. Introduction

The goal of congestion control is to maximize the use of the bandwidth of the network link, avoid performance loss caused by network congestion, and achieve higher transmission efficiency. Since the 1980s, the transmission control protocol (TCP) has been proposed in order to avoid congestion while improving bandwidth utilization so that the Internet is no longer affected by continuous overload. Delay-based congestion control algorithms represented by Vegas [1] will be completely overwhelmed when competing with loss-based algorithms. Loss-based algorithms such as NewReno [2] and CUBIC [3] have been widely used. However, driven by the advancement of transmission technology and the development of network equipment, the increase in network transmission capacity has caused existing loss-based algorithms to experience buffer bloat [4] and unexpected performance degradation. With the expansion of the queue depth of network intermediate devices, the loss-based algorithm is no longer suitable for today's network environment. The congestion avoidance phase of such algorithms will gradually increase the sending window until the bottleneck queue is filled. Even if the link starts to get congested, the sending window will not decrease. The algorithm can only adopt a speed reduction strategy to relieve congestion when a packet loss event occurs. But at this time, it has already caused great fluctuations in network delay and affected the overall network throughput. Different from loss-based and delay-based algorithms, BBR [5] tries to achieve high link utilization by estimating available Bottleneck link bandwidth (BtlBw) and round-trip propagation delay (RTprop) to calculate bandwidth-delay product (BDP) and also avoid creating queues in bottleneck buffers. BBR is also widely deployed on Google's own production platforms such as the B4 WAN and YouTube. BBR is also open source and integrated into the Linux kernel.

In today's network transmission, most Internet traffic uses TCP as the control protocol of the transport layer to achieve reliable transmission. But TCP is difficult to meet the user scenarios that require higher and higher transmission performance with the development of the Internet. The QUIC [6] launched by Google in 2012 has the characteristics of multiplexing, mandatory encryption, avoiding head-of-line blocking, and can

establish connections with less round-trip time (RTT), which solves the most urgent problem of TCP and helps network services to further improve the user experience. Although most of the research related to transport protocols in the past dozen years has focused on TCP, this exploration is tedious and slow to develop. The emergence of the QUIC protocol has broken this rigid situation. It uses UDP as the basic protocol and has the characteristics of flexible deployment. It does not require the support of the operating system and the kernel and realizes the reliable transmission, congestion control, and flow control of TCP at the application level. QUIC can maintain optimal network conditions when the number of users, network traffic, transmission content size, and network interactive services increase, providing different but more accurate and effective congestion control for each user, and meeting the high real-time, high bandwidth utilization, and low latency transmission requirements of Internet applications.

Both QUIC and BBR have been reported by Google as significant performance improvements and have attracted widespread interest. QUIC and BBR have been studied and evaluated in different network scenarios, respectively [7, 8]. They are both developed by Google and currently used together in some commercial products, such as Chrome browser and YouTube. According to [6, 9], as of 2018, 7% of total Internet traffic and 30% of Google egress traffic are generated by QUIC connections. At present, CUBIC is still the default congestion control in QUIC implementations, but BBR is also optional congestion control in QUIC. BBR uses a feedback-driven autonomous adjustment mechanism to keep the initial value of the congestion window consistent with the capacity of the network so that the network maintains a state of high throughput and low latency. BBR does not use packet loss events as a signal of congestion. It controls the sending rate, further reducing the risk of congestion. Combined with QUIC's flexible congestion control mechanism, the algorithm can be optimized at the application layer and have lower transmission delay. BBR with QUIC is relatively accurate for bandwidth estimation, which helps to shorten transmission delay. It can cope with various types of network packet loss, maintain a high transmission speed when the packet loss rate increases, and cope with network changes well.

However, BBR also has some deficiencies and defects. In transmission links with high latency and bandwidth, the characteristics related to the BBR pacing rate and RTT will cause excessive data transmission rate during the transmission process, resulting in packet loss that seriously affects the quality and efficiency of the transmission. In a multiuser scenario, the network delay is variable due to multiple concurrent data senders in the same network. When the network encounters congestion, the congestion control algorithm must control the transmission rate of the data flow while ensuring the bandwidth utilization, to reduce the delay as much as possible and maintain a high throughput. In this situation where multiple streams compete in the network link, the streams will interact with each other, and the network transmission performance of the multiple streams will be affected. Moreover, BBR determines the pacing rate based on the bandwidth-delay product (BDP), and the transmission rate is maintained at high throughput. When there is a burst of traffic in the network, it will inevitably cause network congestion, and it may restart from the initial state of BBR with a low recovery speed and the cache queue is also cleared.

Combining the characteristics of loss-based and delay-based algorithms, this paper proposes an enhanced BBR congestion control (eBCC) algorithm, which reduces the packet loss rate and packet retransmission chance, while also trying to improve the fairness of the algorithm in multi-stream transmission. At the same time, when the link suffers from a higher delay, eBCC has proved to be able to increase the transmission throughput by reducing the sending rate and packet loss.

The rest of this article is organized as follows. Section 2 introduces the related research of QUIC protocol and BBR algorithm. Section 3 analyzes the original BBR algorithm and elaborates the proposed eBCC algorithm in this paper. In Section 4, we evaluate the proposed algorithm in different network scenarios and perform the analysis of the experimental results. Finally, the algorithm and experimental results are concluded in section 5.

2. Related Work

2.1. Research on Congestion Control Handling. In order to meet the increasing demand for Internet services and solve the problem of rapid growth of network traffic, multiple servers are usually used to improve network performance [10]. However, the problems of network congestion and overloaded servers still arise. Therefore, a load balancer is needed to overcome these problems, by distributing the load of requests and traffic among multiple resources such as servers and network links, in order to improve the overall network performance. Software-defined networking (SDN) is considered to solve the problems of traditional load balancers and plays an important role in network optimization and performance improvement. Hamed et al. [11] implement the "CPU-based" and "CPU-Memory-based" load balancing techniques and evaluate their performance compared to the static round-robin and random-based load balancing techniques using Ryu OpenFlow controller. The results show that the proposed schemes achieve more reliability and higher resource utilization than the round-robin and random-based load balancing strategies. In addition, they have a lower response time and higher transaction rate and throughput. The proposed schemes exhibit more scalability and low-cost characteristics.

In addition to SDN for congestion control and load balancing, there are also congestion control algorithms that can be used to improve network transmission performance. Congestion control algorithms protect the Internet from continuous traffic by adjusting the sending rate, reducing congested sending, and improving bandwidth utilization. The classic congestion control algorithms include delay-based and loss-based algorithms.

Vegas [1] was the first implementation of congestion control algorithm using delay as a congestion signal. The Vegas algorithm defines BaseRTT and sets it to the minimum of all measured RTT. When receiving a duplicate ACK, Vegas checks whether the difference between the current time and the recorded timestamp is greater than BaseRTT. If it is, the packet is retransmitted immediately without waiting for a

third duplicate ACK. When acknowledging a retransmitted packet, Vegas reduces the congestion window by a factor of 3/4. When a non-duplicated ACK is received and it is the first or second acknowledgment after data retransmission, Vegas will check again whether the data transmission interval exceeds BaseRTT. If Vegas exceeds BaseRTT, it will also perform data retransmission operation.

Among the congestion control algorithms based on packet loss, NewReno and CUBIC are typical algorithms that are often used for analysis and examples. NewReno [2] adopts the well-known Additive Increase Multiplicative Decrease (AIMD) mechanism. During slow start, its congestion window (cwnd) increases exponentially with each RTT until the slow start threshold ssthresh is reached. After that, cwnd enters the congestion avoidance phase, adding one packet per RTT cycle. If three duplicate ACKs are received, the fast recovery phase is entered, and the new ssthresh is set to cwnd/2. Then, the cwnd is halved to enter the congestion avoidance phase. However, if a timeout occurs, it will go into slow start and start over. NewReno can distinguish one-time congestion from multiple-time congestion, improving the robustness and throughput after packet loss.

BIC [12] calculates the maximum link capacity and maximizes the congestion window. It can reduce the calculation of large link capacity by binary search and quickly find the optimal congestion window size. CUBIC [3] is an enhanced version of BIC that improves TCP friendliness and RTT fairness and simplifies BIC window control. After a packet loss event occurs, the window growth function of CUBIC is controlled by the cubic function of time and provides good stability and scalability. Compared with AIMD based on cubic function, CUBIC adopts a different mechanism. After cwnd is decreased, cwnd rises in a concave shape until it reaches the value of cwnd before the decrease. After that, CUBIC increases its growth rate and rises in a convex shape. CUBIC is currently the default congestion control algorithm in Linux.

2.2. Research on QUIC Protocol. QUIC is developed based on UDP, and its design philosophy enables it compatible with the safety and reliability of TCP and the fast speed of UDP. Kharat et al. [13] explored the QUIC function using an experiment implemented with a local test bench connected to a wireless access point in the campus network environment. The experimental results show that QUIC maintains excellent performance in the form of throughput and TCP/IP acceleration. The fairness of QUIC under competing traffic conditions was also checked, and it was found that it performed well in long-lived traffic.

Transmission protocols continue to evolve to meet the emerging needs of users and new services. Corbel et al. [14] specifically analyzed the protocol fairness when TCP and QUIC streams coexist on the wireless link of the mobile network. The results show that when the loss rate of the mobile network is low, the emulation of multiple TCP connections, the limitation of the congestion window size, and the use of hybrid start (hystart) have a great impact on fairness. The incorrect setting of the default parameters of these mechanisms or the activation of the hystart option will affect the performance of the transmission protocol and therefore also affect fairness.

With the increasing wide-ranging interest in the flexibility and rich features of QUIC, Biasio et al. [15] demonstrated the native implementation of QUIC in ns-3 and illustrated the implemented functions, the main assumptions, differences related to QUIC Internet-Drafts, and a set of examples. Our paper uses this implementation of QUIC in ns-3 to conduct our experiments.

Soni et al. [7] conducted an in-depth study of the QUIC protocol from the perspective of its implementation and application. The authors used Amazon AWS services for test bench implementation to evaluate the performance of the QUIC protocol against TCP and UDP protocols. They found that QUIC performed better than TCP in terms of throughput and data retrieval time. However, its performance is between TCP and UDP. In addition, it provides the fast data transmission of UDP and the reliability of TCP. At the same time, various problems in the QUIC protocol were discovered, such as forward secrecy, replay attacks, and denial of service attacks.

2.3. Related Research on BBR Congestion Control Algorithm. Scholz et al. [8] proposed a publicly available framework for repeatable TCP measurement based on network simulation, designed to analyze the TCP BBR algorithm. They reproduced and confirmed the weaknesses of the current BBR implementation and provided further insights. The two main problems were discovered: One is that bandwidth may be shared unfairly and the second problem is that it takes too long before the bandwidth balance is restored. The behavior on the synchronization between BBR streams was analyzed, showing that it reached a fair balance of long-lived streams.

Zhang et al. [16] revealed that the aggressiveness of BBR would reduce the performance of CUBIC and the entire Internet transmission. They proposed a simple and effective solution based on BBR, Modest BBR. The core idea of Modest BBR is to reduce retransmission and certain aggression by sacrificing a small amount of bandwidth, to obtain better fairness through loss-sensitive methods. Through a large number of test bench experiments and Mininet simulations, it is found that Modest BBR can maintain high throughput and short convergence time when coexisting with Cubic while improving overall performance and achieving better fairness for loss-based solutions.

Kim and Cho [17] proposed a delay-aware BBR (DA-BBR) congestion control algorithm to alleviate the RTT fairness problem among streams using BBR. A network simulation was carried out using Mininet; the results showed that DA-BBR increased the fairness index by 1.6 times of the original BBR, and the number of packet retransmissions was greatly reduced. Even in competition with RTT five times higher, DA-BBR streams with short RTT exhibit fair throughput.

To improve the fairness between the streams using TCP-BBR congestion control algorithm and the ones using delay-based congestion control algorithm, Jia et al. [18] proposed a TCP-BBR-based congestion control algorithm, which has moderate fairness and is called Modest Fairness BBR (MFBBR). The simulation results on Mininet showed that the algorithm can improve the fairness of BBR when coexisting with Westwood

(a congestion control algorithm based on time delay) [19] and it has better fairness than the delay-based congestion control algorithm.

Song et al. [20] proposed a BBR congestion window scaling (BBR-CWS) scheme, which uses a loss-based congestion control algorithm to improve the interprotocol fairness of BBR. The results of Mininet experiments can confirm that the fairness between BBR-CWS and CUBIC is improved by 73% and has a value of 0.9 or higher in most bottleneck buffer scenarios. In addition, compared with the original BBR, the number of packet retransmissions is reduced by 96%.

2.4. Research on BBR Based on QUIC. In the mobile network environment, due to the common but uncertain fluctuations in round-trip time and random loss events in the air, the congestion window growth is unexpectedly hindered. The single-connection strategy still leads to degraded and highly variable completion time interfaces. To maintain a flexible congestion window for networks with such fluctuations, Qian et al. [21] proposed an intelligent connection management scheme based on QUIC. According to the performance evaluation results obtained from the LTE-A/Wi-Fi test network, the proposed multi-QUIC scheme can effectively overcome the existing limitations in congestion control algorithms such as NewReno, CUBIC, and BBR. The median completion time of a piece of web content can be improved up to 59.1%, and the 95th percentile completion time is improved by up to 72.3%. The significance of this work is to achieve highly robust short-term content download performance in response to the uncertainty of various network conditions and different congestion control schemes.

Wang et al. [22] conducted a preliminary evaluation of the performance of QUIC and BBR congestion control algorithm through GEO satellite Internet access on a private network simulation testbed. The obtained results and analysis confirm that compared with the classic CUBIC, the performance of the new satellite Internet using QUIC with BBR is improved.

Due to the strong interest of the ns-3 community in the QUIC module, Paro et al. [23] proposed some extensions of the ns-3 QUIC module to make it more flexible. Integrate BBR into the QUIC module, and implement the necessary pacing and rate sampling mechanisms, as well as a new scheduling interface with three different scheduling styles. Use the network traffic model in the literature to test the new features to verify whether their performance meets expectations. Our paper uses the code of the extension module BBR when conducting the BBR experiment based on QUIC.

Haile et al. [24] used the scalability of QUIC to enhance BBR; instead of using the ACK rate observed at the sender, it applied a more desirable transfer rate calculated at the receiver. Simulation experiments based on 5G tracking in CloudLab proved that the modified QUIC can significantly reduce latency without any significant impact on throughput. In particular, a 39% reduction in round-trip time (RTT) can be observed in some cases, and the throughput is also reduced by 2.7% in the worst case.

The improvement of the BBR algorithm proposed above is an improvement in the fairness of BBR. The eBCC algorithm proposed in this study can achieve higher throughput and lower packet loss rate than the BBR algorithm in the case of low bandwidth in both single and multiple streaming.

3. Proposed Congestion Control Algorithm

3.1. Calculation of Fairness Index. Since multiple links in the network share the network bandwidth, when congestion occurs, the fairness between various congestion control algorithms and the fairness within the same congestion control algorithm is very important. At present, the standard for judging whether the network is fair has not been uniformly stipulated. There are two mainstream standards for judging whether the network is fair. The first standard is that for links with different round-trip delays or using different congestion control algorithms, each link should occupy the same throughput so that it is fair among the links. Another standard is that each link should have the same network throughput when competing, and sources of the same level should get the same amount of network resources, such as response time, throughput, bandwidth, and cache. This standard uses the Jain's fairness index [25] as an evaluation matrix, which is adopted by this paper as a fairness index. The formula is as follows:

$$Jain(x) = \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{\left(n \times \sum_{i=1}^{n} x_i^2\right)}, \tag{1}$$

where n represents n senders and x_i is the throughput of the i-th link. When multiple data streams compete for bandwidth, the result ranges from 1/n to 1. When all connections are competing for network bandwidth, $Jain(x) = 1$ indicates the fairest situation, while $Jain(x) = 1/n$ reaches the minimum, that is, the most unfair situation. And when all users share the same allocated bandwidth, it is the maximum value.

The fundamental reason for the need to pay attention to fairness is that congestion will inevitably lead to data packet loss and thus cause competition among data streams for limited network resources, and data streams with weaker competitiveness will suffer further degradation in QoS. For example, if a congestion control algorithm is based on packet loss when the sender detects that one data packet is lost, it considers that the network is congested and immediately adopts a congestion avoidance strategy to halve the sending window, actively reducing the sending rate of data packets to avoid network congestion. When there are multiple senders and receivers on the transmission link using traditional congestion control algorithms, multiple data streams with different round-trip times compete for the limited bandwidth of the bottleneck link. The duration of the streams with long round-trip time occupying the bottleneck bandwidth is significantly lower than the streams with short round-trip time. This will result in unfair use of network resources. The BBR congestion control algorithm is the opposite of the traditional congestion control algorithm. The calculation of the sending window of the BBR algorithm is related to RTT. Different RTT streams will lead to different sending windows. Long RTT streams are more aggressive than short RTT streams, and they can send more data in a longer period than short RTT streams,

occupying a lot of bandwidth during transmission. This feature enables it to steal bandwidth from other streams, seize more link bandwidth, impair the transmission performance of short RTT streams simultaneously transmitted on the same path, and in the end, result in a decrease in fairness within the protocol.

3.2. BBR Congestion Control Algorithm. Congestion often occurs in the smallest bandwidth section of a link. This smallest bandwidth is the bottleneck bandwidth of the link, and its size determines the maximum transmission throughput. Packet loss and RTT are not considered congestion control factors in BBR. In BBR, the network link is regarded as a "pipe." The diameter and length of this "pipe" are represented by BtlBw and RTprop, respectively, to estimate the maximum bandwidth and minimum delay of the current network. The volume is BDP (the product of BtlBw and RTprop). When the total amount of data in the link equals BDP, that is, when the sending window is equal to BDP, the throughput is maximized. When the sending window is smaller than BDP, the RTT is a fixed value because the buffer is not occupied at this time. So there is no queuing delay, and the throughput will increase as the sending window increases. When the sending window exceeds BDP, it starts to fill the buffer. At this time, the RTT will increase when the sending window increases, and the throughput reaches the maximum and stop increasing. When the buffer is full, the excess data packets will be discarded. The BBR algorithm always maintains the sending window size near the BDP value, which can guarantee higher throughput and lower delay. However, when the buffer is full, it is already congested before the buffer overflows, which will cause greater delay and introduce a large number of packet retransmissions [5, 26]. Such a large number of retransmissions will also occupy link bandwidth and cause waste of bandwidth resources. It will not only reduce its transmission efficiency but also bring great damage to the flow of loss-based congestion control algorithms.

Multiple QUIC connections of a single user may be initiated from a host, and when they send data at the same time, this may squeeze the bandwidth of the same shared link. Multi-streaming methods designed to improve throughput usually increase queuing delay and packet loss rate. The increased queuing delay leads to a higher BDP value and thus gives more advantage in its bandwidth competition. The retransmission caused by packet loss is a waste of bandwidth. Both of the above will damage other streams on the bottleneck of the shared link. What's more, in the multiuser scenarios, the streams from different users and multi-QUIC connections from the same user are overlapped and compete in the link, which will cause the loss of fairness not only among streams but also among users [27].

3.3. Enhanced BBR Congestion Control Algorithm. The essence of a loss-based congestion control algorithm is still to attribute packet loss to network congestion. To achieve the maximum throughput, these algorithms continuously increase cwnd until the link buffer is filled, causing packet loss, and then reduce the cwnd to perform congestion control. The delay-based congestion control algorithms act based on queues. When the queue exceeds the set threshold, cwnd is processed to reduce the growth rate of cwnd. If packets are still lost after adaptation, the size of cwnd is reduced.

The compound TCP (CTCP) congestion control algorithm [28] combines loss-based and delay-based congestion control methods, allowing the algorithm to quickly increase the sending rate while obtaining high bandwidth scalability and improved TCP fairness. CTCP is designed and usually used in high bandwidth environment, and it is disabled by default [29]. eBCC adjusts the algorithm transition state and congestion control window using loss and queuing status as decision factors. Instead, CTCP runs the legacy TCP's AIMD algorithm and a delay-based high speed congestion control algorithm. The idea of the eBCC algorithm is to use both packet loss and delay as rate control signals. It is different from the CTCP algorithm for the operation after packet loss and the way of judging the delay as a congestion signal. Packet loss is used as the primary factor for adaptation decisions. The gain of eBCC's cwnd is no longer a fixed value of 2 but is adaptively changed according to whether the packet is lost and whether the queue exceeds the threshold. When there are other streams in the link, eBCC will not always fill the bottleneck link and the buffer. It will reduce the sending rate promptly to reduce the packet loss rate. The role of the queuing status is to judge the buffer. eBCC judges the queuing status by calculating the sending window and RTprop. This is to avoid inaccurate measurement of RTprop when the buffer is being occupied by other streams, preventing the depletion of the buffer. In multiuser/multi-QUIC streaming cases, eBCC makes full use of bandwidth when it is radical. When competing with other streams or having congestions, eBCC will respond in time to reduce the sending rate to relieve congestion, reduce packet loss, and sacrifice a small amount of throughput to improve fairness. Enhanced BBR congestion control algorithm proposed by this article is described in the Algorithm 1. The parameters and symbols used in the algorithm are explained in Table 1.

When there is no packet loss and the queue is empty, the growth of the congestion window refers to the idea of the binary search congestion control algorithm [12]. Because all link connections wish to obtain bandwidth resources quickly and fairly, a faster convergence speed is required. The binary search method can converge faster when it approaches the optimal adaptation point after experiencing packet loss. The size of cwnd can be calculated as follows:

$$cwnd = (cwnd + targetCwnd)/2, \tag{2}$$

where $targetCwnd$ indicates the maximum limit of the congestion window.

The CTCP algorithm is deduced by Vegas, using baseRTT as the estimated value of the packet transmission delay on the network path, and baseRTT will be updated according to the minimum RTT value measured during the transmission process. The following algorithm can be used to estimate the number of data that are waiting to be sent in the queue:

Input: t, *Time, targetCwnd.*
Output: *Throughput, delay.*
Time: the total time of transmission;
diff: the extra packets in the queue;
isLost: the packet loss signal;
Initialize $t \longleftarrow 0$
while $t <= Time$ **do**
  if $diff < \gamma$ **and** $isLost$ = false **and** in ProbeBW state **then**
    keep in ProbeBW state
    $cwnd = (cwnd + targetCwnd)/2$
  elseif $diff >= \gamma$ **and** $isLost$ = false **and** in ProbeBW state **then**
    keep in ProbeBW state
    $cwnd = (pfCwnd + cwnd)/2$
  elseif $diff >= \gamma$ **and** $isLost$ = true **and** in ProbeBW state **then**
    change to ProbeRTT state
  elseif $diff >= \gamma$ **and** $isLost$ = true **and** in ProbeRTT state **then**
    $cwnd = cwnd(1 - \beta)$
  end if
end while
return *Throughput, delay.*

Algorithm 1: Enhanced BBR Congestion Control Algorithm.

Table 1: Description of parameters and symbols.

Parameter/symbol	Description
n	Number of senders
x_i	Throughput of the i-th link (mbps)
$diff$	The extra packets in the queue
γ	The compromise number of packets in the queue
$baseRTT$	The estimated value of the packet transmission delay (ms)
$targetCwnd$	The maximum limit of the congestion window
$pfCwnd$	The size of the last congestion window
β	The proportion of cwnd size decrease
access_bandwidth	The bandwidth of the transmission link
bottleneck_bandwidth	The bandwidth of the bottleneck link
access_delay	The delay of the transmission link
bottleneck_delay	The delay of the bottleneck link

$$Expected = \frac{cwnd}{baseRTT}, \tag{3}$$

$$Actual = \frac{cwnd}{RTT}, \tag{4}$$

$$diff = (Expected - Actual) * baseRTT. \tag{5}$$

where RTT is the currently measured RTT value and $diff$ is the size of the data packet in the queue. *Expected* refers to the throughput estimate obtained without exceeding the network path, and *Actual* represents the real throughput.

γ is a compromise between fairness and throughput with a value of 30 according to [26+2]. If $diff < \gamma$, the network link is considered to be underutilized. If $diff > = \gamma$, eBCC perceives that the network is fully utilized and the number of data packets in the queue exceeds the specified threshold, and the data transmission rate needs to be reduced accordingly.

$$pfCwnd = 2cwnd - targetCwnd, \tag{6}$$

$$cwnd = (pfCwnd + cwnd)/2. \tag{7}$$

Equations (6) and (7) change the size of the current congestion window to the average value of current *cwnd* and *pfCwnd*. *pfCwnd* is the size of the last congestion window, that is, the size of *cwnd* when the queue does not exceed the specified threshold. This enables the *cwnd* to increase smoothly when data packets are accumulated in the queue without packet loss.

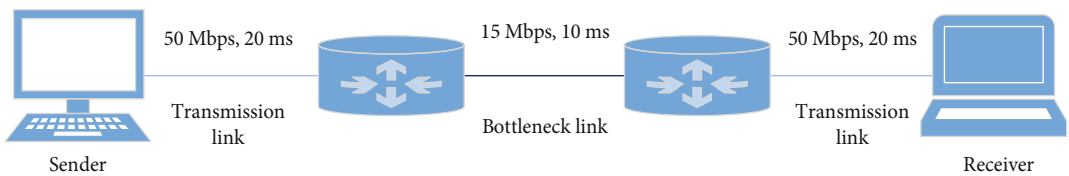


Figure 1: Network simulation for the single-stream scenario.

When a packet loss is detected, the transmission enters the congestion avoidance phase. When the BBR is still in the stable state of ProbeBW, the ProbeBW state is transferred to ProbeRTT. If the BBR is already in the ProbeRTT state, adjust the congestion window size:

$$cwnd = cwnd(1 - \beta), \tag{8}$$

where the value of β is 0.5 according to TCP Reno and NewReno algorithms so that the congestion window can converge to the optimal point faster while the transmission rate is reduced.

4. Experimental Results and Analysis

4.1. General Experimental Setup. The general experimental setup of this paper uses a virtual machine of the Ubuntu operating system with a Linux kernel version of 20.04. We mainly use the extension modules of QUIC [15] and BBR [23] in the network simulator version 3 (ns-3) to simulate the network environment [30]. Two network topologies are created to study the performance of the proposed eBCC algorithm in single-stream and multi-stream scenarios, respectively.

Associated with the Ethernet protocol, the maximum transmission unit in TCP is 1500 bytes, which determines the maximum size of each packet in any transmission. The packet size of the data stream is set to 1400 Bytes, and the transmission interval is set to 100 ms. The total transmission duration is 20 seconds. The transmission starts at end of the first second, which means that the entire process of data streaming lasts for 19 seconds.

4.2. Experiment 1: Varying CWND, RTT, and Throughput of a Single Stream. We set the transmission bandwidth and delay for transmission link and bottleneck link to 50 Mbps and 15 Mbps and 20 ms and 10 ms, respectively. The simulated network scenario of the experiment is shown in Figure 1.

Experiment 1 measured the cwnd, RTT, and transmission throughput of several different congestion control algorithms over time. The measurement results are shown in Figure 2, Figure 3, and Figure 4.

It can be seen from Figure 2 that CUBIC and NewReno regard packet loss events as serious matters. Once the packet loss occurs, an adaption decision to reduce the congestion window will be made. The congestion control window of the BBR algorithm maintains the maximum detected congestion window value most of the time. BBR adjusts the congestion window in the retransmission state when it restores the original maximum detection window value and then exits the retransmission state. In addition to adopting the aggressive transmission strategy of BBR, eBCC considers packet loss when making decisions to reduce the congestion control window.

In Figure 3, NewReno's RTT trend is most stable, followed by CUBIC with two sawtooth changes. And RTT of the above two algorithms is low for a relatively long duration. RTT of BBR and eBCC fluctuates greatly. BBR fluctuates around larger RTT for a long duration, and eBCC fluctuates in a sawtooth shape. The sawtooth peak value of eBCC is close to the larger RTT value maintained by BBR for a long time.

In Figure 4, the throughput variations of the four algorithms can be mapped with cwnd changes in Figure 2. NewReno reduces the congestion control window because of packet loss in its congestion control mechanism. The throughput curve encounters twists and turns around at the 3rd second, and there is a certain drop. The CUBIC algorithm is also for the same reason, because it encounters packet loss and reduces the congestion window to reduce the rate, which causes the throughput curve to twist or the curve has a downward trend at 8th second. The reason for the drop in throughput of BBR and eBCC is because it detects serious congestion and restarts directly from the initial state. BBR is more aggressive in the early stages, but eBCC recovers much faster than BBR when dealing with congestion and has higher average throughput.

4.3. Experiment 2: Varying Link Bandwidth and Delay of a Single Stream. Before conducting multi-stream experiments, two preliminary studies are performed to understand the impact of varying network bandwidth and delay of transmission links and the bottleneck link on the streaming performance when using different congestion control algorithms.

First, in the case of varying bandwidth, we have managed to carry out the experiment using a low bandwidth link. According to the related research on BBR in [8], the difference in bandwidth does not affect the validity of the results. The experiments with high bandwidth can use physical hardware devices in the future so that high-definition video or virtual reality video transmission can be performed. This is considered to be the focus of our following work. Using the same topology illustrated in Figure 1, the transmission link and the bottleneck link were tested with four sets of parameters: 75 Mbps and 17.5 Mbps, 50 Mbps and 15 Mbps, 30 Mbps and 10 Mbps, and 15 Mbps and 5 Mbps, respectively. The delay of the two links is set to 20 ms and 10 ms, respectively.

Second, in the case of varying delay, 50 Mbps and 15 Mbps are selected as the bandwidth of the transmission link bandwidth and the bottleneck link, respectively. The delay values of the transmission link and the bottleneck link are divided into three groups: 20 ms and 10 ms, 15 ms and

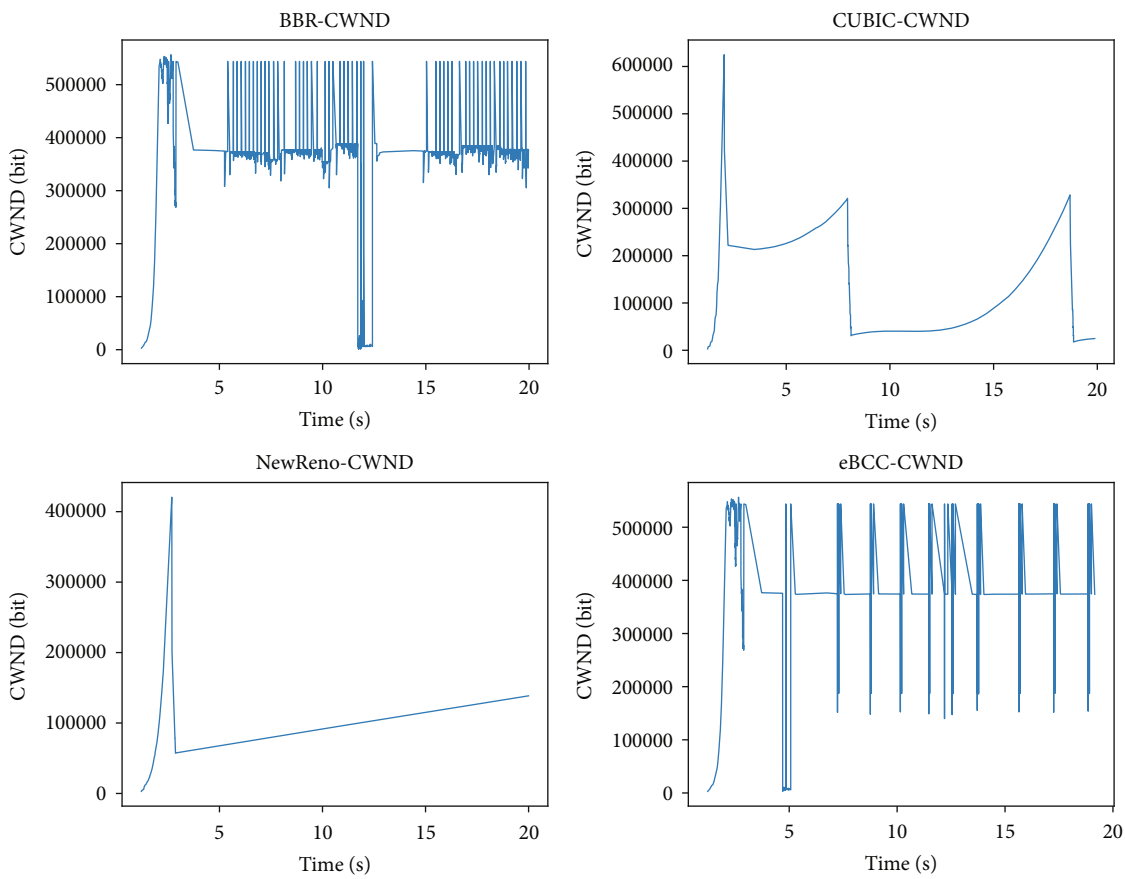


Figure 2: Varying congestion window size during transmission.

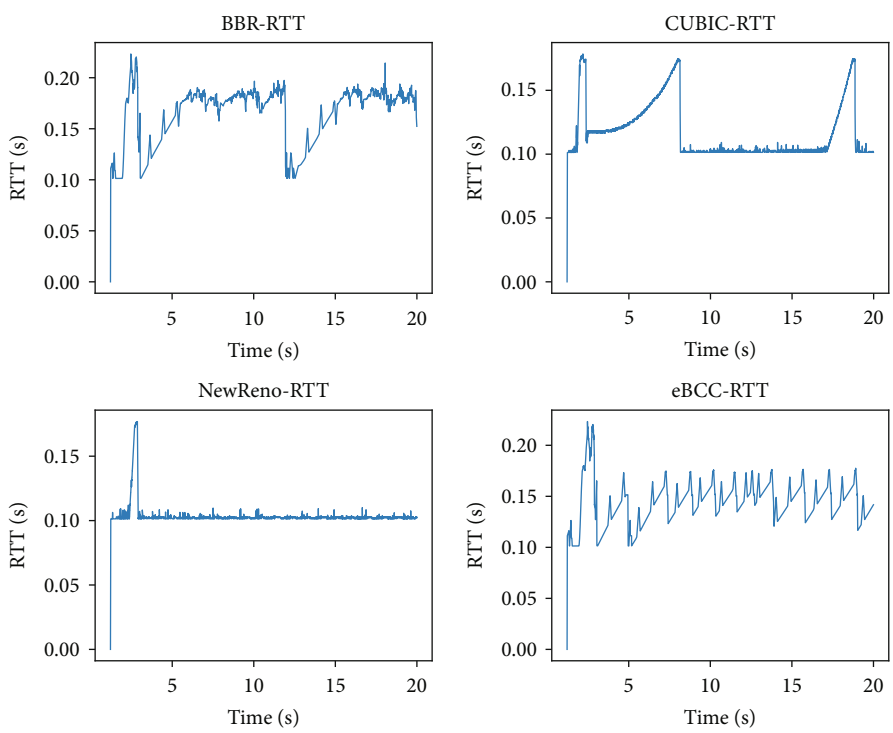


Figure 3: Varying RTT during transmission.

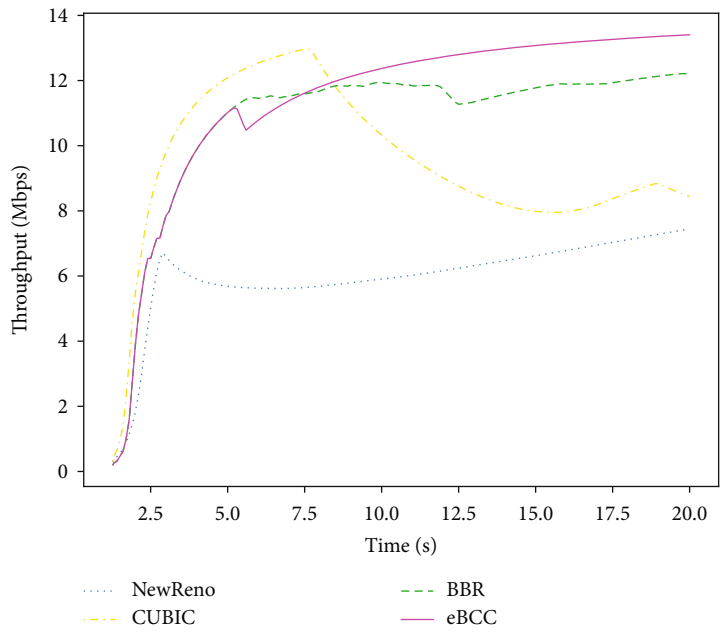


Figure 4: Variation of throughput in real-time transmission.

Table 2: Packet sending/receiving rate (Mbps) for varying bandwidth of different links.

Varied bandwidth(Mbps)	NewReno	CUBIC	BBR	eBCC
access_bandwidth =15; bottleneck_bandwidth =5	4.220/4.188	3.245/3.160	4.776/4.697	4.776/4.697
access_bandwidth =30; bottleneck_bandwidth =10	6.845/6.801	6.509/6.386	9.661/8.870	9.576/9.2662
access_bandwidth =50; bottleneck_bandwidth =15	7.743/7.692	8.840/8.672	14.731/12.465	14.353/13.820
access_bandwidth =75; bottleneck_bandwidth =17.5	8.769/8.708	10.055/9.855	17.413/14.631	16.634/16.087

Table 3: Packet loss rate (%) for varying bandwidth of different links.

Varied bandwidth (Mbps)	NewReno	CUBIC	BBR	eBCC
access_bandwidth =15; bottleneck_bandwidth =5	1.0617	4.8268	2.0855	2.0855
access_bandwidth =30; bottleneck_bandwidth =10	1.008	3.4845	11.1037	4.9245
access_bandwidth =50; bottleneck_bandwidth =15	1.2216	3.4133	19.7338	5.6094
access_bandwidth =75; bottleneck_bandwidth =17.5	1.2343	3.4354	19.9143	5.0625

7.5 ms, and 10 ms and 5 ms, respectively. Both of these experiments were performed under a low link bandwidth and different delay environments, and the experimental results were obtained after repeated measurements. The results are averaged from repeated experiments. The packet sending rate, packet receiving rate, and packet loss rate are shown in Table 2 and Table 3:

Access_bandwidth in Tables 2 and 3 represents the transmission link bandwidth, and bottleneck_bandwidth is the bottleneck link bandwidth. It can be seen from Table 2, Table 3, and Figure 5 that the sending rate of BBR is the highest, but its throughput is not as good as that of eBCC. Table 3 shows that BBR suffers from great packet loss that leads to a large number of packet retransmission and thus its received throughput is lower than the proposed eBCC.

From Figure 5(a) and Table 2, it can be seen that in the four cases, the throughput of the four algorithms all increase as the access_bandwidth increases, and eBCC achieves the highest. The throughput of CUBIC is the lowest when the access_bandwidth is low. When the access_bandwidth increases, the throughput of NewReno becomes the worst among the four algorithms. From Figure 5(b) and Table 3, it can be seen that in terms of packet loss rate performance, NewReno's congestion control is conservative in a rate increase, so packet loss rate is the lowest among the four algorithms, remaining at about 1%. The CUBIC algorithm has the worst packet loss performance

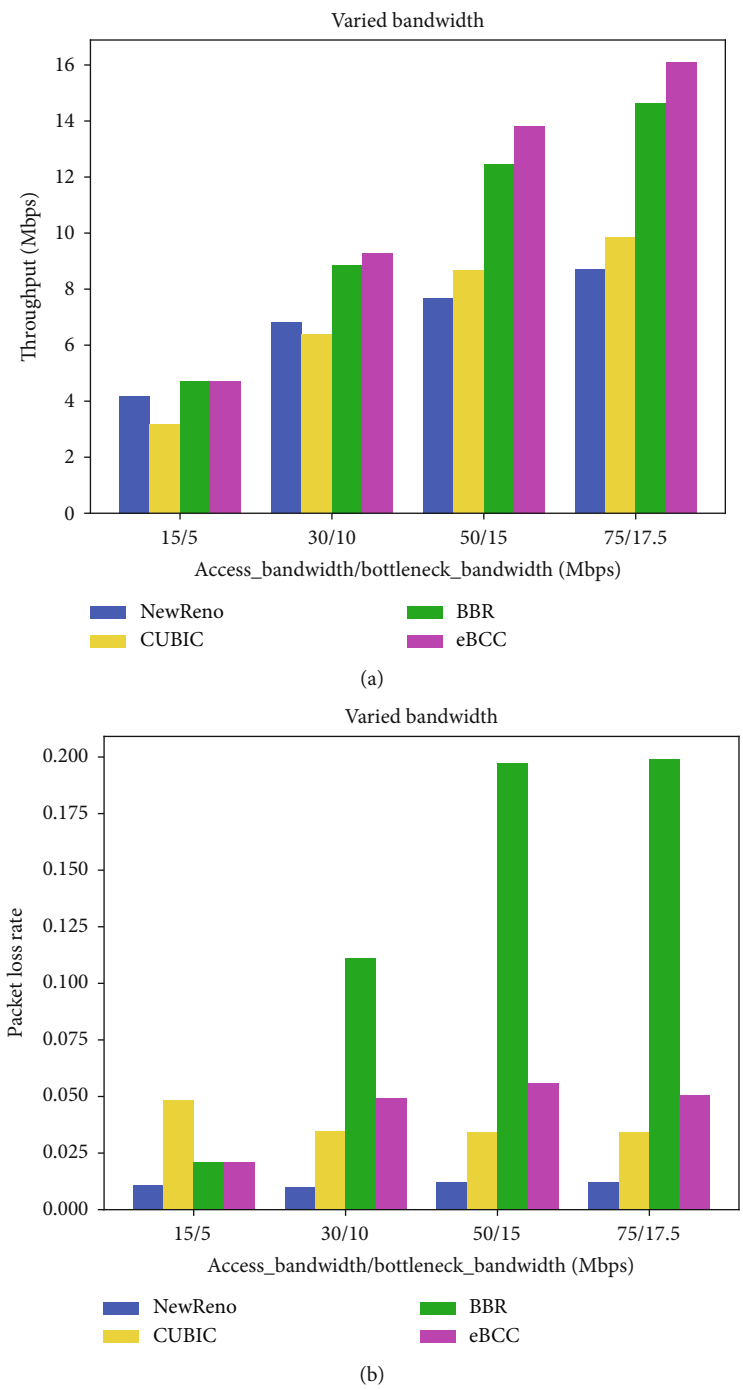


Figure 5: Network performance test results for different link bandwidths and bottleneck link bandwidths. (a) Received throughput. (b) Packet loss rate.

Table 4: Packet sending/receiving rate (Mbps) for different link delays and bottleneck link delays.

Varied delay (ms)	NewReno	CUBIC	BBR	eBCC
access_delay =20; bottleneck_delay =10	7.743/7.692	8.840/8.672	14.126/12.037	14.353/13.823
access_delay =15; bottleneck_delay =7.5	11.737/11.678	10.717/10.593	14.650/13.099	14.469/14.036
access_delay =10; bottleneck_delay =5	12.860/12.827	11.154/11.058	14.448/14.290	14.220/14.079

Table 5: Packet loss rate (%) for different link delays and bottleneck link delays.

Varied delay(ms)	NewReno	CUBIC	BBR	eBCC
access_delay =20; bottleneck_delay =10	1.222	3.217	18.928	5.609
access_delay =15; bottleneck_delay =7.5	0.661	2.118	14.023	4.617
access_delay =10; bottleneck_delay =5	0.456	1.573	1.712	1.744

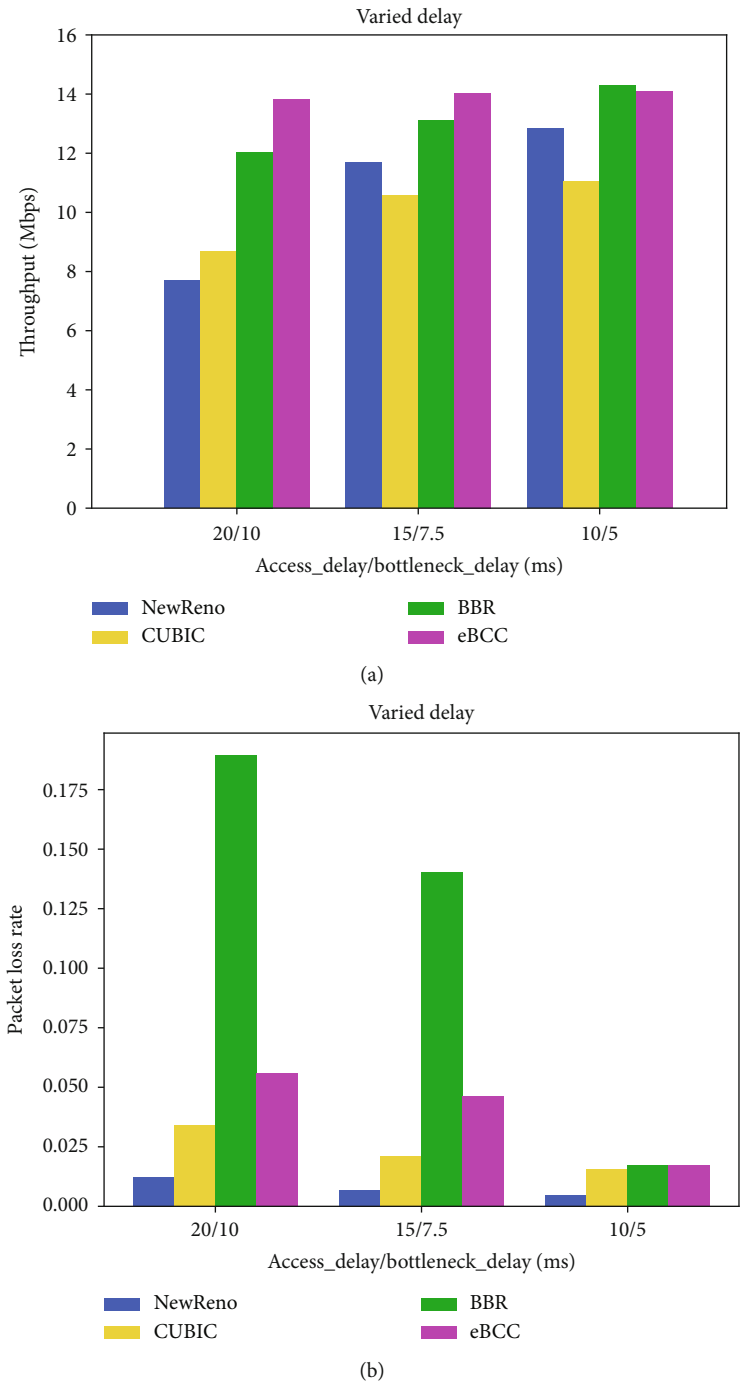


Figure 6: Network performance test results for different link delays and bottleneck link delays. (a) Throughput. (b) Packet loss rate.

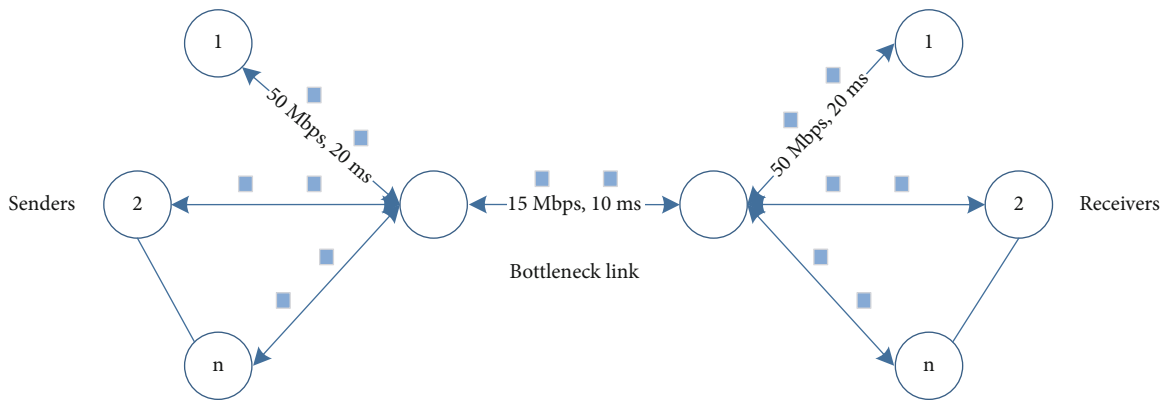


Figure 7: QUIC multi-stream transmission dumbbell network topology.

when the access_bandwidth is low. But when the access_bandwidth increases, the packet loss of CUBIC is stable, maintaining around 3.4%. The packet loss rate of BBR and eBCC is better than CUBIC only when the bandwidth is low, i.e., 15 Mbps and 5 Mbps. In other cases, the packet loss rate of both is relatively high. BBR has the worst performance in terms of packet loss, more than twice that of eBCC.

In Tables 4 and 5, access_delay refers to the transmission link delay, and bottleneck_delay refers to the bottleneck link delay. It can be noted from Tables 4 and 5 and Figure 6 that when the transmission access_delay and bottleneck_delay are 10 ms and 5 ms, BBR has the best throughput performance. eBCC has lower throughput and higher packet loss rate than BBR. When the transmission access_delay and bottleneck_delay are 20 ms and 10 ms and 15 ms and 7.5 ms, the throughput of eBCC is the highest among several algorithms, and the performance of eBCC's packet loss rate is also better than BBR. In these three cases, NewReno maintains the lowest packet loss rate.

Tables 4 and 5 record the packet sending/receiving rate and packet loss rate of NewReno, CUBIC, BBR, and eBCC for different transmission link delays and bottleneck link delays. Inferring from the packet loss rate, in the two sets of experiments with higher delay, the throughput of BBR is lower than eBCC due to retransmission, but in the group with low delay, the performance gap between the throughput and the packet loss rate of the two is not significant.

As shown in Figure 6, Table 4, and Table 5, as the access_delay and bottleneck_delay decrease, the throughput of the four algorithms has different degrees of improvement. The throughput of NewReno and CUBIC is worse than that of BBR and eBCC. Among them, CUBIC performs better than NewReno when the access_delay is 20 ms and the bottleneck_delay is 10 ms, and the throughput of NewReno is higher when the delay is lower. In terms of packet loss, NewReno and CUBIC are lower than BBR and eBCC but also have lower throughput. NewReno performs particularly well in terms of packet loss (around 0.5%) when the delay is low. In the case of higher latency, eBCC reduces cwnd when packet loss is detected to reduce unnecessary data packet retransmissions by slightly slowing the sending rate, which is not implemented by BBR. This not only reduces the packet loss rate but also improves the overall transmission throughput. In the case of lower latency, BBR can maintain high throughput, and the utilization of network links has reached a high peak. In this case, high throughput is a reward for a high packet loss rate risk.

Table 6: Average packet sending/receiving rate (Mbps) per flow in multi-stream transmission.

Number of flows	NewReno	CUBIC	BBR	eBCC
3 flows	4.234/4.209	4.511/4.452	5.245/4.076	5.005/4.527
5 flows	2.796/2.779	2.602/2.554	3.281/2.415	3.105/2.677

Table 7: Average packet loss rate (%) in multi-stream transmission.

Number of flows	NewReno	CUBIC	BBR	eBCC
3 flows	0.857	2.392	28.135	14.488
5 flows	0.929	2.721	35.430	20.443

Table 8: Average fairness index in multi-stream transmission.

Number of flows	NewReno	CUBIC	BBR	eBCC
3 flows	0.999	0.962	0.951	0.996
5 flows	0.994	0.991	0.910	0.986

From Figures 5–6, the delay has a greater impact on packet loss rate, and bandwidth has a greater impact on throughput performance. In the case of higher bandwidth and higher latency, eBCC performs better in throughput, and the performance of packet loss is much better than BBR.

4.4. Experiment 3: Performance Study of Multi-Stream Transmission. The multi-stream transmission experiment sets up multiple nodes in a dumbbell network topology. The experimental network topology is shown in Figure 7. Multiple pairs of clients and servers are defined to simulate different numbers of streams sending and receiving data packets simultaneously. Four congestion control algorithms

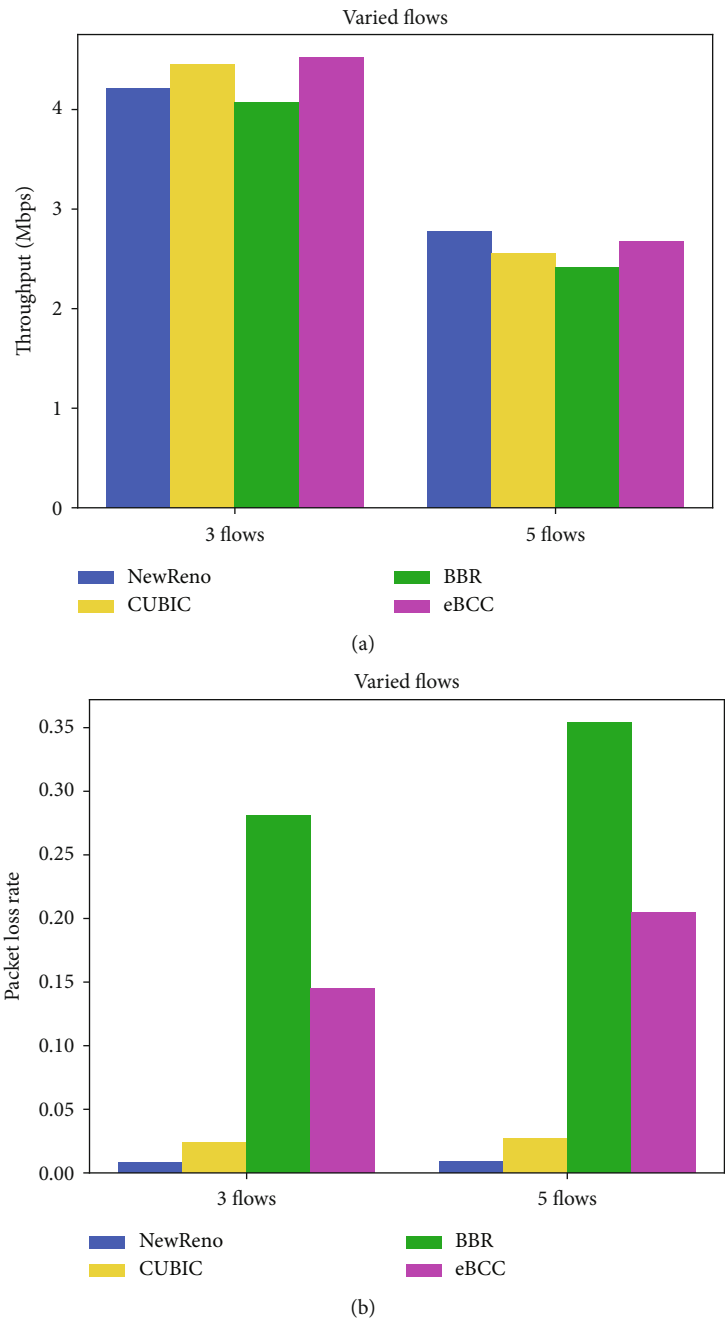


Figure 8: Multi-stream transmission experimental test results. (a) Throughput. (b) Packet loss rate.

are adopted independently to study their performance. The link bandwidth of transmission and bottleneck link is set to 50 Mbps and 15 Mbps, respectively. The experiment is repeated to conduct 3 and 5 different numbers of flows in the network transmission to be compared against a single-stream scenario in Experiments 1 and 2. The experimental results are also the average value obtained through repeated experiments. The fairness index in the multi-stream experiment is calculated by Equation (1).

Table 6, Table 7, and Table 8 show the average sending rate, average packet loss rate, and average fairness index of NewReno, CUBIC, BBR, and eBCC congestion control algorithms. When the packet loss rate is always low, NewReno has good throughput performance in 5 streams. CUBIC has a higher packet transmission rate and throughput in the 3 streams case, but the fairness is the second-lowest. In the case of the 5 streams case, its packet sending rate and throughput are relatively low, but the fairness performance is better than that in the case of 3 streams. The packet sending rate of BBR and eBCC are both very high, but the high packet loss rate damages BBR's throughput significantly. eBCC achieves higher throughput performance in all cases including the single-streaming case compared with BBR by reducing the packet loss rate. While NewReno has the

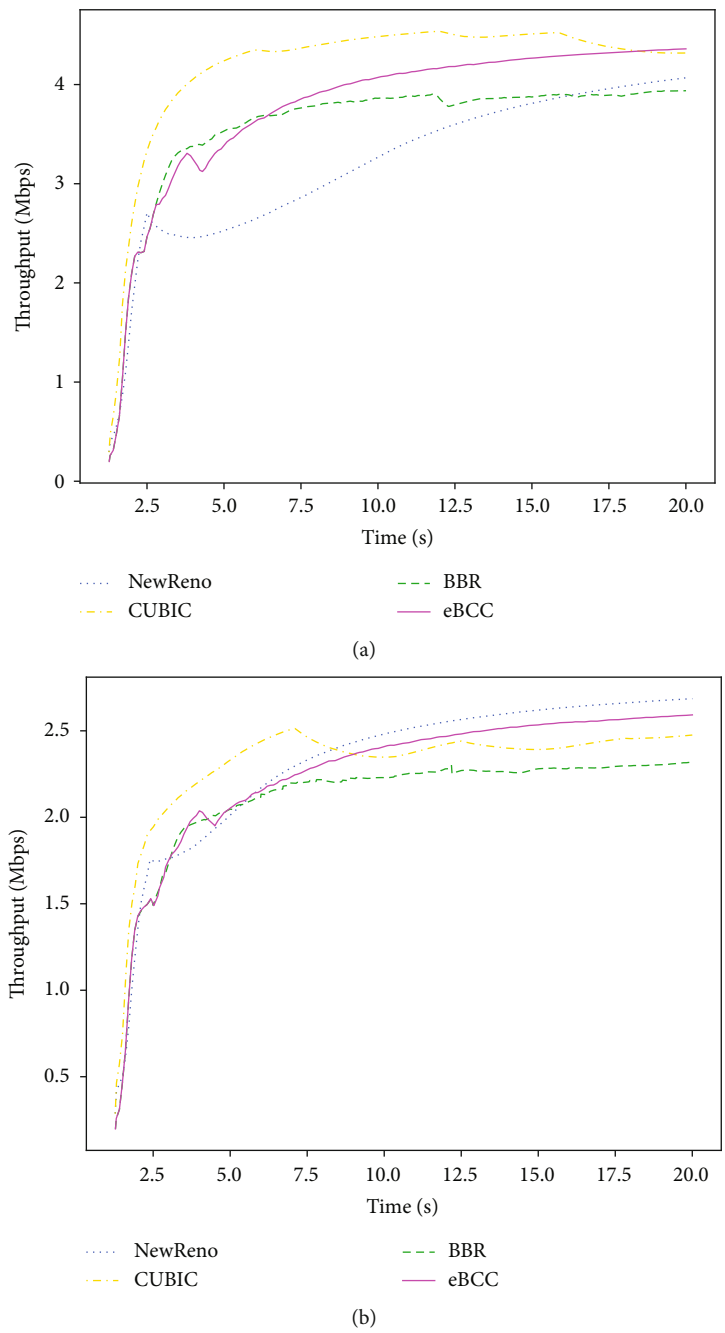


Figure 9: The real-time change of the average throughput in the multi-stream transmission experiment. (a) Three streams case. (b) Five streams case.

highest fairness in both multi-streaming cases, the fairness of BBR is the lowest as a consequence of its radical transmission strategy.

Figure 8 demonstrates a visual performance comparison of these four congestion control algorithms. In multi-stream transmission, the throughput of BBR has always been the lowest. The packet loss rate is the highest among all algorithms. NewReno has always been the holder of the lowest packet loss rate (around 1%). In Figure 8(a), eBCC has the best throughput in the case of 3 streams, and its Jain fairness index is only smaller than NewReno. In Figure 8(b), in the case of 5 streams, NewReno has the highest throughput and the lowest packet loss rate. The throughput of eBCC is only lower than NewReno and higher than CUBIC and BBR. The reason that eBCC's throughput is not as good as NewReno's is due to the high packet loss rate of eBCC. In this case, eBCC's sending rate has not been greatly improved, which leads to a slight gap in the final throughput of eBCC compared to NewReno.

From Figure 9, although the throughput curves of NewReno, BBR, and eBCC have twists and turns, the overall trend is a gradual increase. After the throughput rises to a

certain value, CUBIC starts to decrease due to its congestion control mechanism. Through the comparison of the average throughput curves of 1, 3, and 5 streams, the average throughput curve of a larger number of streams is more aggregated, and the throughput curve of a single stream is more scattered. The throughput curve of NewReno has a smaller decline and a greater rise as the number of streams increases. The throughput of BBR only fluctuates around the value after it reaches a certain value, which is consistent with its control method for obtaining BDP as the sending rate. The throughput curves of eBCC are similar varying from single stream to 3 and 5 streams. In a 5-stream scenario, 5 streams are competing in the link, and aggressive BBR flows can easily cause packet loss, which greatly increases the number of lost packets in the link and ultimately leads to a decrease in throughput performance. Compared with the BBR algorithm, the eBCC algorithm reduces packet loss and improves the throughput performance. The NewReno algorithm does not have a high utilization rate of the link, so it is easy to lead to waste of bandwidth resources. However, the simultaneous transmission of 5 streams makes up for this shortcoming. Therefore, the NewReno algorithm achieves the best throughput performance. The CUBIC algorithm is more aggressive than the NewReno algorithm, and it is easier to lose packets in the competition, resulting in performance loss. Its throughput degrades in the presence of massive packet loss due to stream contention.

Base on the analysis of the three experiments, we noticed that the eBCC algorithm improves the BBR algorithm in two aspects: (1) 10.87% higher throughput in experiment 1 and 74.58% lower packet loss rate in experiment 2 in the low-bandwidth scenario and (2) 8.39% higher fairness in experiment 3 in the multi-stream scenario. In comparison to NewReno and Cubic algorithms, eBCC outperforms both in throughput in most cases while maintaining a high level of fairness. The eBCC algorithm provides better quality for transmission in low bandwidth scenarios, which is applicable IoT communications.

5. Conclusions

In this paper, we proposed the eBCC algorithm under the QUIC protocol. Several experiments were performed to compare its performance against BBR, CUBIC, and NewReno algorithms in multiple network scenarios. In the low-bandwidth link scenario, through changing the bandwidth and delay on both transmission link and bottleneck link, we measured the transmission throughput and packet loss rate of four algorithms. In the simulation, the transmission throughput, packet loss rate, and fairness of the four algorithms were measured. The simulation results show that in the experiment of changing the bandwidth and the delay of the low-bandwidth network link scenario, eBCC achieves the best transmission throughput among the four algorithms. Although its packet loss rate is slightly higher than CUBIC and NewReno, but significantly lower than that of BBR. In the 3-stream scenario under the multi-stream transmission simulation, eBCC also gains the highest throughput, and second-highest fairness, only slightly lower than that of NewReno. In the 5-stream scenario, the throughput performance of the four algorithms is not much different. eBCC is slightly worse than NewReno in throughput performance and still performs better than BBR in fairness, throughput, and packet loss. In view of the above advantages, eBCC can obtain better QoS quality in the transmission of low-bandwidth links in IoT communications.

The experiment of the high bandwidth link will be carried out in the future. At the same time, our multi-stream experiment in this article is only an experiment between the same congestion control algorithm and the same RTT flow, and it does not compare the network performance between different congestion control algorithm flows and different RTT flows. Moreover, the experiment in this article is carried out in a wired dumbbell network simulated by ns-3 and has not been tested in a wireless network, a wide area network, or a cellular network. It is expected that these parts of the experiment will be completed in our following work.

Data Availability

The simulation 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

This work was supported by a grant from the National Natural Science Foundation of China (Grant No. 61801341). This work was also supported by the Research Project of Wuhan University of Technology Chongqing Research Institute and the Science Foundation Ireland (SFI) Industry Fellowship Programme under Grant Number 19/IFA/7445(T).

References

[1] L. S. Brakmo and L. L. Peterson, "TCP Vegas: end to end congestion avoidance on a global internet," *IEEE Journal on Selected Areas in Communications*, vol. 13, no. 8, pp. 1465–1480, 1995.

[2] S. Floyd, T. Henderson, and A. Gurtov, "The NewReno modification to TCP's fast recovery algorithm," *Technical Report*, 2004.

[3] S. Ha, I. Rhee, and L. Xu, "CUBIC: a new TCP-friendly high-speed TCP variant," *ACM SIGOPS operating systems review*, vol. 42, no. 5, pp. 64–74, 2008.

[4] J. Gettys and K. Nichols, "Bufferbloat: dark buffers in the internet," *Communications of the ACM*, vol. 55, no. 1, pp. 57–65, 2012.

[5] N. Cardwell, Y. Cheng, C. S. Gunn, S. H. Yeganeh, and V. Jacobson, "BBR: congestion-based congestion control: measuring bottleneck bandwidth and round-trip propagation time," *Queue*, vol. 14, no. 5, pp. 20–53, 2016.

[6] A. Langley, A. Riddoch, A. Wilk et al., "The QUIC transport protocol: design and internet-scale deployment," in *Proceedings of the Conference of the ACM Special Interest Group on Data Communication*, pp. 183–196, CA, Los Angeles, USA, 2017.

[7] S. Mukesh and B. S. Rajput, *Security and Performance Evaluations of QUIC Protocol*, Data Science and Intelligent Applications. Springer, Singapore, 2021.

[8] D. Scholz, B. Jaeger, L. Schwaighofer, D. Raumer, F. Geyer, and G. Carle, "Towards a deeper understanding of TCP BBR congestion control," in *IFIP Networking Conference (IFIP Networking) and Workshops*, pp. 1–9, Zurich, Switzerland, 2018.

[9] J. Iyengar and M. Thomson, "QUIC: a UDP-based multiplexed and secure transport," *Internet Engineering Task Force, Internet-Draft draft-ietf-quic-transport-27*, 2020.

[10] S. Kaur and J. Singh, "Implementation of server load balancing in software defined networking," in *Information Systems Design and Intelligent Applications*, pp. 147–157, Springer, New Delhi, 2016.

[11] M. I. Hamed, B. M. ElHalawany, M. M. Fouda, and A. S. T. Eldien, "A novel approach for resource utilization and management in SDN," in *2017 13th International Computer Engineering Conference (ICENCO)*, pp. 337–342, Cairo, Egypt, 2017.

[12] L. Xu, K. Harfoush, and I. Rhee, "Binary increase congestion control (BIC) for fast long-distance networks," *IEEE Conference on Computer Communications (INFOCOM)*, vol. 4, 2004.

[13] P. K. Kharat, A. Rege, A. Goel, and M. Kulkarni, "QUIC protocol performance in wireless networks," in *IEEE International Conference on Communication and Signal Processing (ICCSP)*, pp. 0472–0476, 2018.

[14] R. Corbel, S. Tuffin, A. Gravey, A. Braud, and X. Marjou, "Impact of QUIC on fairness in mobile networks," in *IEEE 10th International Conference on Networks of the Future (NoF)*, pp. 82–89, 2019.

[15] A. De Biasio, F. Chiariotti, M. Polese, A. Zanella, and M. Zorzi, "a QUIC Implementation for ns-3," *Proceedings of the 2019 Workshop on ns-3*, 2019.

[16] Y. Zhang, L. Cui, and F. P. Tso, "Modest BBR: enabling better fairness for BBR congestion control," in *IEEE Symposium on Computers and Communications (ISCC)*, pp. 00646–00651, Natal, Brazil, 2018.

[17] G. H. Kim and Y. Z. Cho, "Delay-aware BBR congestion control algorithm for RTT fairness improvement," *IEEE Access*, vol. 8, pp. 4099–4109, 2020.

[18] M. Jia, W. Sun, Z. Wang, Y. Yaohua, Q. Hongyu, and M. Kelong, "MFBBR: an optimized fairness-aware TCP-BBR algorithm in wired-cum-wireless network," in *IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, pp. 171–176, Toronto, ON, Canada, 2020.

[19] S. Mascolo, C. Casetti, M. Gerla, M. Sanadidi, and R. Wang, "TCP Westwood: bandwidth estimation for enhanced transport over wireless links," in *Proceedings of the 7th annual international conference on Mobile computing and networking*, 2001.

[20] Y. J. Song, G. H. Kim, and Y. Z. Cho, "BBR-CWS: improving the inter-protocol fairness of BBR," *Electronics*, vol. 9, no. 5, p. 862, 2020.

[21] P. Qian, N. Wang, and R. Tafazolli, "Achieving robust mobile web content delivery performance based on multiple coordinated QUIC connections," *IEEE Access*, vol. 6, pp. 11313–11328, 2018.

[22] Y. Wang, K. Zhao, W. Li, J. Fraire, Z. Sun, and Y. Fang, "Performance evaluation of QUIC with BBR in satellite internet," in *6th IEEE International Conference on Wireless for Space and Extreme Environments (WiSEE)*, pp. 195–199, Huntsville, AL, USA, 2018.

[23] U. Paro, F. Chiariotti, A. A. Deshpande, M. Polese, A. Zanella, and M. Zorzi, "Extending the ns-3 QUIC module," in *Proceedings of the 23rd International ACM Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems*, pp. 19–26, Alicante, Spain, November 2020.

[24] H. Haile, K. J. Grinnemo, S. Ferlin, P. Hurtig, and A. Brunstrom, "WIP: leveraging QUIC for a receiver-driven BBR for cellular networks," in *22nd IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM)*, pp. 252–255, Pisa, Italy, 2021.

[25] K. R. Jain, W. M. D. Chiu, and W. R. Hawe, "A quantitative measure of fairness and discrimination," *Eastern Research Laboratory, Digital Equipment Corporation, Hudson, MA*, vol. 21, 1984.

[26] M. Hock, R. Bless, and M. Zitterbart, "Experimental evaluation of BBR congestion control," in *IEEE 25th International Conference on Network Protocols (ICNP)*, pp. 1–10, Toronto, ON, Canada, 2017.

[27] G. Kim, I. Mahmud, and Y. Cho, "Fairness improvement of BBR congestion control algorithm for different RTT flows," in *International Conference on Electronics, Information, and Communication (ICEIC*, pp. 1-2, Auckland, New Zealand, 2019.

[28] K. Tan, J. Song, Q. Zhang, and M. Sridharan, "A compound TCP approach for high-speed and long distance networks," in *IEEE Conference on Computer Communications (INFOCOM)*, 2006.

[29] X. Wu, M. C. Chan, A. L. Ananda, and C. Ganjihal, "Sync-TCP: a new approach to high speed congestion control," in *17th IEEE International Conference on Network Protocols*, pp. 181–192, 2009.

[30] T. Bi and G. Muntean, "Location-aware network selection mechanism in heterogeneous wireless networks," in *IEEE conference on computer communications workshops (INFOCOM WKSHPS)*, Atlanta, GA, USA, May 2017.

Hindawi
Wireless Communications and Mobile Computing
Volume 2022, Article ID 7600669, 13 pages
https://doi.org/10.1155/2022/7600669

Research Article

MISD-SLAM: Multimodal Semantic SLAM for Dynamic Environments

Yingxuan You,[1] **Peng Wei**,[1] **Jialun Cai**,[1] **Weibo Huang**,[1] **Risheng Kang**,[2] **and Hong Liu**[1]

[1]*Key Laboratory of Machine Perception, Shenzhen Graduate School, Peking University, Beijing 100871, China*
[2]*Department of Mechanical Engineering, KU Leuven, Leuven 3000, Belgium*

Correspondence should be addressed to Hong Liu; hongliu@pku.edu.cn

Received 19 October 2021; Revised 26 February 2022; Accepted 14 March 2022; Published 5 April 2022

Academic Editor: Chi-Hua Chen

Simultaneous localization and mapping (SLAM) is one of the most essential technologies for mobile robots. Although great progress has been made in the field of SLAM in recent years, there are a number of challenges for SLAM in dynamic environments and high-level semantic scenes. In this paper, we propose a novel multimodal semantic SLAM system (MISD-SLAM), which removes the dynamic objects in the environments and reconstructs the static background with semantic information. MISD-SLAM builds three main processes: instance segmentation, dynamic pixels removal, and semantic 3D map construction. An instance segmentation network is used to provide semantic knowledge of surrounding environments in instance level. The ORB features located on the predefined dynamic objects are removed directly. In this way, MISD-SLAM effectively reduces the impact of dynamic objects to provide precise pose estimation. Then, combining multiview geometry constraint with K-means clustering algorithm, our system removes the undefined but moving pixels. Meanwhile, a 3D dense point cloud map with semantic information is reconstructed, which recovers the static background without the corruptions of dynamic objects. Finally, we evaluate MISD-SLAM by comparing to ORB-SLAM3 and the state-of-the-art dynamic SLAM systems in TUM RGB-D datasets and real-world dynamic indoor environments. The results indicate that our method significantly improves the localization accuracy and system robustness, especially in high-dynamic environments.

1. Introduction

Recently, robot technology has been developed rapidly with the wide range applications of the Internet of Things (IoT). Simultaneous localization and mapping (SLAM) is an essential technology for most mobile robots. SLAM system, using the data of its on-board sensors, constructs a map of unknown environment and simultaneously estimates its pose within the map. The on-board sensors carried by the robot to perceive surrounding environments can be divided into two categories, camera and lidar. Visual SLAM, whose main sensor is the camera, has received considerable attention and research efforts in the last few decades. An increasing number of excellent visual SLAM systems have been proposed, such as MonoSLAM [1], PTAM [2], LSD-SLAM [3], and ORB-SLAM1-3 [4–6]. Most of the visual SLAM systems can build 3D geometric map and estimate pose precisely [7, 8] and serve as the baseline for both indoor and outdoor SLAM systems [9]. Moreover, with the development of deep neural networks (DNN) in recent years, many people have begun to integrate visual SLAM with DNN to achieve object detection and semantic segmentation [10–16], which makes the systems able to understand the surrounding environments in semantic level.

Despite the progress of visual SLAM systems, the robustness of SLAM system in dynamic scenes is still a challenge. "Dynamic" means there are dynamic objects in the scenes. According to the motion state, objects can be divided into five cases: (1) immovable objects, such as the wall. (2) Objects with motion properties are moving, such as a moving car or a walking person. (3) Objects with motion properties are in the stationary state, such as a parking car on the

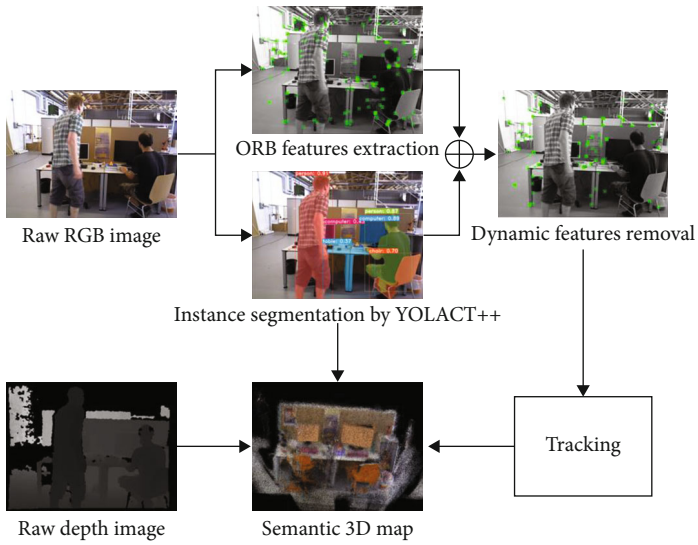


Figure 1: The overview of MISD-SLAM. The raw RGB image is input to perform ORB feature extraction and instance segmentation for semantic information simultaneously. Then, the system removes dynamic features and continues tracking thread to calculate camera pose. Finally, semantic 3D map is built using the data of depth image, camera pose, and semantic information in an independent thread.

side of the road. (4) Objects without motion properties are in the stationary state, such as a static desk. (5) Objects without motion properties are moved, such as a door being opened or closed and tables and chairs being moved. Among these cases, objects with motion properties and objects without motion properties but being moved are defined as dynamic objects in our work. The dynamic objects may lead to the pose estimation inaccurate or failed and make the map corrupted. However, many SLAM works are based on the common assumption that the environments are static. If there are dynamic objects, the motion of objects will be computed into the motion of the camera. Therefore, any dynamic objects in the frame may reduce the accuracy of camera pose estimation or even lead to localization and mapping failure. In this paper, we are focusing on semantic understanding and dynamic robustness in visual SLAM. We propose a semantic visual SLAM system with high performance of accuracy and robustness in dynamic indoor environments, which removes the dynamic objects and reconstructs the static background with semantic information. The overview of the proposed system is shown in Figure 1.

The main contributions of this paper are as follows:

(1) A multimodal semantic visual SLAM system for indoor dynamic environments (MISD-SLAM) is proposed, which significantly increases the accuracy of pose estimation and works more robust in indoor dynamic scenes

(2) A real-time instance segmentation module is proposed to provide semantic knowledge for dynamic objects detection and semantic map reconstruction

(3) A robust tracking strategy is proposed by detecting and removing dynamic features based on the semantic information, which not only reduces the impact of dynamic objects to improve the accuracy of pose evaluation but also remains static features as many as possible to improve the robustness in dynamic environments. Then, the method of multiview geometry constraint removes other dynamic pixels and provides static pixels for map reconstruction without the corruptions of dynamic objects

(4) The high performance of MISD-SLAM in accuracy and robustness is evaluated by the comparison with the state-of-the-art visual SLAM systems on TUM RGB-D datasets [17] and real-world dynamic environments

The rest of this paper is structured as follows: Section 2 discusses an overview of various related work in the fields of visual SLAM with semantic mapping in dynamic environments. Section 3 demonstrates the method of our system in detail. In Section 4, MISD-SLAM is evaluated and compared with the state-of-the-art SLAM systems, DS-SLAM [18], DynaSLAM [19], DetectSLAM [20], SOF-SLAM [21], and SaD-SLAM [22]. And an experiment in real-world environments is carried out to evaluate the performance of the system in real scenes. Finally, Section 5 concludes with a brief conclusion.

2. Related Work

2.1. Semantic Visual SLAM. Traditional visual SLAM mainly focus on geometric information without semantic knowledge of the surrounding environments, which limits the capabilities of robots for high-level tasks. In last few years, with the significant development of deep neural networks (DNN), integrating DNN into visual SLAM to build both geometric and semantic maps has become an important research direction. There are many DNN frameworks. SSD [10] and YOLO [11] can detect objects in boxes. PSPNet [23], SegNet [16], and DeepLab [24–27] are capable to segment objects in pixel level. Moreover, Mask-RCNN [13]

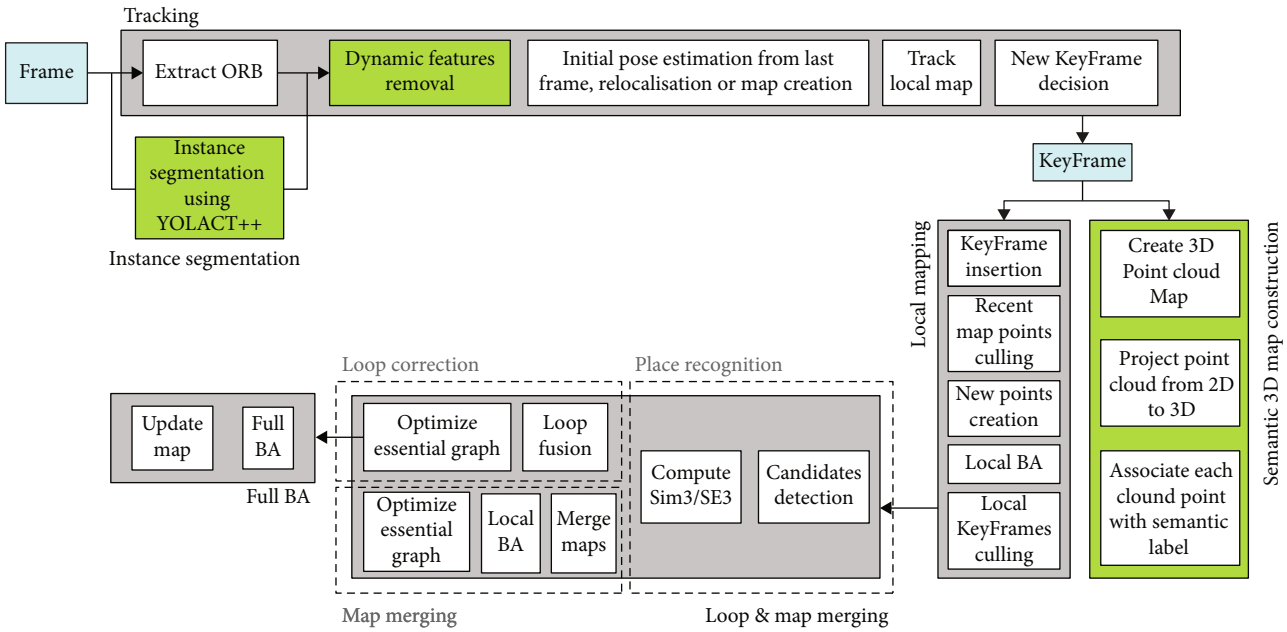


Figure 2: The main architecture of MISD-SLAM. Our system is built on ORB-SLAM3. Modules in green color are new created including instance segmentation, dynamic pixels removal, and semantic 3D map construction, by which MISD-SLAM system could obtain semantic knowledge, reduce the impact of dynamic objects, and build a semantic map.

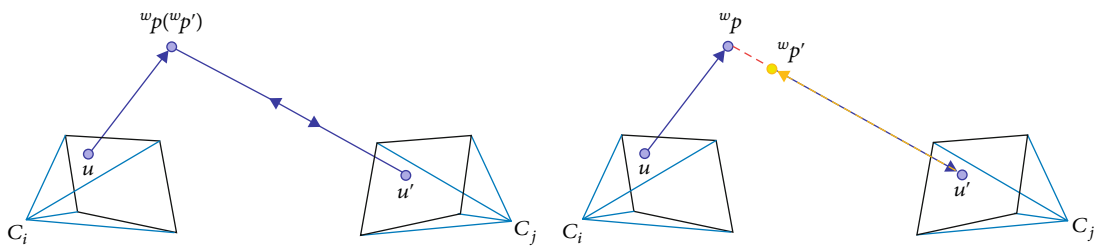
and YOLACT [14, 15] can further distinguish different instances of the same object in pixel-level. The main usages of semantic knowledge obtained from DNN can be divided into two categories, moving dynamic objects and building semantic maps.

2.2. Visual SLAM in Dynamic Scenes. The majority of visual SLAM systems are based on the common assumption that the environments are static, while the real world is changeable and dynamic. In recent years, several dynamic SLAM methods have been proposed. DS-SLAM [18] combines dynamic object detection and moving consistency check to remove the feature points located on dynamic objects. But the categories it can detect is only 20, which limits its application in complex scenarios. DynaSLAM [19] integrates Mask-RCNN [13] and multiview geometry for motion segmentation, which performs well in dynamic environments. However, it removes all potentially moving objects, such as cars parked on the side of the road, which may lead to too few feature points and impact pose estimation. VDO-SLAM [28] maximizes the number of feature points on dynamic objects using the method of dense optical flow and gets impressive results. But it is complex for real-time operation. SaD-SLAM [22] proposes a RGB-D SLAM system based on ORB-SLAM2 [5], which uses epipolar constraint of feature points in two adjacent frames to detect the static feature points in dynamic environments. But the semantic segmentation has to be processed offline that limits its application in real world. PoseFusion [29] combines human detection method, OpenPose [30], and the dense RGB-D SLAM framework, ElasticFusion [31]. However, it is limited to human detection and may not work well if the human is incomplete in the input image. StaticFusion [32] proposes a method of static and dynamic segmentation to reconstruct the background structure and applies K-means clustering algorithm to reduce the computational complexity. But it will fail if the initial images have more than 30% moving objects. Co-Fusion [33] is a model-based method, which combines object segmentation method and the dense reconstruction framework of ElasticFusion [31]. However, the map of static environment is required to be reconstructed as the precondition for tracking, segmentation, and fusion of dynamic objects, which limits its application. If two or more objects move together, they are represented by the same model until they separate. FlowFusion [34] is a flow-based method, which proposes an optical flow residual base dynamic segmentation and dense RGB-D SLAM method. It can distinguish dynamic and static clusters by setting the thresholds for high and low residuals. But it is not sensitive to the slight motions and may fail in very fast motions.

In this paper, we propose a multimodal semantic visual SLAM system for dynamic environments (MISD-SLAM) based on ORB-SLAM3 [6], which can reduce the impact of dynamic objects to evaluate accurate poses and reconstruct a semantic 3D dense map of static background. Different from the prior works, MISD-SLAM combines multiview geometry constraint method and K-means clustering algorithm to reduce the impact of dynamic pixels for map reconstruction of static background. The experiments in both public datasets and real-world environments demonstrate that our method has high performance of accuracy and robustness in dynamic indoor scenes.

3. System Overview

In this section, we will present the technical details about MISD-SLAM. Figure 2 presents the architecture of the system. We build MISD-SLAM on ORB-SLAM3 [6], which is one of the most novel feature-based visual SLAM systems



(a) The situation of static point (b) The situation of dynamic point

Figure 3: Multiview geometry constraint. C_i and C_j are two consecutive frames, u and u' are corresponding image pixels, and ^wp and $^wp'$ are their back-projection points.

that performs impressively in many datasets as well as real-world scenarios. MISD-SLAM builds three new processes:

(i) Instance segmentation, based on a pretrained network, detects and segments different instances to provide semantic knowledge of surrounding environments

(ii) Dynamic pixel removal removes the ORB features located in predefined dynamic objects which are detected by instance segmentation network then combines multiview geometry constraint with K -means clustering algorithm to remove the undefined but moving pixels to improve the accuracy and robustness in changing environments

(iii) Semantic 3D map construction combines semantic knowledge obtained from instance segmentation network with geometric structure to construct a semantic 3D dense point cloud map in global

3.1. Instance Segmentation. Under the demand of dynamic pixel detection and semantic mapping, we adopt a deep learning-based network to provide instance segmentation and semantic labels in pixel-level. MISD-SLAM utilizes the network of YOLACT++ [15] that is pretrained on MS COCO datasets [35] and can segment 80 classes. The semantic knowledge of the surrounding environments has two purposes. On the one hand, it serves as a prior information for dynamic features removal. We predefined person as a dynamic object in indoor environments. System then removes the ORB features located on the predefined objects, which improves the accuracy of pose evaluation in tracking thread and remains static ORB features as many as possible to improve the robustness in dynamic environments. On the other hand, the semantic knowledge of the pixel is integrated into corresponding 3D point to reconstruct semantic dense point cloud map in the thread of map reconstruction.

3.2. Dynamic Pixels Removal. Although prior semantic information can filter out the predefined dynamic objects in images, there may be some missing detections due to image blurring, incomplete observation, and the moving of not predefined objects. Therefore, the method of multiview geometry constraint is applied to detect the real motion of the remaining image pixels.

As shown in Figure 3, a method of multiview geometry constraint [36], which is based on the relationship of corresponding points in two consecutive frames, can be used to detect whether a pixel is static or dynamic. For current frame i and the last frame j, firstly, the pixel u in frame i is back-projected to 3D world coordinate as a point $^w\mathbf{p}$ using the information of camera pose $^w_i\mathbf{T}$ of current frame and its depth value z from the depth image:

$$^w\mathbf{p} = {}^w_i\mathbf{T}\pi^{-1}(\mathbf{u}, z), \tag{1}$$

where π^{-1} denotes the function of back-projection which depends on the camera types.

Then, the 3D point in world coordinate $^w\mathbf{p}$ is projected to the image pixel u' of the last frame j:

$$\mathbf{u}' = \pi\left(\left({}^w_j\mathbf{T}\right)^{-1} {}^w\mathbf{p}\right), \tag{2}$$

where π denotes the function of perspective projection and $^w_j\mathbf{T}$ is the camera pose of frame j estimated in tracking thread.

Furthermore, the 3D point $^w\mathbf{p}'$ in the world coordinate of the pixel u' in the last frame j can be rebuilt as:

$$^w\mathbf{p}' = {}^w_j\mathbf{T}\pi^{-1}\left(\mathbf{u}', z'\right), \tag{3}$$

where z' denotes the depth value of pixel u' in frame j.

If the point is static in both current frame i and the last frame j, as shown in Figure 3(a), the points in 3D world coordinate $^w\mathbf{p}$ and $^w\mathbf{p}'$ is pretty close to each other or even overlap. Otherwise, if the point is dynamic, the distance d between the two points $^w\mathbf{p}$ and $^w\mathbf{p}'$ in 3D world coordinate is large, as shown in Figure 3(b). Therefore, a threshold d_{th} is set to judge the dynamic points and static points. Due to the depth error increases with distance, the threshold d_{th} is set to linearly grow with the depth z:

$$d_{\text{th}} = d_{\text{base}} + kz, \tag{4}$$

where d_{base} is the base value of distance and k is the scaling factor of depth z. We set $d_{\text{base}} = 0.2$ and $k = 0.025$.

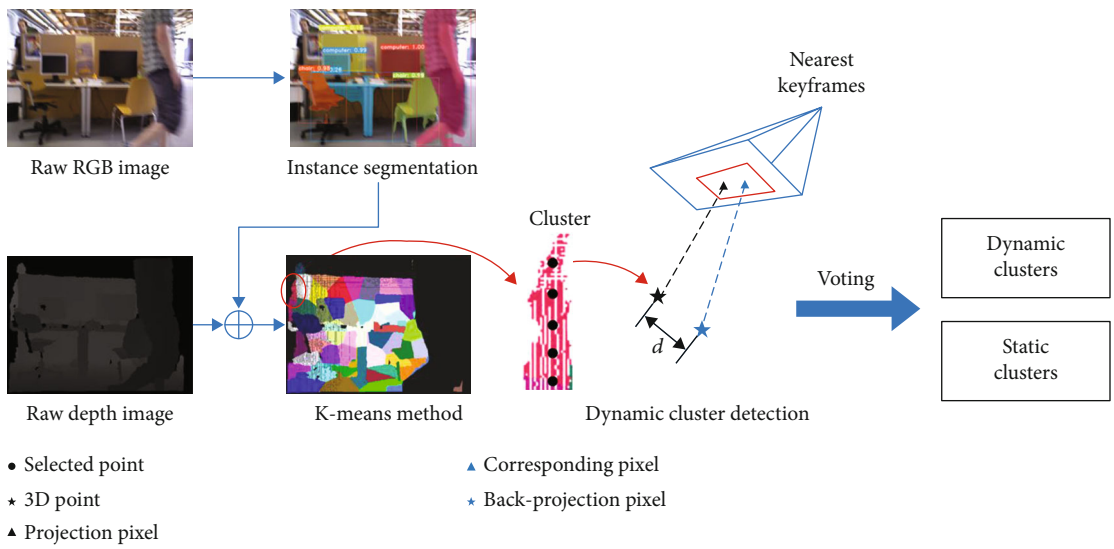
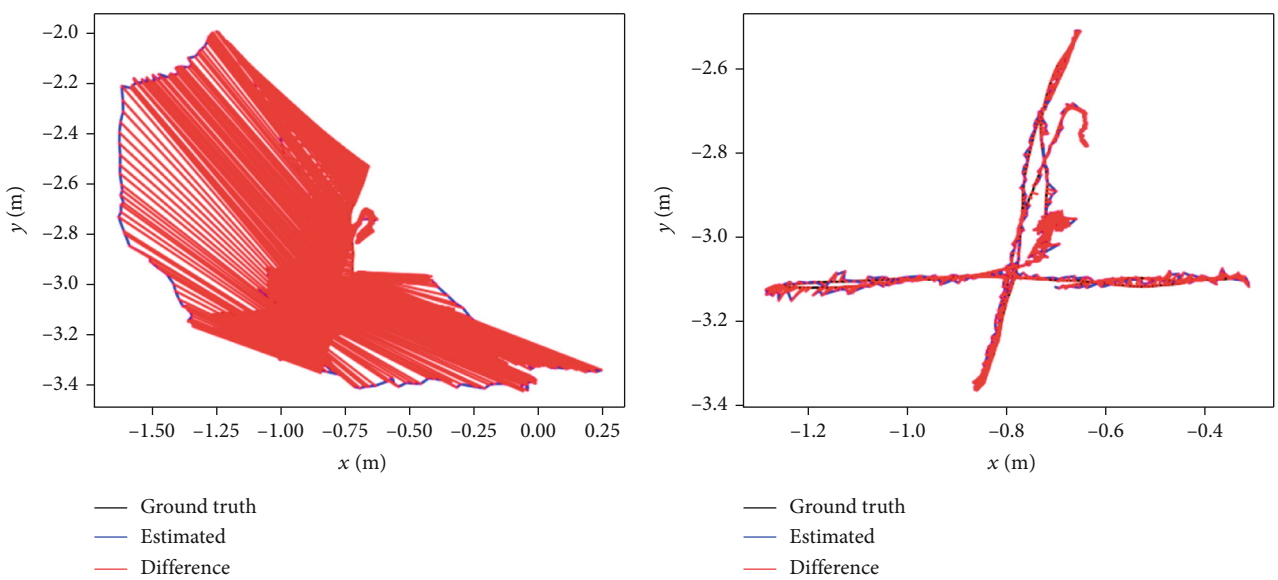


Figure 4: The flowchart of dynamic pixel removal. The raw RGB image is firstly utilized to instance segmentation to remove the pixels located on dynamic objects. Then, the remaining pixels would be combined with the depth image to generate clusters using K-means method. Under multiview geometry constraint, points can be divided into dynamic and static. And the motion state of each cluster is determined by voting method.



(a) Trajectory from ORB_SLAM3

(b) Trajectory from MISD_SLAM

Figure 5: The trajectory difference between two systems and ground truth in high-dynamic sequence *walking_xyz*. The blue line represents the trajectory estimated by the respective SLAM system, and the red line is the difference between estimated trajectory and the ground truth.

If the distance d between ${}^{w}\mathbf{p}$ and ${}^{w}\mathbf{p}'$ is larger than the threshold d_{th}, then the image pixel of current frame is determined as dynamic. Otherwise, the pixel is static.

Furthermore, in order to reduce the calculation and running time, K-means clustering algorithm and voting method are proposed to detect dynamic pixels, as shown in Figure 4. Combing with the depth image and the camera pose estimated from the tacking thread, the remaining pixels of the RGB image are back-projected to 3D points in the world coordinate to create a point cloud. The 3D point cloud are divided into k clusters by K-means clustering algorithm, where k is calculated by the number of remaining pixels/2000. In each cluster, 100 points are randomly selected. If there are not 100 points in one cluster, then all of the points are selected. The motion property of each selected point is determined by the method of multiview geometry constraint method. Then, voting method is used to determine the motion property of each cluster according to the majority motion properties of its selected points:

$$\text{Motion}_i = \begin{cases} \text{dynamic}, & \text{num}_{\text{dynamic}} > \text{num}_{\text{static}}, \\ \text{static}, & \text{num}_{\text{dynamic}} <= \text{num}_{\text{static}}, \end{cases} \tag{5}$$

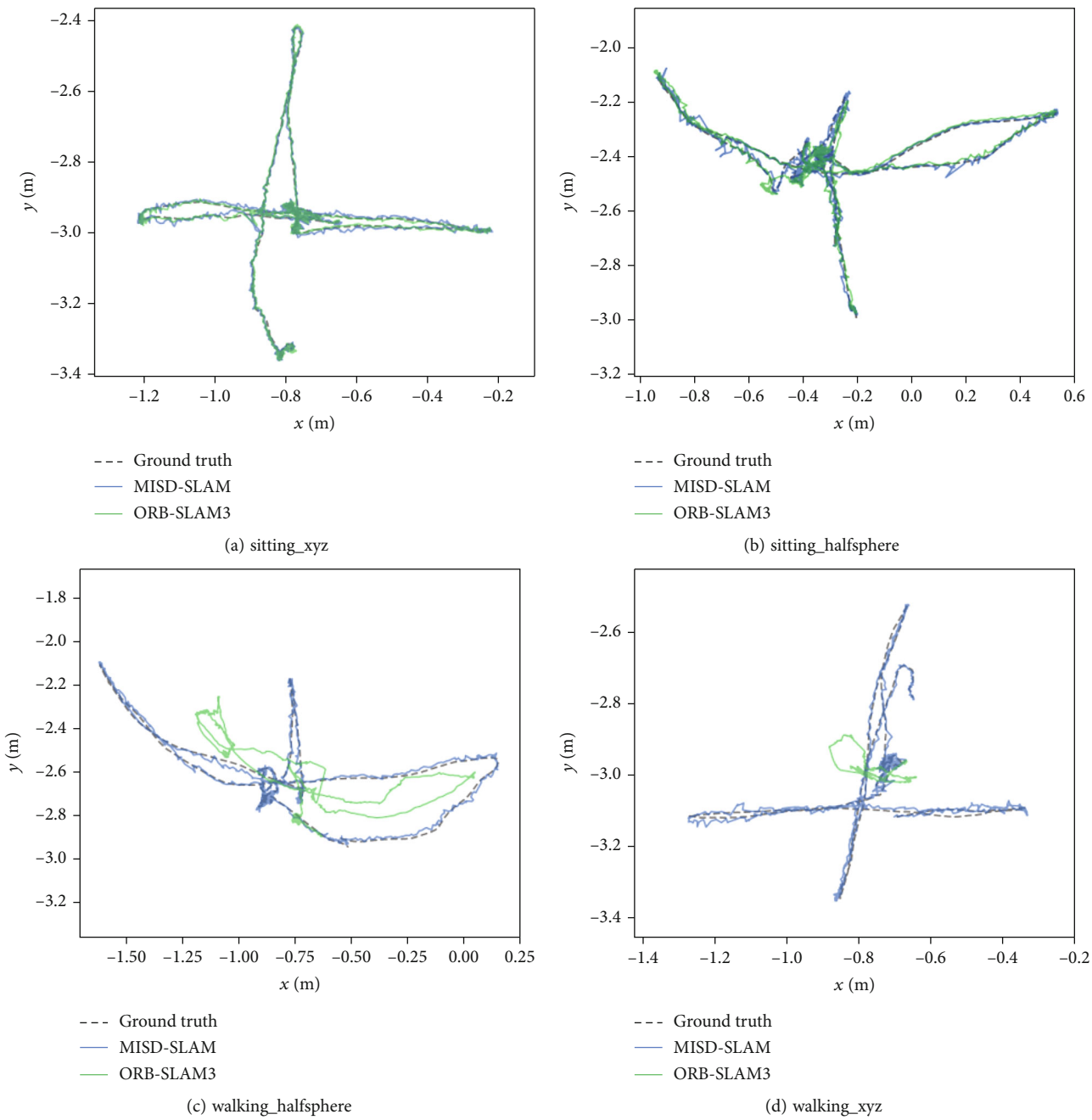


Figure 6: Trajectory comparison in dynamic sequences. Sitting series is in low-dynamic, and walking series is in high-dynamic. The green line is the trajectory of ORB-SLAM3, the blue line is the trajectory of MISD-SLAM, and the black dotted line is the ground truth.

where Motion_i denotes the motion property of the i_{th} cluster, $\text{num}_{\text{dynamic}}/\text{num}_{\text{static}}$ denotes the number of dynamic/static points among the selected points in i_{th} cluster. If the number of dynamic points is larger than the number of static points, then the cluster is dynamic; otherwise, the cluster is static. Finally, the corresponding image pixels of the points in dynamic clusters are determined as dynamic.

Combining the predefined dynamic objects in instance segmentation and the dynamic pixels in motion detection, the total dynamic pixels are obtained. In the following process of map reconstruction, the dynamic pixels are removed, and the static pixels are used to reconstruct the map of static background.

3.3. Semantic 3D Map Construction. The thread focuses on reconstruct the static background of keyframes. The map type is 3D point cloud with semantic labels. As shown in Figure 1, the inputs of map reconstruction thread are the RGB-D image pair, semantic image, and camera pose estimated in the tracking thread. After the process of dynamic pixels removal, the total dynamic regions in RGB image are obtained, including the predefined dynamic objects

Table 1: Results of absolute trajectory error (ATE) and improvements of MISD-SLAM compared to ORB-SLAM3 in dynamic sequences.

Sequence	ORB-SLAM3 [6]			MISD-SLAM (ours)			Improvements		
	RMSE	Mean	S.D.	RMSE	Mean	S.D.	RMSE	Mean	S.D.
Sitting_static	0.0067	0.0055	0.0037	0.0059	0.0052	0.0028	11.94%	5.45%	24.32%
Walking_static	0.0248	0.0199	0.0148	0.0091	0.0079	0.0036	63.31%	60.30%	68.92%
Walking_xyz	0.2895	0.2620	0.1232	0.0129	0.0112	0.0063	95.54%	95.73%	94.89%
Walking_rpy	0.1655	0.1396	0.0889	0.1252	0.0904	0.0866	24.35%	35.24%	2.59%
Walking_halfsphere	0.3305	0.2977	0.1434	0.0168	0.0149	0.0077	94.92%	94.99%	94.63%

Table 2: Results of absolute trajectory error (ATE) and improvements of processed PL-SVO compared to original PL-SVO in dynamic sequences.

Sequence	PL-SVO [37]			Processed PL-SVO			Improvements		
	RMSE	Mean	S.D.	RMSE	Mean	S.D.	RMSE	Mean	S.D.
Sitting_static	0.0084	0.0077	0.0033	0.0064	0.0056	0.0031	23.81%	27.27%	6.06%
Walking_static	0.0075	0.0065	0.0037	0.0057	0.0048	0.0030	24.00%	26.15%	18.92%
Walking_xyz	0.2811	0.2409	0.1363	0.0230	0.0199	0.0116	90.96%	90.74%	91.49%
Walking_rpy	0.1770	0.1541	0.0872	0.0357	0.0275	0.0228	79.83%	82.15%	73.85%
Walking_halfsphere	0.1149	0.1112	0.0289	0.0401	0.0320	0.0242	65.10%	71.22%	16.26%

detected in instance segmentation and the dynamic pixels determined by their motion properties. To reduce the impact of the region edges, the dynamic region edges in the image are expanded 20 pixels. The depth values of the expanded dynamic regions in the depth image are set to zero. Except the image pixels located in the expanded dynamic regions, the other image pixels are static pixels. However, the static pixels are not all reconstructed to the map, because the points of some static pixels may have existed in the map, which are called redundant points. To avoid redundant points, before back-projecting the static image pixels into 3D points in the local point cloud, we perform an operation of pixel-point association. As for each point in the global point cloud, it is projected to a pixel in current image based on the camera pose and camera internal parameters, so that we can get the position (x, y) and depth value z of the projected pixel. Then, according to the position (x, y) of the projected pixel, we obtain four pixels around it in current image. If the minimum depth difference between the four image pixels and the projected pixel is smaller than the threshold (0.02 in our work), which indicates that the image pixel has been reconstructed in the map, then the image pixel of the minimum depth difference will not be used to be back-projected into the 3D space. We set the depth value of this pixel to zero. Then, the image pixels with depth value in the range of $z_{\min}$ to $z_{\max}$ are back-projected into 3D points to generate the local point cloud of current keyframe, based on the camera internal parameters and the camera pose that is estimated in the tracking thread. We set the threshold values $z_{\min} = 0.2$ and $z_{\max} = 8.0$. In this way, the dynamic pixels and the redundant pixels in the image are not reconstructed in the point cloud because their depth values were set to zero. For semantic labels, a color attribute of point is applied to represent its category according to the semantic image. For example, the point classified to chair is labeled in orange color, and the keyboard is purple color. As for the point without semantic information, its color attribute is set the corresponding pixel value in RGB image. Then, the local static point cloud with semantic labels of current keyframe is fused into the global point cloud in the world coordinate, which reconstructs the map of static background incrementally.

4. Experiments

In this section, we demonstrate our MISD-SLAM system in public TUM RGB-D datasets [17] and real-world scenes to evaluate its performance of accuracy and robustness in dynamic environments. First, MISD-SLAM system is compared with original ORB-SLAM3 [6] to verify the improvement of performance. Then, we replace ORB-SLAM3 to another backbone, PL-SVO [37] to validate the effectiveness of the proposed method. In addition, our MISD-SLAM system is compared with the state-of-the-art SLAM systems in dynamic environments. Besides, the semantic 3D dense point cloud maps and the time performance are presented. Finally, an experiment in real-world environments is carried out to evaluate the performance of the system in real scenes. All the experiments run on a computer with Intel E5-2683 CPU and Nvidia GTX 1080 GPU. The GPU is only used for instance segmentation.

4.1. Experiments in TUM RGB-D Datasets. The TUM RGB-D datasets [17] provide video sequences of indoor scenes recorded by Microsoft Kinect at the frame rate of 30 Hz. The datasets include RGB images and depth images with 640×480 resolution, as well as ground truth trajectories. We select the sequences of dynamic scenes to evaluate our MISD-SLAM system. In the sequences of *sitting* series, there are two people sitting on chairs in front of a desk and talking with each other. These sequences represent low-dynamic environments. In the sequences of *walking* series, people

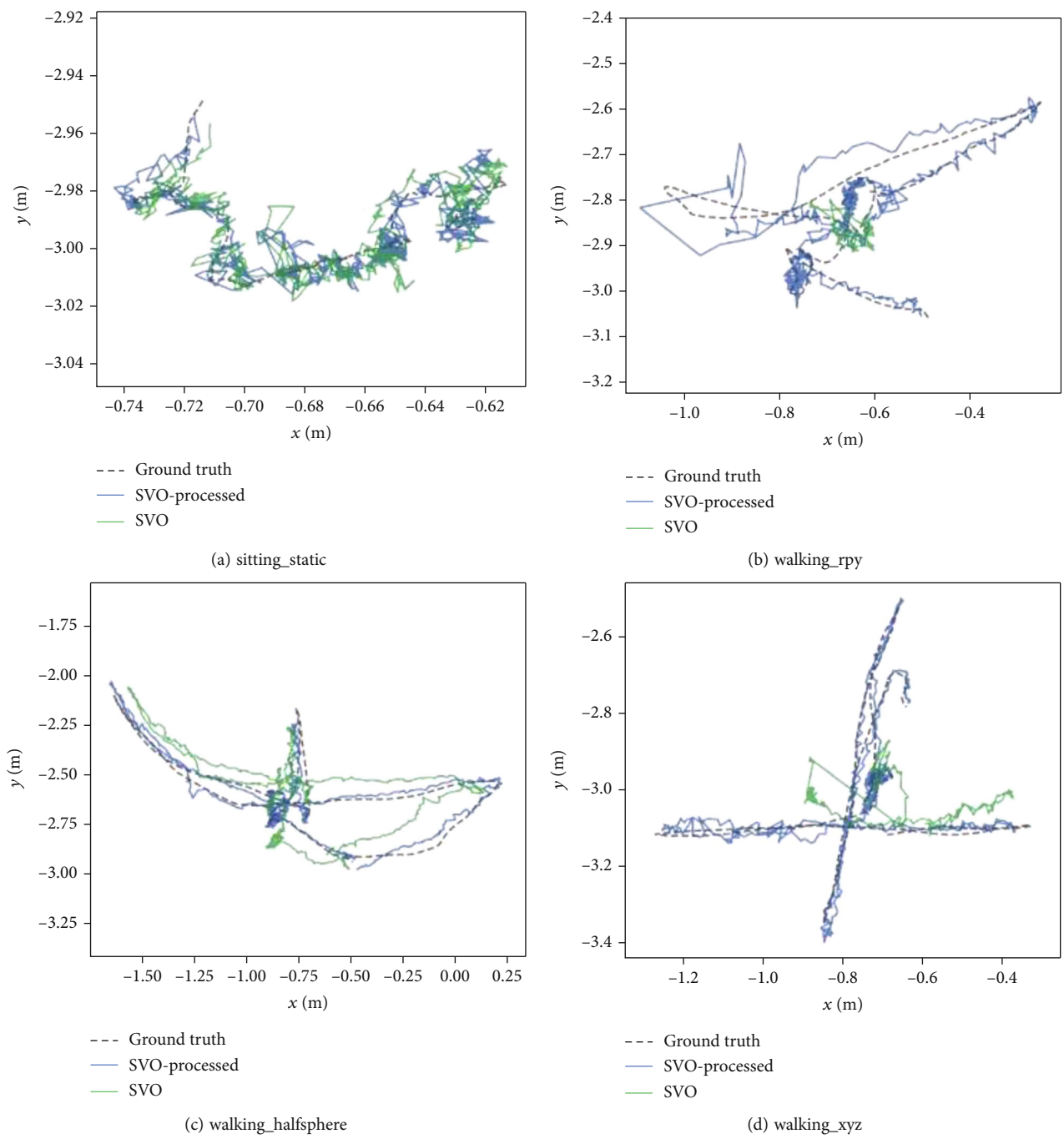


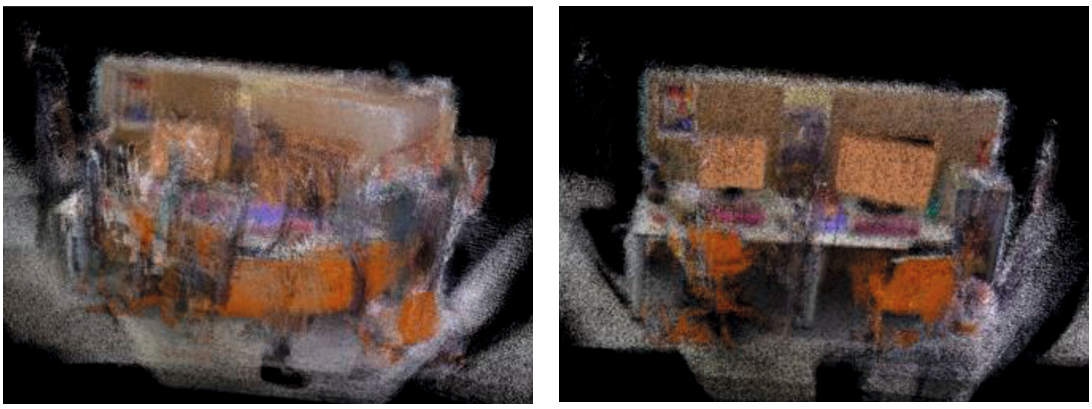
Figure 7: Trajectory comparison of original PL-SVO [37] and processed PL-SVO in dynamic sequences. Sitting series is in low-dynamic, and walking series is in high-dynamic. The green line is the trajectory of original PL-SVO, the blue line is the trajectory of processed PL-SVO, and the black dotted line is the ground truth.

walk most of the time. These sequences are in high-dynamic, which would seriously impact the accuracy and robustness of ordinary SLAM systems. ORB-SLAM3 [6] is a state-of-the-art visual system and serves as the backbone of our MISD-SLAM, so we firstly compare these two systems and make a quantitative evaluation.

We compare ORB-SLAM3 [6] and MISD-SLAM in dynamic sequences of TUM RGB-D datasets [17], which are composed of four patterns of camera ego-motions including keeping still in one place (static), moving along three directions (*xyz*), rotating along the principle axes (*rpy*) and moving on a small half sphere (halfsphere). The comparison results are presented in Figure 5, where the two camera trajectories obtained from these systems in sequence *walking_xyz* are, respectively, plotted with the ground truth trajectory. The difference between trajectories of ORB-SLAM3 and ground truth is apparent in Figure 5(a), while in Figure 5(b), the two trajectories of MISD-SLAM and ground truth are very close, which shows the robustness and accuracy of our system. The reason is

Table 3: Comparison results of absolute trajectory RMSE (m) against the state-of-the-art dynamic SLAM systems.

Sequence	DS-SLAM [18]	DynaSLAM [19]	Detect-SLAM [20]	SOF-SLAM [21]	SaD-SLAM [22]	MISD-SLAM (ours)
Sitting_static	0.0065	—	—	0.010	0.0060	0.0059
Walking_static	0.0081	0.006	—	0.007	0.0166	0.0091
Walking_xyz	0.0247	0.015	0.0241	0.018	0.0167	0.0129
Walking_rpy	0.4442	0.035	0.2959	0.027	0.0318	0.1252
Walking_halfsphere	0.0303	0.025	0.0514	0.029	0.0257	0.0168



(a) map without dynamic points removal (b) map with dynamic points removal

Figure 8: The semantic 3D dense point cloud maps built by MISD-SLAM. (a) is the map without dynamic point removal, and (b) is the map with dynamic point removal.

Table 4: Time evaluation of MISD-SLAM in dynamic sequences of TUM RGB-D datasets (ms). The results of each row are the running time of corresponding modules in each sequence. The last row is the average time of above sequences in corresponding modules.

Sequence	ORB feature extraction (ms)	Instance segmentation (ms)	Multiview geometry (ms)	Semantic map construction (ms)
Sitting_static	12.981	36.465	144.085	121.388
Walking_static	14.575	36.109	205.484	296.667
Walking_xyz	14.146	36.150	180.533	338.903
Walking_rpy	12.345	36.640	200.591	274.928
Walking_ halfsphere	12.484	36.536	161.401	244.241
Average time	13.306	36.380	178.419	255.225

that the dynamic features detected by ORB-SLAM3 are assumed as static features to estimate the trajectory, which leads the wrong results. While in MISD-SLAM system, these dynamic features are removed and the impact on trajectory prediction are reduced significantly.

Furthermore, we compare ORB-SLAM3 [6] and our MISD-SLAM system in other dynamic sequences. The results are shown in Figure 6, where the trajectory of ORB-SLAM3 [6] is green line, the MISD-SLAM is blue line, and the ground truth is black dotted line. In Figures 6(a) and 6(b), the three trajectories are very close which indicates the accuracies of both systems are high in low-dynamic sequences. However, in Figure 6(c) and 6(d), the trajectories of ORB-SLAM3 [6] are deformed seriously, while the trajectories of MISD-SLAM are still close to the ground truth. We perform a quantitative evaluation of the two SLAM systems in different sequences in Table 1, using the values of root mean squared error (RMSE), mean error, and standard deviation (S.D.) of absolute trajectory. It can be seen that due to the removal of dynamic feature points, MISD-SLAM significantly reduce the impact of dynamic objects and effectively

Table 5: Comparison of the time performance to other methods. The total time includes the time of feature extraction, semantic segmentation, and dynamic objects removal but without the time of map construction.

Methods	Total time (ms)
DS-SLAM [18]	76.46
DynaSLAM [19]	286.47
Detect-SLAM [20]	340.00
MISD-SLAM (ours)	228.11

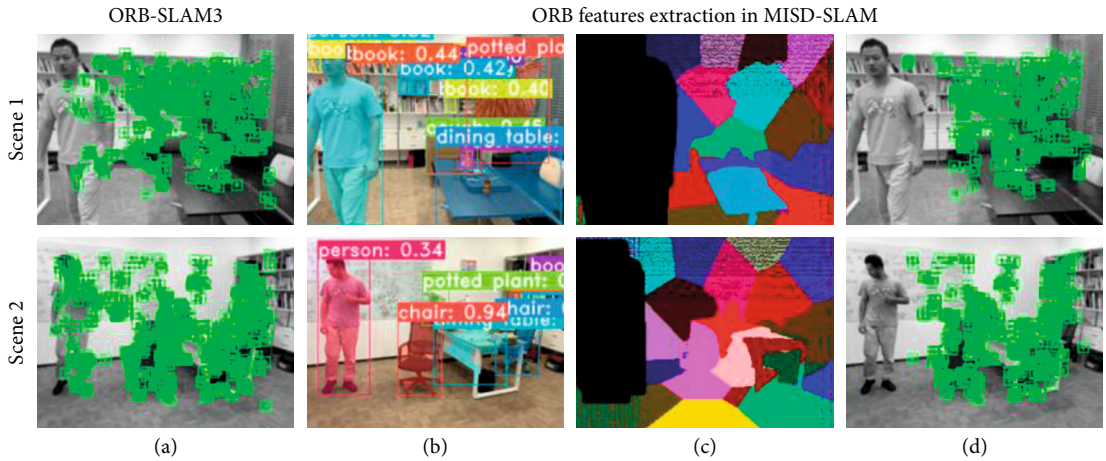


Figure 9: Experiments in real-world environments and comparison of ORB features extraction results between ORB-SLAM3 [6] and MISD-SLAM. From left to right, the first column (a) is the ORB features extraction results in ORB-SLAM3 [6]. (b)–(d) are the processes of MISD-SLAM, which are, respectively, the instance segmentation, clustering results, and ORB features extraction results after dynamic features removal in MISD-SLAM.

estimate the correct trajectory, which demonstrates high robustness and correctness of the system.

To validate the effectiveness and application of the proposed method, we replace ORB-SLAM3 [6] to another backbone, PL-SVO [37], and evaluate its performance of trajectory accuracy. PL-SVO [37] is a monocular visual odometry algorithm combining point features and line features with RGB image inputs, which is appropriate for this experiment of backbone replacement. We processed PL-SVO [37] by adding the modules of instance segmentation and dynamic features removal to remove the point features and line features located in the dynamic objects. We compare original PL-SVO and processed PL-SVO in the dynamic sequences of TUM RGB-D datasets. The quantitative results are shown in Table 2. It can be seen that the performance of accuracy is improved in all of the dynamic sequences. Especially, the improvement in sequence *walking_xyz* is over 90%. Figure 7 shows the qualitative results by visualizing their trajectories, where the green line is the trajectory of original PL-SVO, the blue line is the trajectory of processed PL-SVO, and the black dotted line is the ground truth. Figures 6 and 7 indicate that the original ORB-SLAM3 and the original PL-SVO are impacted by the dynamic objects in the environments, and their trajectory accuracy is low compared to the ground truth, especially in high-dynamic sequences. Processed by the proposed method, the dynamic features are removed, and the accuracy performance is improved significantly, which validates the effectiveness and application of the proposed method. Although we provide the experiment using another backbone, our goal is not to explore the effect of different backbones. We focus our attention on the following experiments based on MISD-SLAM with the backbone of ORB-SLAM3.

4.2. Comparison with Other Visual SLAM Systems. In this part, we adopt dynamic sequences of TUM RGB-D datasets [17] to compare our MISD-SLAM against the five state-of-the-art visual SLAM systems, DS-SLAM [18], DynaSLAM [19], Detect-SLAM [20], SOF-SLAM [21], and SaD-SLAM [22] which have been proposed for dynamic environments and semantic tasks in last three years. The results are shown in Table 3, and the results of the five SLAM systems come from published papers. The results of our MISD-SLAM gained after running on the datasets for five times and taking the average values. The sequence *sitting_static* is in low-dynamic, and *walking* series is in high-dynamic. MISD-SLAM performs better in sequences of *sitting_static*, *walking_xyz*, and *walking_halfsphere*.

The experiment results indicate the high performance of our system. MISD-SLAM removes dynamic features according to the result of instance segmentation, then remaining pixels with potential movement are detected and removed through multiview geometry constraint method. After these two steps, the moving image pixels are deleted, and the impact of dynamic objects is reduced.

However, if there are too few features, the system may estimate wrong pose, or even track failure. Compared with the other four systems, MISD-SLAM reduces the impact of dynamic objects to improve the accuracy of pose evaluation and remains static features as many as possible to improve the robustness in dynamic environments, which improves the performance of accuracy and robustness.

4.3. Semantic 3D Maps. This part presents the semantic 3D dense point cloud maps built by MISD-SLAM system. Figure 8 compares two maps built in the high-dynamic sequence of TUM RGB-D datasets [17], *walking_xyz*. Figure 8(a) is reconstructed without dynamic pixels removal. It can be seen that the dynamic persons reduce the accuracy of camera pose estimation, which make the things misplaced. Besides, the moving persons are modeled in the map. Therefore, Figure 8(a) is corrupted and difficult to use. Figure 8(b) is constructed after dynamic pixel removal, which reduces the influence of dynamic objects, so that the static background can be reconstructed with the accurate camera pose. The map with dynamic pixels removal is

Figure 10: The semantic 3D point cloud maps of real environments. The figures in (a, b) are the semantic 3D point cloud maps. The figures in (c, d) are the color images of experimental scenes.

capable to provide a precise and semantic map of static background in dynamic indoor scenes.

4.4. Evaluation of Time Performance. Time performance is another important indicator to evaluate the proposed method. We evaluate the time performance in four major modules: ORB feature extraction, instance segmentation, dynamic pixels removal, and semantic map construction. The results shown in Table 4 are the running time of corresponding modules in each sequence and the average time of all the sequences. The time performances of ORB feature extraction and instance segmentation are achieved in real time. The most time-consuming module is the semantic map construction module. But it only operates in keyframes, which is selected from input frames so the number of keyframes is less than the number of input frames. And the semantic map construction module operates in parallel with other modules. Therefore, the time cost of semantic map construction module affects little to the whole process. The multiview geometry constraint method slowdowns the process compared to other modules in MISD-SLAM. However, compared to the modules with similar function, MISD-SLAM has higher time performance than DynaSLAM [19] (333.68 ms in sequence *walking_halfsphere* and 235.98 ms in sequence *walking_rpy*) and Detect-SLAM [20] (310 ms), due to its reduction of computational complexity by K-means clustering algorithm and voting method.

Furthermore, Table 5 shows the comparison of the time performance to other methods, including DS-SLAM [18], DynaSLAM [19], and Detect-SLAM [20]. Given that some methods do not provide the time of map construction in their papers and the large variability of map construction time in different hardware conditions of computing, rendering, and displaying, it is more fair to compare the total time except map construction. Table 5 lists the total time except map construction to represent the time performance, which is the sum of the average time of feature extraction, semantic segmentation, and dynamic objects removal. Among these methods, DS-SLAM [18] is optimized for real time, and the other three methods are not optimized for real time. It can be seen that MISD-SLAM has higher time performance among the not optimized methods and achieves comparable performance to the optimized method. The comparison results of time performance indicates that the proposed method plays an important role in reducing the computational complexity and achieves high time performance.

4.5. Experiments in Real-World Environments. Experiments in real-world environments are carried out to evaluate the performance of the MISD-SLAM system in real scenes.

The images of RGB and depth are captured by iPad Pro with 320×240 resolution. The experiment scene is an office, where a person is walking around, and the camera is doing translational motion.

In Figure 9, the images from left to right are raw ORB feature extraction image of ORB-SLAM3 [6] and images in our system including instance segmentation image, clustering image, and ORB features extraction image after dynamic feature removal. It can be seen that the moving people are detected in the instance segmentation image. The dynamic features located in the dynamic objects are removed. The removal of dynamic features reduces the influence of dynamic objects for better camera pose estimation, and semantic mapping. Furthermore, after K-means clustering and multiview geometry constraint, dynamic pixels in the scene are removed significantly.

Figure 10 shows the 3D semantic map constructed by MISDSLAM system in these two real scenes. Because the dynamic features are removed, the camera pose can be correctly estimated. The system back-projects the static image pixels into 3D space based on the camera pose to build a static point cloud map of the real scene. 2D map with semantic information generated by 3D semantic point cloud map can be applied in navigation and planning tasks [18, 38].

5. Conclusions

In this paper, we propose a novel multimodal semantic SLAM system (MISD-SLAM), which could perform robustly in dynamic environments and build a semantic 3D point cloud map. MISD-SLAM builds three main processes: instance segmentation, dynamic pixel removal, and semantic 3D map construction. An instance segmentation network [15] is introduced to provide semantic knowledge of surrounding environments. The ORB features located on the predefined dynamic objects are removed directly. In this way, MISD-SLAM effectively reduces the impact of dynamic objects to provide precise pose estimation. Then, combining multiview geometry constraint with K-means clustering algorithm, our system removes the undefined but moving pixels. Meanwhile, a 3D dense point cloud map with semantic information is reconstructed. Moreover, experiments are carried out on challenging sequences of TUM RGB-D datasets [17] as well as the real-world scenes to evaluate the performance of MISD-SLAM. Compared to original ORB-SLAM3 [6] and the state-of-the-art SLAM systems, the results indicate that our method significantly improves the localization accuracy and system robustness, especially in high-dynamic environments.

However, there exists some limitations in MISD-SLAM. First, the process of dynamic objects is not flexible enough, because the objects may be static in some frames and be dynamic in other frames. Second, the depth range of the RGB-D camera is restricted, which limits its application in larger scenes. In the future, the developments of MISD-SLAM will focus on optimizing the strategy of dynamic objects removal in the reconstructed map and improving the real-time performance. Furthermore, we will adopt inertial measurement unit (IMU) to expand the scope of application in larger environments.

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 no conflicts of interest.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (no. 62073004) and Science and Technology Plan of Shenzhen (no. JCYJ20190808182209321).

References

[1] A. J. Davison, I. D. Reid, N. D. Molton, and O. Stasse, "MonoSLAM: real-time single camera SLAM," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 6, pp. 1052–1067, 2007.

[2] G. Klein and D. Murray, "Parallel tracking and mapping for small AR workspaces," in *IEEE and ACM International Symposium on Mixed and Augmented Reality*, pp. 225–234, 2007.

[3] J. Engel, T. Schöps, and D. Cremers, "LSD-SLAM: large-scale direct monocular SLAM," in *Europeon Conference on Computer Vision (ECCV)*, pp. 834–849, 2014.

[4] R. Mur-Artal, J. M. M. Montiel, and J. D. Tardós, "ORB-SLAM: a versatile and accurate monocular SLAM system," *IEEE Transactions on Robotics*, vol. 31, no. 5, pp. 1147–1163, 2015.

[5] R. Mur-Artal and J. D. Tardós, "ORB-SLAM2: an open-source SLAM system for monocular, stereo, and RGB-D cameras," *IEEE Transactions on Robotics*, vol. 33, no. 5, pp. 1255–1262, 2017.

[6] C. Campos, R. Elvira, J. J. G. Rodriguez, J. M. M. Montiel, and J. D. Tardos, "ORB-SLAM3: an accurate open-source library for visual, visual–inertial, and multimap SLAM," *IEEE Transactions on Robotics*, vol. 37, no. 6, pp. 1874–1890, 2021.

[7] J. Engel, V. Koltun, and D. Cremers, "Direct sparse odometry," *IEEE Transactions on Pattern Analysis and M achine Intelligence*, vol. 40, no. 3, pp. 611–625, 2018.

[8] M. Labbe and F. Michaud, "Online global loop closure detection for large-scale multi-session graph-based SLAM," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 2661–2666, 2014.

[9] Z. Zhao, Y. Mao, Y. Ding, P. Ren, and N. Zheng, "Visual-based semantic SLAM with landmarks for large-scale outdoor environment," in *China Symposium on Cognitive Computing and Hybrid Intelligence (CCHI)*, pp. 149–154, 2019.

[10] W. Liu, D. Anguelov, D. Erhan et al., "SSD: single shot multibox detector," in *Europeon Conference on Computer Vision (ECCV)*, pp. 21–37, 2016.

[11] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: unified, real-time object detection," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 779–788, 2016.

[12] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: towards real-time object detection with region proposal networks," *Advances in Neural Information Processing Systems*, vol. 28, 2015.

[13] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *IEEE International Conference on Computer Vision (ICCV)*, pp. 2961–2969, 2017.

[14] D. Bolya, C. Zhou, F. Xiao, and Y. J. Lee, "YOLACT: real-time instance segmentation," in *IEEE International Conference on Computer Vision (ICCV)*, pp. 9157–9166, 2019.

[15] D. Bolya, C. Zhou, F. Xiao, and Y. J. Lee, "YOLACT++: better real-time instance segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 2, pp. 1108–1121, 2022.

[16] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: a deep convolutional encoder-decoder architecture for image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 12, pp. 2481–2495, 2017.

[17] J. Sturm, N. Engelhard, F. Endres, W. Burgard, and D. Cremers, "A benchmark for the evaluation of RGB-D SLAM systems," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 573–580, 2012.

[18] C. Yu, Z. Liu, X. Liu et al., "DS-SLAM: a semantic visual SLAM towards dynamic environments," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 1168–1174, 2018.

[19] B. Bescos, J. M. Fácil, J. Civera, and J. Neira, "DynaSLAM: tracking, mapping, and inpainting in dynamic scenes," *IEEE Robotics and Automation Letters*, vol. 3, no. 4, pp. 4076–4083, 2018.

[20] F. Zhong, S. Wang, Z. Zhang, C. Chen, and Y. Wang, "Detect-SLAM: making object detection and SLAM mutually beneficial," in *IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 1001–1010, 2018.

[21] L. Cui and C. Ma, "SOF-SLAM: a semantic visual SLAM for dynamic environments," *IEEE Access*, vol. 7, pp. 166528–166539, 2019.

[22] X. Yuan and S. Chen, "SAD-SLAM: a visual SLAM based on semantic and depth information," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 4930–4935, 2020.

[23] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2881–2890, 2017.

[24] L. C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Semantic image segmentation with deep convolutional nets and fully connected CRFs," 2014, http://arxiv.org/abs/1412.7062.

[25] 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. 40, no. 4, pp. 834–848, 2018.

[26] L. C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," 2017, http://arxiv.org/abs/1706.05587.

[27] L. C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Europeon Conference on Computer Vision (ECCV)*, pp. 801–818, 2018.

[28] J. Zhang, M. Henein, R. Mahony, and V. Ila, "VDO-SLAM: a visual dynamic object-aware SLAM system," 2020, http://arxiv.org/abs/2005.11052.

[29] T. Zhang and Y. Nakamura, "PoseFusion: dense RGB-D SLAM in dynamic human environments," in *International Symposium on Experimental Robotics*, pp. 772–780, Springer, Cham, 2020.

[30] Z. Cao, T. Simon, S. E. Wei, and Y. Sheikh, "Realtime multi-person 2D pose estimation using part affinity fields," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 7291–7299, 2017.

[31] T. Whelan, R. F. Salas-Moreno, B. Glocker, A. J. Davison, and S. Leutenegger, "Elasticfusion: real-time dense SLAM and light source estimation," *The International Journal of Robotics Research*, vol. 35, no. 14, pp. 1697–1716, 2016.

[32] R. Scona, M. Jaimez, Y. R. Petillot, M. Fallon, and D. Cremers, "Staticfusion: background reconstruction for dense RGB-D SLAM in dynamic environments," in *IEEE International Conference on Robotics and Automation (ICRA)*, pp. 3849–3856, 2018.

[33] M. Rünz and L. Agapito, "Co-fusion: real-time segmentation, tracking and fusion of multiple objects," in *IEEE International Conference on Robotics and Automation (ICRA)*, pp. 4471–4478, 2017.

[34] T. Zhang, H. Zhang, Y. Li, Y. Nakamura, and L. Zhang, "Flow-Fusion: dynamic dense RGB-D SLAM based on optical flow," in *IEEE International Conference on Robotics and Automation (ICRA)*, pp. 7322–7328, 2020.

[35] T. Y. Lin, M. Maire, S. Belongie et al., "Microsoft COCO: common objects in context," in *Europeon Conference on Computer Vision (ECCV)*, pp. 740–755, Springer, Cham, 2014.

[36] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision, 2nd Edn. Cambridge University Press*, Cambridge University Press, 2000.

[37] R. Gomez-Ojeda, J. Briales, and J. Gonzalez-Jimenez, "PL-SVO: semi-direct monocular visual odometry by combining points and line segments," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 4211–4216, 2016.

[38] N. Sünderhauf, F. Dayoub, S. McMahon et al., "Place categorization and semantic mapping on a mobile robot," in *IEEE International Conference on Robotics and Automation (ICRA)*, pp. 5729–5736, 2016.

Hindawi
Wireless Communications and Mobile Computing
Volume 2023, Article ID 9808517, 1 page
https://doi.org/10.1155/2023/9808517

Retraction

Retracted: Antialiasing Attention Spatial Convolution Model for Skin Lesion Segmentation with Applications in the Medical IoT

Wireless Communications and Mobile Computing

Received 10 October 2023; Accepted 10 October 2023; Published 11 October 2023

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] P. T. Le, C. Chang, Y. Li, Y. Hsu, and J. Wang, "Antialiasing Attention Spatial Convolution Model for Skin Lesion Segmentation with Applications in the Medical IoT," *Wireless Communications and Mobile Computing*, vol. 2022, Article ID 1278515, 15 pages, 2022.

Hindawi
Wireless Communications and Mobile Computing
Volume 2022, Article ID 1278515, 15 pages
https://doi.org/10.1155/2022/1278515

Research Article

Antialiasing Attention Spatial Convolution Model for Skin Lesion Segmentation with Applications in the Medical IoT

Phuong Thi Le,[1] Ching-Chun Chang,[2] Yung-Hui Li,[3] Yi-Chiung Hsu,[1] and Jia-Ching Wang[4]

[1]*Department of Biomedical Sciences and Engineering, National Central University, Taoyuan, Taiwan*
[2]*Department of Computer Science, University of Warwick, Coventry, UK*
[3]*AI Research Center, Hon Hai Research Institute, Taipei, Taiwan*
[4]*Department of Computer Science and Information Engineering, National Central University, Taoyuan, Taiwan*

Correspondence should be addressed to Yi-Chiung Hsu; syic@ncu.edu.tw

Received 28 October 2021; Accepted 11 January 2022; Published 24 February 2022

Academic Editor: Chi-Hua Chen

This study presents a noninvasive visual sensing enhancing system for skin lesion segmentation. According to the Skin Cancer Foundation, skin cancer kills more than two people every hour in the United States, and one in every five Americans will develop the disease. Skin cancer is becoming more popular, so the need for skin cancer diagnosis is increasing, particularly for melanoma, which has a high metastasis rate. Many traditional algorithms, as well as a computer-aided diagnosis tool, have been implemented in dermoscopic images for skin lesion segmentation to meet this need. However, the accuracy of the model is low, and the prognosis time is lengthy. This paper presents antialiasing attention spatial convolution (AASC) to segment melanoma skin lesions in dermoscopic images. Such a system can enhance the existing Medical IoT (MIoT) applications and provide third-party clues for medical examiners. Empirical results show that the AASC performs well when it is able to overcome dermoscopic limitations such as thick hair, low contrast, or shape and color distortion. The model was evaluated strictly under many statistical evaluation metrics such as the Jaccard index, Recall, Precision, F1 score, and Dice coefficient. The performance of the AASC was trained and tested. Remarkably, the AASC model yielded the highest scores in both three databases compared with the state-of-the-art models across three datasets: ISIC 2016, ISIC 2017, and PH2.

1. Introduction

The advances in the Internet of Things and biomedical signal processing have spurred the development of the Medical Internet of Things (Medical IoT or MIoT). More and more healthcare monitoring and diagnosis rely on MIoT devices. Recent machine learning techniques such as deep learning have subsequently enhanced the practicability of the MioT. Medical researchers and scientists can utilize such techniques to discover hidden factors, subsequently helping more patients. In 2020, there were nearly 10 million deaths from cancer according to the WHO (World Health Organization). In the US, more than 600000 deaths in 1.8 million diagnoses are estimated [1]. The fees for cancer treatment have nearly doubled in the last two decades in the US [2]. This could say that cancer is one of the leading causes of death worldwide. All of these statistics compel experts to find a way to reduce cancer risks. In truth, they have statistics on five stages of cancer, including stage 0, stage I (the early stage), stage II, stage III, and stage IV. In addition, the experts clearly show that cancer victims have a high survival rate if they are properly diagnosed as soon as possible in the early stages when the cancer is small and only in one area. Most researchers then started on early cancer detection as the first priority to control cancer.

To find solutions for the most dangerous cancers, the American Cancer Society compiled a list of cancer incidences last year, and the most common type of cancer is melanoma with 100,350 new cases and 6.850 deaths. In fact, the Skin Cancer Foundation predicted one out of every five

Americans by the age of 70 will develop skin cancer. Every hour, more than two people in the United States die from skin cancer. However, melanoma has a 99 percent five-year survival rate when detected early. Thus, melanoma detection is critical for decreasing the threat to patients with skin cancer. There is a popular method to examine the skin through skin surface microscopy called dermoscopy which is mainly applied to the evaluation of pigmented skin lesions. Dermatologists, based on selected information from dermoscopy, can diagnose melanoma easily. To generate a high efficiency for this method, magnifying lenses and lights must be of sufficient quality because different light powers or hand-held devices may return unexpected image quality in a dermoscopy process, such as blur, loss of features, and so on. Furthermore, only trained physicians could analyze precisely the dermoscopy dataset because it is entirely dependent on the visual acuity of the practitioner as well as their specialized knowledge. Overall, dermoscopy only could be used efficiently if it satisfies both conditions about technique equipment, lights, and experts.

Computer-aided diagnosis (CAD), a new solution, could detect automatically and diagnose melanoma efficiently without the experienced hands. CAD integrates elements of artificial intelligence and computer vision with image processing in radiology and pathology to improve radiologist performance [3]. Recent sensing technology [4–6], such as the medical Internet of Things and body sensor networks, also enhanced CAD. Piccolo et al. demonstrated that CAD was a useful tool for diagnosing melanoma compared with an inexperienced clinician [7]. More carefully, their study also exemplified sensitivity evaluation for the CAD, which achieved an accuracy rate of 92% compared with 69% in the evaluation for inexperienced clinicians. Because of the convenience and high accuracy, some algorithms based on the CAD were public for predicted diseases.

In this article, we propose the antialiasing attention spatial convolutional model (AASC) to segment automatic melanoma for skin lesions. A representation of the model is described in Figure 1. The AASC consists of the encoder and decoder sides. To indicate the location and strength of input features at the encoder, we designed a layer with double convolution that could simultaneously learn a huge number of filters from the input dataset automatically. Additionally, before the downsampling step, an attention module is added to the encoder to remark the signature features in the input dataset and remind the model to save these features during the training time. Antialiasing technique is proposed to reduce the dimension of the image and maintain shift-equivariance. At the decoder, the Pyramid Max Pooling Module (PMP) is considered a highlight for improving accuracy of the model. In fact, the module separates each input feature into four different sizes to decide the most important features and forward them to the next step. Furthermore, a skip connection is used to minimize the loss of information during down- and upsampling. To evaluate loss of the model, binary cross entropy is applied during training and testing. The preprocessing assistance is helpful in increasing the performance and removing the overfitting problem. For the preprocessing, we resized the input image dimension first and then used blur Gaussian to make the calculation process easier. Simultaneously, to overcome the constraint of the number of input images, horizontal and vertical flips and random rotation were used. As a result, the number of images has increased fourfold. Some optimization parameters were also checked and set up in the AASC model, such as a weight decay of 0.0005, a kernel regularizer of 0.0006, and a learning rate of 0.003. Finally, to check the true efficiency of the AASC model, we run the model in three different databases, namely ISIC 2016, ISIC 2017, and PH2, and evaluate the result under a variety of metrics such as recall, precision, accuracy, F1 score, Dice coefficients, and Jaccard indexes.

The rest of this paper is organized as follows. Section 2.1. gives an overview of the traditional algorithms for skin segmentation. Section 2.2. then describes the CAD systems. Section 2.3. introduces the proposed attention spatial model. Next, Section 3 summarizes the performance of the proposed method and the analytic results. Conclusion is finally drawn in Section 4.

2. Related Work

2.1. Traditional Algorithms for Skin Segmentation. In the early days, Principal component analysis, Markov Chains, or Otsu Algorithm, K-means clustering, Fuzzy C-means clustering were used for skin segmentation. Firstly, principal component analysis (PCA) applies some sort of transformation on a large set of variables of the original image data to condense information at a new set of fewer variables [8]. The main advantage of such a technique is that details not apparent in false color composite images can be highlighted in one of the component images that result. OLugbara [9] revealed that the skin lesion could be identified correctly through the PCA. However, the range between lesion and background was unclear, which caused mistakes in the diagnosis.

In view of the limitations of PCA, a Markov chain (MC) has been proposed for segmenting features of interest and shapes [10]. In comparison to conventional methods, the MC technique develops novel efficient methods for shape and texture segmentation resulting in higher accuracy and economical solutions. Although the MC algorithm outperformed the PCA algorithms, the segmentation generated by the MC algorithm has heterogeneous areas and fuzzy borders that highlight a part. As a result, healthy skin may be segmented as a skin lesion.

Continually, Otsu Algorithms [11], K-means clustering [12], and Fuzzy C-means clustering [13] are closely related for binary segmentation, but their performance for skin lesion segmentation is poor in the condition of a variety of skin types and minimal healthy skin. General cons of these algorithms must set parameters independently in each dataset, resulting in a limited application range.

2.2. CAD Systems. The CAD was presented in the introduction part and is highly recommended for skin segmentation in general and for melanoma segmentation in

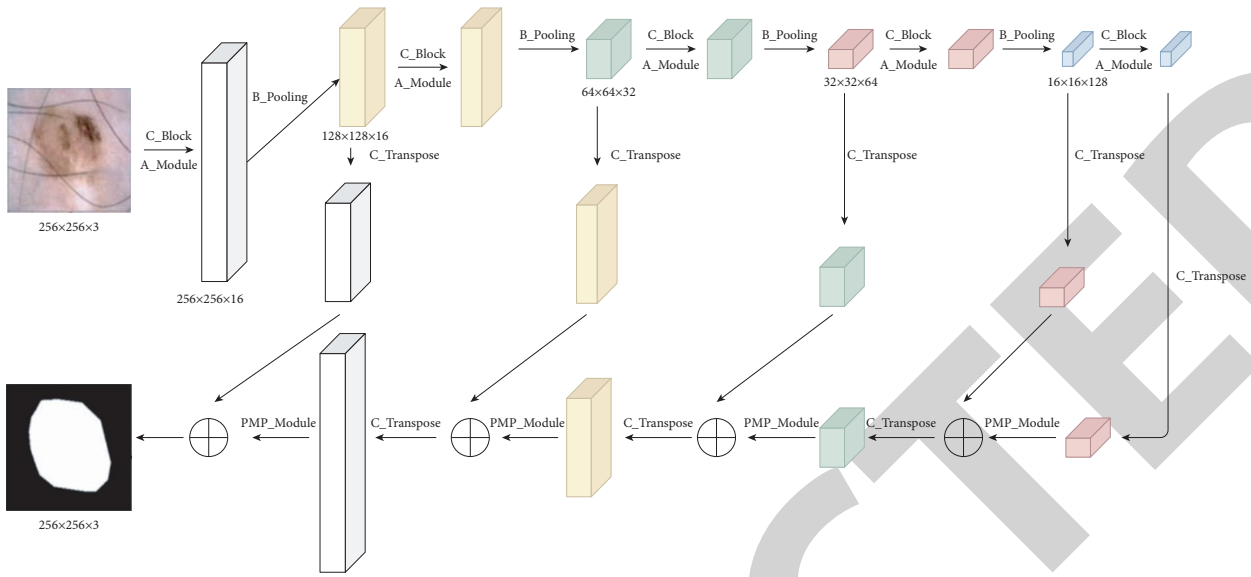


Figure 1: The pictorial representation of the proposed AASC model.

particular. The CAD system is mainly based on computer vision algorithms such as classification [14], detection [15, 16], and recognition [17–19]. At the International Skin Imaging Challenge (ISIC) of recent years, many methods based on the CAD were designed and evaluated as the top leader board for skin cancer segmentation. For example, in the ISIC 2016 dataset, the performance of Inception-v3 and Vgg-16 for skin lesion segmentation was evaluated, with the highest performance being around 61.6 percent and 69.3 percent testing accuracy, respectively [20]. Besides, Unet model can be run with fewer layers (total 23 convolutional layers) and training samples while still producing accurate segmentation results, and then it quickly became popular with many updated versions. The combination of Unet and Recurrent Residual Convolutional Neural Network (RRCNN) for skin cancer segmentation in ISIC 2017 achieved higher performance than SegNets and Residual Unet (ResUNets). In the following year, U-net34 ran on ISIC dataset 2018 for Melanoma segmentation with the average Jaccard index of 85,39%, and this result compared against the top-ranked team of 76,5% [21]. The U-net34 combines insight Unet decode and a pretrained Resnet34 as the Unet encode. The Resnet34 is made up of the initial convolutional layer, 16 blocks, and a fully connected layer. It is noticeable that the pretrained Resnet34 significantly improved the performance of the model. Another version of Unet was named LadderNet including a number of encoder-decoder paths [22]. In addition, a skip connection has been built into the adjacent decoder to save information from the encoder to the decoder.

In 2019, FucusNet presents the other Unet version, which includes multiple Unet models running in parallel, with the feature maps from the first decoding unit in the Unet associated with the components of the second encoding unit of Unet [23]. This model outperformed the Unet and ResUnet models in the 2017 skin cancer segmentation challenge. Last year, Kashan Zafar introduced UResNet-50 with 50 layers which contained ResNet architecture at the contracting path and Unet architecture at the expensive path [24]. The UResNet-50 performed well, with Jaccard Indexes of 77.2 and 85.4 percent on the ISIC 2017 and PH2 datasets, respectively, when compared to other architectures like the Mask-RCNN [25] and Deep labV3+ [26]. Actually, the Mask-RCNN is highly recommended for image segmentation because it includes an additional brand for predicting masks pixel by pixel and three outputs such as object segmentation, a class name, and a bounding box. Deeplabv3 is impressive by combining Atrous Spatial Pyramid Pooling (ASSP) for encoding multiscale contextual information and Encoder-Decoder Architecture for recovering both location and spatial information. Regrettably, Mask-RCNN and Deeplabv3 only attained the Jaccard indexes of 83% and 81.4% in the PH2 dataset, lower than that of UResNet-50. Based on previous research, we can conclude that the Unet model, i.e., its modified models, distributed significantly to segment skin cancer. However, some models were implemented without performing prepost processing on the input images causing a lack of responsiveness to sensitivity metric evaluation. Furthermore, running a parallel model may easily result in an overfitting problem. Overfitting is regarded as a major issue in medical databases due to the use of fewer datasets on the deeper model.

2.3. Antialiasing Attention Spatial Convolution Model (AASC). The preceding analysis demonstrates that many previous architectures, such as Unet, Fusion net, and Res-Unet, were successful for skin image segmentation, which includes two paths: encoding and decoding. In this study, AASC was also designed with an encoder and decoder approach. Instead of using available convolutional layers as in the previous versions, this encoding unit consists of reconstruction of convolution, attention module, and subsampling operations. Furthermore, the decoder unit was

implemented with a combination of atrous convolution layers at the PPM module in different sizes, convolution transpose, and several convolutional layers with a resolution of 256×256 pixels set as the input images. The output of the network is binary segmentation masks such as melanoma areas and backgrounds. For this purpose, the AASC model was trained and evaluated in three databases, namely ISIC 2016, ISIC 2017, and PH2. The AASC model architecture is shown in Figure 1.

On the encoding path, after receiving the input dataset, the convolutional block (C_Block) with two convolutional layers available in Keras library (Conv2D) set up corresponding parameters such as a stride (s) of 1, weight decay (w) of 0.0005, kernel regularizer (r) of 0.0006, a kernel (k) of 3×3 as shown in Figure 2.

Next, the attention module (A_module) is applied to concentrate on the highlight features as shown in Figure 3. Spatial information is complementary to channel information based on the A_module. It is meaningful to emphasize the position and the information of objects. Applying Global Average Pooling and Global Max Pooling along the channel axis and concatenating them generates the spatial attention map which shows the position of objects. The channel information map uses the two pooling operations resulting in the Global Average Pooling feature maps and the Global Max Pooling feature maps.

The main task of the encoding is to reduces the input size in order to make calculations simpler and to mark necessary features. However, using pooling layers is the reason for the variance problem. For instance, if we use Max-pooling with kernel of 2 and stride of 2 for [0,0,1,1,0,0,1,1] as input signal, the result will be [0,1,0,1]. This step significantly affects lost shift-equivariance. Shifting the input and output of a function is shift equivariant ($\widetilde{F}$) as defined in (1) if the input and output are shifted equally, so shifting and feature extraction are commutative. In contrast, shift invariance is shown in (2).

$$\text{Shift}_{\Delta h,\Delta w}(\widetilde{F}(X)) = \widetilde{F}\left(\text{Shift}_{\Delta h,\Delta w}(X)\right) \forall \Delta h, \Delta w, \quad (1)$$

$$\widetilde{F} = \widetilde{F}\left(\text{Shift}_{\Delta h,\Delta w}(X)\right) \forall \Delta h, \Delta w. \quad (2)$$

In which, R and W represent resolutions of an image. $X \in R^{R \times W \times 3}$ is input image and $\widetilde{F}(X) \in R^{R \times W \times 3}$ is the feature maps that could be rescaled to the original resolution.

To overcome these cons, the antialiasing technique was proposed in lieu of Max-pooling. Blur-pooling [27] is a new pooling technique that reduces the image sizes through two steps instead of one step, as in the traditional max-pooling operation. The blur pooling is described in Figure 4.

In the first step, the max operation is performed densely, which includes the Max-pooling layer with a stride of 1 and the Blur-pooling (B_Pooling) with a stride of 2 instead of the Max-pooling layer with a stride of 2. The second step integrates an antialiasing filter with subsampling.

In truth, each time the Blur-pooling layer is applied, the input dimension could be halved. Besides, output features at each Blur-pooling layer were implemented in two tasks. In the first task, these features are used in the next step with C_Block and A_module. The other task is to double the size of these features through convolution transpose (C_Transpose), which were considered as the important features in the decoder step. This step repeats four times during the training process, but it is notable that the values of the number in the filters of the attention module at each downsampling block are different, such as 16, 32, 64, 128. As a result, the number of channels decreases significantly, as does the calculating process and training time. Simultaneously, features ranging from low-level (such as colouring, contours, texture, and so on) to high-level (the entire shape of the object) are thoroughly learned in this encoder side.

The inputs of the decoder side are received from the output features of the encoder. Thus, the dimension range at the final encoded layer is 16×16, which is considered as the input of the first layer at the decoder. The primary function of the decoder is to increase the size of images. This is also the concerning root cause of the loss of important features when the dimension increases significantly and rapidly, as mentioned in previous studies. A solution is to give out a PMP module (Figure 5) in the decoder that divides each input into four parts so that features can be learned carefully before increasing the dimension. The mask and background are better and more accurately segmented through this step.

Overall, the AASC model provides superior performance, which has already been demonstrated against Unet, Res-Unet, Mask-R CNN, DeeplabV3+.

3. Experiment

3.1. Database. In this study, three available datasets were used: ISIC 2016 [28], ISIC 2017 [29], and PH2 [30]. Some examples of original datasets are shown in Figure 6.

The International Skin Imaging Collaboration (ISIC) includes expert-labeled digital images for melanoma and other cancer diagnostics. Every year, this organization launches skin lesion challenges in order to improve diagnosis. In 2016, 1279 images were public with corresponding masks by this organization for melanoma skin lesion segmentation. We split 1279 images into two sets: 900 images for training and 379 images for testing. The following year, ISIC 2017, contains 215 images, with 2000 and 150 images in training, and testing, respectively. The third dataset, PH2, consists of 40 images of melanoma and 160 images of common and atypical nevus. The total images were divided into two parts: 150 images for training and 50 images for testing.

3.2. Preprocessing Step. Lack of input dataset is one of the main factors affecting nuclei segmentation. Furthermore, the skin images were selected from various positions under different light conditions and equipment. Some skin areas are covered by arm and leg hair. It reduces segmentation performance. Thus, the preprocessing step is necessary to improve performance efficiency. Firstly, we resized the original images to the same size (256×256). And then, some augmentation techniques were suggested to increase the number of images.

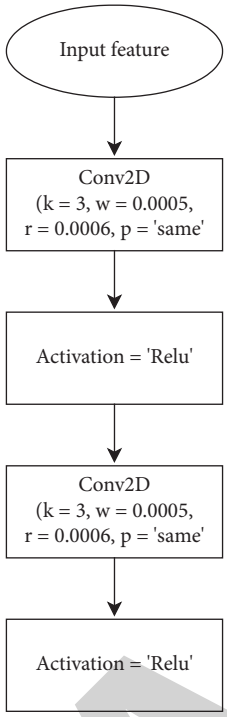


Figure 2: The variant of the convolutional layer.

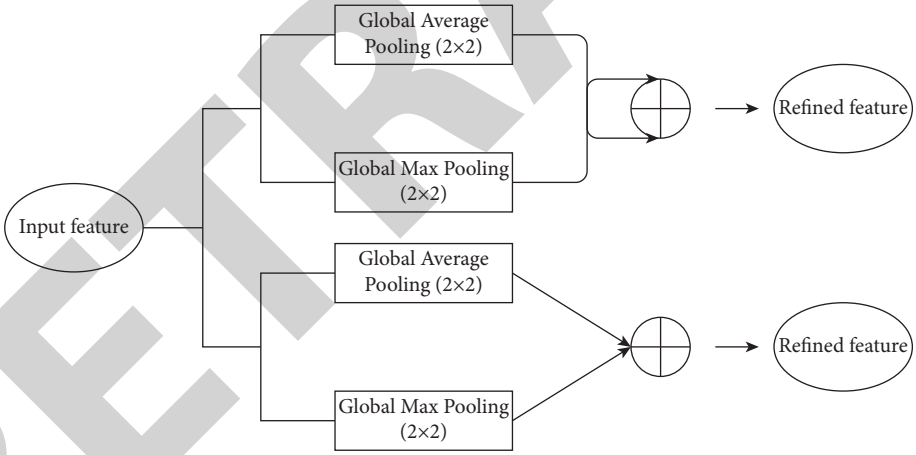


Figure 3: The variant of the convolutional layer.

Image Resizing: In truth, the original images have different sizes, such as 767×575, 1022×767, 1504×1129, or 2048×1536 that require additional computational time and affect the efficiency as the accuracy was low. Therefore, the input images were resized to new sizes like 256×256.

Gaussian Blur [31]: both objects and background are blurred before putting it into the next steps. This reduces the range of pixel values in the input images to simplify the computation and avoid distraction in training.

Horizontal flip [32]: the horizontal flip considers the simple and quick step to increase the number of images through flipping images around a horizontal centerline. It solutes the biggest problem in almost biomedical mathematics as the limitation of the dataset.

Vertical flip [32]: instead of flipping images around a horizontal line, the vertical flip turns around the vertical center flip. These techniques could create many images from one image under different views.

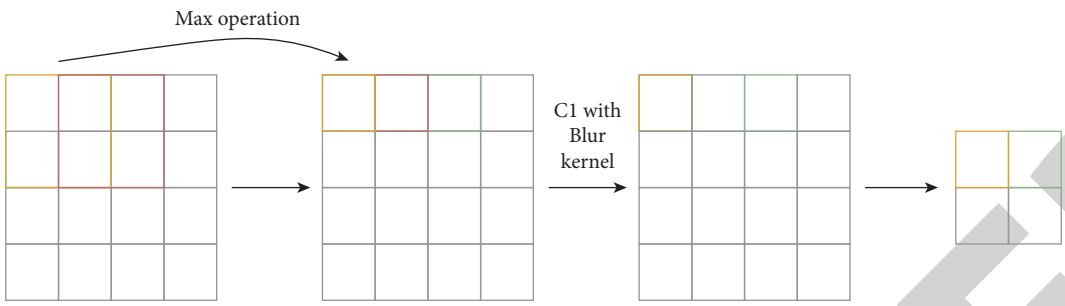


Figure 4: The blur-pooling.

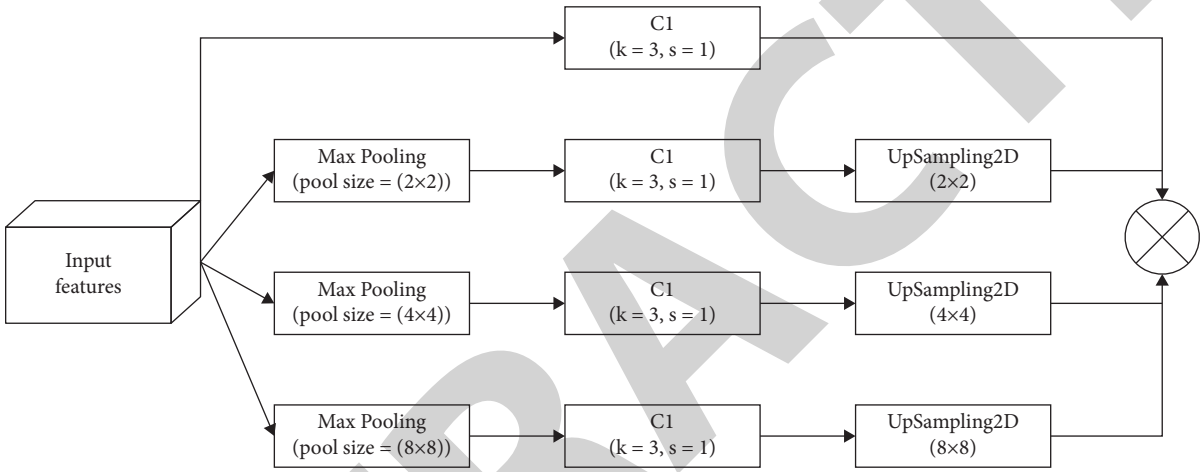


Figure 5: The pyramid max-pooling module.

Both flip [32]: both flip turns images around both horizontally and vertically at the same time.

Random rotation [33]: based on setting the rotation, an image could be randomly rotated to create a new image with the same content but different shapes. Random rotation also is a simple way to make variety in the number of images.

3.3. Network Training. The AASC model was trained for melanoma skin segmentation with hyperparameters as described in Table 1. We trained the model for 150 epochs. During training, data augmentation positively affected the performance of the model because of the increased number of samples. In addition, early stopping is configured to terminate the training model if the loss value does not decrease after 15 epochs. The loss and accuracy curves of the training and validation in ISIC 2016, ISIC 2017, and PH2 with the AASC method are shown in Figures 7–9.

3.4. Model Evaluation. In this paper, to assess the performance of the proposed model, we used six statistical metrics, namely precision, recall, accuracy, F1 score [34], Jaccard index (IoU), and the dice coefficient [35].

Precision is the ratio of correctly segmented skin lesion pixels to the total number of pixels. The ratio of correctly segmented skin lesion pixels to the total number of skin lesion pixels is defined as recall.

The F1 score (F1) is a test accuracy metric. It is derived from precision and recall. The ratio of the total number of correctly segmented pixels to the total number of skin lesions and background pixels is represented as accuracy.

The dice coefficient (Dice) is the ratio between the ground truth and the prediction. These evaluation metrics are based on the parameters listed below.

True positive (TP) refers to the number of skin lesion pixels that were correctly segmented as skin lesion pixels.

False negative (FN) refers to healthy skin pixels that are predicted as skin lesion pixels by the model.

False positive (FP) computes the statistics when the ground truth is skin lesion pixels, and the model predicts the healthy skin pixels.

True negative (TN) is the number of correctly segmented healthy skin pixels.

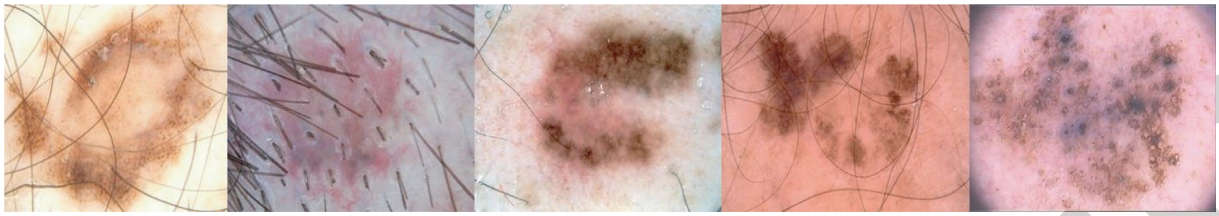


Figure 6: Examples of dermoscopic images in ISIC 2016, 2017, PH2.

Table 1: Hyperparameters used in the training process.

Parameters	Value
Input	$256 \times 256 \times 3$
Weight decay	0.0005
Kernel-regularizer	0.0006
Batch size	32
Learning rate	0.003
Optimizer	Adam
Epoch	150

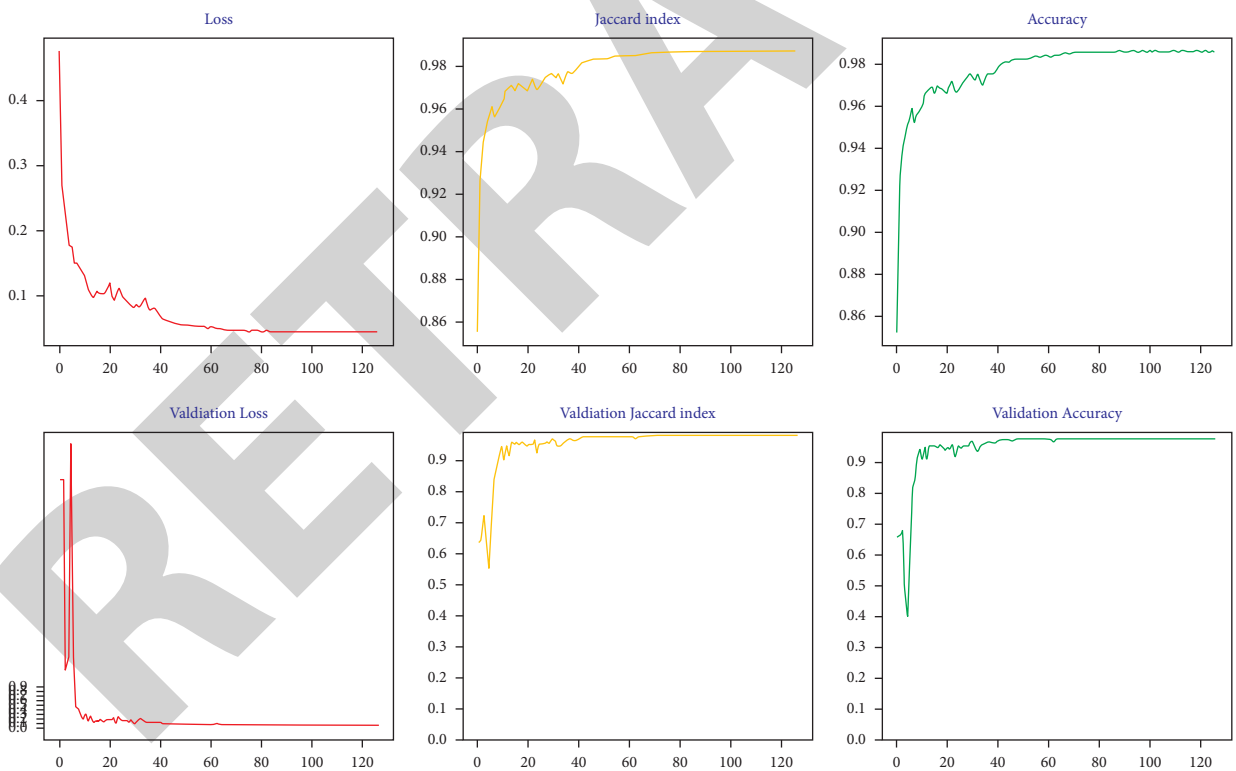


Figure 7: The loss and accuracy curves of training and validation in ISIC 2016.

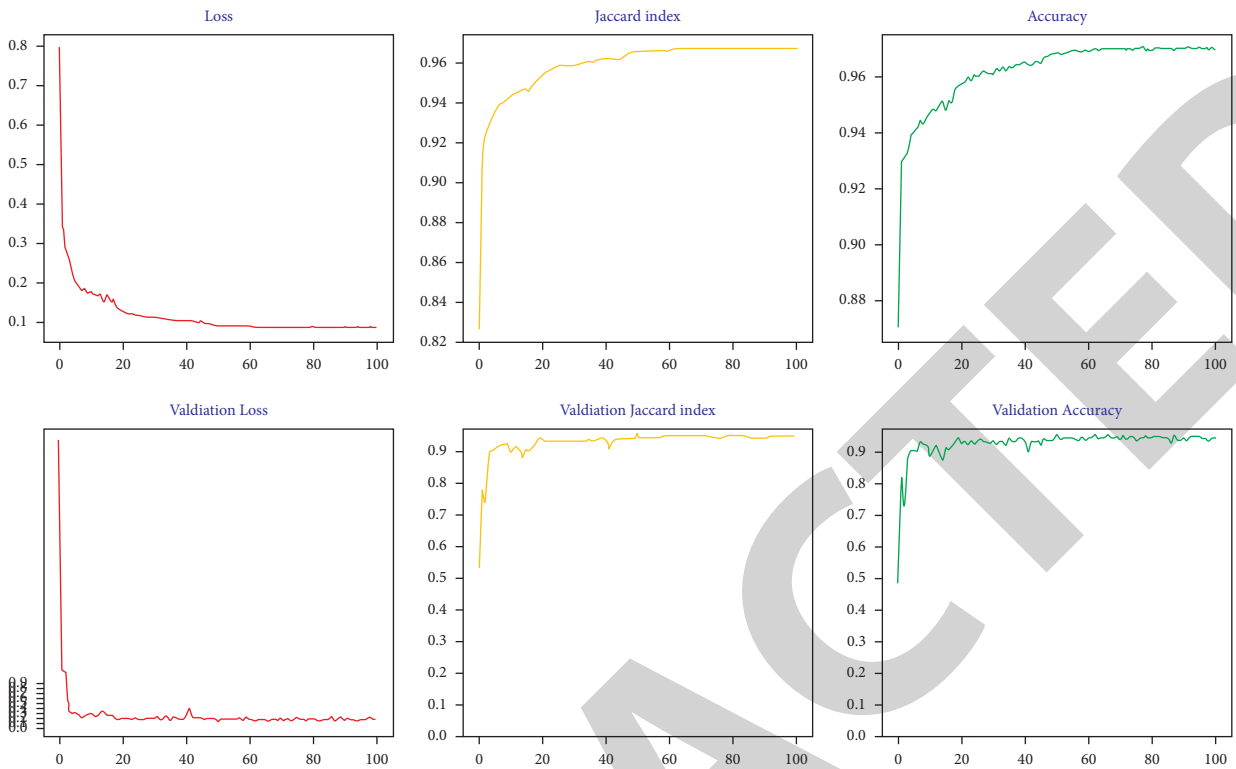


Figure 8: The loss and accuracy curves of training and validation in ISIC 2017.

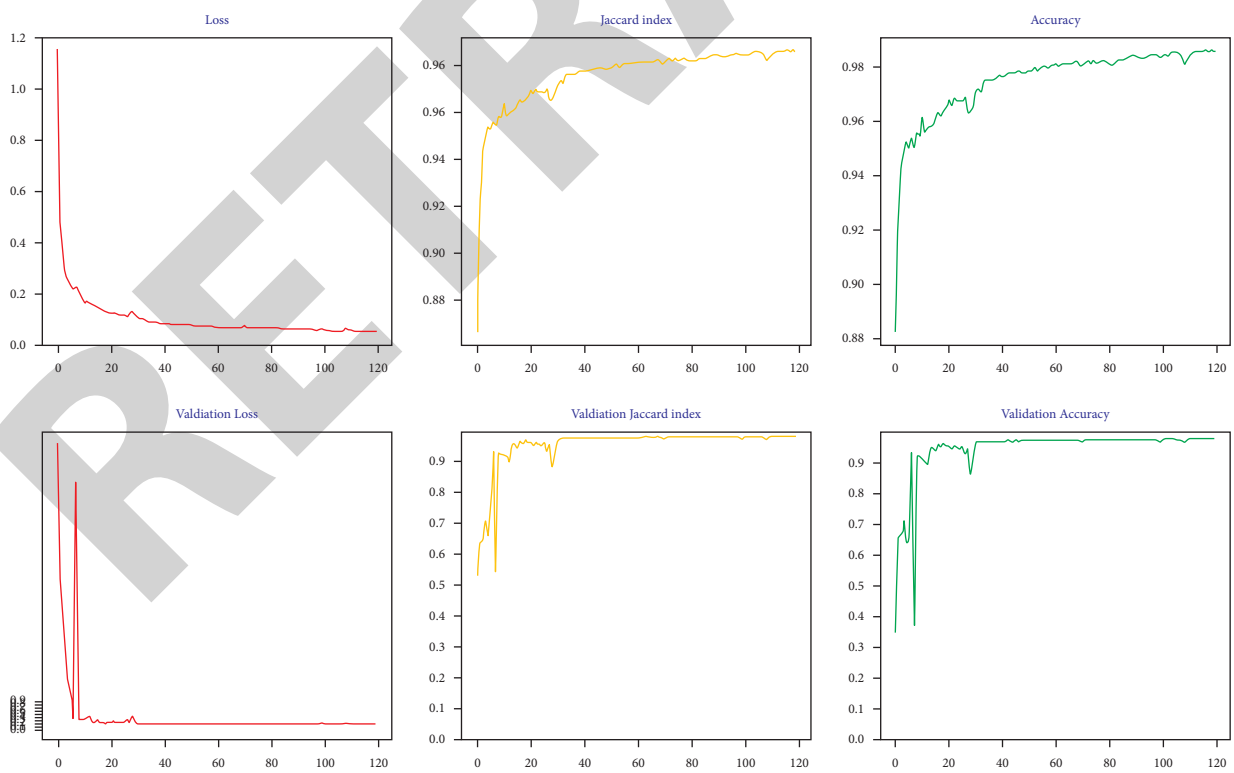


Figure 9: The loss and accuracy curves of training and validation in PH2.

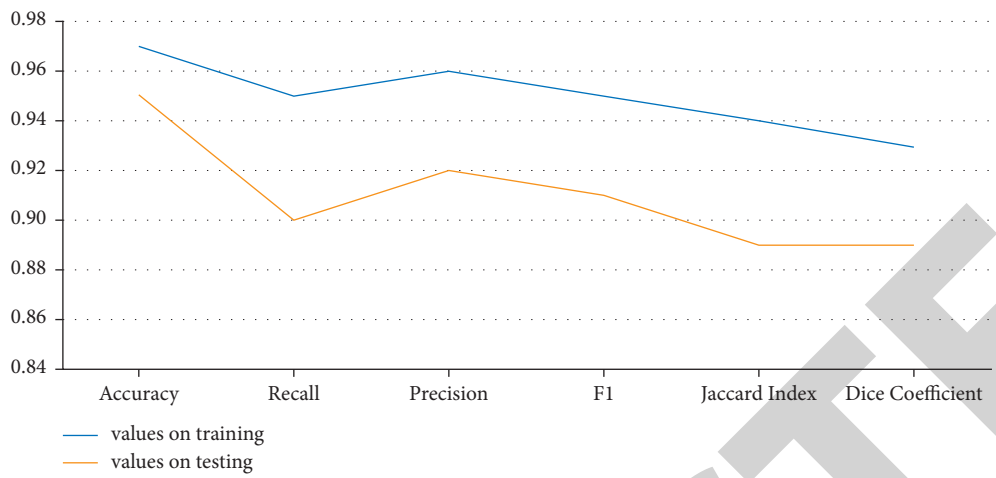


Figure 10: The performances of the proposed model were evaluated by the six metrics in both training and testing in ISIC 2016.

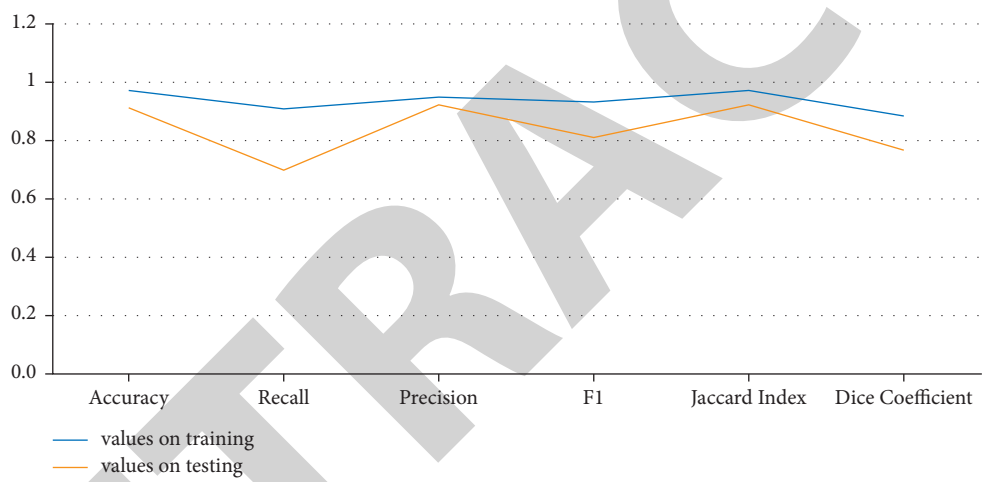


Figure 11: The performances of the proposed model were evaluated by the six metrics in both training and testing in ISIC 2017.

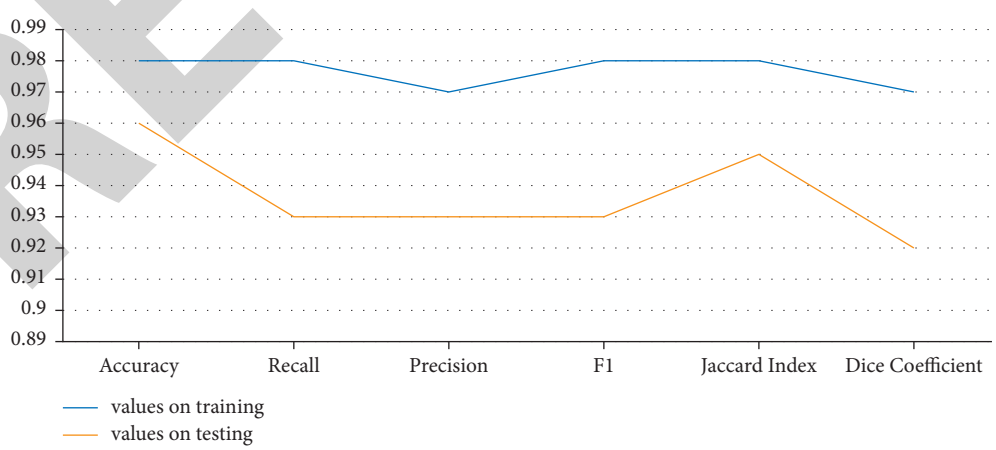


Figure 12: The performances of the proposed model were evaluated by the six metrics in both training and testing in PH2.

Table 2: Performance evaluation of the state-of-the-art segmentation methods in the ISIC 2016 dataset.

Model name	Jaccard index	Accuracy	Recall	Precision	F1
Unet [36]	0.76	0.94	0.90	0.89	0.89
Attention Unet [37]	0.81	0.94	0.90	0.89	0.89
Unet++ [38]	0.81	0.94	0.90	0.90	0.90
Recurrent U-Net [39]	0.79	0.94	0.89	0.88	0.87
Proposed model	0.89	0.96	0.90	0.92	0.91

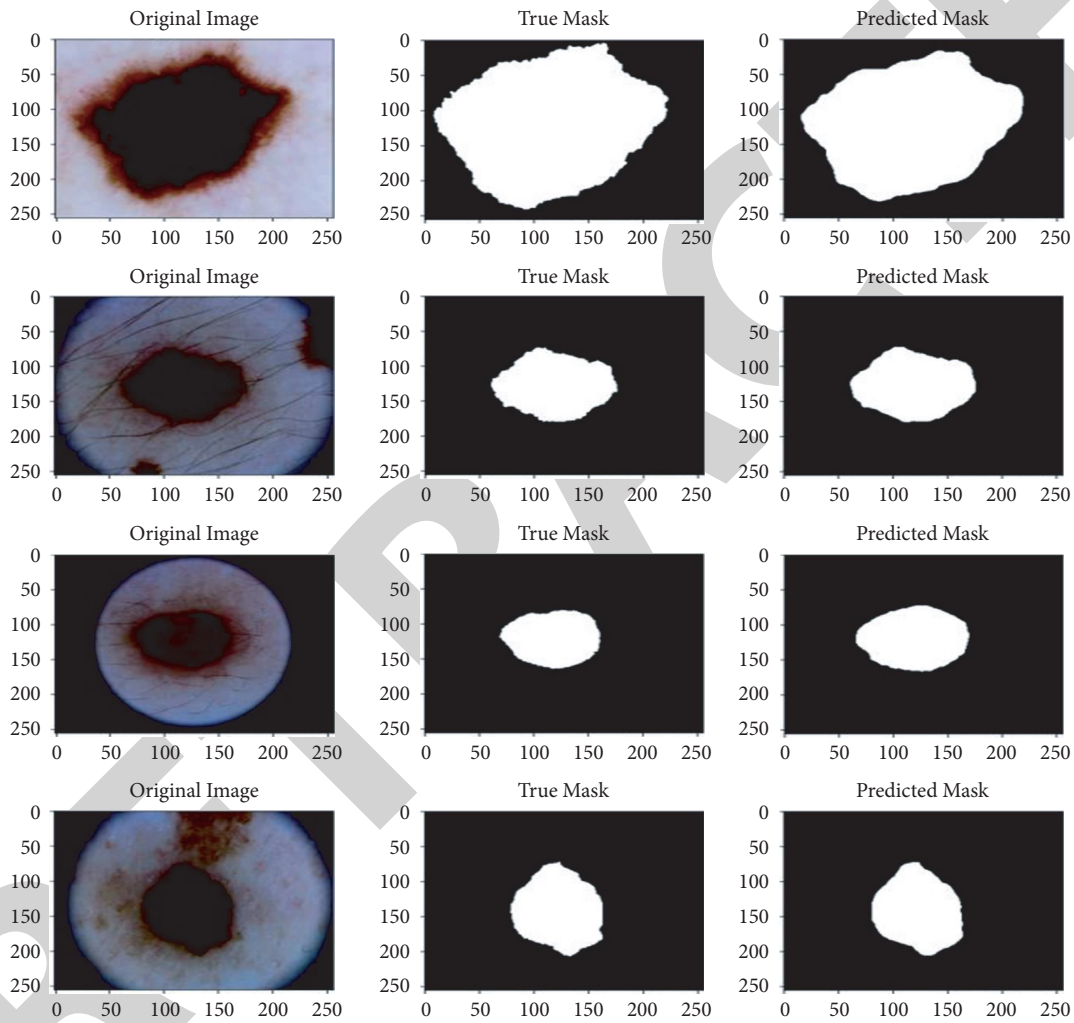


Figure 13: The visual segmentation of the AASC model in the ISIC 2016. From left to right is the visualization of the original images, true mask, and predicted mask.

Table 3: Performance evaluation of the state-of-the-art segmentation methods in the ISIC 2017 dataset.

Model name	Jaccard index	Accuracy	Recall	Precision	F1
Unet [36]	0.68	0.91	0.76	0.89	0.78
Attention Unet [37]	0.69	0.91	0.76	0.89	0.78
Unet++ [38]	0.68	0.91	0.75	0.90	0.77
Recurrent U-Net [39]	0.64	0.91	0.82	0.78	0.75
Proposed model	0.92	0.95	0.91	0.95	0.90

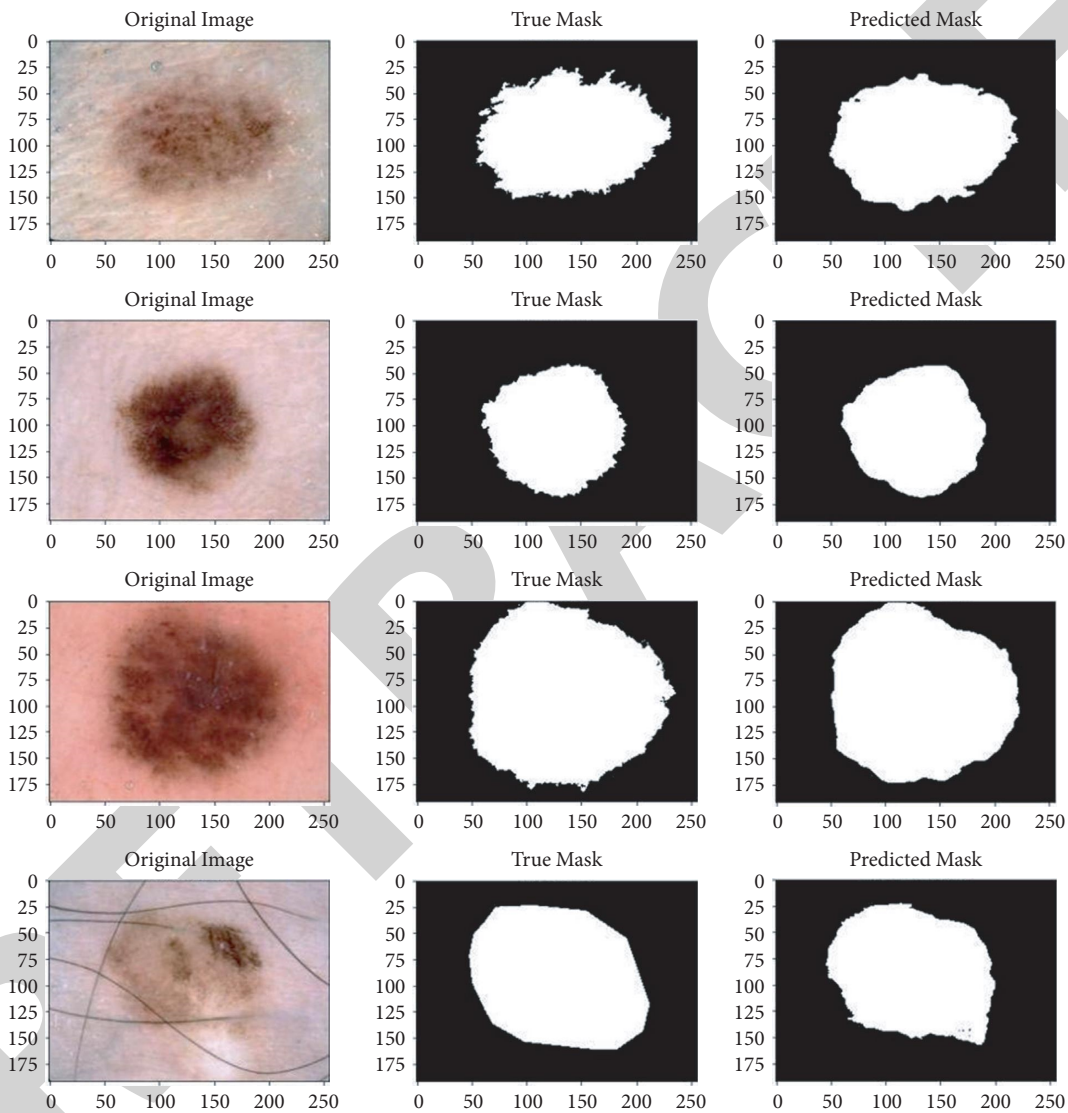


Figure 14: The result segmentation of the AASC model in the ISIC 2017. From left to right is the visualization of the original images, true mask, and predicted mask.

Table 4: Performance evaluation of the state-of-the-art segmentation methods in the PH2 dataset.

Model name	Jaccard index	Accuracy	Recall	Precision	F1
Unet [36]	0.79	0.91	0.89	0.89	0.87
Attention Unet [37]	0.80	0.92	0.90	0.89	0.88
Unet++ [38]	0.79	0.91	0.88	0.90	0.87
Recurrent U-Net [39]	0.80	0.92	0.93	0.87	0.88
Mask-RCNN [25]	0.84	0.94	0.93	—	—
DeepLab V3+ [26]	0.81	0.92	0.94	—	—
Proposed model	0.95	0.96	0.95	0.93	0.94

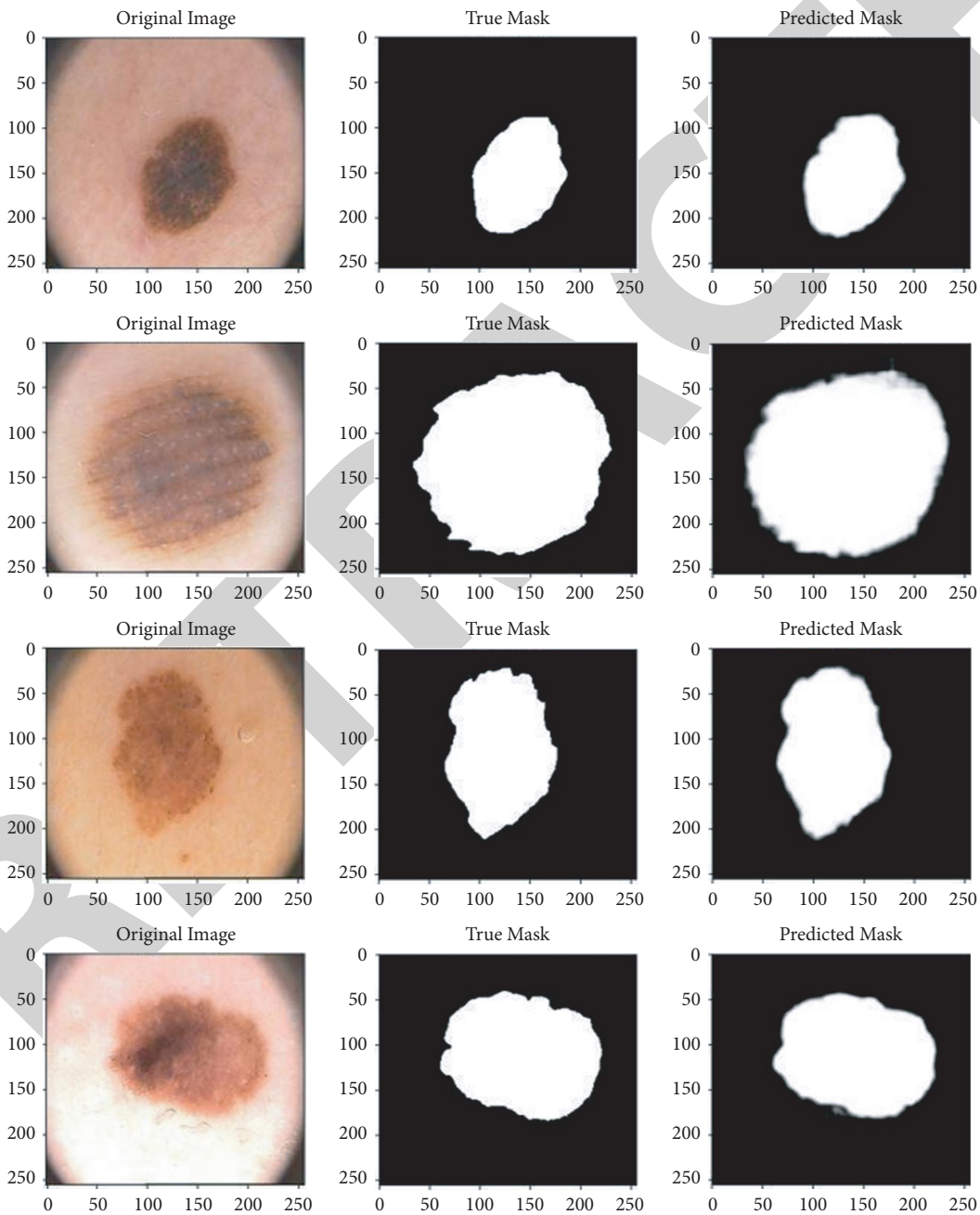


Figure 15: The result segmentation of the AASC model in the PH2. From left to right is a visualization of the original images, true mask, and predicted mask.

The mathematical definitions of these measures are as follows:

$$\begin{aligned}
\text{accuracy (ACC)} &= \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}},\\
\text{recall} &= \frac{\text{TP}}{\text{TP} + \text{FN}},\\
\text{precision} &= \frac{\text{TP}}{\text{TP} + \text{FP}},\\
\text{F1 score (F1)} &= \frac{2 * (\text{recall} + \text{precision})}{(\text{recall} + \text{precision})},\\
\text{Jaccard Index} &= \frac{|\text{GT} \cap \text{SR}|}{|\text{GT}| + |\text{SR}|},\\
\text{Dice Coefficient (Dice)} &= \frac{2 * \text{TP}}{2 * \text{TP} + \text{FN} + \text{FP}}.
\end{aligned} \tag{3}$$

where GT and SR are two sets. Based on these six evaluated metrics, we compared the performance of the proposed method in training and testing three datasets as shown in Figures 10–12. We observed that the training and testing performances are dependent on all three datasets, indicating a robust generalized model without overfitting.

3.4.1. Comparative Experiment in the ISIC 2016 Dataset. We compared our results to the state-of-the-art methods to show the feasibility and high reliability of our proposed model in three datasets. During the comparison, five measures were used: Jaccard Index, Accuracy, Recall, Precision, and F1. The five indicators in our suggested network were a strong sign that the AASC was effective for skin lesions. As a result, our performance consistently ranks first in all comparisons.

Clearly, Table 2 highlights the quantitative findings of the AASC and the existing approaches like Unet [36], Unet attention [37], Unet++ [38], and Recurrent-Unet [39] in the ISIC 2016 database. Overall, the Jaccard index achieved 89%. The accuracy rate reached as high as 96%. Continually, precision achieved 92%. F1 also arrived at 92%. Finally, 90% is the result evaluated by Recall. All evaluations demonstrated the effectiveness of AASC in the ISIC 2016. The visualization of melanoma skin segmentation using the AASC model in this dataset is shown in Figure 13.

3.4.2. Comparative Experiment in the ISIC 2017 Dataset. The AASC model was subsequently used to benchmark the ISIC 2017 to demonstrate the efficacy of our approach. The quantitative results of our model and the state-of-the-art models are shown in Table 3. The suggested network achieved satisfactory results under the five statistical metrics. Figure 14 shows the ISIC 2017 segmentation results.

3.4.3. Comparative Experiment in the PH2 Dataset. Furthermore, the PH2 database is a well-known database for melanoma skin lesions. Because the PH2 dataset has only 200 images, a recent study used the ISIC 2017 dataset for the model training to enhance the segmentation capability of their model before performing skin lesion segmentation in the PH2. We proposed training and testing in the PH2 dataset and compared their results in this work. Our proposed approach was performed, as illustrated in Table 4. The results are given in Figure 15.

4. Conclusion

Skin lesion segmentation is critical in the evolution of a computer-aided skin cancer diagnosis system. The AASC model was successfully developed in this paper, which focuses on the impressive features and then zooms in and out to evaluate these features under different views, ensuring that both low and high levels of information for skin segmentation in dermoscopic images can be learned thoroughly. Moreover, the preprocessing step enhanced the model performance, reduced shift-invariance loss, and removed overfitting. This study demonstrated strongly that the lightweight model (AASC model) could perform well without the dataset. The AASC algorithm has been tested in three databases, namely the ISIC 2016, ISIC 2017 challenge, and PH2 dataset. Jaccard index, Recall, Precision, F1 score, and Dice coefficient are famous statistical evaluation metrics used to evaluate and compare the efficiency of the AASC model and the other state-of-the-art models. The experimental result shows that the AASC model achieves the highest accuracy in melanoma skin lesion segmentation compared to the existing methods in the literature. The empirical result also demonstrates transparently that the presence of noises from input images such as shape distortion, thick hair, or low contrast was successfully removed by the proposed method. Future work will apply the AASC model to other applications, for example, melanoma classification in dermoscopic images. In addition, we try to promote the applicability of the model through its use in a variety of data.

Data Availability

The ISIC 2016, ISIC 2017, and PH2 databases used to support the findings of this study have been deposited in the International Skin Imaging Collaboration (ISIC) and Automatic computer-based Diagnosis system for Dermoscopy Images (ADDI) repository (https://challenge.isic-archive.com/data/ and https://www.fc.up.pt/addi/ph2).

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] T. Gansler, P. A. Ganz, M. Grant et al., "Sixty years of CA: a cancer journal for clinicians," *CA: A Cancer Journal for Clinicians*, vol. 60, no. 6, pp. 345–350, 2010.

[2] H. K. Weir, T. D. Thompson, A. Soman, B. Møller, and S. Leadbetter, "The past, present, and future of cancer

incidence in the United States: 1975 through 2020," *Cancer*, vol. 121, no. 11, pp. 1827–1837, 2015.

[3] B. Halalli and A. Makandar, "Computer aided diagnosis-medical image analysis techniques," in *Breast Image*, pp. 85–109, IntechOpen, London, UK, 2018.

[4] B.-W. Chen and W.-C. Ye, "Low-error data recovery based on collaborative filtering with nonlinear inequality constraints for manufacturing processes," *IEEE Transactions on Automation Science and Engineering*, vol. 18, no. 4, pp. 1602–1614, 2020.

[5] B.-W. Chen, K.-L. Hou, P.-H. Wu, W.-C. Ye, and J.-Y. Huang, "Heterogeneous multiview crowdsensing based on half quadratic optimization for the visual internet of things," *IEEE Wireless Communications*, vol. 28, no. 4, pp. 19–25, 2021.

[6] B.-W. Chen, "Novel kernel orthogonal partial least squares for dominant sensor data extraction," *IEEE Access*, vol. 8, pp. 36131–36139, 2020.

[7] D. Piccolo, A. Ferrari, K. Peris, R. Daidone, B. Ruggeri, and S. Chimenti, "Dermoscopic diagnosis by a trained clinician vs. a clinician with minimal dermoscopy training vs. computer-aided diagnosis of 341 pigmented skin lesions: a comparative study," *British Journal of Dermatology*, vol. 147, no. 3, pp. 481–486, 2002.

[8] R. Margolin, A. Tal, and L. Zelnik-Manor, "What makes a patch distinct?" in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1139–1146, Portland, Oregon, USA, June 2013.

[9] O. O. Olugbara, T. B. Taiwo, and D. Heukelman, "Segmentation of melanoma skin lesion using perceptual color difference saliency with morphological analysis," *Mathematical Problems in Engineering*, vol. 2018, pp. 1–19, 2018.

[10] B. Jiang, L. Zhang, H. Lu, C. Yang, and M. Yang, "Saliency detection via absorbing Markov chain," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1665–1672, Sydney, New South Wales, Australia, December 2013.

[11] Z. Qu and L. Zhang, "Research on image segmentation based on the improved Otsu algorithm," in *Proceedings of the Second International Conference on Intelligent Human-Machine Systems and Cybernetics*, pp. 228–231, Nanjing, China, August 2010.

[12] R. Melli, C. Grana, and R. Cucchiara, "Comparison of color clustering algorithms for segmentation of dermatological images," *SPIE Medical Imaging*, vol. 6144, Article ID 61443S, 2006.

[13] H. Castillejos, V. Ponomaryov, L. Nino-de-Rivera, and V. Golikov, "Wavelet transform fuzzy algorithms for dermoscopic image segmentation," *Computational and Mathematical Methods in Medicine*, vol. 2012, Article ID 578721, 2012.

[14] K. L. Hua, C. H. Hsu, S. C. Hidayati, W. H. Cheng, and Y. J. Chen, "Computer-aided classification of lung nodules on computed tomography images via deep learning technique," *OncoTargets and Therapy*, vol. 8, pp. 2015–22, 2015.

[15] K.-Y. Lung, C.-R. Chang, S.-E. Weng, H.-S. Lin, H.-H. Shuai, and W.-H. Cheng, "ROSNet: robust one-stage network for CT lesion detection," *Pattern Recognition Letters*, vol. 144, pp. 82–88, 2021.

[16] J. Sanchez-Riera, S. Kathiravan, H. Kai-Lung, C. Wen-Huang, M. Anwar Hossain, and M. F. Alhamid, "Robust RGB-D hand tracking using deep learning priors," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 28, no. 9, pp. 2289–2301, 2017.

[17] T.-H. Tsai, W.-H. Cheng, C.-W. You, M.-C. Hu, A. W. Tsui, and H.-Y. Chi, "Learning and recognition of on-premise signs from weakly labeled street view images," *IEEE Transactions on Image Processing*, vol. 23, no. 3, pp. 1047–1059, 2014.

[18] J. Sanchez-Riera, K.-L. Hua, Y.-S. Hsiao, T. Lim, S. C. Hidayati, and W.-H. Cheng, "A comparative study of data fusion for RGB-D based visual recognition," *Pattern Recognition Letters*, vol. 73, pp. 1–6, 2016.

[19] L. Lo, H. X. Xie, H.-H. Shuai, and W.-H. Cheng, "Facial chirality: using self-face reflection to learn discriminative features for facial expression recognition," in *Proceedings of the IEEE International Conference on Multimedia and Expo (ICME)*, Shenzhen, China, July 2021.

[20] M. Goyal, A. Oakley, P. Bansal, D. Dancey, and M. H. Yap, "Skin lesion segmentation in dermoscopic images with ensemble deep learning methods," *IEEE Access*, vol. 8, pp. 4171–4181, 2019.

[21] F. Guth and T. E. deCampos, "Skin lesion segmentation using U-Net and good training strategies," 2018, https://arxiv.org/abs/1811.11314.

[22] J. Zhuang, "LadderNet: multi-path networks based on U-Net for medical image segmentation," 2018, https://arxiv.org/abs/1810.07810.

[23] C. Kaul, S. Manandhar, and N. Pears, "Focusnet: an attention-based fully convolutional network for medical image segmentation," in *Proceedings of the IEEE International Symposium on Biomedical Imaging*, pp. 455–458, Venice, Italy, April 2019.

[24] K. Zafar, S. O. Gilani, A. Waris et al., "Skin lesion segmentation from dermoscopic images using convolutional neural network," *Sensors*, vol. 20, no. 6, p. 1601, 2020.

[25] J. W. Johnson, "Adapting mask-rcnn for automatic nucleus segmentation," 2018, https://arxiv.org/abs/1805.00500.

[26] Y. Wang, "Skin lesion segmentation using atrous convolution via DeepLab V3," 2018, https://arxiv.org/abs/1807.08891.

[27] R. Zhang, "Making convolutional networks shift-invariant again," in *Proceedings of the International Conference on Machine Learning*, pp. 7324–7334, Long Beach, CA, USA, June 2019.

[28] N. C. D Gutman, E. Celebi, B. Helba, M. Marchetti, and N. Mishra, "Skin lesion analysis toward melanoma detection: a challenge at the international symposium on biomedical imaging (ISBI) 2016, hosted by the international skin imaging collaboration (ISIC)," 2016, https://arxiv.org/abs/1605.01397.

[29] D. Gutman, "Skin lesion analysis toward melanoma detection: a challenge at the 2017 international symposium on biomedical imaging (ISBI), hosted by the international skin imaging collaboration (ISIC)," in *Proceedings of the International Symposium on Biomedical Imaging*, pp. 168–172, Washington, DC, USA, April 2018.

[30] T. Mendonça, P. M. Ferreira, J. S. Marques, A. R. S. Marcal, and J. Rozeira, "PH2-a dermoscopic image database for research and benchmarking," in *Proceedings of the IEEE Engineering in Medicine and Biology Society*, pp. 5437–5440, Osaka, Japan, July 2013.

[31] E. S. Gedraite and M. Hadad, "Investigation on the effect of a Gaussian Blur in image filtering and segmentation," in *Proceedings of the International Symposium on Electronics in Marine*, pp. 393–396, Zadar, Croatia, September 2011.

[32] M. Nixon and A. Aguado, *Feature Extraction and Image Processing for Computer Vision*, Academic Press, Cambridge, MA, USA, 2019.

Hindawi
Wireless Communications and Mobile Computing
Volume 2022, Article ID 3213311, 11 pages
https://doi.org/10.1155/2022/3213311

Research Article

Time-Driven Scheduling Based on Reinforcement Learning for Reasoning Tasks in Vehicle Edge Computing

Bing Lin,[1] **Qiaoxin Chen**,[1] **and Yu Lu**[2]

[1]*College of Physics and Energy, Fujian Normal University, Fuzhou 350117, China*
[2]*Concord University College, Fujian Normal University, Fuzhou 350117, China*

Correspondence should be addressed to Yu Lu; fzluyu@163.com

Received 19 October 2021; Revised 13 January 2022; Accepted 22 January 2022; Published 23 February 2022

Academic Editor: Chi-Hua Chen

Significant challenges for reasoning tasks scheduling remain, including the selection of an optimal tasks-servers solution from the possible numerous combinations, due to the heterogeneous resources in edge environments and the complicated data dependencies in reasoning tasks. In this study, a time-driven scheduling strategy based on reinforcement learning (RL) for reasoning tasks in vehicle edge computing is designed. Firstly, the reasoning process of vehicle applications is abstracted as a model based on directed acyclic graphs. Secondly, the execution order of subtasks is defined according to the priority evaluation method. Finally, the optimal tasks-servers scheduling solution is chosen by Deep Q-learning (DQN). The extensive simulation experiments show that the proposed scheduling strategy can effectively reduce the completion delay of reasoning tasks. It performs better in algorithm convergence and runtime compared with the classic algorithms.

1. Introduction

In recent years, Internet of Vehicles (IoV) has become a research hotspot for the Intelligent Transportation System (ITS) [1]. The autonomous driving of IoV not only improves the driving safety, but also solves the problem of traffic inefficiency and lane congestion. It is a challenge for the autonomous driving to complete the target application under strict time constraint and restricted computing resources. The current work for autonomous driving mostly focuses on how to design the specific functions, such as traffic recognition, into reasoning tasks [2, 3]. Less attention is paid to how to schedule these reasoning tasks of autonomous driving to the appropriate computing nodes with low latency. Fortunately, IoV in Mobile Edge Computing (MEC) could schedule the real-time tasks from vehicles to the Road Side Units (RSU) with powerful computing resources, alleviating the task execution delay. Besides, a reasonable scheduling for reasoning tasks in MEC can effectively reduce both the execution latency of tasks and the workload of vehicles [4–11]. However, due to the heterogeneous resources in edge environments and the complicated data dependencies in reasoning tasks, significant challenges for reasoning tasks scheduling remain, including the selection of an optimal tasks-servers solution from the possible numerous combinations.

Existing studies are mainly done subject to task scheduling and task coordination through heuristic algorithms [12–15], such as Particle Swarm Optimization (PSO), Colony Algorithm (CA), and Genetic Algorithm (GA). Although these works could obtain the feasible solutions while satisfying different constraints, they fail to predict the deviation between the feasible and optimal solutions in advance, which makes their solutions easily fall into the local optimum. Several studies have been devoted to task scheduling using reinforcement learning (RL) algorithm [16–26], which can not only correct the deviation between the feasible and optimal solutions, but also accelerate the convergence of perfect results. Specifically, Lin et al. [23] proposed a time-driven scheduling strategy based on Q-learning algorithm for reasoning tasks of autonomous driving in IoV. The experimental results demonstrated that the performance of RL algorithms based on simulated annealing was better than other classic algorithms. This work

is instructive for our work. Zhao et al. [20] put forward a distribution scheduling algorithm based on DQN to achieve the best balance between latency, computational rate, and energy consumption, for an edge access network of IoV. They prioritized the tasks of different vehicles according to the analytic hierarchy process (AHP). The experimental results showed that the average task processing delay of the proposed method could effectively improve the task offload efficiency. However, the priority between tasks has not been scientifically calculated and weighted, but only evaluated by experts based on their experience. The current work [27, 28] for priority evaluation is mostly subjective by experts. There are great achievements in multivehicle task collaborative scheduling [5, 7, 11, 20]. However, the time-driven scheduling for single-vehicle reasoning tasks with data dependencies is still an open issue.

In response to this issue, two research questions are considered: (1) how to design a model for the reasoning tasks with data dependencies to evaluate the latency caused by task execution and data transmission? (2) How to develop an efficient and reliable scheduling strategy to reduce the latency during vehicle driving? To solve the above questions, we design time-driven scheduling based on RL for reasoning tasks in vehicle edge computing, which considers the differences of heterogeneous real-time reasoning tasks and optimizes the completion latency of reasoning tasks.

The main contributions of this paper are concluded as follows:

(1) A latency model is designed for reasoning tasks with data dependencies, which considers the latency caused by task execution and data transmission

(2) The scheduling for reasoning tasks in MEC is defined as a Markov Decision Process (MDP), which models the scheduling strategy for a reasoning task as the state, the resource allocation decision for each subtask as the action, and the completion latency of a reasoning task as the reward

(3) A time-driven scheduling strategy based on DQN is designed to explore an optimal tasks-servers solution from the possible numerous combinations in vehicle edge computing

The remaining part of the paper proceeds as follows. We review the related work in Section 2. Section 3 introduces the problem definitions of reasoning tasks scheduling. Section 4 describes the proposed reasoning tasks scheduling strategy in detail. Section 5 conducts the comparative experiments and analyzes the performance of the proposed strategy. Finally, Section 6 summarizes the work of this paper and looks forwards to the future research directions.

2. Related Work

Task scheduling in MEC has been extensively studied [4–10]. In general, task scheduling approaches mainly include the methods based on heuristic algorithms [12–15] and reinforcement learning [16–26].

2.1. Methods Based on Heuristic Algorithms. Xie et al. [12] proposed a novel Directional and Non-local-Convergent Particle Swarm Optimization (DNCPSO) to address workflow scheduling in cloud-edge environment, which could reduce the make span and cost dramatically and work well for task scheduling in complex applications. Wu et al. [13] studied how to dynamically partition a given application effectively into local and remote parts while reducing the total cost in cloud-edge environment. They proposed a Min-Cost Offloading Partitioning (MCOP) algorithm, which could significantly reduce the execution time and energy consumption by optimally distributing tasks between mobile devices and servers. Lin et al. [15] proposed a linear-time rescheduling algorithm for the task migration in MCC environment. The algorithm started from a minimal-delay scheduling solution and subsequently performed energy reduction by migrating tasks among the local cores and the cloud.

The methods based on heuristic algorithms can easily fall into the local optimal solution, which usually fails to get a good result. Moreover, the time required to process reasoning tasks of IoV is usually strict. The methods based on heuristic algorithms are not suitable for such problem due to their long algorithm execution time.

2.2. Methods Based on Reinforcement Learning. To adapt the scheduling strategies for dynamic scenarios, Deep Reinforcement Learning (DRL) has been widely applied to the task scheduling problems in MEC systems in recent years.

Chen et al. [16] designed a double DQN-based computation scheduling policy for a virtual MEC system. Numerical experiments showed that their proposed policy could achieve a significant improvement in computation scheduling performance. Xiong et al. [17] proposed an improved DQN algorithm to minimize the long-term weighted sum of average completion time of jobs and average number of requested resources in IoT edge computing system. Simulation results showed that the proposed algorithm has a better performance than the original DQN algorithm. Wang et al. [18] proposed a new DRL-based scheduling framework to address the challenges of task dependency and adapting to dynamic scenarios in the MEC system. The proposed DRL solution could automatically discover the common patterns behind various applications so as to infer an optimal scheduling policy in different scenarios. Rjoub et al. [21, 26] proposed four deep and RL-based scheduling approaches to automate the process of scheduling large-scale workloads onto cloud computing resources, while reducing both the resource consumption and task waiting time. These approaches derived an appropriate task scheduling mechanism that could minimize both tasks' execution delay and cloud resources utilization. Qi et al. [22] firstly proposed a multitask DRL approach for scalable parallel task scheduling (MDTS) in IoV. For avoiding the curse of dimensionality when coping with complex parallel computing environments and jobs with diverse properties, they extended the action selection in DRL to a multitask decision, where the output branches of

multitask learning were fine-matched to parallel scheduling tasks. Huang et al. [24] proposed a DRL-based Online Offloading (DROO) framework to optimally adapt task scheduling decisions and wireless resource allocations to the time-varying wireless channel conditions in a wireless powered MEC network. Numerical results showed that the proposed framework could achieve near-optimal performance while significantly decreasing the computation time.

RL-based methods mostly assume that the scheduling problem is a learning task. Through preliminary training, an effective scheduling policy for the task can be quickly formed by a reasonable designed RL algorithm. Note that current work for IoV mostly focuses on multivehicle collaborative scheduling, but the time-driven scheduling for single-vehicle reasoning tasks with data dependencies is still an open issue.

3. Problem definition

Table 1 shows the notations used in this paper.

Figure 1 gives an example of reasoning tasks scheduling in vehicle edge computing. This example considers autonomous driving reasoning system [2, 3], which consists of applications such as emergency rule inference engine and security operations. The user equipment (UE) makes scheduling decision for those applications according to the status in edge environments and application profiles; thus some of them are executed locally on the vehicle (i.e., UE) while others are scheduled to the edge by wireless channels. In this work, we consider that the edge environment is composed of m RSUs providing resources including computing, communications, and storage to UE in each time-slot, which is expressed as $\mathbf{F} = \{f_j \mid 1 < j < m\}$. The computation capacities of vehicle and RSU are denoted as f_{vehicle} and f_{RSU}.

In time-slot i, a reasoning task can be expressed as a directed acyclic graph (DAG) $\mathbf{G} = \langle \mathbf{N}, \mathbf{E} \rangle$ as in Figure 2, where $\mathbf{N} = \{n_i \mid 1 < i < z\}$ is a set of z subtasks and $\mathbf{E} = \{\mathbf{e_{u,k}} \mid 1 \le u, k \le z, u \ne k\}$ is the data dependencies between subtasks. The data dependency $\mathbf{e_{u,k}}$ indicates that there is a directed arc between subtasks n_u and n_k, and task n_k cannot start until task n_u has been completed. The set of direct precursors of task n_k can be expressed as $\mathbf{R_k} = \{n_k \mid 1 \le k \le z, k \ne j, \mathbf{e_{u,k}}\}$. A task cannot be executed until all of its direct precursors are completed.

In vehicle edge computing, a subtask in the reasoning task can be either offloaded to the edge or executed locally on the vehicle. If the offloading occurs for a reasoning task, the process delay will be related to the subtask profile and the environment state. The subtask profile includes the required CPU cycles for running the subtasks c_i, data size of the subtasks data_i, and the tolerable delay of the subtasks t_d^i. Besides, the environment state contains the transmission rate of wireless channel $v_{\text{transmission}}^{m_i}$. Therefore, the transmission latency $t_{\text{transmission}}^i$ for subtasks executing on edge nodes can be calculated as (1) and (2). If subtasks are executed locally on the vehicle, there is only execution latency on the user equipment, which can be obtained by t_{vehicle}^i.

$$t_{\text{transmission}}^i = \frac{\text{data}_i}{v_{\text{transmission}}^{m_i}}, \tag{1}$$

$$t_{\text{RSU}}^i = \frac{c_i}{f_{\text{RSU}}}, t_{\text{vehicle}}^i = \frac{c_i}{f_{\text{vehicle}}}. \tag{2}$$

The scheduling plan for a reasoning task is denoted as distribution relationship matrix as (3). If $a_{x,y} = 1$, it means that subtask n_x is offloaded to edge node m_y; otherwise, it means that subtask n_x is executed locally. When the edge node is running normally, the execution latency of the reasoning tasks can be expressed by (4), where $t_{process}$ means the processing latency of the reasoning tasks. If there is no available edge node in the edge environment, all subtasks will be executed serially on the vehicle, where m_i is set to 0. When the worst scheduling occurs, the completion latency of the reasoning tasks is described as in equation (5):

$$\mathbf{A}_{\mathbf{m_i} \times \mathbf{z_i}} = \begin{Bmatrix} 1 & \dots & a_{1,z_i} \\ \dots & \dots & \dots \\ 0 & \dots & a_{m_i,z_i} \end{Bmatrix}, \tag{3}$$

$$t_{\text{all}} = t_{\text{process}} + \text{MAX}\left(t_{\text{transmission}}^i\right), \tag{4}$$

$$t_{\text{worst}} = \sum_{i=0}^{z} t_{\text{vehicle}}^i. \tag{5}$$

To make better use of computing resources in different edge environments, we assume that edge nodes should satisfy the following processing principles:

(1) A subtask is processed by only one corresponding edge node, which is formally defined as (6).

(2) After all subtasks are assigned to the corresponding edge nodes, the edge nodes begin to process the subtasks.

(3) The subtasks on different edge nodes without data dependencies can be processed in parallel.

(4) The subtasks on the same edge node are processed according to the data dependencies. Otherwise, they are processed according to their corresponding priorities.

(5) The tolerable delay of the subtasks on the edge node is not greater than the execution latency of the corresponding edge node, which is formally defined as (7):

$$\sum_{y=1}^{z_i} a_{x,y} \le 1, \tag{6}$$

$$\begin{aligned} t_{\text{RSU}}^i &\le t_d^i, \\ t_{\text{vehicle}}^i &\le t_d^i. \end{aligned} \tag{7}$$

The reasoning task scheduling discussed in this paper can be summarized as follows: in various time-slots, a reasoning task on a vehicle can be decomposed into several

Table 1: Notations and descriptions.

Notation	Description
$\mathbf{G}$, $\mathbf{N}$, $\mathbf{E}$	A reasoning task, a set of subtasks, and the data dependencies between subtasks
$\mathbf{A}_{\mathbf{m_i}\times\mathbf{z_i}}$	Computational scheduling plan for n tasks
c_i, data_i, t_d^i	Task size, computation intensity, tolerable delay of subtask n_i
t_{vehicle}^i, t_{process}	Computation latency in vehicle and RSU
f_{vehicle}, f_{RSU}	Local computing capacity and scheduling capacity
$t_{\text{transmission}}^i$	Transmission latency of subtasks
t_{all}	Completion latency of a reasoning task
s_t, a_t, r_t	State, action, and reward of an MDP at time step t
α, γ	Parametrized policy and value function for computation scheduling

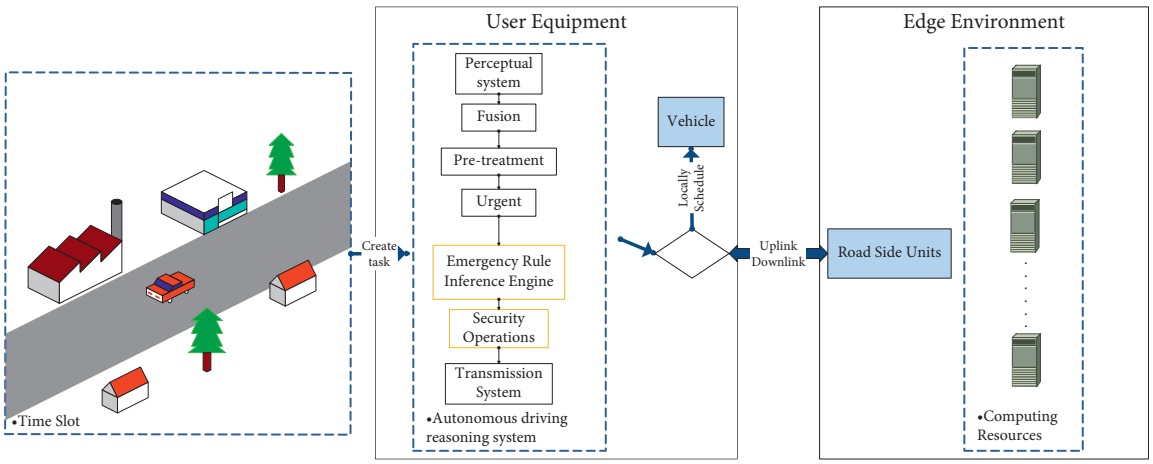


Figure 1: An example of reasoning tasks scheduling in vehicle edge computing.

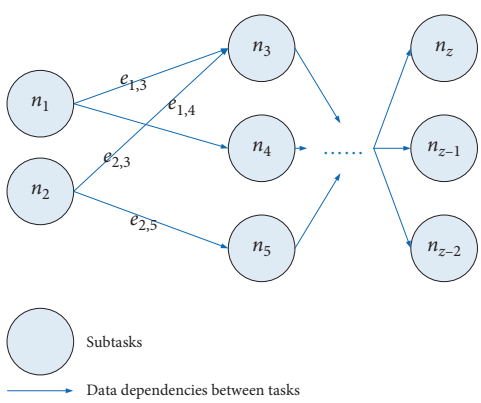


Figure 2: A reasoning task.

subtasks, and these subtasks can be reasonably scheduled to the edge nodes for processing based on a specific scheduling algorithm. The scheduling algorithm proposed in this paper can minimize the execution latency of reasoning tasks, which can be expressed as

$$\text{Minimize } t_{\text{all}}, t_{\text{all}} \leq t_{\text{worst}}. \tag{8}$$

4. Algorithm Design

In this section, we first describe the priority evaluation for subtasks in a reasoning task, which is employed to determine the order of execution for the subtasks without data dependencies. And then give an overview of our proposed scheduling algorithm. Finally, we introduce the implementation of the scheduling algorithm in detail.

4.1. Priority Evaluation for Each Subtask. It is difficult to estimate the execution time of a reasoning task that defines the execution sequence of subtasks. Fuzzy analytic hierarchy process (FAHP) [27–29] is usually employed to analyze multiobjective problems, which decomposes the problem hierarchically according to its feature and overall goal, forming a bottom-up gradient hierarchy. In this work, FAHP is commonly used to measure the subtask weight, which can determine the order of execution for the subtasks without data dependencies. Each subtask weight is modified by calculating the information entropy of objective factors (i.e., each subtask's own parameters) [30, 31]. The pseudocode of the priority evaluation for each subtask is described as follows:

$$p_{i,j} = \begin{cases} 0, s_i < s_j, \\ 0.5, s_i = s_j, \\ 1, s_i > s_j, \end{cases} \tag{9}$$

$$\mathbf{R} = \left(r_{i,j}\right)_{\alpha\cdot\alpha}, r_{i,j} = r_{i,k} - r_{j,k} + 0.5, \tag{10}$$

Input: computational complexity a_1, the amount of data a_2, the tolerable delay a_3
Output: the priority of subtask z

```
Sort subtask's factor according to equation (9) and construct matrix P
for i ⟵ 0 to maximum rows of P do
    r_i = 0
    for j ⟵ 0 to maximum columns of P do
        r_i = r_i + p_{i,j}
    end for
end for
r_i are transformed through equation (12) to obtain R
for i ⟵ 0 to maximum rows of R do
    w_i = 0
    for j ⟵ 0 to maximum columns of R do
        update w_i via equation (13)
    end for
end for
calculate the information entropy δ_i via equations (14) and (15)
obtain w'_i via (16)
z = w'_1 · a_1 + w'_2 · a_2 + w'_3 · a_3
```

Algorithm 1: Priority evaluation for each subtask.

$$r_i = \sum_{k=1}^{\alpha} r_{i,k}, \quad i = 1, 2, \ldots, \alpha, \tag{11}$$

$$r_{ij} = \frac{r_i - r_j}{2 \cdot \alpha} + 0.5, i = 1, 2, \ldots, \alpha, \tag{12}$$

$$w_i = \frac{\sum_{k=1}^{\alpha} r_{i,k}}{\sum_{i=1}^{\alpha} \sum_{j=1}^{\alpha} r_{i,j}} = \frac{2}{\alpha^2} \sum_{k=1}^{\alpha} r_{i,k}, i = 1, 2, \ldots, \alpha, \tag{13}$$

$$\gamma_{ij} = \frac{\max(a_i) - a_{i,j}}{\max(a_i) - \min(a_i)}, \tag{14}$$

$$\delta_i = -\frac{1}{\ln n} \sum_{j=1}^{\alpha} \beta_{ij} \ln \beta_{ij}, \quad 0 \le \delta_i \le 1, \ \beta_{ij} = \frac{\gamma_{ij}}{\sum_{j=1}^{\alpha} \gamma_{ij}}, \tag{15}$$

$$\begin{cases} g_i = \dfrac{1 - \delta_i}{\sum_{i=1}^{\alpha} \delta_i}, \\ w_i' = \dfrac{w_i \cdot g_i}{\sum_{i=1}^{\alpha} w_i \cdot g_i}, \end{cases} \tag{16}$$

s_i and s_j represent the relative importance of subtask factors. α and δ are the number of factors and the information entropy of them.

4.2. Scheduling Algorithm. MDP is a basic model of the RL in this paper. The scheduling algorithm can be simplified according to the MDP property, which means that the next state is only related to the current state as Figure 3. In Figure 3, each state represents a corresponding allocation strategy for real-time vehicle tasks in different edge environment and corresponds to a specific reward. Each action is calculated by the agent (neural network), and it is used to guide the current state to a better direction.

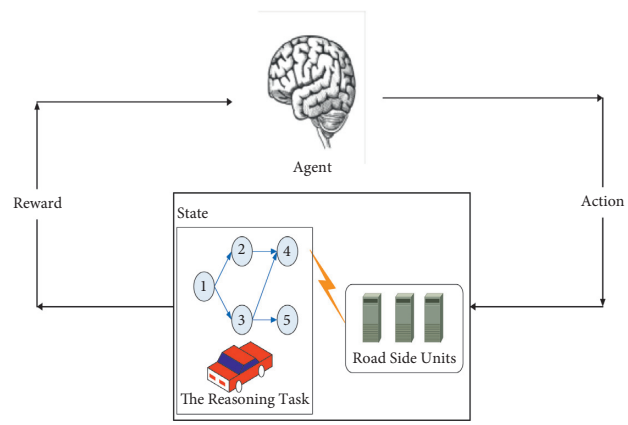


Figure 3: The MDP model of scheduling algorithm.

$$\begin{cases} Q_{k+1}(s_t, a_t, \theta_t) = Q_k(s_t, a_t, \theta_t) + \alpha_k \cdot R_k, \\ R_k = r_t + \gamma \cdot \max_{a \in a} Q(s', a', \theta_t') - Q(s_t, a_t, \theta_t), \end{cases} \tag{17}$$

The model characteristics of the discussed problem in this paper are described as follows:

(1) State space: the number of states for the feasible solutions is not constant. They can change dynamically as the change of the number of subtasks after decomposition and the changed distribution of edge nodes in various time-slots.

(2) Action space: the number of optional actions in the action space is equal to the number of subtasks. Action selection means scheduling the corresponding subtasks in current state to the specific edge nodes.

(3) Reward value: this work tries to minimize the completion latency of the reasoning task, so the reward value is set to $r_t = (1/t_{\text{all}})$.

Input: initial state, maximum number of rounds, maximum number of iterations in a single round
Output: the scheduling strategy for reasoning tasks
Initialize the experience pool of constant storage space, the action-value function $Q_\theta(s_t, a_t)$ with random weight θ and the corresponding target $-Q_\theta(s_t, a_t)$
for $i \leftarrow 0$ to Maximum number of rounds **do**
  $s_t \leftarrow$ initial state
  for $i \leftarrow 0$ to maximum number of iterations in a single round **do**
  Choose the action $a_t = \text{argmax}Q(s_t, a, \theta)$ with the largest historical reward with possibility ε, otherwise choose a random action
  Execute action a_t to get the next state s_{t+1} and use Algorithm 2 to calculate the reward r_t
  Store (s_t, s_{t+1}, r_t, a_t) in the experience pool
  $s_t \leftarrow s_{t+1}$
    Random sampling (s_j, s_{j+1}, r_j, a_j) from the experience pool
    Construct an error function according to equation (17), and back-propagation to update the parameters θ
  Update target $-Q_\theta(s_t, a_t) = Q_\theta(s_t, a_t)$ per few steps
  If s_{t+1} satisfies the termination state, the current iteration is ended
 end for
end for

Algorithm 2: Scheduling algorithm.

The scheduling strategy is based on the DQN algorithm. It can be abstracted as a function fitting problem when the discrete tangent dimension of the state and action space are high. The pseudocode of our scheduling algorithm is described in Algorithm 2.where α_k and γ represent the learning rate and discount factor, respectively. s' is the state after executing the action a_t in the k_{th} iteration.a' represents the action of the largest reward in state s', and R_k represents the accumulated reward during the iterations.

4.3. Algorithm Implementation. In various time-slots, reasoning tasks and edge environments can change dynamically. These changes are summarized as follows:

(1) The topological structure of reasoning tasks and the number of nodes in edge environments
(2) The computational complexity, the datasets between subtasks, and the tolerable delay of subtasks in various environments
(3) The transmission latency and execution latency of subtasks

The algorithm implementation will calculate the completion latency of reasoning tasks in edge environments. The pseudocode of the algorithm implementation is described in Algorithm 3. Figure 4 presents the calculation process of execution latency, which includes the following steps.

Step 1: initiate the parameters of Algorithm 3, including the subtask queue Q and the set of predecessor nodes R. Next, a reasoning task is expressed as a specific directed acyclic graph.

Step 2: Q is used to sort the subtasks by the topology of the reasoning task.

Step 3: calculate the task execution time according to the specific strategy derived from Algorithm 2.

5. Simulation Experiment and Analysis

5.1. Experimental Parameter Settings. The simulation experiments are implemented with the Python 3.7 and conducted on a 64-bit Win10 system, which is configured with Inter(R) Core (TM) i7-7700HQ CPU and 16 GB RAM. Our proposed scheduling algorithm is DQN, and Q-learning algorithm [23] and GA-PSO [32] are introduced as the comparison algorithms. Based on the effects of adjusting parameters in many experiments, the corresponding parameters of DQN and Q-learning [23] are set as $\alpha = 0.005$, $\gamma = 0.9$, and $\varepsilon = 0.9$. The corresponding parameters of GA-PSO [32] are set as $w_{\max} = 0.9$, $w_{\min} = 0.4$, $C_1^{\text{start}} = 0.9$, $C_1^{\text{end}} = 0.2$, $C_2^{\text{start}} = 0.9$, and $C_2^{\text{end}} = 0.4$. In addition, the number of rounds is set as 100 and the number of iterations is set as 1000 for DQN, Q-learning, and GA-PSO, respectively.

All the algorithms try to find the optimal scheduling result with the shortest completion latency of reasoning tasks in edge environments.

UEs have different reasoning tasks with various topologies and task number, and the topological structure of reasoning tasks is shown in Figure 5. The related parameters for the vehicle edge computing environment are set according to the IEEE 802.11p [33], and other parameters are set as Table 2.

5.2. Analysis of Results. Table 3 shows the completion latency of different reasoning tasks in various edge environments with our proposed scheduling algorithm, where m and n denote the number of edge nodes and subtasks in each experiment. Note that $n = 6$ corresponds to the "Topology I," $n = 9$ corresponds to the "Topology II," and $n = 12$ corresponds to the 'Topology III' in Figure 5. Each grid in Table 3 corresponds to an experiment with different reasoning tasks with specific topologies and different edge nodes in edge environments. In addition, the execution order of subtasks is

Input: m_i, z_i, G_i $A_{z_i \times 3}$, $B_{m_i \times z_i}$, $H_{m_i \times z_i}$
Output: t^i_{all}
Initialization: set the array I, the subtask queue $\boldsymbol{Q}$ and the set of predecessor nodes $\boldsymbol{R}$ to $\varnothing$
Use the constraint relationship G_i to set the array $I(i)$
Enqueue the ith subtask with $I(i) = 0$ to $\boldsymbol{Q}$ and set the number of traversed subtasks $u = 0$, the number of subtasks in the current layer k to the current queue size
while $\mathbf{Q}! = \varnothing$ **do**
  if $u = k$ **then**
    $u = 0, k = \text{size}(\mathbf{Q})$.
end if
  The subtask is dequeued, and the task is expressed as v, $u+ = 1$
  for $\mathbf{i} \longleftarrow 0$ to $z_{\mathbf{i}}$ **do**
    if there exists a directed edge of **v** to ***i*** **then**
  Add the **v**th subtask and its predecessor node set **R(v)** to ***R*(i)** $I(i)- = 1$
      if $I(i) = 0$ **then**
  enqueue the ith subtask to $\boldsymbol{Q}$
  end if
    end if
end for
end while
According to $B_{m_i \times z_i}$, the subtasks are assigned to edge nodes.
Initialization: set the subtask completion list O to $\varnothing$, set the remaining execution latency of subtasks Y by $C_{m_i \times z_i}$, and set the current running time $h = 0$
while $O < z_i$ **do**
  Determine the subtask to be assigned to each edge node, which satisfies the direct predecessor set is subset of O
  Find the minimum execution latency w from the currently executed subtasks in parallel
  $Y(i) - = w$, when $Y(i) = 0$, add the ith subtask to O and set $h + = w$
end while
return h

Algorithm 3: Algorithm implementation.

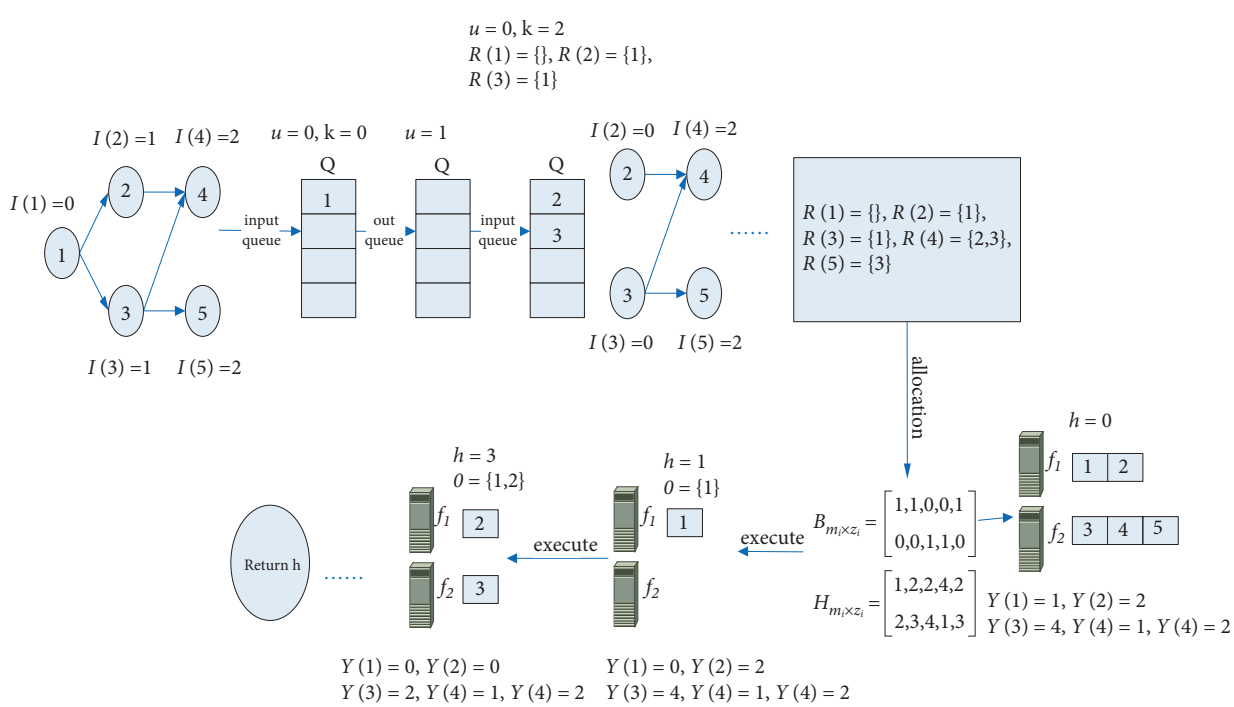


Figure 4: The calculation process of execution latency.

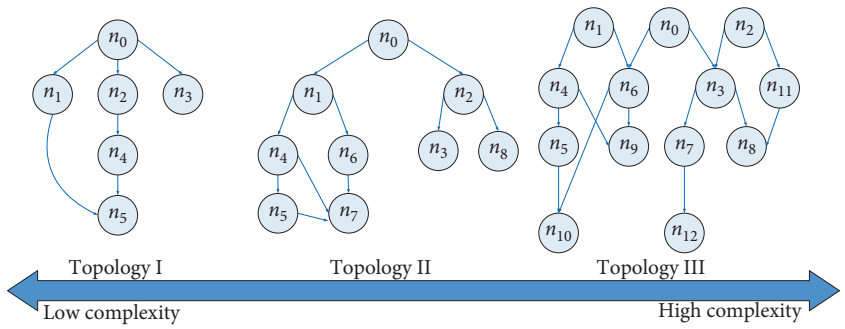


Figure 5: Topological structure of reasoning tasks.

Table 2: Constraint parameters of simulation experiment.

Parameter	Value
Task data volume	3–5Mbits
Task complexity	1200–3000 megacycles
Vehicle computing power	1-2GPOS
Edge node transmission rate	100–120Mbits/s
Edge node computing power	10–15GPOS
Number of scene edge nodes	1–4

Table 3: Completion latency (ms) of different reasoning tasks in various edge environments.

	Traditional rule			Priority rule		
	$n = 6$	$n = 9$	$n = 13$	$n = 6$	$n = 9$	$n = 13$
$m = 1$	6276	6891	7687	6276	6891	7687
$m = 2$	1025	1329	1849	1025	1329	1849
$m = 3$	840	862	1119	668	845	1078
$m = 4$	728	793	949	668	751	800

based on two rules: traditional rule and priority rule. Traditional rule executes the subtasks according to their corresponding topology depths [23] and priority rule executes the ones according to the priority evaluation for each subtask discussed in Section 4.1.

From Table 3, we find that the completion latency of reasoning tasks reduces as the number of edge nodes increases. Under the same circumstances, the priority rule for subtask execution can effectively reduce the completion latency of reasoning tasks compared to the traditional rule. As the topology complexity of reasoning tasks increases, this gap will become larger. This is because that the maximum number of parallel subtasks in the same time is limited by the number of edge nodes, and the priority rule can increase the upper limit number of the parallel subtasks.

Figure 6 shows the average completion latency of different scheduling algorithms (i.e., GA-PSO, Q-learning, and DQN) with different seasoning tasks in various edge environments, where m denotes the number of edge nodes in each experiment. In each subgraph, we record the completion latency of reasoning tasks with different topologies for 100 rounds and display the average completion latency of reasoning tasks for every 10 rounds. From Figure 6(a), we find that GA-PSO is difficult to converge, although it can get a feasible solution with shorter completion latency of reasoning tasks. In contrast, DQN can not only get a feasible solution with shorter completion latency, but also convergence well. From Figure 6(b) and Figure 6(c), the performance of Q-learning is similar to that of DQN when the numbers of subtasks and edge nodes are both small. However, the convergence performance of Q-learning will decrease as the topology complexity of reasoning tasks increases. The main reason for the different scheduling results with various algorithms is that the increase in the number of subtasks has brought about the multiplication of the number of solutions in searching space. For GA-PSO, finding the optimal solution mainly relies on randomness and fitness function. Therefore, when the number of feasible solutions in searching space is huge, GA-PSO is easy to fall into the local optimal solution. Q-learning is difficult to build the Q list and converge also due to the huge number of feasible solutions in searching space. However, DQN converts the Q list to the Q value function by neural network, which can solve the problem with a huge number of states (i.e., feasible solutions) and make it easier to converge.

Table 4 shows the average runtime (s) of different algorithms with different seasoning tasks in various edge environments. Each grid in Table 4 is the average of the runtime of 100 rounds for different algorithm. From Table 4, we find that the runtime of GA-PSO is relatively stable with different seasoning tasks in various edge environments. This is because that the runtime of GA-PSO mainly depends on

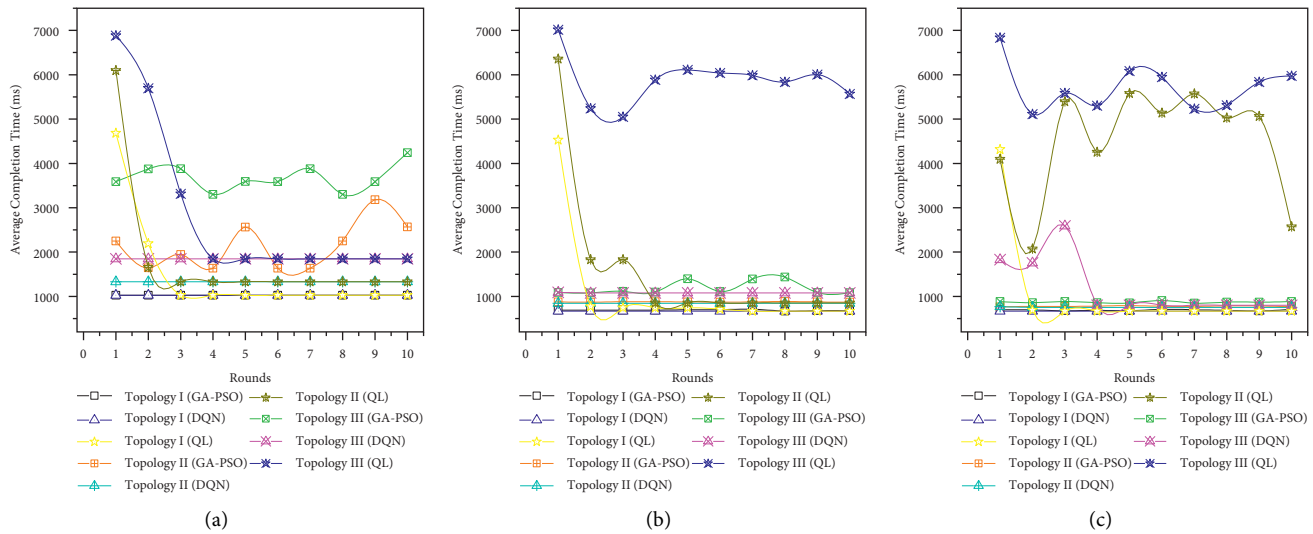


Figure 6: Average completion latency of different scheduling algorithms with different seasoning tasks in various edge environments. (a) $m = 2$. (b) $m = 3$. (c) $m = 4$.

Table 4: Average runtime (s) of different algorithms with different seasoning tasks in various edge environments

Number of edge nodes	Algorithm	Topology I ($n = 6$)	Topology II ($n = 9$)	Topology II ($n = 13$)
$m = 2$	GA-PSO	0.53	0.76	2.02
	DQN	**0.02**	**0.05**	**0.08**
	Q-learning	**0.02**	0.06	0.44
$m = 3$	GA-PSO	0.58	0.72	1.19
	DQN	**0.04**	**0.12**	**0.28**
	Q-learning	0.10	0.58	11.36
$m = 4$	GA-PSO	0.61	0.83	0.93
	DQN	**0.07**	**0.13**	**0.92**
	Q-learning	0.19	7.77	11.29

the number of particles used in the update process, which is relatively stable even if the edge environments change during the scheduling process. The average runtime of DQN and Q-learning is better than that of GA-PSO, and DQN performs best with different seasoning tasks in various edge environments. This is because the runtime of RL algorithms will decrease as the number of feasible solutions learned increases. In addition, the architecture of the neural network used in DQN is more suitable for reasoning tasks scheduling in vehicle edge computing, compared with the Q list used in Q-learning.

6. Conclusions

This paper proposes a scheduling strategy based on DQN for reasoning tasks in vehicle edge computing, which aims to reduce the completion latency of reasoning tasks. The extensive simulation experiments show that the proposed strategy can achieve the superior performance compared to other classic methods. Our strategy and other classic methods all perform well when the structure of reasoning tasks is simple, while GA-PSO has poor convergence. Specially, our strategy has better performance and convergence than any other classic methods when the structure of reasoning tasks is complex.

In the future, we will improve the scheduling algorithm through optimizing the training efficiency of the neural network to fit the wireless channel fluctuations and radio interference in vehicle edge computing. In addition, we will further consider a multivehicle collaborative scheduling strategy to alleviate uneven resources allocation for multivehicle tasks in edge environments.

Data Availability

The data used to support the findings of this study are included within the article.

Disclosure

This work was presented in part at the 2019 IEEE Intl Conf on Parallel and Distributed Processing with Applications (ISPA) with the title "A Time-Driven Workflow Scheduling Strategy for Reasoning Tasks of Autonomous Driving in Edge Environment."

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was partly supported by the Intelligent Computing and Application Research Team of Concord University College, Fujian Normal University under Grant no. 2020TD001 Natural Science Foundation of China under Grant no. 62072108; Project on the Integration of Industry and Education of Fujian Province under Grant no. 2021H6026 Natural Science Foundation of Fujian Province under Grant nos. 2019J01286 and 2019J01427; and Young and Middle-Aged Teacher Education Foundation of Fujian Province under Grant no. JT180098.

References

[1] J. Cheng, G. Yuan, M. Zhou et al., "Accessibility analysis and modeling for IoV in an urban scene," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 4, pp. 4246–4256, 2020.

[2] J. Wang, Y. Qiao, and H. Hong, "Graph-based auto-driving reasoning task scheduling," *Journal of Computer Research and Development*, vol. 8, pp. 68–77, 2017.

[3] Z. Gao, T. Sun, and L. He, "Causal reasoning decision-making for vehicle longitudinal automatic driving," *Journal of Jilin University(Engineering and Technology Edition)*, vol. 49, no. 5, pp. 1392–1404, 2019.

[4] R. Xie, X. Lian, and Q. Jia, "Survey on computation offloading in mobile edge computing," *Journal on Communications*, vol. 39, no. 11, pp. 138–155, 2018.

[5] J. Xie, Y. Jia, Z. Chen, Z. Nan, and L. Liang, "Efficient task completion for parallel offloading in vehicular fog computing," *China Communications*, vol. 16, no. 11, pp. 42–55, 2019.

[6] Z. Xiao, F. Li, H. Jiang et al., "A joint information and energy cooperation framework for CR-enabled macro-femto heterogeneous networks," *IEEE Internet of Things Journal*, vol. 7, no. 4, pp. 2828–2839, 2020.

[7] Z. Xiao, X. Dai, H. Jiang et al., "Vehicular task offloading via heat-aware MEC cooperation using game-theoretic method," *IEEE Internet of Things Journal*, vol. 7, no. 3, pp. 2038–2052, 2020.

[8] Y. Mao, C. You, J. Zhang, K. Huang, and K. B. Letaief, "A survey on mobile edge computing: the communication perspective," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 4, pp. 2322–2358, 2017.

[9] P. Mach and Z. Becvar, "Mobile edge computing: a survey on architecture and computation offloading," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 3, pp. 1628–1656, 2017.

[10] J. Fang, Z. Zhang, and X. Zhang, "Cloud workflow scheduling algorithm based ON trade-off fitness," *Computer Applications and Software*, vol. 36, no. 5, pp. 255–261, 2019.

[11] A. Ghorbannia Delavar and Y. Aryan, "HSGA: a hybrid heuristic algorithm for workflow scheduling in cloud systems," *Cluster Computing*, vol. 17, no. 1, pp. 129–137, 2014.

[12] Y. Xie, Y. Zhu, Y. Wang et al., "A novel directional and non-local-convergent particle swarm optimization based workflow scheduling in cloud-edge environment," *Future Generation Computer Systems*, vol. 97, pp. 361–378, 2019.

[13] H. Wu, W. J. Knottenbelt, and K. Wolter, "An efficient application partitioning algorithm in mobile environments," *IEEE Transactions on Parallel and Distributed Systems*, vol. 30, no. 7, pp. 1464–1480, 2019.

[14] T. Q. Thinh, J. Tang, and Q. D. La, "Offloading in mobile edge computing: task allocation and computational frequency scaling," *IEEE Transactions on Communications*, vol. 65, no. 8, pp. 3571–3584, 2017.

[15] X. Lin, Y. Wang, Q. Xie, and M. Pedram, "Task scheduling with dynamic voltage and frequency scaling for energy minimization in the mobile cloud computing environment," *IEEE Transactions on Services Computing*, vol. 8, no. 2, pp. 175–186, 2015.

[16] X. Chen, H. Zhang, C. Wu, S. Mao, Y. Ji, and M. Bennis, "Optimized computation offloading performance in virtual edge computing systems via deep reinforcement learning," *IEEE Internet of Things Journal*, vol. 6, no. 3, pp. 4005–4018, 2019.

[17] X. Xiong, K. Zheng, and L. Lei, "Resource allocation based on deep reinforcement learning in IoT edge computing," *IEEE Journal on Selected Areas in Communications*, vol. 99, p. 1, 2020.

[18] J. Wang, J. Hu, G. Min, W. Zhan, Q. Ni, and N. Georgalas, "Computation offloading in multi-access edge computing using a deep sequential model based on reinforcement learning," *IEEE Communications Magazine*, vol. 57, no. 5, pp. 64–69, 2019.

[19] C. Wang, F. R. Yu, C. Liang, Q. Chen, and L. Tang, "Joint computation offloading and interference management in wireless cellular networks with mobile edge computing," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 8, pp. 7432–7445, 2017.

[20] T. Zhao, T. Zhang, and Y. Chen, "Task distribution offloading algorithm of vehicle edge network based on DQN," *Journal on Communications*, vol. 41, no. 10, pp. 172–178, 2020, in chi.

[21] G. Rjoub, J. Bentahar, and O. Abdel Wahab, "Deep and reinforcement learning for automated task scheduling in large-scale cloud computing systems," *Concurrency and Computation: Practice and Experience*, vol. 33, no. 23, 2020.

[22] Q. Qi, L. Zhang, J. Wang et al., "Scalable parallel task scheduling for autonomous driving using multi-task deep reinforcement learning," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 11, pp. 13861–13874, 2020.

[23] K. Lin, B. Lin, and X. Chen, "A time-driven workflow scheduling strategy for reasoning tasks of autonomous driving in edge environment," in *Proceedings of the 2019 IEEE Intl Conf on Parallel & Distributed Processing with Applications*, ISPA, Shenzhen, China, December 2019.

[24] L. Huang, S. Bi, and Y.-J. A. Zhang, "Deep reinforcement learning for online computation offloading in wireless powered mobile-edge computing networks," *IEEE Transactions on Mobile Computing*, vol. 19, no. 11, pp. 2581–2593, 2020.

[25] M. Chen, U. Challita, and W. Saad, "Artificial neural networks-based machine learning for wireless networks: a tutorial," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 4, pp. 3039–3071, 2017.

[26] G. Rjoub, J. Bentahar, and O. A. Wahab, "BigTrustScheduling: trust-aware big data task scheduling approach in cloud computing environments," *Future Generation Computer Systems*, vol. 110, pp. 1079–1097, 2020.

[27] H. Ghunaim and J. Dichter, "Applying the FAHP to improve the performance evaluation reliability of software defect classifiers," *IEEE Access*, vol. 7, pp. 62794–62804, 2019.

[28] Z. Geng, Z. Wang, C. Peng, and Y. Han, "A new fuzzy process capability estimation method based on kernel function and FAHP," *IEEE Transactions on Engineering Management*, vol. 63, no. 2, pp. 177–188, 2016.

[29] Y. Lu, B. Lin, and M. Liu, "The performance evaluation of university scientific research Project management based on

the FAHP," *Journal of Digital Information Management*, vol. 12, no. 1, pp. 18–25, 2014.

[30] F. Xiao, "EFMCDM: evidential fuzzy multicriteria decision making based on belief entropy," *IEEE Transactions on Fuzzy Systems*, vol. 28, no. 7, pp. 1477–1491, 2020.

[31] T. Li and L. She, "Retail location decision-making based on the combination of AHP and entropy weight," in *Proceedings of the 2010 Third International Joint Conference on Computational Science and Optimization*, Huangshan, China, May 2010.

[32] B. Lin, F. Zhu, J. Zhang et al., "A time-driven data placement strategy for a scientific workflow combining edge computing and cloud computing," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 7, pp. 4254–4265, 2019.

[33] Y. Liu, L. Liu, and L. Zheng, "Study on the downlink performance of roadside unit in vehicular ad-hoc networks," *Journal of Software*, vol. 26, no. 7, pp. 1700–1710, 2015.

Hindawi
Wireless Communications and Mobile Computing
Volume 2023, Article ID 9761846, 1 page
https://doi.org/10.1155/2023/9761846

Retraction

Retracted: Finite-Time Attitude Cooperative Control of Multiple Unmanned Aerial Vehicles via Fast Nonsingular Terminal Sliding Mode Control

Wireless Communications and Mobile Computing

Received 10 October 2023; Accepted 10 October 2023; Published 11 October 2023

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. Han, X. Tuo, Y. Tang, and P. He, "Finite-Time Attitude Cooperative Control of Multiple Unmanned Aerial Vehicles via Fast Nonsingular Terminal Sliding Mode Control," *Wireless Communications and Mobile Computing*, vol. 2022, Article ID 4324626, 11 pages, 2022.

Hindawi
Wireless Communications and Mobile Computing
Volume 2022, Article ID 4324626, 11 pages
https://doi.org/10.1155/2022/4324626

Research Article

Finite-Time Attitude Cooperative Control of Multiple Unmanned Aerial Vehicles via Fast Nonsingular Terminal Sliding Mode Control

Qiang Han,[1,2] **Xianguo Tuo**,[1,2] **Yuxin Tang**,[2] **and Ping He**[3]

[1]*Robot Technology Used for Special Environment Key Laboratory of Sichuan Province, Southwest University of Science and Technology, Mianyang 621010, Sichuan, China*
[2]*Artificial Intelligence Key Laboratory of Sichuan Province, Sichuan University of Science & Engineering, Zigong 643000, Sichuan, China*
[3]*College of Engineering, Huazhong Agricultural University, Wuhan 430070, Hubei, China*

Correspondence should be addressed to Ping He; pinghecn@qq.com

Received 20 October 2021; Revised 20 December 2021; Accepted 4 January 2022; Published 23 February 2022

Academic Editor: Chi-Hua Chen

This paper discusses the attitude cooperative control of multiple unmanned aerial vehicle systems (MUAVs) with unknown dynamics and external disturbances. Distributed fast nonsingular terminal sliding mode control (FNTSMC) is used with quaternion-description dynamical systems. The dynamics and external disturbances are changed into lumped disturbances by formula transformation. A robust nonlinear disturbance observer is proposed to estimate the lumped disturbances in finite time. Then, combining the fast terminal sliding mode control, distributed FNTSMC controllers are designed under the directed topology. Based on the Lyapunov stability theory and graph theory, convergence stability of the nonlinear systems is strictly proved, and the tracking errors between the leader and the followers approach to a small residual set. Finally, the simulation example is presented to illustrate the effectiveness and advantage of the proposed controllers.

1. Introduction

In recent years, formation flight control (FFC) for multiple unmanned aerial vehicle systems has attracted more and more attention owing to its various applications in both military and civilian areas [1–4]. Attitude cooperative is required to complete the formation flight mission. However, attitude dynamics of MUAVs are highly nonlinear, such as inertia uncertainties and external disturbances. How to develop a robust attitude control law is challenging work. Sliding mode control (SMC) has been widely used in attitude control for unmanned aerial vehicles (UAVs) [5]. Because of its robustness and insensitivity to disturbances and uncertainties, $\mathscr{H}^{\infty}$ controller was proposed for unmanned helicopters with uncertain models [6]. Although most of the existing robust control methods provide sufficient stability and reliability, they can only guarantee the asymptotic convergence as time goes into infinity.

Finite-time control (FTC) is more desirable for attitude synchronization of spacecraft or some other flight vehicles due to its rapid maneuverability highly demanded in some real-time actions [7]. Besides high precision performance and fast convergence rate, finite-time control can also provide better disturbance rejection properties and better robustness against uncertainties, and it has been gradually implemented in spacecraft attitude stability and tracking control [7, 8]. Terminal sliding mode control (TSMC) has been proposed, which can significantly improve the transient performance substantially in finite time [9]. Adaptive sliding mode control method was proposed for quadrotors to handle the parametric uncertainties [10]. However, disadvantages also exist in TSMC. One of the disadvantages is the singularity problem, and the other is that TSMC has slower convergence to the equilibrium point than the traditional SMC when the system state is far away from the equilibrium point. Hence, a nonsingular TSMC (NTSMC)

scheme was proposed, which has been used in multiple agent systems while maintaining the advantages of the conventional TSMC [11], the NTSMC can eliminate the singularity problem, and fast TSMC (FTSMC) was proposed to provide fast convergence rate for the slow convergence rate problem [12]. To overcome the two disadvantages of the SMC, fast nonsingular TSMC (FNTSMC) has been used for attitude synchronization of spacecrafts [13]. Recently, FNTSMC combined with event-triggered scheme was designed for distributed attitude coordination with hyperbolic tangent function for multiple agent systems [14]. However, inertia uncertainties were not considered, which is not practical in dynamical environments in the mentioned work [13]. Based on the robust decentralized attitude control, finite-time attitude synchronization mechanism was proposed for flying vehicles; however, the precise knowledge of the inertial matrix should be known, which is not practical because of its time-varying [15–18]. Neural network (NN) and fuzzy logic systems (FLSs) can approximate any smooth functions over a compact set to arbitrary accuracy [19]. Based on Chebyshev NN and TSMC, the attitude tracking problem of spacecrafts is studied, and a switching is designed between the adaptive NN controller and robust controller [20], finite-time formation control for multiple helicopters was proposed [21], and the disturbances were estimated by the NN technique without considering the uncertain parameters. However, the mentioned works [20, 21] estimated the disturbances by NN and FLS only effective on some compact sets.

In this paper, the FNTSMC method is used for the attitude cooperative control for MUAVs, considering the uncertain inertia part and external disturbances. It is significantly noted how to deal with lumped disturbances in cooperative attitude control. However, adaptive control or SMC scheme cannot reconstruct information about the lumped disturbances. Nonlinear disturbances observer (NDO) can eliminate the problem and reconstruct the lumped disturbances without an additional sensor [22]. The NDO technique has attracted much attention and has been widely used in unmanned aerial vehicle (UAV) attitude control and mobile robots [21, 23]. A composite disturbance observer-based control (DOC) and TSMC were proposed for a class of spacecraft formation [24]. A composite control based on NDO was proposed to compensate the disturbances through feedforward [25]. The NDO technique combined with an event-triggered scheme has been used in multiple agent systems [26]. Recently, the observer schemes for different objects have been investigated [21, 27–32]. A nonlinear internal model-based observer is proposed to estimate the nonlinear signals with parametric uncertainty. The NDO-based controller is proposed to compensate the disturbances for the MUAVs or single quadrotor [29–31]. A distributed antidisturbance method combining with Nussbaum function, adaptive NNs, and disturbances observer for attitude tracking of MUAVs was proposed [31]; however, the abovementioned works used the NDO not considering the finite-time property. In order to obtain high precision performance and fast convergence rate, the NDO scheme was proposed, which could obtain the lumped disturbances information in finite time [32]. Decentralized FNTSMC combined with terminal sliding mode disturbance observer was proposed for quaternion-description attitude synchronization of multiple rigid spacecrafts [33]. However, these works have their own limitations. The authors only consider the attitude cooperative problem under the undirected communication topology [33]. However, cooperative control for attitude cooperative of MUAVs under directed communication topology is more challenging.

Motivated by the aforementioned analysis, this paper aims to consider a more interesting cooperative attitude control problem for MUAVs with inertia uncertainties and external disturbances. The contributions of this paper are summarized as follows:

(i) Compared with the attitude cooperative under undirected topology [33], our analysis is based on the directed topology, and it is more challenging and practical

(ii) A robust NDO is proposed to estimate the lumped disturbances, which can obtain the information of the lumped disturbances in finite time

(iii) Based on the directed communication topology, combining the NDO and FNTSMC, distributed cooperative attitude controller for MUAVs is proposed, which can guarantee that the tracking errors converge to the regions containing the origin in finite time

The rest of this paper is organized as follows. In Section 2, preliminaries of graph theory, quaternion-based attitude dynamics of MUAVs, and some useful lemmas are given. Section 3 gives main results. The effectiveness of the controller is illustrated by a numerical simulation example in Section 4. Finally, the conclusion is given in Section 5.

Notation. For convenience, the following notations are used throughout this paper. All matrices are assumed to have compatible dimensions. R^n and $R^{n\times m}$ denote the n-dimensional Euclidean space and the set of all $n \times m$ real matrices, respectively. I_n is a $n \times n$ dimensional identity matrix. A^T denotes the transpose of vector or matrix A, $\mathrm{tr}(\cdot)$ is the trace of a given matrix, $\otimes$ stands for Kronecker product, and $|\cdot|$ refers to the absolute value. $\|\cdot\|$ refers to the Euclidean vector norm or the induced matrix 2-norm, and $\|B\| = (\mathrm{tr}(B^T B))^{1/2}$. The superscript '$T$' denotes matrix transposition. I_n denotes the $n \times n$ identity matrix, and $I_n \in R^n$ $(0_n \in R^n)$ is a column vector with each entry being $I(0)$. $\lambda_{\max}(M)$ be the largest eigenvalue of matrix M. $\mathrm{diag}(\cdot)$ represents the diagonal matrix, and $\mathrm{sgn}(\cdot)$ denotes the sign function. $\mathrm{sig}^r(\cdot) = \mathrm{sgn}(\cdot)|\cdot|^r$.

2. Preliminaries and Problem Formulation

In this section, basic algebraic graph theory and notations are introduced. Then, the system description and attitude tracking problems of MUAVs are described.

2.1. Communication Algebraic Graph Theory. Graph theory is a useful tool to solve the cooperative problems among the MUAVs. In this paper, we consider the directed communication topology. A weighted directed graph is denoted by

$$\mathcal{G} = (v, \xi, \mathcal{G}), \tag{1}$$

where $v = \{1, 2, \ldots, n\}$ stands for the single UAV and $\xi = v \times v$ denotes the edges set. If there exists an edge between two nodes, the two nodes are called adjacent. $A = (a_{ij})_{N\times N} \in R^{n\times n}$ is weighted matrix which is symmetric, the symmetric adjacency matrix A is defined as $a_{ii} = 0$, and $a_{ij} = a_{ji} \geq 0$ means that agent v_i and v_j are connected by an edge. The Laplacian matrix corresponding to the weighted undirected graph is defined as $L = [l_{ij}]_{n\times n}$, which has the property $L = D - A$, where $D = \mathrm{diag}\{d_1, d_2, \ldots, d_n\}$ is the out-degree of the i th node $d_i = \sum_{j=1}^{n} a_{ij}$.

In this paper, we consider the MUAVs composed of one leader and N followers. Without loss of generality, let agent 0 be the leader, and the N followers are labeled as $i\,(i = 1, 2, \ldots, n)$. The communication topology of MUAVs between the followers is characterized by a digraph $\mathcal{G}$ with the Laplacian matrix L. The communication between the followers and the leader is represented by a diagonal matrix $B = \mathrm{diag}\{b_1, b_2, \ldots, b_n\}$. If the follower agent i has access to the information of leader 0, then $b_i > 0$; otherwise, $b_i = 0$. Besides, the communication topology between followers and the leader is characterized by $\mathcal{G}$. Obviously, the graph $\overline{\mathcal{G}}$ is the subgraph of $\overline{\mathcal{G}}$.

Assumption 1. The augmented digraph $\overline{\mathcal{G}}$ contains a directed spanning tree.

Lemma 1 (see [34]). *If the digraph $\overline{\mathcal{G}}$ has a directed spanning tree, then the matrix $L + B$ is invertible.*

In this paper, we consider that the directed communication topology is firmly connected. If there at least one UAV can receive the information from the leader, which implies that $b_i \neq 0$, then $L + B$ is dominant and irreducible diagonal matrix. By Lemma 1, we can obtain that $L + B$ is invertible, and the matrix $(L + B) \otimes I_3$ is invertible.

2.2. Dynamic Model of MUAVs. Compared with the rotation matrix for attitude description, the quaternion-based attitude description has more advantages because of providing simplified design and analysis of control systems. Hence, the attitude dynamics equations of the i th UAV in terms of quaternion-based are given as follows:

$$\begin{cases} \dot{Q}_i = \dfrac{1}{2} T(Q_i)\omega_i, \\ \\ J_i\dot{\omega}_i = u_i - S(\omega_i)J_i\omega_i + d_i, \end{cases} \tag{2}$$

where $Q_i = \left[\, q_i \;\; q_{i0} \,\right]^T$ represents the attitude of the i th UAV, $q_i \in \mathbb{R}^3, q_{i0} \in \mathbb{R}, Q_i \in \mathbb{R}^4, |Q_i| = 1$, and $\omega_i \in \mathbb{R}^3$ is the angular velocity. J_i is the symmetric positive definite constant moment of the inertial matrix of the i th UAV; u_i represents the control torque of the rotor thrust for each UAV; $T(Q_i)$ is given by $T(Q_i) = \begin{pmatrix} q_{i0}I_3 + S(q_i) \\ -q_i^T \end{pmatrix}$, and $S(\omega_i) = \begin{pmatrix} 0 & -\omega_{i3} & \omega_{i2} \\ \omega_{i3} & 0 & -\omega_{i1} \\ -\omega_{i2} & \omega_{i1} & 0 \end{pmatrix}$; $R(Q_i)$ denotes the rotation matrix associated to $Q_i = \left[\, q_i \;\; q_{i0} \,\right]^T$ that brings the inertial frame into the body frame, and $R(Q_i)$ is given as $R(Q_i) = (2q_{i0}^2 - 1)I_3 + 2q_iq_i^T - 2q_{i0}S(q_i)$; d_i denotes the external disturbances.

The multiplication between two unit quaternions is given by

$$Q_1 \odot Q_2 = \begin{pmatrix} q_{10}q_2 + q_{20}q_1 + S(q_1)q_2 \\ q_{10}q_{20} - q_1^T q_2 \end{pmatrix}, \tag{3}$$

where $Q_1 = \left[\, q_1 \;\; q_{10} \,\right]^T$ and $Q_2 = \left[\, q_2 \;\; q_{20} \,\right]^T$.

In this paper, the control objective for the MUAVs is that the followers can track the desired attitudes of the leader. Based on the mentioned relevant knowledge of multiplication, the tracking errors of the i th UAV can be formulated as

$$\begin{cases} \tilde{Q}_{i\,d} = Q_i^{-1} \odot Q_d, \\ \tilde{\omega}_i = \omega_i - R\left(\tilde{Q}_{i\,d}\right)\omega_d, \end{cases} \tag{4}$$

where $Q_d \triangleq [q_d, \eta_d]^T$ and ω_d are the desired attitude and desired angular velocity. $\tilde{Q}_{i\,d}$ and $\tilde{\omega}_i$ are the attitude tracking error and angular velocity error, respectively.

Based on equations (2)–(4), the attitude error systems can be obtained:

$$\begin{aligned} J_i\dot{\tilde{\omega}}_i = &-S(\omega_i) \cdot J_i\omega_i + u_i \\ &+ J_i \cdot \left[S(\tilde{\omega}_i) \cdot R\left(\tilde{Q}_{i\,d}\right) \cdot \omega_d - R\left(\tilde{Q}_{id}\right) \cdot \dot{\omega}_d\right] + d_i, \end{aligned} \tag{5}$$

$$\dot{\tilde{Q}}_{id} = \frac{1}{2} T\left(\tilde{Q}_{i\,d}\right) \cdot \tilde{\omega}_i. \tag{6}$$

Assumption 2. The desired angular velocity ω_d and its derivatives are bounded and satisfied. The tracking error $\tilde{\omega}_i$ and its derivatives are bounded.

Assumption 3. The inertial matrix J^* is known and nonsingular. ΔJ denotes the uncertainties satisfying $\|\Delta J\| \leq J_\theta$ with J_θ as a positive constant.

Lemma 2 (see [35]). *For the system given by $\dot{x} = f(x), f(0) = 0, x \in \mathbb{R}^n$. There exists a function $V(x) \in \mathbb{C}^1$ defined on a neighborhood of the origin such that $V(x)$ is positive definite; $\dot{V}(x) + c_1V^\eta(x) + c_2V(x) \leq 0$, where $\eta \in (0, 1)$, $c_1, c_2 > 0$.*

Then, the origin is locally finite-time stable, and the finite time T depending on the initial state $x(0)$ is

$$T(x(0)) \leq \frac{\ln\left[\left(c_2V^{1-\eta}(x(0)) + c_1\right)/c_1\right]}{c_2(1-\eta)}. \tag{7}$$

2.3. Control Objective. In the presence of inertia uncertainties and external disturbances, we aim at designing FNTSMC such that the closed-loop systems (5) and (6) can reach the sliding mode surface (SMS) in finite time. Then, $\tilde{\omega}_i$ and $\tilde{q}_i$ can converge to a residual set in finite time, respectively.

TSMC can effectively achieve high performance of finite-time stability and high robustness in the presence of model dynamics and environmental disturbances. However, it has some drawbacks, such as chattering and singularity. The modified finite-time nonsingular sliding mode surface for the MUAVs is designed to avoid the chattering and singularity problems, and the sliding mode surface is as follows:

$$\begin{aligned} s_i = {} & b_i \cdot J_i^* \left(\tilde{\omega}_i + k_1 \cdot \tilde{q}_i + k_2 \cdot h_i \cdot (\tilde{q}_i)\right) \\ & + \sum_{j=1, j\neq i}^{n} a_{ij} \cdot \Big[\left(J_i^* \tilde{\omega}_i - J_j^* \tilde{\omega}_j\right) \\ & + \left(k_1 \cdot J_i^* \cdot \tilde{q}_i - k_1 \cdot J_j^* \cdot \tilde{q}_j\right) \\ & + \left(k_2 \cdot J_i^* \cdot h_i \cdot (\tilde{q}_i) - k_2 \cdot J_j^* \cdot h_j \cdot \left(\tilde{q}_j\right)\right)\Big], \end{aligned} \tag{8}$$

where $h_i(\tilde{q}_i) = [h_{i1}(\tilde{q}_{i1}) h_{i2}(\tilde{q}_{i2}) h_{i3}(\tilde{q}_{i3})]^T$ is a function and k_1 and k_2 are positive constants.

$$h_{ij}\left(\tilde{q}_{ij}\right) = \begin{cases} \left|\tilde{q}_{ij}\right|^r \cdot \operatorname{sgn}\left(\tilde{q}_{ij}\right), & \text{if } \bar{s}_{ij} = 0 \text{ or } \bar{s}_{ij} \neq 0, \left|\tilde{q}_{ij}\right| > g, \\ \delta_1 \cdot \tilde{q}_{ij} + \delta_2 \cdot \left|\tilde{q}_{ij}\right|^2 \cdot \operatorname{sgn}\left(\tilde{q}_{ij}\right), & \text{if } \bar{s}_{ij} \neq 0, \left|\tilde{q}_{ij}\right| \leq g, \end{cases} \tag{9}$$

where $i = 1, 2, \ldots, n$, $j = 1, 2, 3$, $\bar{s}_i = \bar{s}_{ij, i=1,2,3} = [\bar{s}_{i1}, \bar{s}_{i2}, \bar{s}_{i3}]^T$, $\bar{s}_i = \tilde{\omega}_i + \delta_1 \cdot \tilde{q}_i + \delta_2 \cdot |\tilde{q}_i|^r \cdot \operatorname{sgn}(\tilde{q}_i)$, $\delta_1 = (2-r) \cdot g^{r-1}$, $\delta_2 = (r-1) \cdot g^{r-2}$, $r = r_1/r_2$. r_1 and r_2 are positive odd integers; $r \in (0, 1)$, g is a small positive constant. Based on the knowledge of graph theory, the parameter a_{ij} is the weighted value, which is the control signal of the attitude cooperative between the i th and j th UAV to keep the formation behavior. b_i is the weighted value denoting the information change between the i th UAV and the leader for tracking the leader's attitude.

Equations (8) can be rewritten by the Kronecker product.

$$s = \left[(L + B) \otimes I_3\right] \cdot J^* \cdot \left[s^*\right], \tag{10}$$

where $s^* = \tilde{\omega} + k_1 \tilde{q} + k_2 h(\tilde{q})$, $s = [s_1, s_2, \ldots, s_n]^T$, $s^* = [s_1^*, s_2^*, \cdots, s_n^*]^T$, $\tilde{\omega} = [\tilde{\omega}_1, \tilde{\omega}_2, \ldots, \tilde{\omega}_n]^T$, $J^* = \operatorname{diag}\{J_1^*, J_2^*, \ldots, J_n^*\}$, $\tilde{q} = [\tilde{q}_1, \tilde{q}_2, \ldots, \tilde{q}_n]^T$, and $h(\tilde{q}) = [h(\tilde{q}_1), h(\tilde{q}_1), \ldots, h(\tilde{q}_n)]^T$.

Using equations (8) and (9), we can obtain the derivative of equation (10) as follows:

$$J_i^* \dot{s}_i^* = x_i + u_i + \Delta M_i, \tag{11}$$

where x_i is the certain dynamics of the MUAVs:

$$\begin{aligned} x_i = {} & -S(\omega_i) \cdot J_i^* \cdot \omega_i + J_i^* \cdot \Big[S(\tilde{\omega}_i) \cdot R\left(\tilde{Q}_{i\,d}\right) \cdot \omega_d \\ & - R\left(\tilde{Q}_{i\,d}\right) \cdot \dot{\omega}_d\Big] + J_i^* \cdot k_1 \cdot \left(\frac{1}{2}\tilde{q}_i + \tilde{q}_{i0} \cdot I_3\right) \cdot \tilde{\omega}_i \\ & + J_i^* \cdot k_2 \cdot \dot{h}(\tilde{q}_i), \end{aligned} \tag{12}$$

where J_i^* and ΔJ_i are the nominal inertial matrix and uncertain inertial matrix of the i th UAV, and the ΔM_i denotes the lumped disturbances.

$$\begin{aligned} \Delta M_i = {} & -S(\omega_i) \cdot \Delta J_i \cdot \omega_i + \Delta J_i \cdot \Big[S(\tilde{\omega}_i) \cdot R\left(\tilde{Q}_{i\,d}\right) \\ & \cdot \omega_d - R\left(\tilde{Q}_{i\,d}\right) \cdot \dot{\omega}_d\Big] + d_i + \Delta J_i \cdot k_1 \cdot \left(\frac{1}{2}\tilde{q}_i + \tilde{q}_{i0} \cdot I_3\right) \cdot \tilde{\omega}_i + \Delta J_i \cdot k_2 \cdot \dot{h}(\tilde{q}_i). \end{aligned} \tag{13}$$

In view of equation (11), the attitude cooperative problem has become the regulation problem, which is to design FNTSMC to compensate the certain dynamics x_i and ΔM_i, and then the variable states satisfy $s \cdot \dot{s} \leq 0$.

Lemma 3 (see [36]). *Consider the error system equations (9)–(11) for sliding surface $\delta_i = \tilde{\omega}_i + c_1 q_i + c_2 q_i^r$, where $0 < r < 1, c_1 > 0, c_2 > 0$, for $i = 1, 2, \ldots, n$. If $\delta_i = 0$, then $\tilde{\omega}_i = 0, q_{i0} = 1$ and $q_i = 0$ can be reached in finite time, respectively.*

Remark 1. Based on Lemma 3 and sliding mode variable structure theorem, the states reach the sliding mode surface $s = [(L + B) \otimes I_3] \cdot J^* \cdot [s^*] = 0$ with the controller to be designed. Based on the definition of matrix B, $[(L + B) \otimes I_3] \cdot J^*$ has full rank, and we can deduce that $s^* = 0$.

3. Main Results

3.1. Nonlinear Distributed Disturbance Observer. In practical environments, disturbances cannot avoid in the dynamics of the MUAVs. In [22], distributed traditional observer has been designed for the multiple agent systems, which is widely used to reconstruct the external disturbances. However, the disturbance cannot compensate in finite time. Inspired by reference [32], a nonlinear disturbance observer is to be designed to compensate the lumped disturbances in equation (11), and we denote $\Delta \widehat{M}_i$ as the output of the observer. The NDO is designed as

$$\Delta \widehat{M}_i = \phi_i + P_i, \tag{14}$$

$$P_i = n_1 \cdot J_i^* \cdot \tilde{\omega}_i + n_2 \cdot \int_0^t \operatorname{sig}^{n_m}\left(\Delta \widetilde{M}_i\right) \mathrm{d}\tau + n_3 \cdot \int_0^t \operatorname{sgn}\left(\Delta \widetilde{M}_i\right) \mathrm{d}\tau, \tag{15}$$

$$\dot{\phi}_i = -n_1 \cdot \left(x_i + u_i + \Delta\widehat{M}_i\right), \tag{16}$$

where n_1, n_2, n_3 and n_m are positive constants and $\Delta\widetilde{M}_i = \Delta M_i - \Delta\widehat{M}_i$ is the estimate error of ΔM_i.

Theorem 1. *Under Assumption 1 and Assumption 2, consider the closed-loop systems (15), the designed NDO (18) guarantees that the estimation of lumped disturbances $\Delta\widetilde{M}_i$ converges to the neighborhoods of ΔM_i in finite time.*

Proof. Construct the following Lyapunov function candidate for the lumped disturbances:

$$V_{\Delta M} = \frac{1}{2} \cdot \Delta\widetilde{M}_i^T \cdot \Delta\widetilde{M}_i. \tag{17}$$

Taking the derivative of $V_{\Delta M}$, $\dot{V}_{\Delta M} = \Delta\widetilde{M}_i^T \cdot (\Delta\dot{\widetilde{M}}_i)$.

Applying equation (15), we can obtain the following inequality:

$$\begin{aligned}\dot{V}_{\Delta M} = \Delta\widetilde{M}_i^T \cdot \Big[&\Delta\dot{M}_i - n_1 \cdot \Delta\widetilde{M}_i - n_2 \cdot \left(\Delta\widetilde{M}_i\right)^{n_m} \\ &\cdot \operatorname{sgn}\left(\Delta\widetilde{M}_i\right) - n_3 \cdot \operatorname{sgn}\left(\Delta\widetilde{M}_i\right)\Big].\end{aligned} \tag{18}$$

Under Assumption 3, the lumped disturbances is bounded, $\|\Delta\dot{M}_i\| \le \rho$, where ρ is a positive constant.

$$\begin{aligned}\dot{V}_{\Delta M} \le \sqrt{2} \cdot &\left(n_3 - \rho\right) \cdot V_{\Delta M}^{1/2} - 2 \cdot n_1 \cdot V_{\Delta M} \\ &- 2^{\left(n_m + 1/2\right)} \cdot n_2 \cdot V_{\Delta M}^{\left(n_m + 1/2\right)}.\end{aligned} \tag{19}$$

By Lemma 3, the estimation error $\Delta\widetilde{M}_i$ will converge to the residual set within a finite time. The residual set of $\Delta\widetilde{M}_i$ is $\Delta\widetilde{M}_i \le \sqrt{2} \cdot \left(\left(\rho - n_3\right)/2^{\left(\left(n_m+1\right)/2\right)} \cdot n_2\right)^{1/n_m}$. □

Remark 2. The parameter n_1 determines the convergence speed of the lumped disturbances. Hence, a trade-off between stability and convergence speed should be considered. Furthermore, we will consider the upper bound of the lumped disturbances and first derivatives unknown in future work.

3.2. Observer-Based FNTSMC Design. Based on the NDO and FTNSMC technique, a decentralized cooperative controller for MUAVs is proposed, which can achieve a fast finite-time convergence result without singularity in the control input.

Theorem 2. *Consider the attitude dynamics (2)–(4) and suppose that Assumption 2 holds. With the proposed NDO (14)–(16), the decentralized finite-time control law (20) guarantees that the tracking errors $\widetilde{q}_i$ and $\widetilde{\omega}_i$ converge to the regions in finite time, respectively.*

The attitude coordinated control law is proposed as

$$\begin{aligned}U = -X - \Delta\widehat{M} + \left[(L+B) \otimes I_3\right]^{-1} \times [-Z \\ \cdot S - \beta \cdot \operatorname{sgn}(S) - \alpha \cdot \operatorname{sgn}(S) \cdot |S|^r],\end{aligned} \tag{20}$$

where $\Delta\widehat{M} = [\Delta\widehat{M}_1, \Delta\widehat{M}_2, \cdots, \Delta\widehat{M}_n]$, $X = [x_1, x_2, \cdots, x_n]$. $X = -S(\omega_i) \cdot J_i^* \cdot \omega_i + J_i^* \cdot [S(\widetilde{\omega}_i) \cdot R(\widetilde{Q}_{i\,d}) \cdot \omega_d - R(\widetilde{Q}_{i\,d}) \cdot \dot{\omega}_d] + J_i^* \cdot k_1 \cdot ((1/2) \cdot \widetilde{q}_i + \widetilde{q}_{i0} \cdot I_3) \cdot \widetilde{\omega}_i + J_i^* \cdot k_2 \cdot \dot{h}(\widetilde{q}_i)$, $S = [s_1, s_2, \cdots, s_n]$, $\beta = \operatorname{diag}\{\beta_1, \beta_2, \cdots, \beta_n\}$, $\alpha = \operatorname{diag}\{\alpha_1, \alpha_2, \cdots, \alpha_n\}$, and $Z = \operatorname{diag}\{z_1, z_2, \cdots, z_n\}$. $z_i, \beta_i, \alpha_i \in R^{3\times3}$ are positive definite matrices, $r \in (0, 1)$. If β_i is selected as $\beta_i > \varepsilon_0$, where β_i is the minimum eigenvalue of β_i, ε_i is the absolute value of estimation error, $\varepsilon_i = \|(L+B) \otimes I_3\| \cdot \Delta\widetilde{M}_i$, $\varepsilon_0 \ge \varepsilon_m$, and $\varepsilon_m = \max\{\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n\}$.

Proof. In the presence of the inertia uncertain matrices and external disturbances, we define the candidate Lyapunov function as follows:

$$V_S = \frac{1}{2} S^T \cdot S. \tag{21}$$

Using equations (10) and (11), taking the derivative of V_s,

$$\begin{aligned}\dot{V}_S &= S^T \cdot \dot{S} \\ &= S^T \cdot \left[(L+B) \otimes I_3\right] \cdot [X + \Delta M + U] \\ &= S^T \cdot \left[(L+B) \otimes I_3 \cdot \Delta\widetilde{M} - Z \cdot S\right. \\ &\quad \left. -\beta_i \cdot \operatorname{sgn}(S) - \alpha \cdot sig^r(S)\right] \\ &\le -S^T \cdot Z \cdot S - S^T \cdot \alpha \cdot sig^r(S) \\ &\le -\lambda_{\min}(Z) \cdot S^T \cdot S - \lambda_{\min}(\alpha) \cdot \|S\|^{r+1} \\ &\le -\mu_1 \cdot V_S - \mu_2 \cdot V_S^{(r+1)/2},\end{aligned} \tag{22}$$

where $\mu_1 = 2 \cdot \lambda_{\min}(Z)$ and $\mu_2 = 2^{((r+1)/2)} \cdot \lambda_{\min}(\alpha)$. From Lemma 2, we can conclude that the sliding mode surface S converges to its origin in finite time:

$$T_r \le \frac{2}{\mu_1(1-r)} \ln \frac{\mu_1 \cdot V_{S(0)}^{(1-r)/2} + \mu_2}{\mu_2}, \tag{23}$$

where $V_{S(0)}$ is the initial value of V_S.

When the tracking error system is on the sliding mode manifold $s_i = 0$, three cases should be considered as follows:

Case 1. If $\bar{s}_{ij} = 0, i = 1, 2, \ldots, n$ and $j = 1, 2, 3$, it means that

$$\widetilde{\omega}_{ij} + \delta_1 \cdot \widetilde{q}_{ij} + \delta_2 \cdot \operatorname{sig}^r\left(\widetilde{q}_{ij}\right) = 0. \tag{24}$$

By Lemma 3, we can obtain $\widetilde{\omega} \longrightarrow 0$ and $\widetilde{q} \longrightarrow 0$, and the closed-loop system can achieve finite-time stability.

Case 2. If $\bar{s}_{ij} \ne 0$ and $|\widetilde{q}_{ij}| > g$, then we can obtain $s_{ij} \ne 0$; obviously, that would not occur.

Case 3. If $\bar{s}_{ij} \ne 0$ and $|\widetilde{q}_{ij}| \le g$, then we can obtain

$$\widetilde{\omega}_{ij} + k_1 \cdot \widetilde{q}_{ij} + k_2\left(\delta_1 \cdot \widetilde{q}_{ij} + \delta_2 \cdot \operatorname{sig}^r\left(\widetilde{q}_{ij}\right)\right) = 0, \tag{25}$$

$\widetilde{\omega}_{ij}$ will converge to the residual set

$$\left|\widetilde{\omega}_{ij}\right| \le k_1\left|q_{ij}\right| + k_2\left|\delta_1 \cdot \widetilde{q}_{ij} + \delta_2 \cdot \operatorname{sig}^r\left(\widetilde{q}_{ij}\right)\right|, \tag{26}$$

in finite time. □

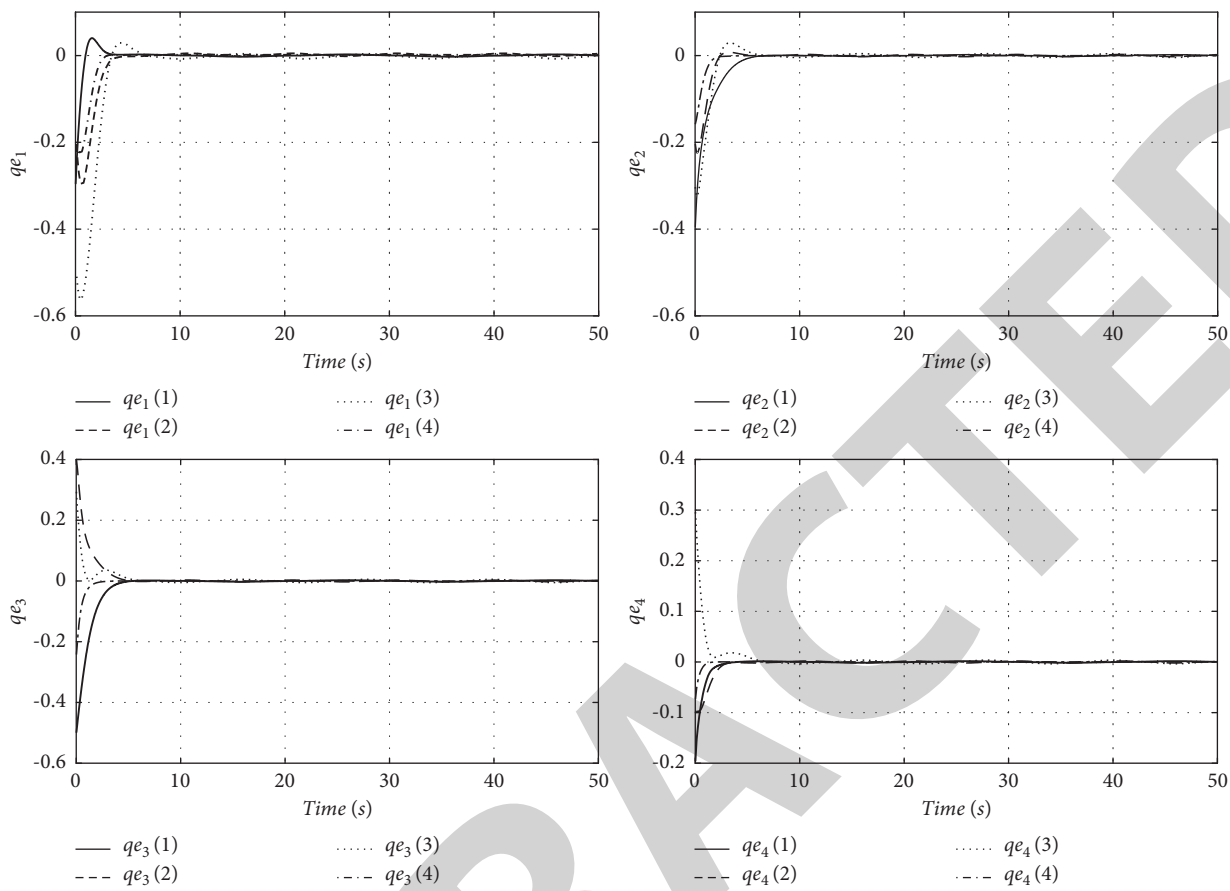


Figure 1: Quaternion errors of each UAV.

Remark 3. Attitude synchronization is the most important in the formation of the MUAVs. In this paper, we only consider the attitude cooperative problem, the position tracking problem is more challenging, and we will develop the work in the future.

Remark 4. The parameter β_i in the control law in designed control law is to compensate the estimation error and gives the robustness of the control law. The robustness parameter β_i can also be modified by the adaptive scheme. If we remove the parameters α_i, then the control law can guarantee that the system is asymptotically stable but have less precision performance compared with the finite-time control law.

Remark 5. This work considers the parameter uncertain and external disturbances case under directed communication topology. It is noted that in the case of the undirected communication topology, $(L+B)\otimes I_3$ does not satisfy Lemma 1, and then the item $[(L+B)\otimes I_3]^{-1}$ cannot be used in the control law. In [33], we note that the item $[(L+B)\otimes I_3]^{-1}$ can be eliminated by constructing the Lyapunov function containing $(L+B)\otimes I_3$.

Remark 6. It is noted that the item $Z\cdot S$ is the feedback control which could make that all the state signals of the MUAVs are uniformly ultimately bounded (UUB), and the term $\alpha\cdot \operatorname{sgn}(S)\cdot |S|^r$ is the nonlinear feedback to make the system achieve the finite-time stability.

4. Numerical Example

In this section, numerical simulations demonstrate the effectiveness of the proposed observer-based FNTSMC scheme. Consider a group of UAV modeled with equations (2) and (3). The corresponding weighted Laplacian matrix is given as

$$L=\begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \\ -1 & 0 & 0 & 1 \end{bmatrix}^T. \tag{27}$$

The initial states of the UAV are arbitrary. The parameters of the MUAVs are as follows: the inertial matrix

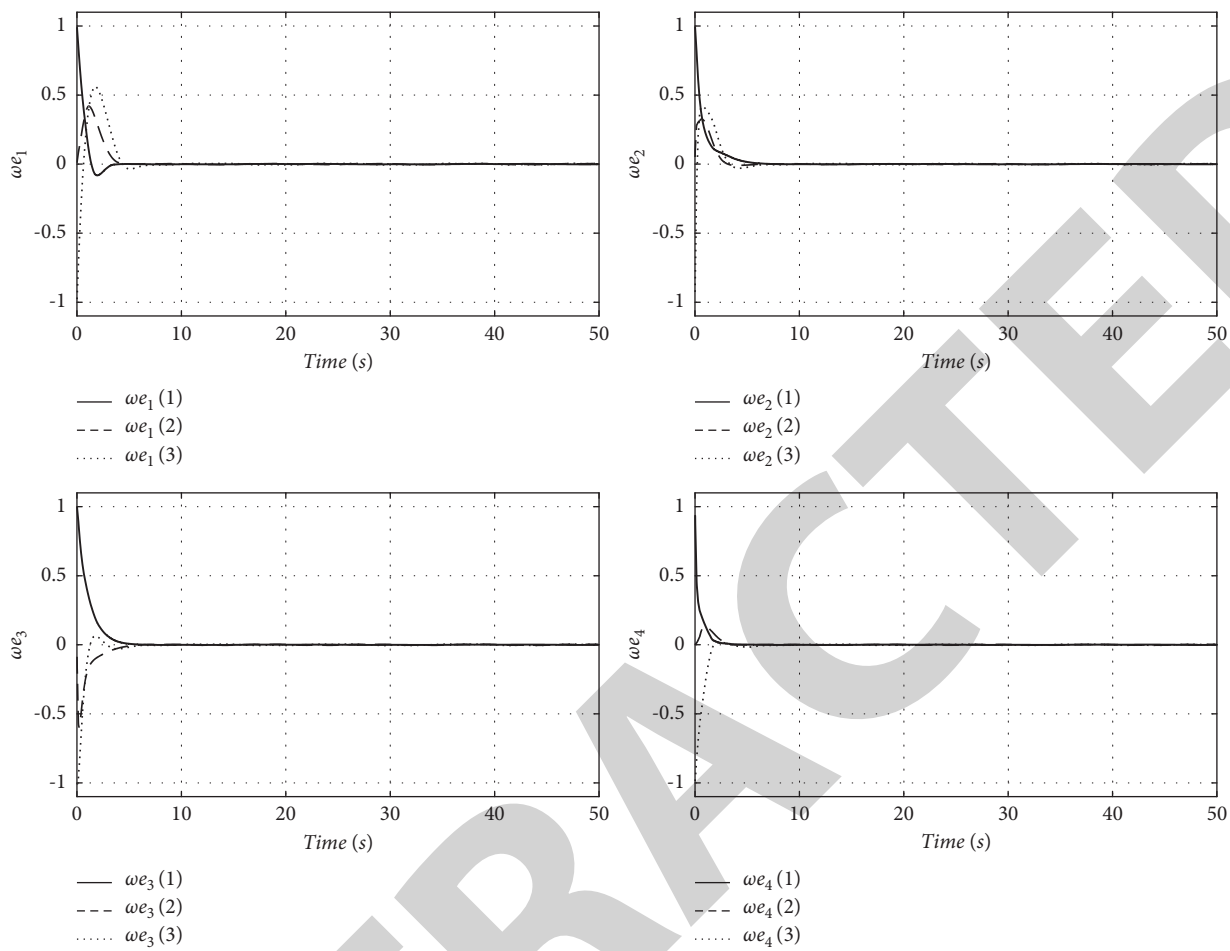


Figure 2: Angular velocity errors of each UAV.

$$J_1 = \begin{bmatrix} 20 & 2 & 0.5 \\ 2 & 22 & 0.9 \\ 0.5 & 0.9 & 30 \end{bmatrix},$$
$$J_2 = \begin{bmatrix} 22 & 1 & 0.9 \\ 1 & 25 & 0.3 \\ 1.5 & 0.7 & 25 \end{bmatrix},$$
$$J_3 = \begin{bmatrix} 23 & 0.5 & 1 \\ 1 & 21 & 1.5 \\ 1 & 1.5 & 22 \end{bmatrix}, \tag{28}$$
$$J_4 = \begin{bmatrix} 25 & 0.5 & 1.5 \\ 0.9 & 28 & 0.5 \\ 0.5 & 0.7 & 25 \end{bmatrix},$$

the nominal inertial matrix $J_1^* = J_2^* = J_3^* = J_4^* = \mathrm{diag}\{20, 20, 20\}\,(\mathrm{kg}\cdot\mathrm{m}^2)$, $Q_d(t) = \begin{bmatrix} 0 & 0 & 0 & 1 \end{bmatrix}^T$, and $\omega_d(t) = \begin{bmatrix} 0 & 0 & 0 \end{bmatrix}^T$, and the parameter of $k_1 = 1$, $k_2 = 0.3$, $\beta = \mathrm{diag}\{0.2, 0.2, 0.2\}$, $n_1 = 20$, $n_2 = n_3 = 0.5$, $n_m = 0.4$, $r = 0.6$, $z_1 = z_2 = z_3 = z_4 = 2$, $\alpha = \mathrm{diag}\{2, 2, 2\}$, $d_1 = [0.5 \sin(\pi t/10), -0.6 \cos(\pi t/5), 0.8 \sin(\pi t/6)]^T$, $d_2 = [0.5 \cos(\pi t/10), 0.6 \cos(\pi t/5), -0.8 \sin(\pi t/6)]^T$, $d_3 = [0.5 \sin(\pi t/10), -0.6 \cos(\pi t/5), 0.8 \cos(\pi t/6)]^T$, and $d_4 = [0.5 \cos(\pi t/10), -0.6 \cos(\pi t/5), 0.8 \sin(\pi t/6)]^T$.

The quaternion $\tilde{q}_i\,(i = 1, 2, 3, 4)$ and angular velocity errors $\tilde{\omega}_i\,(i = 1, 2, 3, 4)$ of each UAV are shown in Figures 1 and 2, respectively. From Figures 1 and 2, we can see the quaternion errors and angular velocity errors converge in finite time, respectively.

For convenience, let RFC represent the designed controller [36]; the response for the RFC and the FNTSMC is shown in Figures 3 and 4. Based on the proposed FNTSMC, the attitude of the MUAVs converges to equilibrium point smoothly with a settling time in less than 10 s. The results show that FNTSMC has faster convergence rate and higher control performance than the RFC controller, indicating that FNTSMC has a better disturbances rejection than the RFC controller.

Figure 5 shows the curves of the lumped disturbances. It is clear that reconstruction of the lumped disturbances can be obtained. The lumped disturbances contain the disturbances and uncertainty. The errors between the estimation and the real values converge to zero not only asymptotically but also achieve finite-time stability. The precise estimation errors showed the high performance of the decentralized finite-time NDO. In the finite-time disturbance observer

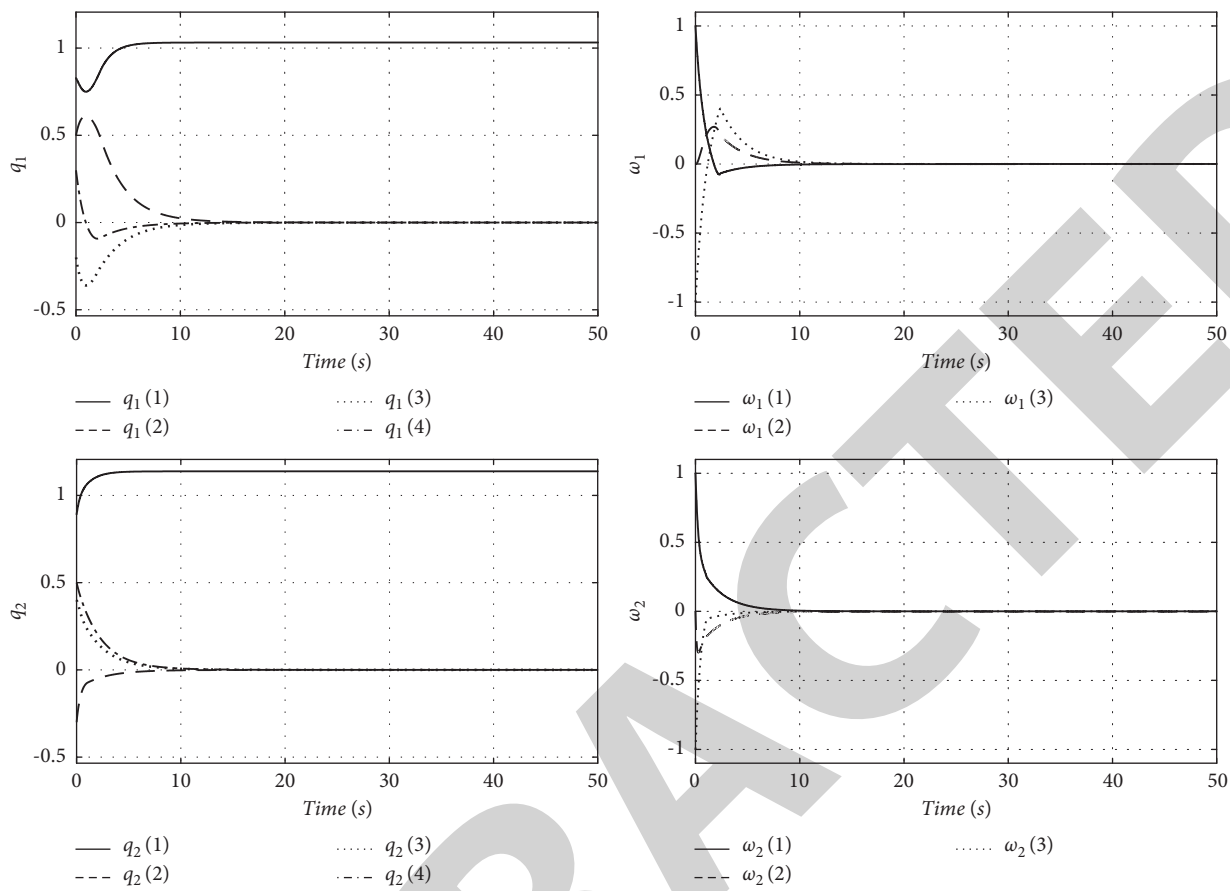


Figure 3: The attitude and angular velocity of each UAV based on RFC.

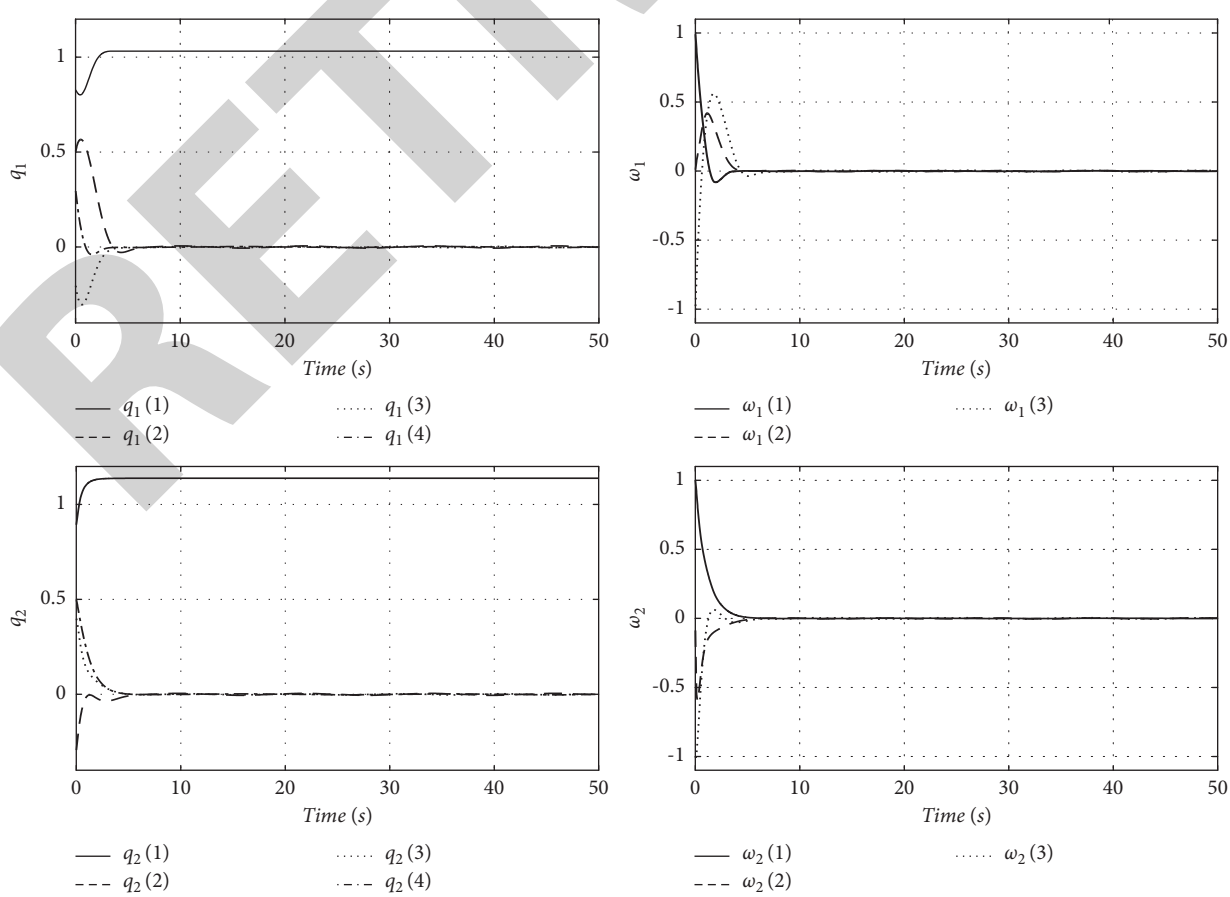


Figure 4: The attitude and angular velocity of each UAV based on FNTSMC.

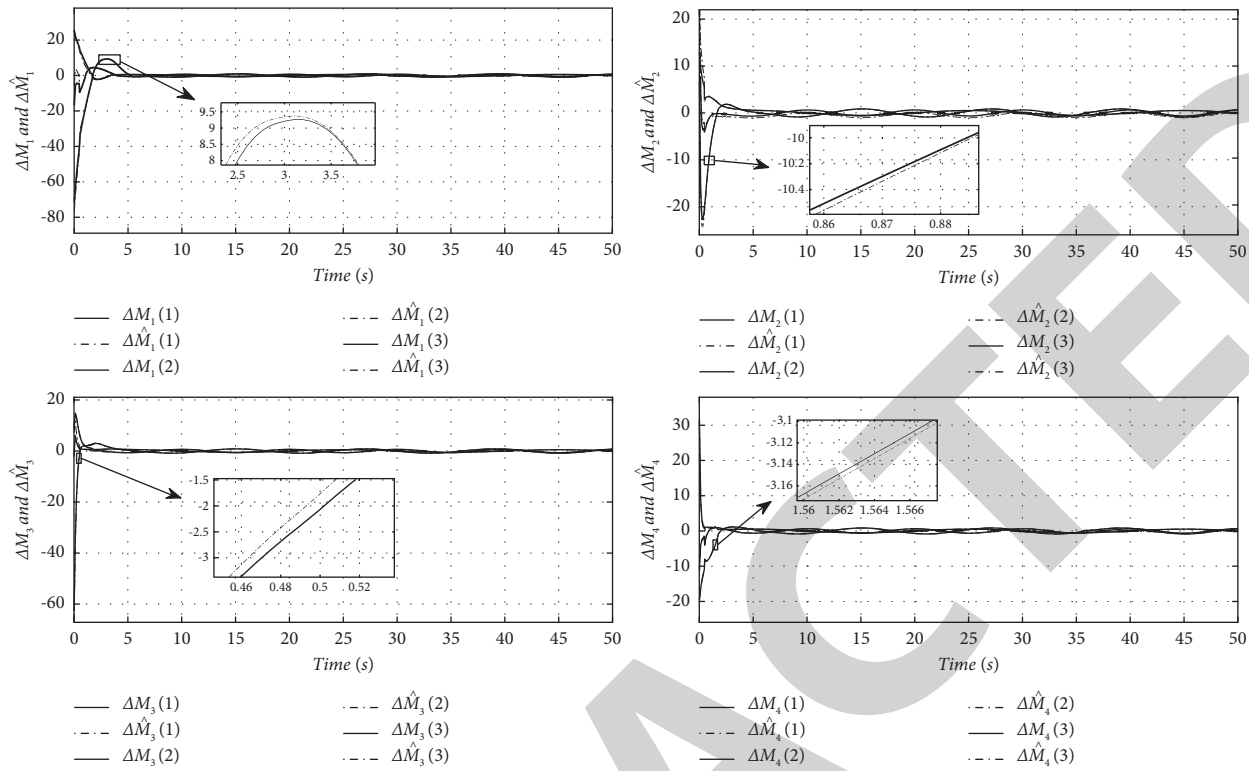


Figure 5: Actual observer output ΔM and estimated observer output $\Delta\widehat{M}$.

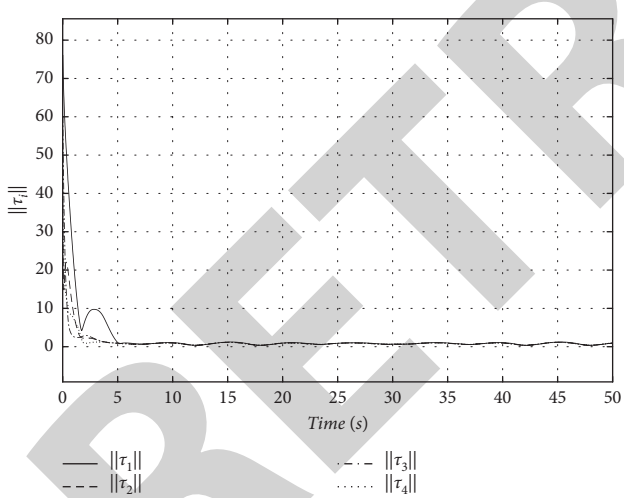


Figure 6: Control torques of each UAV.

technique, the parameter n_1 determines the convergence rate of observer errors of lumped disturbances. Based on the feedforward compensation, the switch gain is greatly reduced and the chattering is effectively reduced. It is noted that the finite-time NDO is given in equations (14)–(16); if we remove the term $n_2 \cdot \int_0^t \mathrm{sig}^{\gamma}(\Delta\widetilde{M}_i)\mathrm{d}\tau$, then we can get the similar observer proposed in [25]. Figure 6 shows the control torques of each UAV. It is observed from the above Figures 5 and 6 that the finite-time attitude cooperative controllers for MUAVs have high control performance.

5. Conclusion

This paper has solved the cooperative problem of multiple unmanned vehicles composed of one leader and many followers under directed topology. A distributed finite-time nonlinear disturbance observer has been proposed to estimate the lumped disturbances containing the inertia uncertainties and external disturbances. Based on the nonlinear observer, decentralized finite-time terminal sliding mode controller has been proposed such that each UAV can track the desired attitude and angular velocity. Future work includes extending the results to the case of communication delays and no angular velocity.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

Acknowledgments

This research was jointly supported by Sichuan Science and Technology Program of China (2020YFH0124), Zigong Key Science and Technology Project of China (2020YGJC01), Robot Technology Used for Special Environment Key Laboratory of Sichuan Province (17kftk05), and Science and Technology Project of Sichuan Province (2021YFS0339).

References

[1] E. T. Alotaibi, S. S. Alqefari, and A. Koubaa, "LSAR: multi-UAV collaboration for search and rescue missions," *IEEE Access*, vol. 7, pp. 55817–55832, 2019.

[2] M. Mozaffari, W. Saad, M. Bennis, and M. Debbah, "Efficient deployment of multiple unmanned aerial vehicles for optimal wireless coverage," *IEEE Communications Letters*, vol. 20, no. 8, pp. 1647–1650, 2016.

[3] C. Yuan, Y. Zhang, and Z. Liu, "A survey on technologies for automatic forest fire monitoring, detection, and fighting using unmanned aerial vehicles and remote sensing techniques," *Canadian Journal of Forest Research*, vol. 45, no. 7, pp. 783–792, 2015.

[4] Q. Zong, D. Wang, S. Shao, B. Zhang, and Y. Han, "Research status and development of multi UAV coordinated formation flight control," *Journal of Harbin Institute of Technology*, vol. 49, no. 3, pp. 1–14, 2017.

[5] F. Xing, A. Wu, Y. Shang, and N. Dong, "A novel sliding mode controller for small-scale unmanned helicopters with mismatched disturbance," *Nonlinear Dynamics*, vol. 83, no. 1, pp. 1053–1068, 2016.

[6] J. Gadewadikar, F. Lewis, K. Subbarao, and B. M. Chen, "Structured h-infinity command and control-loop design for unmanned helicopters," *Journal of Guidance, Control, and Dynamics*, vol. 31, no. 4, pp. 1093–1102, 2008.

[7] A.-M. Kumar and K. Dev Kumar, "Neural network-based distributed attitude coordination control for spacecraft formation flying with input saturation," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 23, no. 7, pp. 1155–1162, 2012.

[8] H. Du and S. Li, "Finite-time attitude stabilization for a spacecraft using homogeneous method," *Journal of Guidance, Control, and Dynamics*, vol. 35, no. 3, pp. 740–748, 2012.

[9] Z. Man, M. O'day, and X. Yu, "A robust adaptive terminal sliding mode control for rigid robotic manipulators," *Journal of Intelligent and Robotic Systems*, vol. 24, no. 1, pp. 23–41, 1999.

[10] O. Mofid and S. Mobayen, "Adaptive sliding mode control for finite-time stability of quad-rotor UAVs with parametric uncertainties," *ISA Transactions*, vol. 72, pp. 1–14, 2018.

[11] Y. Feng, X. Yu, and Z. Man, "Non-singular terminal sliding mode control of rigid manipulators," *Automatica*, vol. 38, no. 12, pp. 2159–2167, 2002.

[12] X. Man Zhihong and Z. Man, "Fast terminal sliding-mode control design for nonlinear dynamical systems," *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, vol. 49, no. 2, pp. 261–264, 2002.

[13] K. Lu and Y. Xia, "Adaptive attitude tracking control for rigid spacecraft with finite-time convergence," *Automatica*, vol. 49, no. 12, pp. 3591–3599, 2013.

[14] J. Liu, Y. Zhang, H. Liu, Y. Yu, and C. Sun, "Robust event-triggered control of second-order disturbed leader-follower MASs: a nonsingular finite-time consensus approach," *International Journal of Robust and Nonlinear Control*, vol. 29, no. 13, pp. 4298–4314, 2019.

[15] H. Liang, Z. Sun, and J. Wang, "Robust decentralized attitude control of spacecraft formations under time-varying topologies, model uncertainties and disturbances," *Acta Astronautica*, vol. 81, no. 2, pp. 445–455, 2012.

[16] L. Yang and J. Yang, "Nonsingular fast terminal sliding-mode control for nonlinear dynamical systems," *International Journal of Robust and Nonlinear Control*, vol. 21, no. 16, pp. 1865–1879, 2011.

[17] N. Zhou, Y. Xia, K. Lu, and Y. Li, "Decentralised finite-time attitude synchronisation and tracking control for rigid spacecraft," *International Journal of Systems Science*, vol. 46, no. 14, pp. 2493–2509, 2015.

[18] J. Zhou, Q. Hu, and M. I. Friswell, "Decentralized finite time attitude synchronization control of satellite formation flying," *Journal of Guidance, Control, and Dynamics*, vol. 36, no. 1, pp. 185–195, 2013.

[19] H. Kurt, M. Stinchcombe, and H. White, "Multilayer feed-forward networks are universal approximators," *Neural Networks*, vol. 2, no. 5, pp. 359–366, 1989.

[20] A.-M. Zou, K. D. Kumar, Z.-G. Hou, and X. Liu, "Finite-time attitude tracking control for spacecraft using terminal sliding mode and Chebyshev neural network," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 41, no. 4, pp. 950–963, 2011.

[21] D. Wang, Q. Zong, B. Tian, S. Shao, X. Zhang, and X. Zhao, "Neural network disturbance observer-based distributed finite-time formation tracking control for multiple unmanned helicopters," *ISA Transactions*, vol. 73, pp. 208–226, 2018.

[22] Y. Hong, G. Chen, and L. Bushnell, "Distributed observers design for leader-following control of multi-agent networks," *Automatica*, vol. 44, no. 3, pp. 846–850, 2008.

[23] M. F. Hassan and M. Hammuda, "Leader-follower formation control of mobile nonholonomic robots via a new observer-based controller," *International Journal of Systems Science*, vol. 51, no. 7, pp. 1243–1265, 2020.

[24] D. Lee, "Nonlinear disturbance observer-based robust control for spacecraft formation flying," *Aerospace Science and Technology*, vol. 76, pp. 82–90, 2018.

[25] Z. Wang and Z. Wu, "Nonlinear attitude control scheme with disturbance observer for flexible spacecrafts," *Nonlinear Dynamics*, vol. 81, no. 1, pp. 257–264, 2015.

[26] J. Duan, H. Zhang, Y. Liang, and Y. Cai, "Bipartite finite-time output consensus of heterogeneous multi-agent systems by finite-time event-triggered observer," *Neurocomputing*, vol. 365, pp. 86–93, 2019.

[27] Y. Dong and J. Chen, "Nonlinear observer-based approach for cooperative control of networked rigid spacecraft systems," *Automatica*, vol. 128, Article ID 109552, 2021.

[28] T. Chen and J. Shan, "Distributed spacecraft attitude tracking and synchronization under directed graphs," *Aerospace Science and Technology*, vol. 109, Article ID 106432, 2021.

[29] F. Wang, H. Gao, K. Wang, C. Zhou, Q. Zong, and C. Hua, "Disturbance observer-based finite-time control design for a quadrotor UAV with external disturbance," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 57, no. 2, pp. 834–847, 2020.

[30] M. Sharma and I. Kar, "Finite time disturbance observer based geometric control of quadrotors," *IFAC-PapersOnLine*, vol. 53, no. 1, pp. 295–300, 2020.

[31] G. Dong, L. Cao, D. Yao, H. Li, and R. Lu, "Adaptive attitude control for multi-MUAV systems with output dead-zone and actuator fault," *IEEE/CAA Journal of Automatica Sinica*, vol. 8, no. 9, pp. 1567–1575, 2020.

[32] S. M. Amrr and M. Nabi, "Finite-time fault tolerant attitude tracking control of spacecraft using robust nonlinear disturbance observer with anti-unwinding approach," *Advances in Space Research*, vol. 66, no. 7, pp. 1659–1671, 2020.

[33] Q. Zong and S. Shao, "Decentralized finite-time attitude synchronization for multiple rigid spacecraft via a novel disturbance observer," *ISA Transactions*, vol. 65, pp. 150–163, 2016.

[34] S. Khoo, L. Xie, and Z. Man, "Robust finite-time consensus tracking algorithm for multirobot systems," *IEEE/ASME*

Hindawi
Wireless Communications and Mobile Computing
Volume 2022, Article ID 3700135, 16 pages
https://doi.org/10.1155/2022/3700135

Research Article

An Effective Task Offloading Method for Separable Complex Mobile Terminal Tasks

Zemin Liu,[1] Na Zhou,[1,2] Yan Wang,[1,2,3] Jian-Tao Zhou,[1,2,3] Haotian Zhang,[1] and Gang Xu[1]

[1]*Inner Mongolia Engineering Lab of Cloud Computing and Service Software, College of Computer Science, Inner Mongolia University, Hohhot, China*
[2]*Ecological Big Data Engineering Research Center of the Ministry of Education, Hohhot, China*
[3]*National and Local Joint Engineering Research Center of Mongolian Intelligent Information Processing, Hohhot, China*

Correspondence should be addressed to Yan Wang; 55234043@qq.com

Received 25 October 2021; Revised 9 January 2022; Accepted 21 January 2022; Published 14 February 2022

Academic Editor: Ting Bi

Due to limited energy and computing power of IoT devices, they cannot handle complex tasks. Edge computing technology effectively solves the requirements of computing power and response delay for complex tasks in devices by migrating computing power to the vicinity of IoT devices. For a separable complex task on IoT terminal, we focus on the effects of data distribution, dependencies, and offloading sequence of subtasks on its total delay when it is offloaded to edge servers. Through comprehensively considering these factors, we study the slicing and choreographing method during the offloading process of a complex task. Firstly, a task slicing method based on hierarchical clustering is presented and an improved hierarchical clustering algorithm is used to obtain the optimal solution of task partitioning. Secondly, a task choreographing method based on overlapping the longest path is presented. Finally, through the simulation experiments of complex tasks with different structures and loads, the effectiveness of our method is verified.

1. Introduction

In recent years, as the mobile Internet industry matures, the rapid explosion of the Internet of Things (IoT) leads the vigorous development of mobile intelligent terminal devices, which are widely used in transportation, health, entertainment, and other fields. At the same time, the applications deployed in IoT devices are becoming more and more complex. For example, they need to deal with large amounts of IoT data and complex processing processes. Limited by their own processing capacity and battery capacity, IoT devices have been unable to meet the needs of these tasks. Cloud computing emerges as a computing mode with unlimited supply of resources and becomes an effective supplement to terminal processing capacity. Mobile devices transfer data to remote terminals, use the resources of cloud data centers to complete efficient operations, and return the final results to users, so as to achieve the goal of fast data exchange. Compared with performing the task directly on the user terminal, the task processing mode of transferring data to the cloud is faster and more efficient, which can support mobile applications to achieve richer functions. However, such mode also has some disadvantages. For example, as users need to transfer amount of data to the cloud center for processing, the data transfer time is too long to exceed the effective delay requirement of the application. In addition, the link distance between mobile devices and cloud center is long, which is prone to interruption or instability, leading to the failure of the feedback results.

The appearance and application of edge computing solve the above problems to a certain extent. Edge computing provides cloud services and IT environment for IoT devices through sinking the processing capacity of the cloud platform to the network edge closer to IoT devices, which makes tasks on IoT devices be offloaded and processed more quickly. Compared with cloud computing, edge computing enables mobile devices to have a shorter data transmission path for offloading tasks, thus reducing the feedback and

transmission delay of data and results, meeting the needs of delay-sensitive tasks on IoT devices. Meanwhile, processing tasks at the edge is also helpful to relieve the traffic pressure on the network.

There are some drawbacks to offloading tasks to edge servers. Due to the limitation of edge server's own processing capacity, the processing time of offloaded tasks will increase when the computing requirement is relatively large. Such processing delay cannot also satisfy a delay-sensitive task, and some complex tasks even exceed the processing capacity of a single edge service. For this, there are two solutions. One way is to offload tasks to multiple edge servers through distributed computing to shorten the time delay of task execution. The other way is to combine the edge computing with the cloud computing to work together to complete the tasks, which can improve the processing capacity and shorten the transmission time. In both cases, we need to divide complex tasks into smaller ones and choreograph them to multiple edge servers or the cloud.

In this paper, we focus on the slicing and choreographing method of a complex delay-sensitive task at edge servers. The main research work and contributions include the two following aspects: (1) Aiming at the problem that the response time of offloading task is too long to meet the delay requirements of IoT devices, the task slicing method based on hierarchical clustering is improved to reduce the communication cost of subtasks on different servers and minimize the time consumption of task workflow while supporting the parallel and distributed execution of subtasks. (2) On the basis of task slicing, a subtasks choreography method combining static, dynamic, and the earliest start time of the task workflow is proposed, and a scheduling algorithm of task workflow on edge servers is designed.

2. Related Works

Mobile edge computing provides a promising solution for mobile terminal devices with limited computing capacity to complete complex, intensive, and sensitive computing tasks. Therefore, in the MEC system, it is very important to optimize the assignment and scheduling of mobile terminal tasks. In recent years, many strategies, methods, and algorithms have been proposed. In this section, we make a detailed introduction and analysis of representative work related to this paper.

Mobile edge computing provides enhanced computing power for mobile devices by deploying edge servers next to the communication base stations. When a mobile terminal receives a task request, the first issue to be decided is whether to offload part or all of the task to an MEC server. Considering the computing needs of different tasks on mobile devices, the optimization method of task partition ratios was proposed to minimize the maximum task latency. In [1], the authors divided multiple parts into a single subtask with prior knowledge and modeled the ordinal number relation of parts to guide the segmentation process in a circular way. In [2], each user could partition their computation task into offloading computing and locally computing parts in multiuser MEC networks. In [3], the offloading location of the task was further extended to the cloud server. Each task on the mobile device can be decided to be processed locally at its mobile device or offloaded to one of the edge servers or a cloud server. When the offloaded task is complex and the computing power of a single edge server is limited, an offload strategy is proposed to divide a task into some subtasks and deploy them in multiple servers. In [4], the authors assumed that each user's tasks were separable and proposed a distributed algorithm to obtain the hierarchical multilevel offloading decisions. In [5], a multiserver system with dynamic speed and power management was modeled as queueing systems, and then the issue of the optimal task dispatching on multiple heterogeneous server systems was addressed. In [6], the authors allocated computing tasks to suitable cores of mobile devices or the cloud in the MCC and proposed an optimization framework to minimize the total energy consumption and maximize the system reliability. For the task offloading problem of a heterogeneous multilayer MEC (HetMEC), the authors designed the latency minimization algorithm by jointly coordinating the resources among the end devices, multilayer MEC servers, and the cloud center in [7]. In [8], the authors designed LL-MLS algorithm to find an optimal partition of a given workload through task scheduling and energy allocation strategies. In [9], the authors proposed an energy-aware cooperative routing (ECoR) scheme for optimal handling of task offloading between source and target UAVs in a gridlocked swarm.

Offloading the task to the edge computing system not only provides the task with expanded processing capacity but also brings with it the transmission delay caused by the offloading process. Therefore, the allocation of wireless resources for offloading task is also the focus of many research works. In [10], the authors transformed the problem of joint task assignment and wireless resource assignment into a mixed-integer nonlinear program (MINLP) and proposed a suboptimal solution algorithm based on relaxation convex problem to reduce time delay for offloading tasks. In [11], the problem of task assignment in the MEC in the return network was solved by a similar method. In [12], an online adaptive task allocation and computing offload strategy was proposed, which coordinated and optimized the wireless and computing resource allocation by considering dynamic wireless conditions and service delay constraints.

In order to efficiently implement the allocation and deployment of multiple tasks in the MEC, some researchers regard the problem as a joint optimization problem considering various offloading conditions. In [13], the authors minimized energy consumption for all devices and their task delay constraints by cooptimizing communication and computing resource allocation on devices and mobile edge servers. Considering the task completion time and the mobile device energy consumption, the authors in [14] proposed a heuristic offloading decision algorithm (HODA), which jointly optimized the offloading decisions, communication, and computing resources to maximize the system utility. In order to reduce the complexity of the joint optimization problem, the original problem was decomposed into two subproblems in [15–21]. In [15], the authors

addressed the resource allocation problem using the convex and quasi-convex optimization techniques and solved the problem of task assignment by a heuristic algorithm. In [16], the task partitioning subproblem was taken as a set of univariate optimization problems, which can be easily solved, and the task scheduling subproblem was solved through a heuristic algorithm. In [17], the problem of resource allocation was further decomposed into two stages: the computing resource optimization and the communication resource allocation. The authors proposed a subchannel allocation scheme, and then the transmission power allocation was considered as a convex optimization problem based on the scheme and was solved by the Lagrange multiplier method. For the resource allocation scheme, the authors in [18] proposed a computing framework based on the weighted sum of task completion time and energy consumption in the MEC system, while the authors in [19] proposed a task shunting and resource allocation algorithm based on Deep-Q network. In [20], the authors considered users' risk-seeking or loss-avoidance behaviors in their final decision. In [21], the authors proposed the energy-efficient multihop communication solution in smart city environment.

In addition to finding a better task offloading strategy by optimizing the allocation and consumption of communication resources and computing resources during task offloading, some other factors, such as shared data among tasks, offloading sequence of tasks, and the mobility of mobile devices, also have an important impact on the efficiency of task offloading. In [22], the authors studied the task assignment algorithm in data shared mobile edge computing systems and proposed three algorithms to deal with holistic tasks and divisible tasks, respectively. In [23], an adaptive slicing method for decentralized workflow based on clustering was proposed. Then a data-related task scheduling algorithm based on the correlation task model was designed, which gave an evaluation function to reduce the intercore communication during the process of task execution by assigning highly the correlated tasks to the same core. In [24], the authors gave full consideration to the mobility of user in the MEC and then proposed a device-to-device (D2D) cooperation method to expedite the task execution of mobile user by leveraging proximity-aware task offloading. In [25], the user mobility and network constraints were considered, and a lightweight heuristic solution was proposed for fast scheduling. In [26], a task allocation solution for optimizing latency and service quality was proposed to support the mobility of vehicles, in which the constraints on service latency, quality loss, fog capacity, stationary task allocation, and mobile fog nodes were taken into account. In [27], the effective task offloading scheme in the MEC was designed, in which the tasks are offloaded to the adjacent servers at the next AP in the direction of vehicle driving. In [28], the authors emphasized the importance of optimizing operation sequence in multiuser MEC system and established a computation offloading model to optimize the task operation sequences and starting times for uploading, executing and downloading, and duration times for uploading and downloading. In [29], a spatiotemporal framework based on stochastic geometry and continuous time Markov chains was proposed. The experimental results showed that the framework can find the optimal number of edge servers for parallel computing of the user task. In [30], the authors studied the scheduling method of parallel tasks merging and scheduling for parallel deep learning applications in the MEC.

When the tasks of mobile terminals are offloaded to the edge computing system, particularly the tasks that can be divided are offloaded to different edge servers, and the execution sequence of tasks on edge servers is the key to determine the actual execution efficiency of tasks. For this, the authors in [31] focused on the problem of providing QoS and performance guarantees to divisible loads and then proposed a linear algorithm for real-time divisible load scheduling by eliminating the need to generate exact schedules in the admission controller. In [32], the authors adopted a Markov decision process to handle the problem of computation task scheduling for MEC systems, where the computation tasks were scheduled based on the queueing state of the task buffer, the execution state of the local processing unit, and the state of the transmission unit. In [33], a partitioned fixed-priority real-time scheduling based on dependent tasks split on homogeneous multicore platform was proposed, which converted dependent tasks into a series of sequential jobs and obtained the interrelated subtasks path as well as synthetic deadlines through the B-tree task model. In [34], the authors creatively proposed a deep learning architecture based on tightly connected network and proposed a corresponding multitask parallel scheduling algorithm. In [35], a peer-to-peer (P2P) enhanced task scheduling framework to minimize the average task duration in device-to-device (D2D) network was proposed. In the framework, an iterative algorithm based on alternating optimization and sorting technology was used to solve the approximate optimal scheduling solution.

To sum up, when a complex task is divided and offloaded to multiple edge servers or cloud, the efficiency of task offloading is affected by many factors. The above studies put forward some effective task offloading strategies from the perspectives of processing capacity of mobile terminals and servers, communication channel allocation in the process of task offloading, and optimization of multitask deployment in multiple servers. However, these studies pay little attention to the effect of data interaction between subtasks and the offloading sequence of subtasks on the execution delay of the whole task after tasks are divided into subtasks. Obviously, these factors also have a great impact on the efficiency of subsequent task scheduling. Although papers [31–35] focus on these two factors to optimize the task offloading process, they are not considered as a whole. However, task slicing is closely related to its offloading sequence, and different slicing schemes should correspond to different offloading sequence to optimize the execution delay of the task to the maximum extent. So we focus on the two following problems in the offloading process: (1) in the distributed deployment of complex tasks on edge servers, the data dependencies among the subtasks are fully considered to minimize the communication delay caused by such data

dependencies during task execution. (2) In the process of task scheduling, the effect of subtask offloading sequence on task execution is considered, and the execution delay of the whole task is minimized by parallel subtask offloading and subtask execution.

3. Problem Definition and Formalization

Here, we consider a mobile task offloading scenario with multiuser, multiedge servers. Users' mobile devices can connect and communicate with base stations covering their signals. Edge servers are uniformly deployed near these base stations. At least one edge server is deployed near each base station. Assume that there are m mobile terminal devices and n edge devices in this scenario, $U = \{u_1, u_2, \ldots, u_m\}$ represents the set of mobile terminal devices, and $S = \{s_1, s_2, \ldots, s_n\}$ represents the set of edge servers. The service request sent by each terminal device can be divided into a series of subtasks. We use a directed acyclic graph (DAG) to represent the tasks offloaded by the mobile terminal and the relationships among them, represented as $G(T, E, C(T), W(E))$, where $T = \{t_1, t_2, \ldots, t_k\}$ represents all subtasks offloaded by the mobile terminal and k is the number of subtasks; $E = \{e_{ij} |$ if the output of subtask t_i is an input to subtasks $t_j\}$ represents data dependencies between subtasks; $C(T) = \{c_1, c_2, \ldots, c_k\}$ represents the workload of each subtask in set T; that is, c_i is the CPU cycle of subtasks t_i; $W(E) = \{w_{ij} |$ if $\exists e_{ij}$, $W(e_{ij}) = w_{ij}\}$ represents the size of the input data from t_i to t_j. An example of a mobile terminal task offloaded on edge servers is shown in Figure 1.

Given $G(T, E, C(T), W(E))$, we need to offload multiple subtasks with dependent relationships to the edge service system. If all subtasks in T are deployed on the same edge server, the computing capacity constraints of a single server and the serial execution of tasks may lead to too long feedback delay of the task to meet its needs. In order to take advantage of edge service system to better meet the demands of mobile terminal, we need to offload the task to multiple edge servers, respectively, make some tasks in parallel execution, and shorten the overall delay of the task. For example, the subtasks t_2, t_3, and t_4 can be executed in parallel in Figure 1. A new challenge is that the data dependencies between subtasks introduce new transmission delays. In particular, when the subtasks with large data dependencies are deployed on different servers, the new latency introduced may even outweigh the time savings in the process of executing the tasks in parallel. To solve this problem, the goal of this paper is to first find a task partitioning scheme based on the dependencies between subtasks; we call it task slicing, which can reduce the introduction of new delay as much as possible while deploying all subtasks in a distributed way. In addition to the task slicing affecting the execution delay of the task, the offloading order of the subtasks also has a certain impact on the feedback delay of the task. However, most of the existing studies mainly ignore this problem. In fact, due to the limited wireless communication resources in the MEC environment, when multiple tasks are offloaded at the same time, each task will receive less wireless resources, which will inevitably increase the transmission delay of offloaded subtasks. In $G(T, E, C(T), W(E))$, the subtasks do not need to be executed at the same time, so it is not necessary to simultaneously offload subtasks to different edge server. We just need to make sure that a subtask is offloaded before it is executed, which can maximize offloading bandwidth allocation of the subtasks so as to shorten their transmission delay. Therefore, we will study the task choreography method based on task slicing to optimize the overall delay of the task. The problem in this paper is formally described as follows:

$$\underset{G}{\text{Min}}\ \text{total Time} = T_{\text{exec}}^{\text{end}}(\text{Max}T(\text{Ord}(\mathbb{S})) - T_{\text{offload}}^{\text{start}}(\text{Min}T(\text{Ord}(\mathbb{S})), \tag{1}$$

where

$$\mathbb{S} = \{TL_1, TL_2, \ldots, TL_{k'}\}, \quad (k' \le k), \tag{2}$$

$$\text{Ord}(\mathbb{S}) = \{\langle TL_i, pr_i \rangle |, \quad i \in [1, k'],\ pr_i \in R,\ pr_i > 0\}, \tag{3}$$

s.t.

$$TL_i \subseteq T, \quad (i \in [1, k']), \tag{4}$$

$$\bigcup_{i \in [1,k']} TL_i = T \text{ and } \bigcap_{i \in [1,k']} TL_i = \Phi, \tag{5}$$

$$\text{T}_{\text{exec}}^{\text{start}}(t_i) \ge \text{Max}\left(\text{T}_{\text{exec}}^{\text{end}}(t_j)\right), \quad (i, j \in [1, k]) \text{ if } t_j \in \text{PRE}(t_i) \text{ and } t_i,\ t_j \in TL_p, \tag{6}$$

$$\text{T}_{\text{exec}}^{\text{start}}(TL_i) \ge \text{Max}\left(\text{T}_{\text{exec}}^{\text{end}}(TL_j) + \text{T}_{\text{tran}}(TL_j, TL_i)\right), \quad \left(i, j \in [1, k'] \text{ and } TL_j \in \text{PRE}(TL_i)\right), \tag{7}$$

$$\text{T}_{\text{offload}}^{\text{end}}(TL_i) \le \text{T}_{\text{exec}}^{\text{start}}(TL_i), \quad (i \in [1, k']), \tag{8}$$

$$T_{\text{offload}}^{\text{end}}(TL_i) \ge \sum_{t_j \in TL_i} T_{\text{offload}}^{\text{end}}(t_j) - T_{\text{offload}}^{\text{start}}(t_j). \tag{9}$$

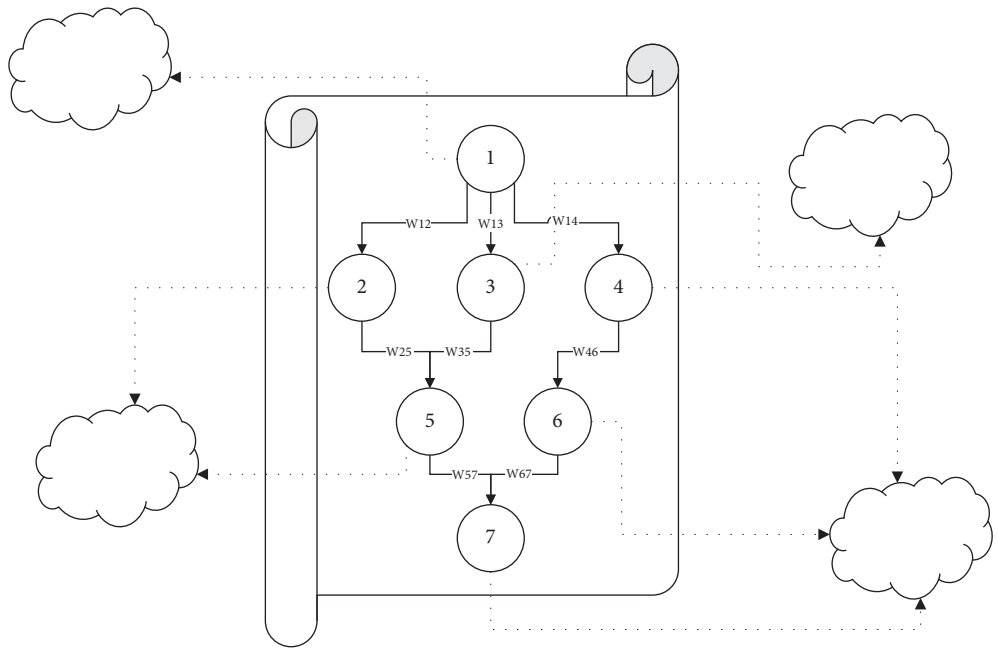


Figure 1: An example of complex task offloading.

Equation (1) is our optimization goal to find a slicing and choreography scheme for task offloading of mobile devices, that is, $\mathbb{S}$ and $\mathrm{Ord}(\mathbb{S})$, which minimizes the response time of the overall task. As shown in equation (2), $\mathbb{S}$ is a partition of T. Each task slicing corresponds to a priority, which represents the order in which subtasks are offloaded, as shown in equation (3). $\mathrm{T}_{\mathrm{exec}}^{\mathrm{start}}(t)$ and $\mathrm{T}_{\mathrm{exec}}^{\mathrm{end}}(t)$ represent the start and end times for executing t, respectively. $T_{\mathrm{offload}}^{\mathrm{start}}(t)$ and $T_{\mathrm{offload}}^{\mathrm{end}}(t)$ represent the start and end times for offloading t, respectively. Min$T$ and Max$T$ are used to obtain the last and first task slices in the choreography scheme, respectively. Equations (4)–(9) represent the constraint conditions that need to be satisfied when offloading the task, where $T_{\mathrm{tran}}(TL_j, TL_i)$ represents the data transmission time between TL_j and TL_i and $PRE(TL_i)$ is used to obtain the preorder task slicing of TL_i. Equations (4) and (5) represent the basic requirements of task slicing. Equation (6) indicates that if there is a dependency relationship between the subtasks in a task slicing, the start time of the subsequent subtasks must be later than the end time of all the preordering subtasks. Equation (7) indicates that if there is a dependency relationship between the subtasks in different task subsets, the subsequent subtask cannot be executed until it has received the output data of all the preordering subtasks to the server where it was deployed. Equation (8) indicates that any task slicing must be offloaded to the corresponding edge server before it begins to execute. Equation (9) indicates that when multiple subtasks need to be offloaded on the same edge server, they must be offloaded in sequence.

4. Task Slicing and Choreographing Model

For complex mobile terminal requests that can be divided into multiple subtasks, we can distribute these tasks on multiple edge servers to improve the processing capacity, which is conducive to reducing the execution delay of terminal tasks. In this section, we will establish a slicing and choreographing model for complex tasks to minimize the transmission and computation delay in the process of task offloading. Here, we name our method SSCS (Slicing Similarity and Choreograph Sequence).

4.1. Task Slicing Method Based on Data Dependencies. The objective of task slicing model is to optimize the parallel execution of subtasks on different edge servers and minimize the delay caused by data transmission between subtasks. We propose a slicing model based on task workflow, in which the concept of task similarity is defined based on the dependency relationship between subtasks. For two subtasks, if there is a large amount of data exchange between them but less contact with other subtasks, they will be divided into the same cluster. The related definitions are given below.

Definition 1. Subtask (t). It istThe basic unit of a task slice, which refers to a task that cannot be divided again, corresponding to an element in the task set T.

Definition 2. Task slice (TL). It is the basic unit of task deployment. After a task is split into multiple subtasks, the similar subtasks will be grouped into the same subtask set, called task slice.

Definition 3. Task hierarchy association matrix (A). It is a hierarchical logic relation of task execution for given $G(T, E, DS(T), W(E))$, represented as $A_{(r*k)}$, where k is the number of subtasks in T and r is the number of logical levels that the task workflow needs to be executed at least, formally

$$a_{ij} = \begin{cases} 1, & \text{if } \lambda(t_j) = \text{NULL at } L(t_j) = i, \\ 0, & \text{otherwise.} \end{cases} \tag{10}$$

$L(t_j) = i$ means t_j is at the i-th layer. $\lambda(t_j) = \text{NULL}$ means that all the preceding subtasks of t_j have been executed, and task t_j can start to execute.

Definition 4. Direct correlation of subtasks (dirCorr). In G, if $e_{ij} \in E$ exists, then the direct correlation degree between task t_i and t_j is represented by

$$\text{dirCorr}_{ij} = w_{ij}. \tag{11}$$

It represents the amount of direct communication data between two subtasks, and the larger the value is, the greater the probability that the two subtasks should be deployed to the same edge server.

Definition 5. Dependency between subtasks (D). In G, the dependencies between subtasks can be divided into three categories: single dependency (D^{single}), split dependency (D^{split}), and joint dependency (D^{join}). D_{ij}^{single} means that task t_j depends entirely on task t_i; that is, $\forall x$, if $x \neq j$, then $e_{ix} \notin E$, and, $\forall y$, if $y \neq i$, then $e_{yj} \notin E$. D_{ij}^{split} means that task t_j depends partly on task t_i, that is, $\exists x$; make $e_{xj} \in E$ and $x \neq j$. D_{ij}^{join} means that a part of the input of task t_j depends on task t_i, that is, $\exists x$; make $e_{xj} \in E$ and $x \neq i$. If D_{ij} = false, then there is no direct dependency between t_i and t_j.

Definition 6. Subtask dependency correlation (depCorr). Based on different types of dependency relationships between subtasks, the dependency correlation degree between t_i and t_j is represented by

$$\text{depCorr}_{ij} = \begin{cases} 1, & \text{if } D_{ij}^{\text{single}} = \text{true}, \\ \dfrac{w_{ij}}{\sum_{i=1}^{k} w_{ix}}, & \text{if } D_{ij}^{\text{split}} = \text{true}, \\ \dfrac{w_{ij}}{\sum_{i=1}^{k} w_{xj}}, & \text{if } D_{ij}^{\text{join}} = \text{true}, \\ 0, & \text{if } D = \text{false}, \end{cases} \tag{12}$$

where wij represents the amount of communication data between t_i and t_j; if there is no dependency relationship between them, then $w_{ij} = 0$.

Definition 7. Subtask distribution correlation (disCorr). It reflects the influence of data distribution between subtasks on task partition results. The larger the data traffic between subtasks, the greater the probability of the subtasks coupling into the same slice. The communication correlation degree between t_i and t_j can be expressed by

$$\text{disCorr}_{ij} = \frac{\left(Corr_{ij}^{\text{in}} + Corr_{ij}^{\text{out}}\right)}{2}, \quad (i, j \in [1, k] \text{ and } i \neq j). \tag{13}$$

From the above equation,

$$\text{Corr}_{ij}^{\text{in}} = \frac{\sum_{x=1}^{k}\left(w(e_{xi}) + w(e_{xj})\right)}{2}, \tag{14}$$

$$\text{Corr}_{ij}^{\text{out}} = \frac{\sum_{x=1}^{k}\left(w(e_{ix}) + w(e_{jx})\right)}{2}. \tag{15}$$

In equation (13), $\text{Corr}_{ij}^{\text{in}}$ and$\text{Corr}_{ij}^{\text{out}}$ are the correlation degrees calculated by the input and output data traffic of t_i and t_j, respectively. $w(e_{ij})$ represents the communication data volume from t_i to t_j. If there is no dependency between them, then $w(e_{ij}) = 0$.

Definition 8. Subtask computation time (exetime). Due to the different sizes of subtasks, their computation time will be different. Assume that the computing power of the CPU is C^{Edge}; $exetime_i$ can be expressed as follows:

$$\text{exetime}_i = \sum \frac{ps_i * Mpc}{C^{\text{Edge}}}, \tag{16}$$

where Mpc represents the CPU cycles required to process a unit of data.

Definition 9. Sequential execution time between subtasks (exetime_{ij}). If there is a sequential dependency between two consecutive subtasks, their sequential execution time is the sum of their respective execution times; otherwise, the value is 0.

$$\text{exetime}_{ij} = \begin{cases} \text{exetime}_i + \text{exetime}_j, & \text{if } D = \text{true}, \\ 0, & \text{if } D = \text{false}. \end{cases} \tag{17}$$

Definition 10. Similarity between subtasks ($T\ D$). It represents the degree of comprehensive correlation between subtasks. The similarity between t_i and t_j is represented by

$$TD_{ij} = \frac{\left(\text{dirCorr}_{ij} + \text{exetime}_{ij}\right) \times \text{depCorr}_{ij} \times \text{dirCorr}_{ij}}{\left(\text{disCorr}_{ij} - \text{dirCorr}_{ij}\right)}. \tag{18}$$

Task similarity comprehensively measures the correlation between two subtasks in the whole workflow system from the direct data dependency between tasks and the relative importance of such dependency in the whole task flow, which serves as the basis for further coupling subtasks. First of all, dirCorr_{ij} is the main factor that determines whether the subtasks can be aggregated into a task slice. In addition, depCorr_{ij} reflects the degree of association between t_i or t_j and other tasks; the lower the degree of association between t_i or t_j and other tasks, the higher the probability that they are aggregated into a task slice. $exetime_{ij}$ represents

the computation time required to complete them. Dividing more subtasks that need to be executed sequentially into a task slice will help to avoid the delays caused by data transfer between subtasks. disCorr_{ij} reflects the degree of correlation between t_i and t_j and their precedence and postorder subtasks. The smaller the degree of correlation, the higher the probability that t_i and t_j are grouped into one task slice.

An example is given in Figure 2, where the weights between nodes represent the reciprocal of similarity between different subtasks. Firstly, starting from the first node of the task workflow, we look for the subtasks that can be grouped into one task slice from top to bottom. The rule of merging is that each subtask is merged with one of the subsequent subtask which has the lowest weight with it until the last node in the workflow. Secondly, starting from the last node of the task workflow, we continue to look for the subtasks that can be merged into one task slice from bottom to top. The rule of merging is that each task is merged with one of the preceding tasks which has the lowest weight with it until the first node in the workflow. For example, in Figure 2, nodes t_1, t_3, and t_7 are merged into one task slice, and nodes t_2, t_5, t_{10}, and t_{11} are merged into one task slice. Then, the minimum weight of the merged subtasks is set as the merge threshold. For those subtasks that are not merged, they are merged when the weight between continuous subtasks is less than the threshold. The threshold value is obtained by comprehensively considering all the similarity of the whole workflow. According to the rule, nodes t_4 and t_9 are merged into one task slice.

In the following, an improved horizontal clustering algorithm is given to solve the slicing scheme based on the similarity between subtasks, as shown in Algorithm 1.

Algorithm 1 provides a method to determine the optimal slicing scheme of task workflow under the premise of a given number of edge servers. The value of SM can be allocated statically according to the resource situation in the MEC system or solved dynamically by optimizing the overall computation delay of the task workflow. In Algorithm 1, lines 6–9 are used to calculate the similarity between the subtasks in T by using equations (10)–(18). Lines 10–18 are used to solve the task slicing scheme based on the improved hierarchical clustering process. Here, the concurrency of task slices is mainly considered, and subtasks at the same level cannot be divided into the same task slice, as shown in line 15.

4.2. Task Slice Choreography Method Based on the Longest Overlapping Path. Next, we need to choreograph these subtasks for offloading. Our goal is to accomplish sequential offloading of task workflow and shorten the task wait delay for executing while offloading. In this paper, we propose a subtask choreography method based on the longest overlapping path by analyzing the logical relationship and execution constraints among subtasks. The definitions are given below.

Definition 11. The computation time of task slice ($D^{\text{comp}}(TL)$). It refers to the sum of execution delays of all subtasks in the task slice.

$$D^{\text{comp}}(TL) = \sum_{t_i \in TL} \text{exetime}_i. \tag{19}$$

Each task slice must wait until all the subtasks on which it depends have been completed before it begins to execute. The earliest start time of a task slice is determined by the longest path from the initial task slice to the last task slice. Therefore, we first carry out static sorting for all tasks slices. Here, we give an example for the task slice hierarchical relationship, as shown in Figure 3.

The following is a brief description of static sorting rules. The node in the first layer is the first task slice of task workflow, its subsequent task slices are in the second layer, and so on. In general, when the task slice at layer i completes, the task slices at layer $i+1$ can start executing. But it is not strict. For example, when TL_3 is complete, TL_5 and TL_6 can be executed regardless of TL_2. If the computation delay of TL_5 or TL_6 is significantly higher than that of TL_4, then whether to offload TL_2 or TL_3 first will have an impact on the overall delay of the task workflow. Obviously, in this case, TL_3 should be preferred. In addition, TL_5 and TL_6, which are deployed on different edge servers, can be executed in parallel. But, compared to TL_6, TL_5 has more subsequent tasks. When the execution delay of TL_5 is not greater than that of TL_6, we should offload TL_5 first, so that it is executed more earlier than TL_6. The offloading order of task slices has an important effect on the computation time of the task slices on the edge servers. Here we define the earliest execution start time for a task slice.

Definition 12. The earliest start time of the task slice ($T_{\text{exec}}^{\text{start}}(TL)$). Each task slice must wait until all the subtasks on which it depends have been executed before it executes. The earliest start time of task slice TL is determined by the longest path from the initial task slice to this task slice, and the calculation formula is represented by

$$T_{\text{exec}}^{\text{start}}(TL_i) = \begin{cases} 0, & \text{if } L(TL_i) = 1, \\ \max\limits_{TL_x \in PRE(TL_i) \text{ and } L(TL_x) = L(TL_i)-1} \left\{ \dfrac{T_{\text{exec}}^{\text{start}}(TL_x) + D^{\text{comp}}(TL_x) + w_{xi}}{B^{\text{StoS}}} \right\}, & \text{if } L(TL_i) > 1, \end{cases} \tag{20}$$

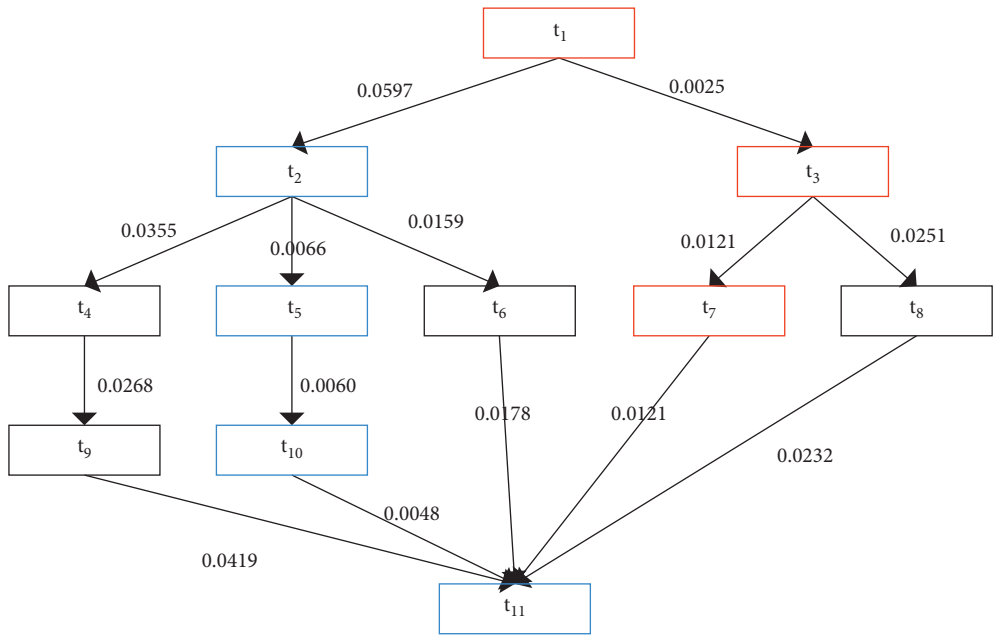


Figure 2: A task workflow example.

```
function [S] = Slice (G, SM, A)
Input: G
      SM//Number of edge servers, 0 < M < m
      A//Logical hierarchy matrix
output: S
InitNum(T)//Initializes the subtask
taskNum = Count(T)//the number of T
for i = 1: taskNum
    TD = TaskSim (G)//TD is sliceNum * sliceNum matrix, and the similarity between sub-tasks is calculated
sliceNum = taskNum
while true
if sliceNum ≤ SM
      break;
Stemp = MaxSim (TD);
if Notlevel (Stemp, A)//Tasks at the same logical level cannot be divided into a task slice
      Cluster = Merge (Stemp)//Task clustering, forming a new task slice division
      sliceNum = Count (Cluster)
S = Cluster
```

Algorithm 1: Task slicing algorithm based on similarity between subtasks.

where the longest path of task slice TL_i is equal to the longest path of all its presequence task slices, its own computation time, and the transmission delay between it and other task slices.

As shown in Figure 4, the subtasks are executed in the order of $TL_1, TL_3, TL_2, TL_5, TL_6, TL_4, TL_7, TL_8$ according to the static sort. However, since the size of each subtask is different, the computation time is also different. Scheduling tasks in a statically sorted manner can cause too much delay in the execution of the overall task workflow. Therefore, we need to combine static sorting with task slice's earliest start time to produce a comprehensive sorting result.

Next, we use the concept of task priority to represent the choreography scheme $\mathrm{Ord}(\mathbb{S})$ of task slice set $\mathbb{S}$. The priority of each task slice is determined by its latest offloading time. Let the priority of the last task slice in the task workflow be 0. The higher the priority of the task slice is, the earlier it should be offloaded. The related definitions are as follows.

Definition 13. Task slice priority ($\mathrm{Prio}(TL)$). It represents the time to offload the task, that is, the latest time for the task slice to be offloaded. The calculation formula of TL_i priority is as follows:

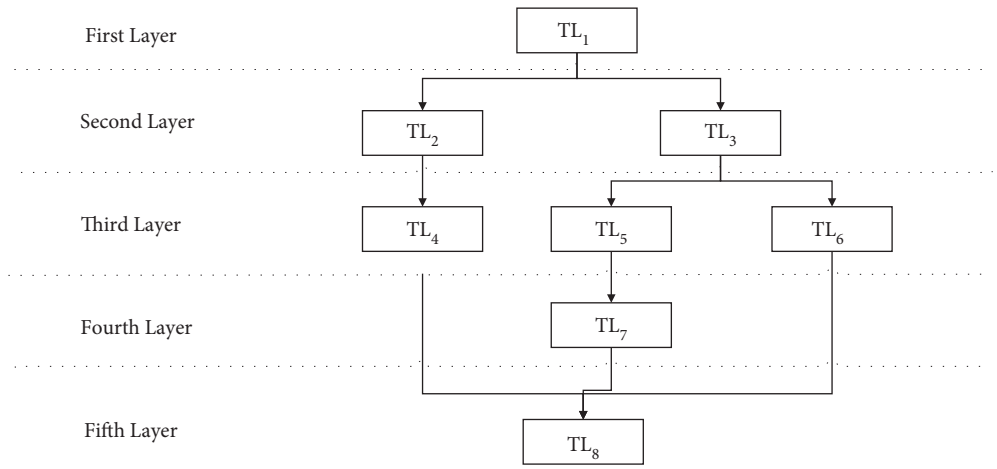


Figure 3: Task slice hierarchical association relationship.

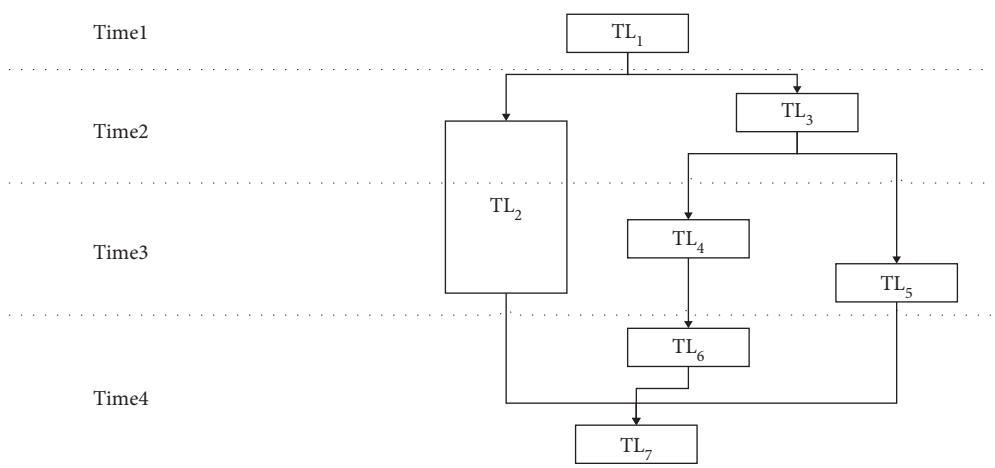


Figure 4: Diagram of the earliest start time of task slices.

$$\mathrm{Prio}\left(TL_i\right) = D^{\mathrm{offload}}\left(TL_i\right) + D^{\mathrm{comp}}\left(TL_i\right) + \max_{TL_x \in SUC\left(TL_i\right) \text{ and } L\left(TL_x\right)=L\left(TL_i\right)+1} \left(\frac{w_{ix}}{B^{\mathrm{StoS}}} + \mathrm{Prio}\left(TL_x\right)\right), \tag{21}$$

where $SUC(TL_i)$ represents the set of subsequent task slices of TL_i, which have direct data dependence between TL_x and TL_i, $\mathrm{Prio}(TL_x)$ is the task priority of TL_x, and B^{StoS} represents the channel bandwidth between edge servers.

pr_i in the optimal choreography $\mathrm{Ord}(\mathbb{S})$ is the optimal start time of the transmission corresponding to TL_i; in order to ensure that each task slice has been offloaded before it is executed on the edge server, $pr_i \leq \mathrm{Prio}(TL_i)$. The task choreography method based on overlapping longest path can solve the best offload time for each task slice in T, so as to maximize the parallelization of the transmission and execution of subtasks, to realize the optimization goal in equation (1).

As shown in Figure 5, the choreography of subtasks is as follows: (1) Firstly, a static sort is done. For example, t_1 is executed in the first sequence, t_2 and t_3 are executed in the second sequence, and subtasks t_4, t_5, t_6, t_7, and t_8 are executed in the third sequence. (2) After that, all the subtasks are choreographed according to the earliest start time. It can be seen that the earliest start time of t_7 and t_8 is earlier than that of t_4, t_5, and t_6, so, in the third execution sequence, t_7 and t_8 should be executed earlier than t_4, t_5, and t_6. (3) Finally, the above two sequences are dynamically sorted according to the following rules: the nodes with more children take precedence or the nodes whose child nodes have high computation time take precedence. In Figure 5, t_4 and t_5 have more child nodes, and the computation delays of their child nodes are longer, so t_4 and t_5 have priority over t_6, t_7, and t_8.

We design a heuristic algorithm to solve the subtask choreography scheme (Algorithm 2).

The input of Algorithm 2 is the output of Algorithm 1. It uses dynamic iterative optimization to solve the optimal choreography scheme $\mathrm{Ord}(\mathbb{S})$, namely, the offloading

```
(1) function [S] = choreography (G, S, En)
(2) Input: G//Task workflow
(3)        S//Task slice scheme
(4)        En//Environmental parameters, including channel bandwidth, server processing capacity, etc
(5) output: Ord(S)
(6) CreTree(G, S)//Build the number of task slice levels
(7) Init(En)//Initialize the offload environment
(8) for i = 1:|S|
(9) T(i) = ExeTime(TL_i)//Calculate the earliest start time of the task slices
(10) KP = LogicP(S, T)//Find the longest path for the task slices to execute
(11) NKP = DelP(S, KP)//Get the task slices not in the longest path
(12) Cons = priority(S, ExeTime)//Obtain scheme constraint
(13) Sord = rand (Popsize, Cons, S)//Program population size
(14) while (k ≤ maxnum)
(15) for i = 1: Popsize
(16)        F(i) = fitness (G, Sord)//Set the optimization target of the heuristic algorithm
(17) [globlalMinT, ordi] = min(F)
(18) for i = 1: Popsize
(19)        Sord = IteV (Cons, KP, NKP)//Optimize the population of the choreography scheme under the constraints
(20) [BestMinT, bestord] = min (global, ordi)
(21) Ord(S) = BestMinT
```

Algorithm 2: Task choreography based on the longest overlapping path.

sequence and timing of each task slice in S. In this algorithm, lines 6–12 are used to calculate all kinds of time delays during the offloading, computation time of each task slice, and the computation time constraints of each task slice according to equations (19)–(21). In lines 13–21, a heuristic optimization algorithm is used to find the task choreographing scheme that minimizes the overall computation delay of the task workflow under the execution constraints. The combination of Algorithms 1 and 2 can achieve the optimal task slicing and choreographing scheme under the condition of a specific number of edge servers. In the case of sufficient resources of the MEC system, we can traverse from 1 to k (number of T subtasks) to find the optimal number of task slices, that is, the optimal number of edge servers for distributed deployment of the whole task workflow.

The computational complexity of an algorithm is determined by the number of basic operations when the input size is N. The SSCS algorithm proposed in this paper is a heuristic algorithm based on discrete optimization. The algorithm consists of generating the initial solutions, generating neighborhoods, judging the infeasible task scheduling list, and removing the infeasible task scheduling list. Since the generation of the initial solutions is constrained by the earliest start time of a task, the computational complexity of the operation is determined by the horizontal clustering result of the task workflow, that is, $O(N\log N)$. Similarly, the computational complexity of generating neighborhoods is determined by the maximum number of parallel tasks at each level, and the size does not exceed $O(C^2_{\log N})$. The computational complexity of judging the infeasible task scheduling list and removing the infeasible task scheduling list is $N^* L$, where L is the length of the infeasible task list. So, the overall complexity of the SSCS algorithm is $O(\text{Max\_Gen}^* N\log N)$, where Max_Gen is the maximum number of iterations.

5. Experiment and Analysis

In order to verify the effectiveness of the task slicing and choreographing method proposed in this paper, we set up two groups of simulation experiments. In the experiment, we first generated the complex task workflow of mobile devices covering the three data dependencies and assigned corresponding parameters to each subtask and MEC environment. The setting range of main parameters is shown in Table 1.

5.1. Verification Experiments of Task Slicing Method. Firstly, in order to prove the advantages of task slicing method, we use HPD (Hierarchical Process Decentralization) algorithm and HIPD (Hierarchical Intelligent Process Decentralization) in papers [36, 37] to generate subtask slicing and then conduct comparative experiments with our algorithm. The HPD algorithm applies breadth-first search/traversal algorithm to find the most relevant, closely related, and parallel activities in the workflow view and then encapsulates the closely related activities in the same broker to reduce the need for interbroker messaging. The HIPD algorithm combines HPD and a frequent path mining algorithm together. In this experiment, the task workflow to be offloaded by the mobile device and the data dependency relationship between subtasks are shown in Figure 5. We use these two methods to generate task slices of different granularity and compare the results with the slicing results of our algorithm. The partitioning results of the three slicing methods for the task workflow are shown in Table 2.

When the task is offloaded to the server, this group compares the advantages and disadvantages of the results of different task slices from three aspects: the overall computation delay of the task workflow, the load of the edge servers

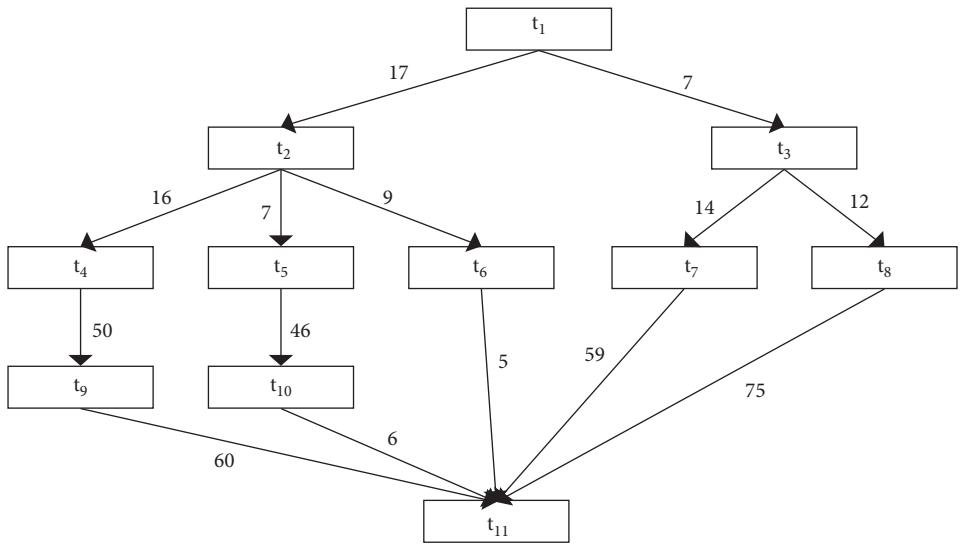


Figure 5: An example of dynamic choreography of tasks.

Table 1: Setting of main experimental parameters.

Parameter name	Parameter value range	Parameter meaning
ps_i $(i \in [1, k])$	1–100 M	The size of the task load
w_{ij} $(i, j \in [1, k])$	1–100 M	The amount of data transferred between tasks
Mpc	10^2 cpu cycles $\cdot$ M^{-1}	CPU cycles per M data to process
C^{Edge}	10^3 cpu cycles $\cdot$ s^{-1}	CPU cycles per second which the edge server can process
B^{StoS}, B^{DtoS}	$10\,\text{M} \cdot \text{s}^{-1}$, $8\,\text{M} \cdot \text{s}^{-1}$	Transfer bandwidth between edge servers and from mobile devices to edge services

Table 2: Slice results for the task workflow by different methods.

Task slicing method	Task slice results
HPD0	$\{t_1, t_2, t_3, t_4, t_5, t_6, t_7, t_8, t_9, t_{10}, t_{11}\}$
HPD1	$\{t_1\}\{t_2, t_4, t_5, t_6, t_9, t_{10}, t_{11}\}\{t_3, t_7, t_8\}$
HPD2	$\{t_1, t_2, t_5, t_8, t_{10}, t_{11}\}\{t_4, t_7, t_9\}\{t_3, t_6\}$
HPD3	$\{t_1, t_2, t_5, t_8, t_{11}\}\{t_4, t_7, t_{10}\}\{t_3, t_6, t_9\}$
HIPD0	$\{t_1, t_2, t_3, t_5, t_6, t_7, t_8, t_{10}, t_{11}\}\{t_4\}\{t_9\}$
HIPD1	$\{t_1, t_9\}\{t_2, t_3, t_5, t_6, t_7, t_8, t_{10}, t_{11}\}\{t_4\}$
HIPD2	$\{t_1, t_4, t_7\}\{t_2, t_5, t_8, t_{10}\}\{t_3, t_6, t_9, t_{11}\}$
HIPD3	$\{t_1, t_2, t_5, t_8, t_{11}\}\{t_4, t_7, t_{10}\}\{t_3, t_6, t_9\}$
SSCS	$\{t_1, t_3, t_6, t_7\}\{t_2, t_5, t_8, t_{10}, t_{11}\}\{t_4, t_9\}$

during the execution process of the task, and the idle time of the edge server. In order to better prove the stability and applicability of our algorithm, in the experiments, we assigned two load schemes under different conditions to the subtasks in Figure 5. In the first case, all the subtasks are executed only once. The experimental comparison results are shown in Figures 6(a)–6(c). In the second case, all the subtasks are executed many times, and some subtasks may not be executed. The results of experimental comparison are shown in Figures 6(d)–6(f). Figures 6(a)–6(c) show the overall workflow computation delay corresponding to different slice results in Table 2, the average data transfer load between different edge servers, and the average idle time of the server itself during the execution process. It can be seen from these three figures that, compared with other methods, the SSCS method has the shortest total delay of task workflow, lower average data transmission between subtasks introduced by task distribution deployment, and the highest server utilization rate. Different task slicing schemes obtained by HPD and HIPD methods are effective and shorten the overall delay of task workflow, but they usually introduce a large amount of task transmission. At the same time, they are not superior to our method in terms of server utilization.

For example, the traffic volume introduced by HPD1, HIPD0, and HIPD1 methods is much more than that of our method. For the HPD2 and HIDP2 algorithms, their goal is to slice tasks more evenly across complex workflows. Although in the long run they can make the load distributed among the different server edges more balanced, in most cases, the total delay of the tasks they generate is higher than that of our method. Figures 6(d)–6(f) show that the SSCS method is applicable to different task loads of different sizes. In this case, the slice deployment using the SSCS method can obtain the lowest task computation time delay. It also has good load balance and server utilization. This lays a good foundation for the choreography of subtasks.

5.2. Verification Experiments of Choreography Method. In this group of experiments, we verified the effect of our choreography method from the overall feedback delay of the task workflow, as well as the number and time of completions of the subtasks on the edge servers at each time period.

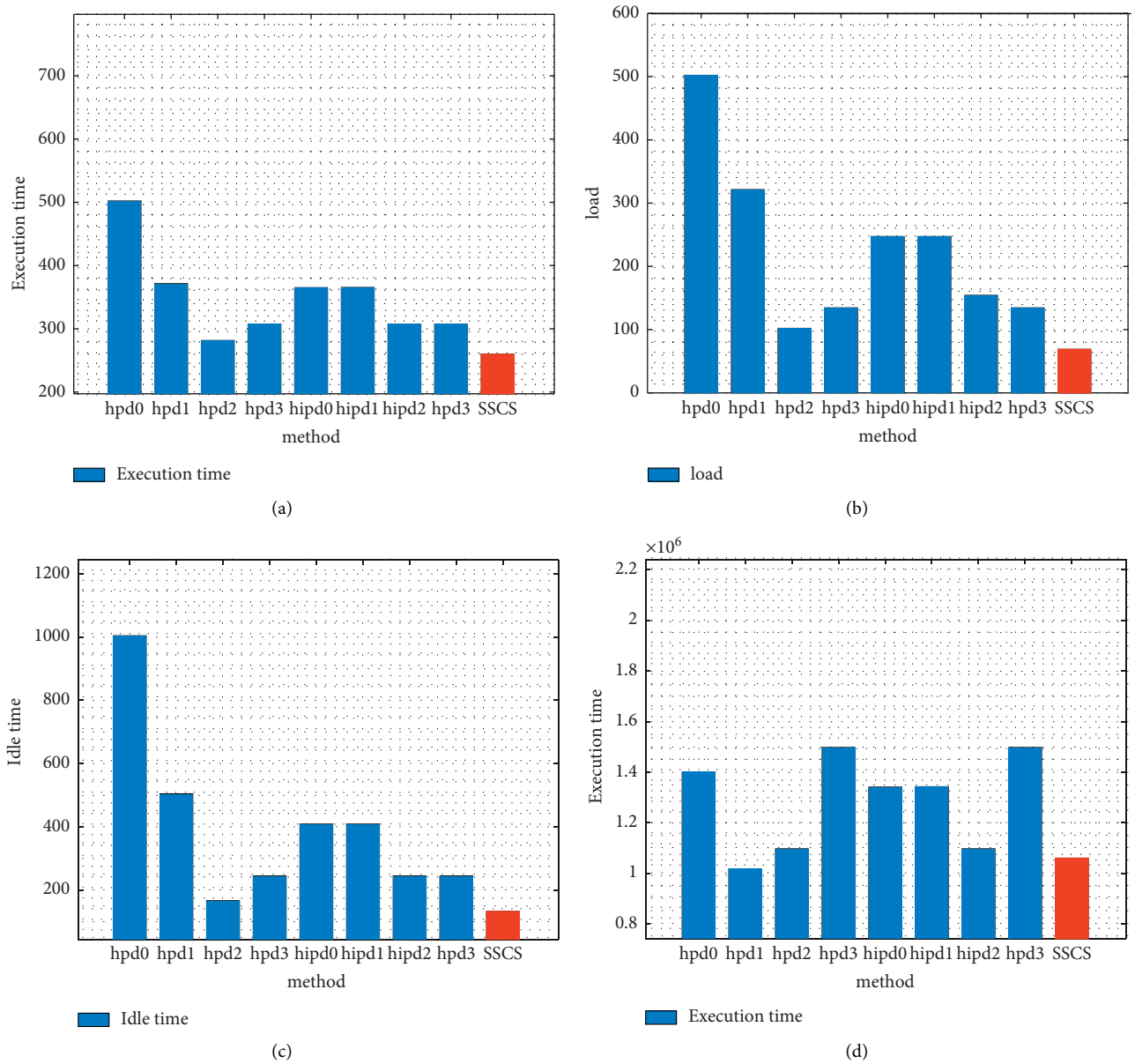


Figure 6: Continued.

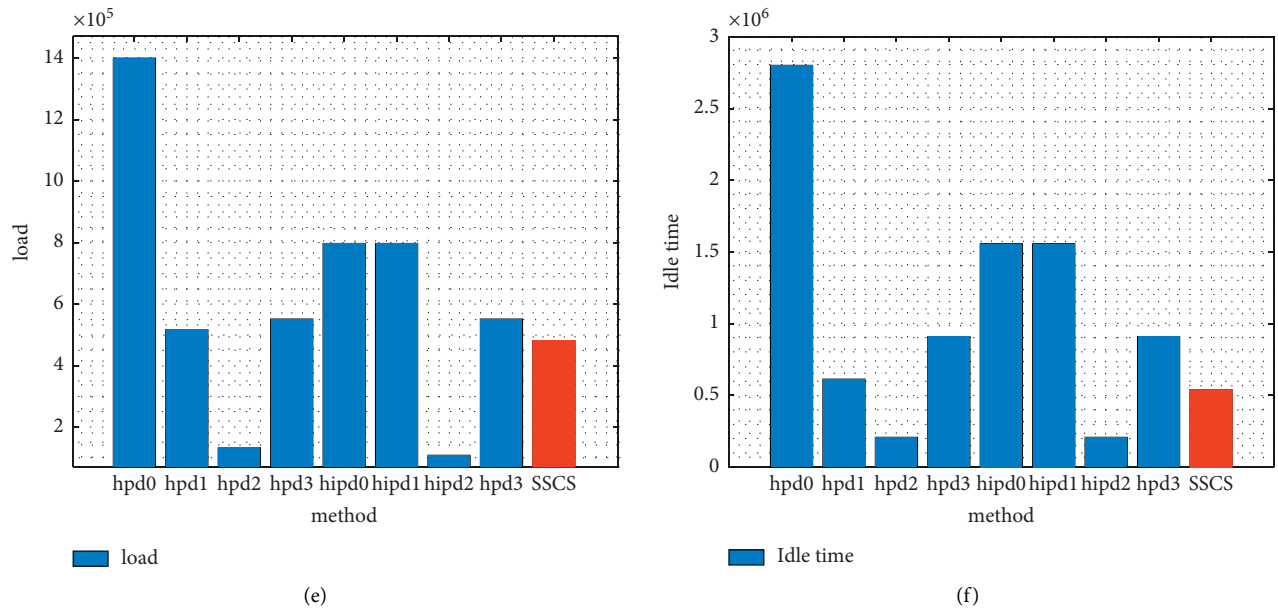


Figure 6: Comparison experiments based on the results of task slices in Table 1 under different task loads. (a) Comparison of execution time of edge server of each method. (b) Load balancing. (c) Edge server free time comparison. (d) Execution time of the task workflow after 10000 executions. (e) Load balance (10000 executions). (f) Idle time on the edge server (10000 executions).

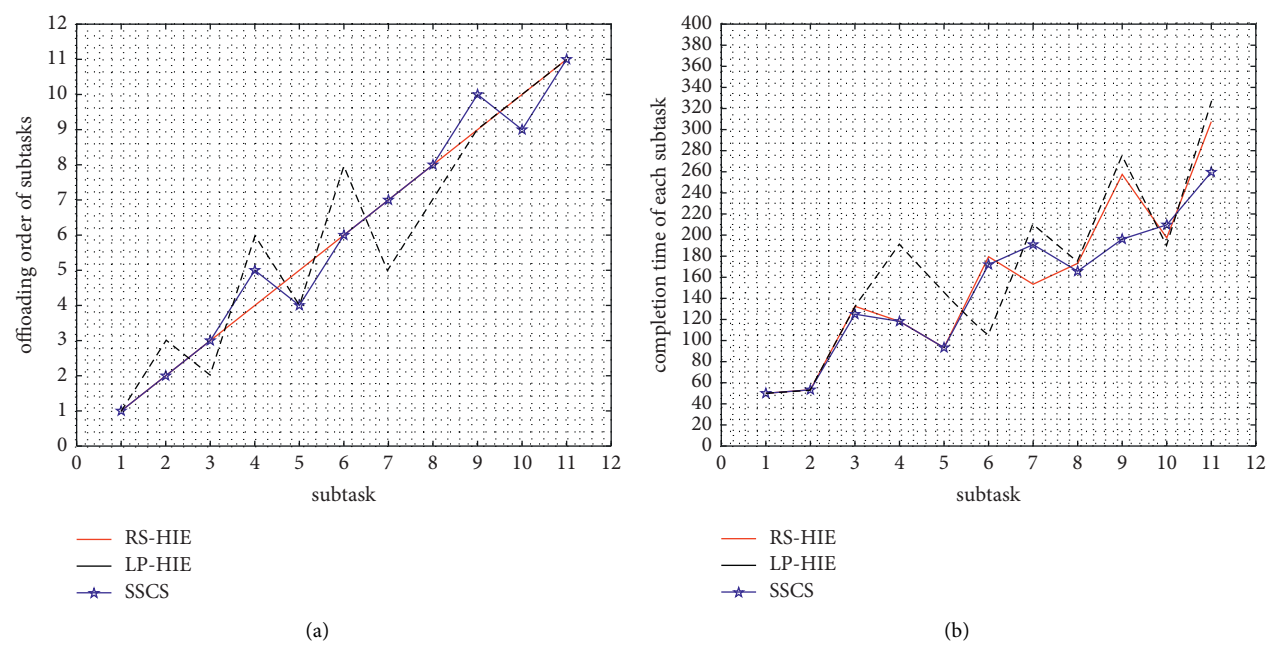
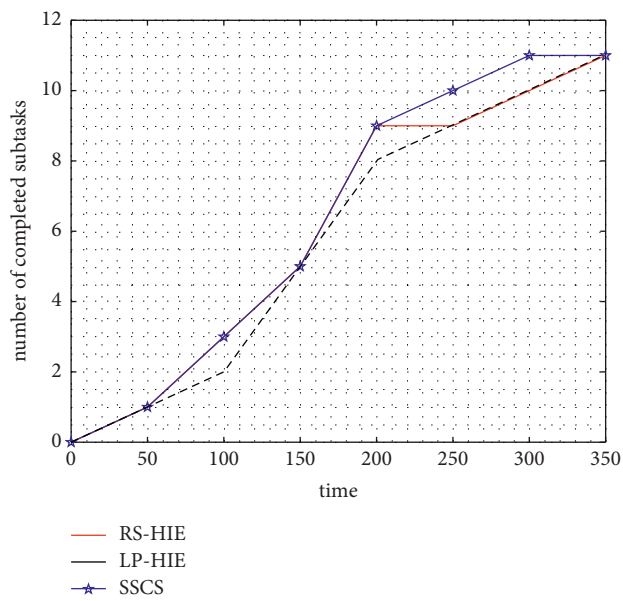


Figure 7: Continued.



(c)

Figure 7: Comparison experiments based on different task choreography methods.

We choose two common task offloading sorting methods. One is according to the hierarchical structure tree of the task workflow; the lower levels of subtasks are first offloaded and the same levels of subtasks are randomly ordered, which is called random sorting based on hierarchy (RS-HIE). The other is similar idea but for the same levels of subtasks which are ordered according to the size of the load, called load prioritization based on hierarchy (LP-HIE). The three offloading sequencing methods are completed based on the task slice results of our method. The experimental results are shown in Figure 7. Figure 7(a) corresponds to the offloading order of subtasks solved by the three methods. It can be seen that the offloading order of subtasks generated by different methods is greatly different. Figures 7(b) and 7(c) show the impact of different offloading sequences on the completion time of subtasks on the edge servers. Figure 7(b) shows the completion time of each subtask in the task workflow on the edge servers, and Figure 7(c) shows the number of subtasks that have been completed on the edge servers in each period. As you can see, the SSCS method can offload the most subtasks per unit time and has the shortest total computation time. It gives full consideration to various key time points affecting the task workflow when solving the choreography scheme, reasonably arranges the offloading time of each subtask, and avoids the total delay caused by waiting for the task to be offloaded as much as possible. Some critical tasks that affect the overall latency of the workflow are offloaded first.

6. Conclusion

To resolve offloading problem of complex tasks in the MEC system, we study the distributed deployment strategy and efficient offloading method of the task workflow. Considering the characteristics of data distribution and logic relationships between subtasks in the process of task offloading and execution, a slicing method of task distribution and deployment is proposed based on similarity between subtasks, and a choreographing method of task offloading sequence is proposed based on the longest overlapping path. Finally, aiming at minimizing the overall delay of task workflow, a heuristic algorithm is designed to solve the approximate optimal solution of slicing and choreographing. Simulation experiments compare our method with other commonly used slicing and choreography methods and prove the effectiveness and advantages of our method in solving task offloading of complex tasks from various angles.

Data Availability

The data involved in this paper include migration algorithm code and simulation data (generated by the simulation algorithm). For copyright protection purposes, data access currently requires contact with institutional authors. If data are needed, the authors should be e-mailed at 55234043@qq.com.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grants 62162047 and 62162046; Natural Science Foundation of Inner Mongolia under Grants 2019ZD15 and 2019MS06029; Inner Mongolia Science and Technology Plan Project under Grants 2021GG0155 and 2019GG372; the Self-topic of Engineering Research Center of Ecological Big Data, Ministry of

Education; the Open-topic of Inner Mongolia Big Data Laboratory for Discipline Inspection and Supervision (IMDBD2020012 and IMDBD2021014); Inner Mongolia Engineering Laboratory for Cloud Computing and Service Software; Inner Mongolia Key Laboratory of Social Computing and Data Processing; and Inner Mongolia Engineering Lab of Big Data Analysis Technology.

References

[1] Y. Zhao, L. Jia, Y. Zhang, Y. Song, and Y. Tian, "Ordinal multi-task part segmentation with recurrent prior generation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 5, pp. 1636–1648, 2019.

[2] M. Fayyaz, B. Cao, W. Almughalles, Y. Li, and L. Ali, "Optimizing task execution for mobile edge computing," in *Proceedings of the 2019 The 8th International Conference on Network, Communication and Computing*, pp. 13065–13076, Luoyang, China, December 2019.

[3] L. Huang, X. Feng, L. Zhang, L. Qian, and Y. Wu, "Multi-server multi-user multi-task computation offloading for mobile edge computing networks," *Sensors*, vol. 19, no. 6, p. 1446, 2019.

[4] T. Mahn, H. Al-Shatri, and A. Klein, "Distributed algorithm for energy efficient joint cloud and edge computing with splittable tasks," in *Proceedings of the IEEE Conference on Wireless Communications and Networking*, pp. 167–182, Marrakesh, Morocco, April 2019.

[5] K. Li, "Optimal task dispatching on multiple heterogeneous multiserver systems with dynamic speed and power management," *IEEE Transactions on Sustainable Computing*, vol. 2, no. 2, pp. 167–182, 2017.

[6] H. Liu, J. Pu, L. T. Yang et al., "A holistic optimization framework for mobile cloud task scheduling," *IEEE Transactions on Sustainable Computing*, vol. 4, no. 2, pp. 217–230, 2017.

[7] P. Wang, Z. Zheng, B. Di, and L. Song, "Joint task assignment and resource allocation in the heterogeneous multi-layer mobile edge computing networks," in *Proceedings of the 2019 IEEE Global Communications Conference (GLOBECOM)*, vol. 69, no. 3, pp. 217–230, IEEE, Waikoloa, HI, USA, December 2019.

[8] K. Li, "Energy-efficient task scheduling on multiple heterogeneous computers: algorithms, analysis, and performance evaluation," *IEEE Transactions on Sustainable Computing*, vol. 1, no. 1, pp. 7–19, 2016.

[9] A. Mukherjee, S. Misra, V. S. P. Chandra, and N. S. Raghuwanshi, "ECoR: energy-aware collaborative routing for task offload in sustainable UAV swarms," *IEEE Transactions On Sustainable Computing*, vol. 5, no. 4, pp. 514–525, 2020.

[10] H. Xing, L. Liu, J. Xu, and A. Nallanathan, "Joint task assignment and wireless resource allocation for cooperative mobile-edge computing," in *Proceedings of the 2018 IEEE International Conference on Communications*, vol. 20, no. 1, pp. 360–374, Kansas City, MO, USA, May 2018.

[11] K. A. Noghani, H. Ghazzai, and A. Kassler, "A generic framework for task offloading in mmWave MEC backhaul networks," in *Proceedings of the 2018 IEEE Global Communications Conference (GLOBECOM)*, Abu Dhabi, UAE, December 2018.

[12] Y. Sun, T. Wei, H. Li, Y. Zhang, and W. Wu, "Energy-efficient multimedia task assignment and computing offloading for mobile edge computing networks," *IEEE Access*, vol. 8, pp. 36702–36713, 2020.

[13] M. Tajallifar, S. Ebrahimi, M. R. Javan, N. Mokari, and L. Chiaraviglio, "Energy-efficient task offloading under E2E latency constraints," *IEEE Transactions on Green Communications & Networking*, 2021.

[14] X. Lyu, T. Hui, C. Sengul, and Z. Ping, "Multiuser joint task offloading and resource optimization in proximate clouds," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 4, pp. 3435–3447, 2016.

[15] T. X. Tran and D. Pompili, "Joint task offloading and resource allocation for multi-server mobile-edge computing networks," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 1, pp. 856–868, 2018.

[16] T. Yang, R. Chai, and L. Zhang, "Latency optimization-based joint task offloading and scheduling for multi-user MEC system," in *Proceedings of the 2020 29th Wireless and Optical Communications Conference (WOCC)*, Newark, NJ, USA, May 2020.

[17] J. Xue, Y. An, and Y. An, "Joint task offloading and resource allocation for multi-task multi-server NOMA-MEC networks," *IEEE Access*, vol. 9, pp. 16152–16163, 2021.

[18] Y. Zeng, W. Chen, Z. Tang, and J. Wu, "Joint proportional task offloading and resource allocation for MEC in ultra-dense networks with improved whale optimization algorithm," *Journal of Physics: Conference Series*, vol. 1646, no. 1, 2020.

[19] L. Huang, X. Feng, C. Zhang, L. Qian, and Y. Wu, "Deep reinforcement learning-based joint task offloading and bandwidth allocation for multi-user mobile edge computing," *Digital Communications and Networks*, vol. 5, no. 1, pp. 10–17, 2019.

[20] P. A. Apostolopoulos, E. E. Tsiropoulou, and S. Papavassiliou, "Risk-aware data offloading in multi-server multi-access edge computing environment," *IEEE/ACM Transactions on Networking*, vol. 28, no. 3, pp. 1405–1418, 2020.

[21] M. Anedda, C. Desogus, M. Murroni, D. D. Giusto, and G. Muntean, "An energy-efficient solution for multi-hop communications in low power wide area networks," in *Proceedings of the 2018 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*, pp. 1–5, Valencia, Spain, June 2018.

[22] S. Cheng, Z. Chen, J. Li, and H. Gao, "Task assignment algorithms in data shared mobile edge computing systems," in *Proceedings of the 2019 IEEE 39th International Conference on Distributed Computing Systems (ICDCS)*, Dallas, TX, USA, July 2019.

[23] D. Nan, N. S. Hang, X. U. Li, and G. Z. Tan, "Data related task scheduling for vehicular ad hoc networks," *Chinese Journal of Computers*, vol. 40, no. 7, pp. 1614–1625, 2017.

[24] U. Saleem, Y. Liu, S. Jangsher, Y. Li, and T. Jiang, "Mobility-aware joint task scheduling and resource allocation for cooperative mobile edge computing," *IEEE Transactions on Wireless Communications*, vol. 20, no. 1, pp. 360–374, 2021.

[25] Z. Wang, Z. Zhao, G. Min, X. Huang, Q. Ni, and R. Wang, "User mobility aware task assignment for mobile edge computing," *Future Generation Computer Systems*, vol. 85, pp. 1–8, 2018.

[26] C. Zhu, J. Tao, G. Pastor et al., "Folo: latency and quality optimized task allocation in vehicular fog computing," *IEEE Internet of Things Journal*, vol. 6, no. 3, pp. 4150–4161, 2019.

[27] C. Yang, Y. Liu, X. Chen, W. Zhong, and S. Xie, "Efficient mobility-aware task offloading for vehicular edge computing networks," *IEEE Access*, vol. 7, pp. 26652–26664, 2019.

[28] X. Wang, Y. Cui, Z. Liu, J. Guo, and M. Yang, "Optimal resource allocation for multi-user MEC with arbitrary task arrival times and deadlines," in *Proceedings of the ICC 2019-2019 IEEE International Conference on Communications (ICC)*, Shanghai, China, May 2019.

[29] M. Emara, H. ElSawy, M. C. Filippou, and G. Bauch, "Spatio-temporal dependable task execution services in MEC-enabled wireless systems," *IEEE Wireless Communications Letters*, vol. 10, no. 2, pp. 211–215, 2021.

[30] X. Long, J. Wu, Y. Wu, and L. Chen, "Task merging and scheduling for parallel deep learning applications in mobile edge computing," in *Proceedings of the 2019 20th International Conference on Parallel and Distributed Computing, Applications and Technologies (PDCAT)*, Gold Coast, Australia, December 2019.

[31] A. Mamat, Y. Lu, J. Deogun, and S. Goddard, "Efficient real-time divisible load scheduling," *Journal of Parallel and Distributed Computing*, vol. 72, no. 12, pp. 1603–1616, 2012.

[32] J. Liu, Y. Mao, J. Zhang, and K. B. Letaief, "Delay-optimal computation task scheduling for mobile-edge computing systems," 2016, https://arxiv.org/abs/1604.07525.

[33] G. Wu, Y. Li, J. Ren, and C. Lin, "Partitioned fixed-priority real-time scheduling based on dependent task-split on multicore platform," in *Proceedings of the 2013 12th IEEE International Conference on Trust, Security and Privacy in Computing and Communications*, Melbourne, Australia, May 2013.

[34] Z. Liu, X. Yang, and J. Shen, "Optimization of multitask parallel mobile edge computing strategy based on deep learning architecture," *Design Automation for Embedded Systems*, vol. 24, no. 3, pp. 129–143, 2020.

[35] Z. Xie, X. Song, and S. Xu, "Peer-to-peer enhanced task scheduling for D2D enabled MEC network," *IEEE Access*, vol. 8, pp. 138236–138250, 2020.

[36] F. Safi Esfahani, M. A. A. Murad, M. N. Sulaiman, and N. I. Udzir, "Using process mining to business process distribution," in *Proceedings of the 2009 ACM Symposium on Applied Computing*, pp. 2140–2145, Honolulu, HI, USA, March 2009.

[37] F. Safi Esfahani, M. A. A. Murad, M. N. B. Sulaiman, and N. I. Udzir, "Adaptable decentralized service oriented architecture," *Journal of Systems and Software*, vol. 84, pp. 159l–1617, 2011.

Hindawi
Wireless Communications and Mobile Computing
Volume 2022, Article ID 1500646, 16 pages
https://doi.org/10.1155/2022/1500646

Research Article

Mutigroup-Based Phasmatodea Population Evolution Algorithm with Mutistrategy for IoT Electric Bus Scheduling

Yunxiang Zhu,[1] Fengting Yan,[1] Jeng-Shyang Pan,[2] Lei Yu,[1] Yuanfei Bai,[1] Weigang Wang,[3] Chunxia He,[4] and Zhicai Shi[1,5]

[1]*Shanghai University of Engineering Science, Shanghai 201620, China*
[2]*Shandong University of Science and Technology, Qingdao, Shandong 266590, China*
[3]*Shanghai Municipal Engineering Design and Research Institute (Group) Co., Ltd., Shanghai 200092, China*
[4]*Shanghai Urban Construction Vocational College, Shanghai 201999, China*
[5]*Shanghai Key Laboratory of Integrated Administration Technologies for Information Security, Shanghai 200240, China*

Correspondence should be addressed to Fengting Yan; yanfengting2008@163.com

Received 29 October 2021; Revised 7 December 2021; Accepted 4 January 2022; Published 31 January 2022

Academic Editor: Feng-Jang Hwang

The Phasmatodea population evolution algorithm (PPE) is a novel metaheuristic algorithm proposed in recent years, which simulates the evolutionary trend of stick insect population. In this article, a multigroup-based Phasmatodea population evolution algorithm with mutistrategy (MPPE) is proposed to further improve the overall performance of PPE. During the initialization period, the stick insect population is divided into multiple groups, and the step factor of the flower pollen algorithm is introduced into the population growth model of no more than half of the groups. This makes the population evolution trend diversified and prevents the algorithm from falling into the local optimal solution to a certain extent. In terms of intergroup communication, two communication strategies are adopted to mutate and replace the inferior particles, respectively, which improves the convergence speed and search ability of the algorithm. In the MPPE performance test, we compared it with PPE, five standard algorithms, and other parallel algorithms in CEC 2013 Benchmark Suite. Finally, this algorithm is applied to the IoT based electric bus scheduling for urban waterlogging situation, and the excellent performance of MPPE is verified comprehensively.

1. Introduction

Metaheuristic algorithms are inspired by relevant experiences, behaviors, rules, and mechanisms in the fields of physics, chemistry, biology, society, and art [1]. It is a method based on computational intelligence to solve complex optimization problems with the best or satisfactory solution [2–5]. Because it does not depend on specific problems, metaheuristic algorithms can be widely used in industrial production, economic life, military, and other fields [6–9].

Metaheuristic algorithms can be roughly divided into three categories according to their sources. The first category is derived from human theory. In recent years, the gaining-sharing knowledge based (GSK) algorithm [10] and the volleyball premier league (VPL) algorithm [11] have been proposed, which simulate the development and activities of human life. The second category comes from biological behavior. Among the novel algorithms are island artificial bee colony (iABC) algorithm [12], monarch butterfly optimization (MBO) [13], moth search algorithm (MSA) [14], Harris hawks optimization (HHO) [15], and so on. The third category is inspired by physical phenomena. Such novel algorithms include thermal exchange optimization (TEO) [16], electrosearch (ES) algorithm [17], and lightning attachment procedure optimization (LAPO) [18]. Since people have obtained a lot of inspiration from the infinite nature, more and more nature-inspired algorithms have been proposed.

The Phasmatodea population evolution (PPE) algorithm is a new metaheuristic algorithm proposed in 2020 [19]. Song et al. [20] simplified Phasmatodea population

evolution algorithm and applied it to engineering optimization problems. This algorithm is inspired by the evolution trend of stick insect population in the natural environment. The characteristic of stick insect population growth is that it can adapt to a dynamic environment and change autonomously. In the initial stage, there is a favourable growth environment for the population, so it grows exponentially. When the population grows to a certain extent, due to the interference of environmental factors, its growth rate tends to slow down and eventually stops growing.

The advantage of the PPE algorithm is that it has relatively good exploratory ability and stability. In terms of application, it is suitable for various engineering optimization problems. So, this algorithm also has a good generalization ability. But it still has some disadvantages. Its evolutionary mechanism will lead to the generation of some poor particles during the early period, and it converges slowly. When solving some problems, its solution tends to be locally optimal rather than globally optimal.

To address the shortcomings of the PPE algorithm, we attempt to use parallel method to improve [21, 22]. Predecessors have done a lot of work in parallel optimization algorithm research over the years, such as parallel particle swarm optimization (PPSO) [23], parallel genetic algorithm (PGA) [24], parallel ant colony algorithm (PACO) [25–27], and parallel differential evolution algorithm (PDE) [28]. In this study, we introduce multiple strategies into the parallel approach. First, the stick insect population is divided into several groups, all of which evolve simultaneously. Secondly, the step factor of the flower pollen algorithm is introduced into the population growth model of some groups, and its evolution trend is changed to some extent. The diversity of population evolution mechanism makes the algorithm more capable of global exploration. Finally, we adopt two strategies for intergroup communication to improve the overall performance of PPE. The first strategy is to mutate the particles with poor fitness during the early stage to speed up the convergence of PPE. The second strategy is to randomly replace the best particle with other random particles in all the groups during the late period. It is intended to improve the global exploration accuracy of the algorithm. We made the above improvements to PPE and named it MPPE. In the comparative experiment of CEC 2013 Benchmark Suite, we verify the excellent performance of MPPE algorithm.

Public traffic optimization is a hot issue in recent years [29–31]. With the rapid development of new energy technology, electric buses have gradually replaced fuel buses in short-distance transportation and have been popularized in many cities. This change promotes the wide application of intelligent transportation combined on Internet of Things (IoT) technology [32], in which the electric bus scheduling problem has become the focus of many scholars [33–35]. Bus scheduling is a multiobjective combination optimization problem, which has a recognized NP-hard problem. The purpose of this problem is to find an optimal solution in a complex objective function, that is, the best scheduling scheme. The metaheuristic algorithm can effectively calculate the optimal solution in a limited time and space and is not restricted by the problem itself. So, it is suitable for solving this kind of large-scale optimization problem. In recent years, more and more researchers have applied metaheuristic algorithms to bus scheduling [36–38], which proves their effectiveness in solving such problems.

However, traditional metaheuristic algorithms often have some limitations for the solution results of the bus scheduling optimization. Novel metaheuristic algorithms are often more suitable for solving such complex optimization problems. In order to verify the effectiveness of the improved PPE algorithm in solving practical applications, we apply the MPPE algorithm to this hot problem for the first time. While solving this problem, we also successfully tried a new method, not limited to classic algorithms. This is to open up new ideas for scholars doing this kind of research. In this article, in addition to comparing MPPE with several original algorithms, we also compare with the recently proposed algorithms that have been used to solve the bus scheduling problem. The final experimental results prove that MPPE we proposed can achieve good results in solving this problem.

The main contributions of this article are as follows:

(1) A parallel method is proposed with a two-stage communication strategy

(2) The step size factor of the flower pollen algorithm was introduced into the growth of the stick insect population to achieve the diversity of evolutionary trends

(3) MPPE algorithm has been successfully implemented for the optimization of the IoT electric bus scheduling on rainy days in city, and the result is satisfactory

The remaining sections of the article are organized as follows. Section 2 introduces the relevant formulas of the IoT electric bus scheduling problem and basic principle of PPE. In Section 3, a mutigroup-based Phasmatodea population evolution algorithm with mutistrategy (MPPE) is proposed. In Section 4, we compared MPPE with other algorithms on the benchmark suite and analyzed the performance of this algorithm. In Section 5, MPPE algorithm is applied to the IoT electric bus scheduling. Section 6 concludes the study.

2. Related Work

2.1. IoT Electric Bus Scheduling Problem. The Internet of Things based electric bus under the intelligent bus system is equipped with multisensor, high precision positioning, and wireless communication technology. This multitechnical fusion can detect the passenger flow and the consumption in the course of driving. In addition, passengers at the station can receive real-time bus scheduling information. This includes the estimated arrival time of next bus and the departure time interval for a certain period of time. When the same type of bus operates on the same section of the road, its total mileage and number of stations are fixed, the fare is uniform throughout the journey, and the time to reach each station during the journey is roughly equal. For these basic

conditions, we can do the following analysis of the bus scheduling problem in this case.

When only one-way operation of the electric bus is considered, the departure time of the first and last bus is fixed. Then, the day is divided into five time periods; each time period of passenger flow is different. Passenger flow is usually calculated by GPS positioning and IC card records. After the first bus leaves, the bus company should consider the departure interval of the next bus, which is different for each time period. The motivation of the IoT electric bus scheduling is to balance the interests of the company and passengers [39], which makes the setting of the departure interval fully consider the company's profit and the patience of passengers. This is a multiobjective minimization problem. Therefore, the usual approach is to apply a weight to the company benefit and passenger benefit coefficient, respectively. In this case, decision-makers can balance the relationship between these two weights according to the actual situation. The objective function model can be represented by the following equation:

$$F(\Delta t_k) = w_1 Q_1 D \sum_{k=1}^{K} \frac{T_k}{\Delta t_k} + w_2 Q_2 \sum_{k=1}^{K} \sum_{j=1}^{J} m_k \left(\frac{u_{kj} \Delta t_k^2}{2} \right), \quad (1)$$

where the first item of the equation represents the cost consumption of the company and the second is the loss factor of passengers. w_1 and w_2 represent two weights, and the sum of these two values is equal to 1. Q_1 is the cost of each bus per kilometer distance, including the electricity consumed, the loss of vehicle equipment, and the salary of bus drivers. Q_2 is the patience loss of a unit of time for each passenger waiting at the station. D is the total distance of one-way bus operation, divided into K time periods. k indicates the current time period, T_k is the length of the current time period, Δt_k is the departure interval of the kth time period, J is the total number of sites, j is the number of the current site, m_k is the total number of departures in the kth time period, and u_{kj} is the arrival rate of passengers at station j in the kth time period.

In solving a practical optimization problem, in addition to the objective function, the corresponding constraints should also be considered. The constraint of this bus scheduling problem is that the bus company must be in a profitable state. In other words, the company's operating income should be greater than the cost, to ensure that the company does not lose money. The equation for its constraint is

$$n \times \frac{\sum_{k=1}^{K} \sum_{j=1}^{J} u_{kJ}}{\sum_{k=1}^{K} (T_k / \Delta T_k) > Q_1 D}. \quad (2)$$

Using the above equation, if the turnover of bus company is below cost, an additional time is added to the bus departure interval as set out in the current system until the turnover is not less than the cost.

According to the above introduction, the complexity of this problem mainly depends on the number of sites and time periods. With a series of data such as cost and passenger flow, this nonlinear problem is dynamic and difficult to solve by conventional methods.

2.2. Phasmatodea Population Evolution Algorithm. The basic principle of the Phasmatodea population evolution algorithm is that the movement of the solution in multidimensional space simulates the evolutionary trend of stick insect population. In a dynamic environment, each stick insect population is self-determining and is assigned two properties: growth rate and size. The growth trend of population is similar to the logistic regression model. During the initialization phase, solutions x of N_p populations are randomly generated, and the size of each x is computed as follows:

$$p_i = \frac{1}{N_p}. \quad (3)$$

In addition, h historically optimal solutions were selected to lead the search of the other solutions, and the number of h is

$$h = \left[\log\left(N_p\right)\right] + 1. \quad (4)$$

These historically optimal solutions are stored in ho, $ho = [x_{k1}, \ldots, x_{ki}, \ldots, x_{kh}]$. The initialized generated N_p solutions select h local optimal values through ho and store their fitness values. These local optimal solutions will direct other solutions throughout the population.

During the iterative update phase, the evolutionary trend of the population is represented by evt, the position of the t-th evolution of the population is x^t, and then the position obtained by the next evolution is x^{t+1}. The position update equation is

$$x^{t+1} = x^t + evt. \quad (5)$$

After the position is updated, the global optimal values Gbest and h optimal solutions ho of the population are calculated. Evolutionary trend evt is divided into three parts. The first two parts use convergent evolutionary and path-dependent population growth models. If the quality of the updated solution is better than before, the first part of the update is adopted, and the update equation is

$$evt^{t+1} = \left(1 - p^{t+1}\right) A + p^{t+1}\left(evt^t + m\right), \quad (6)$$

where p^{t+1} represents the population size of the next iteration, and its update formula is

$$p^{t+1} = a^{t+1} p^t \left(1 - p^t\right), \quad (7)$$

where a represents growth rate, which is set to 1.1 at the beginning, and A indicates the degree of closest to the optimal value, calculated as follows:

$$A = \left(s\left(ho, x^t\right) - x^t\right) \cdot c, \quad (8)$$

where s (ho, x^t) is to find the closest historically optimal solution x^t in ho. c is a coefficient of the extent to which the most recent optimal solution affects the current population, ranging from 0 to 1. m represents the population mutation in a partial dimension.

If the quality of the solution after the position update is worse, the update method of Part 2 will be used instead of

Part 1. The second part is called convergence evolution, which tends to the most recent local optimal solution. The update equation is

$$evt^{t+1} = \text{rand} \cdot A + st \cdot N, \tag{9}$$

where N is the n-dimensional random factor generated by the normal distribution. The value of st is set to $0.1 \times (Ub - Lb)$.

The third part describes the model of population competition under environmental intervention. When populations x_i and x_j are competing with each other, the population size of x_i is

$$p_{\text{i}} = p_i + a_i p_i \left(1 - p_i - \frac{f(x_j)}{f(x_i)} p_j\right), \tag{10}$$

where x_j is randomly selected and different from x_i. Calculate the distance between x_i and x_j; then compare it to the set threshold TH. TH is calculated as follows:

$$\text{TH} = (Ub - Lb) \times 0.1 \times \frac{\text{MaxGen} + 1 - t}{\text{MaxGen}}, \tag{11}$$

where Ub and Lb represent upper and lower boundaries. MaxGen represents the maximum number of iterations. t is the current iteration. If the distance between x_i and x_j is less than TH, the two populations begin to compete each other, and the evolutionary trend update equation is

$$evt^{t+1} = evt^{t+1} + \frac{f(x_j) - f(x_i)}{f(x_j)} (x_j - x_i). \tag{12}$$

Among them, if the population is too small or it is growing abnormally, the solution will be replaced by a new population.

According to the formulas above, the pseudocode of PPE is shown in Figure 1.

```
Input: Np, d, MaxGen;
Output: Gbest and minimum fitness fmin
  Initialize Np population, p, evt, h using Eq. (3), (4);
  Initialize fitness f(x), ho and Gbest;
  for iter = 2 to MaxGen
    Update x to xnew using Eq. (5);
    Calculate f(xnew), update Gbest and ho;
    for n = 1 to Np
    if f(x)≥f(xnew) then
      x = xnew, update f(x);
      Update p_i using Eq. (7);
      Update evt using Eq. (6), (8);
    else if rand < p_i then
      x = xnew, calculate f(x);
      Update p_i using Eq. (7);
    end if
    Update evt using Eq. (9);
    end if
    select x_j randomly, (i ≠ j);
    if distance (x_i, x_j) < TH then
    Update p_i using Eq. (10), update evt using Eq. (12);
    end if
    end for
  end for
```

Figure 1: The pseudocode of PPE.

3. Proposed Method

In this study, we propose multigroup-based Phasmatodea population evolution algorithm with multiple strategies, including a two-stage parallel communication strategy and the introduction of step factor, which improve the PPE algorithm in many aspects.

3.1. Parallel Strategy. In the previous research on the improvement of optimization algorithm, the parallel method is often taken into account [40]. The parallel method can effectively improve the global exploration ability of the algorithm through population division. In this study, we divide stick insect populations into multiple groups, make each group evolve in parallel, and implement a two-stage intergroup communication strategy at each iteration. The first stage of communication is to accelerate the convergence of PPE by mutating the worst particles in each group to the best particles in some dimensions. The second stage is to improve the search accuracy of PPE by randomly replacing the best particles with random particles from each group. To balance the convergence speed and global exploration accuracy, we use the first strategy for the first half of the iterations and the second strategy for the second half.

In the first communication strategy, the particles with the worst fitness values in each group are mutated individually, and the mutation is close to the current global optimal particles. The equation for the mutation is as follows:

$$X_d = \text{totalbest}_d \times (m + n \times \text{rand}), \tag{13}$$

where X_d represents the position of the particle after the d-dimensional mutation, totalbest_d is the position of the current global optimal particle of the d-dimensional, rand is a random number of 0-1. m and n are parameters that control the degree to which the mutant particle approaches the global optimal particle. In general, $m, n \in (0, 1)$, and the sum of m and n is 1. In order to make mutant particles close to the current global optimal particles, the value of m should be set larger. In this article, we set m to 0.8 and n to 0.2. In this way, the positions of mutated particles are randomly distributed near the current global optimal position.

In the second communication strategy, some random particles are replaced with the current global optimal particles. The specific operation is to randomly generate indexes of the positions of some particles ranging from 1 to the total number of particles per group, and then replace these particles with the current global optimal particles in each group corresponding to these indexes. Finally, the fitness values of the replaced particles are updated.

Figure 2 shows the above introduction of proposed two-stage parallel communication strategy. This parallel strategy combines two different methods, which leads to closer communication between groups. Therefore, what we expect is to make the algorithm more stable in the process of search.

3.2. Step Factor of the Flower Pollen Algorithm. The flower pollen algorithm simulates the pollination of flowering plants [41–43], which is divided into global pollination and local pollination. This study was inspired by global pollination, which simulates the process by a flying insect carries pollen from one flowering plant over a distance to another. The equation of pollen position movement for global pollination is

$$x_i^{t+1} = x_i^t + \text{Levy} \times \left(x_i^t - g^t\right), \tag{14}$$

where g^t is the current global optimum. Levy represents the flight of flying insects in global pollination, which is calculated as follows:

$$\text{Levy} = \frac{\sin(\pi r/2) \times \lambda \Gamma(r)}{\pi \times s^{i+r}}, \tag{15}$$

where $\Gamma(r)$ is the gamma function, the distribution is valid for the large steps of $s > 0$. In this article, we set r to 1.5. As the position index i changes, so does the value of Levy, which indicates the degree of influence of global pollination can be seen as a step factor. In order to allow the population of stick insect to evolve with different growth trends and prevent algorithms from tending to locally optimal solutions when solving some optimization problems, we add the step factor Levy to the population growth model of stick insect in some groups. This method mainly changes the degree of influence of convergence evolution in (6), so that the original influence factor c was replaced by a changing step factor Levy. The equation after the change of A is

$$A = \left(s\left(ho, x^t\right) - x^t\right) \cdot \text{Levy} \cdot \gamma, \tag{16}$$

where γ indicates the scaling factor that controls step Levy. If the value of Levy is too large, it will inevitably break the evolutionary rule of the population, at which point the γ will narrow the Levy when it exceeds a certain range.

3.3. Multigroup-Based Phasmatodea Population Evolution Algorithm with Multistrategy. According to the introduction of the above two-stage communication strategy and step factor of the flower pollen algorithm, we propose the multigroup-based Phasmatodea population evolution algorithm with multistrategy and named it as MPPE. The flowchart for this method is shown in Figure 3. Figure 4 shows the pseudocode of MPPE. In the initialization stage, the stick insect population is divided into several groups, and each group has the same population size. Then, the initial solution and position of each group were generated according to the initialization method of PPE. In the update iteration stage, no more than half of the groups are introduced into the population growth model with step factor *Levy*, and the remaining groups are updated in the original method. After each group completes an iteration update at the same time, the two-stage intergroup communication strategy described above will be carried out according to the number of iterations. The solution obtained after each intergroup communication will be the current global optimal solution.

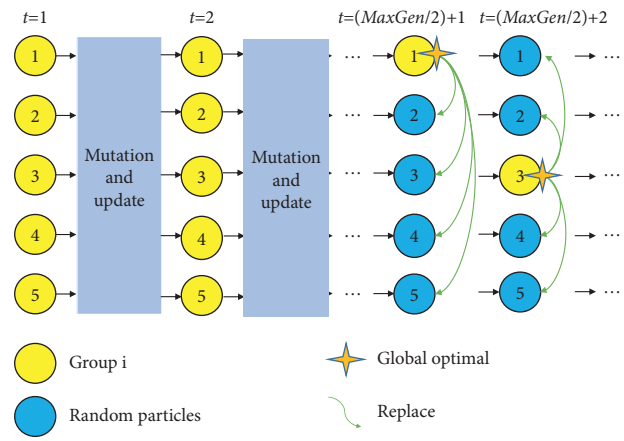


Figure 2: Two-stage communication strategy.

This research is based on a parallel method, and multiple strategies are implemented on the basis of multiple groups in parallel. The diversity of population evolution trends can be used to find the best particles through communication between groups. In general, the limitations of a single strategy cannot achieve the effect of algorithm improvement. Only the combination of multiply strategies can work together.

4. Experiment and Discussion

To test the optimization effect of MPPE, we conduct two comparative experiments. The first set of experiments is compared with PPE algorithm and five standard metaheuristic algorithms. The second is compared with other parallel metaheuristic algorithms. In this article, we select CEC 2013 Benchmark Suite to test the optimization problems of all algorithms. CEC 2013 Benchmark Suite is composed of 28 classic functions and has been widely used in experiments to test the performance of algorithms, where f_1~f_5 are single-mode functions, f_6~f_{20} are basic multimode functions, and f_{21}~f_{28} are combination functions. These optimization problems have a common search range of −100 to 100, with minimum values increasing from −1400 to 1400.

All experiments in this article are completed in MATLAB R2020b. The experimental equipment is a laptop computer with 2.60 GHz CPU, 16 g memory, and 64-bit Win10 system. Generally speaking, a larger population size and number of iterations can make the optimization result of the algorithm better. To ensure fair comparison and fully reflect the performance of these algorithms, we set the population of all algorithms to 100, the dimension to 30, and the number of function evaluations to 200,000. Each algorithm is independently experimented with 20 averages as an optimization result.

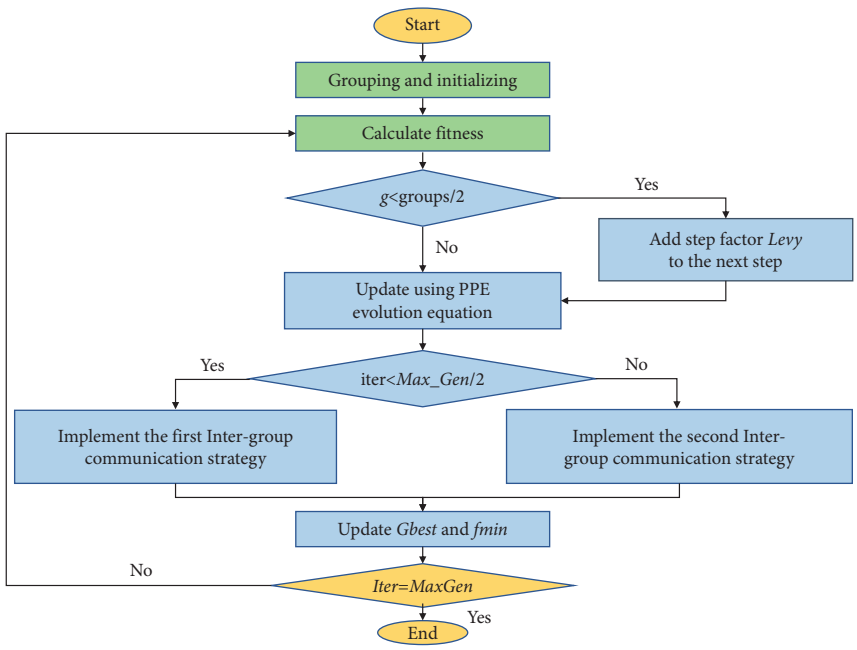


Figure 3: The entire phases of the proposed method.

These experiments mainly compare the exploration accuracy and convergence performance of algorithms and implemented appropriate nonparametric hypothesis tests [44, 45]. In order to further analyze the proposed algorithm, we discuss the time complexity of MPPE in the last part of this section.

4.1. Compared to Standard Algorithms. In this part, we compared MPPE with PPE, PSO [46], SCA [47], WOA [48], GWO [49], and GSK [10]. Table 1 lists the parameter settings for these algorithms. Table 2 shows the comparison results in 28 benchmark functions. It shows that MPPE can achieve satisfactory results in most optimization problems. Among them, the optimal value obtained by the optimization of f_1, f_5, and f_{10} are their actual minimum value. In comparison with PPE, except for f_1 and f_5, which have achieved the minimum of the function together, optimal values are better in 18 functions and equal in 7 functions, and only f_{13} is worse. In addition, the optimization effects of MPPE on single-mode problems and multimode problems are significantly improved. It can be inferred that these improvement strategies make the algorithm no longer tend to the local optimal solution when solving most problems. Compared with PSO, the optimization effect of PSO is not worse than that of MPPE in some functions. However, as a classic swarm intelligence algorithm, PSO has been recognized as a drawback of premature convergence, so it is easier to tend to local optimal solutions when solving practical problems. For example, its optimization effects in f_4, f_{15}, and f_{23} are far behind that of MPPE, and its stability is not good. In comparison with SCA and WOA, MPPE has great advantages, with better results in almost all function optimizations than these two algorithms. In the comparison of GWO, MPPE algorithm is superior in the optimization of most single-mode and multimode problems, but poor in the optimization of combinatorial problems. This is because GWO algorithm has a better result in solving these problems from its clustering cooperation mechanism. Overall, MPPE is better than GWO at solving most problems. In comparison with GSK, it can be seen that the overall optimization effect of GSK is slightly better than that of MPPE. But in the optimization of some multimode problems, MPPE performs significantly better.

The comparison of these seven algorithms shows that MPPE can achieve better or equal results than other six algorithms in the optimization results of 14 functions. Table 3 counts the number of the best results obtained by each of the seven algorithms on three types of functions and their rankings. As can be seen from the rankings, the optimization performance of MPPE in this benchmark suite is second only to GSK. MPPE is overall better than other five classic algorithms. It should be emphasized that MPPE performs best in multimode functions.

Table 4 shows the results of Wilcoxon signed rank-sum test between MPPE and other six algorithms at a significant level of 0.05. The "+" symbol in the table indicates the optimization results in 28 functions of MPPE is better than that algorithm. The meaning of the "−" symbol is the optimization results of MPPE is worse than that algorithm. The "≈" symbol indicates that the optimization results almost have no difference. R^+ and R^- both represent the sum of ranks calculated by comparing all differences. p value records the significance of MPPE compared to each algorithm. "Sig." represents the significance level of MPPE compared to each algorithm. It can be concluded that the optimization results of MPPE is only significantly better than SCA, and there is no significant difference from the other five algorithms. From the perspective of rank-sum, in the comparison between MPPE and PPE, SCA, and WOA, R^+ is

```
Input: Np, d, MaxGen, groups;
Output: Gbest and minimum fitness fmin
Initialize Np population, evt, p, h of each group using Eq. (3), (4);
Initialize fitness f(x), ho, Gbest of each group;
for iter = 2 to MaxGen
    for g = 1 to groups do
        Update x to xnew using Eq. (5);
          Calculate f(xnew), ho, Gbest;
        for i = 1 to Np
            if f(x)≥f(xnew) then
                x = xnew, update f(x);
                Update p_i using Eq. (7);
                If g ≤ groups/2 then
                    Update evt using Eq. (6), (8);
                else Update evt using Eq. (6), (16);
                end if
            else if rand < p_i then
                x = xnew, update f(x);
                Update p_i using Eq. (7);
            end if
                Update evt using Eq. (9);
            end if
                select x_j randomly, (i ≠ j);
            if distance (x_i, x_j)< TH then
                Update p_i using Eq. (8), evt using Eq. (12);
            end if
        end for
    end for
    for g = 1 to groups do
        Compare, find Gbest and fmin;
    end for
    for g = 1 to groups do
        if iter < MaxGen/2 do
          Use communication strategy 1 and Eq. (13) in each group;
          else use communication strategy 2 in each group;
        end if
    end for
end for
```

Figure 4: The pseudocode of MPPE.

significantly greater than R^-. Compared with GSK, R^+ is worse than R^-.

Figure 5 shows the convergence performance of seven algorithms in different benchmark functions. We select the experimental results of two single-mode functions, two multimode functions, and two combination functions to compare the convergence performance of these algorithms from multiple perspectives. To make the comparison more obvious, we set the minimum fitness ordinate to log scale. It shows that MPPE has a fast convergence speed and is not prone to premature convergence. Compared with PPE, the convergence curve of MPPE has a faster downward trend during the early period, and the convergence in the later period improves the search accuracy more obviously. The convergence speeds of GWA, WOA, and SCA in the optimization of these functions are slow. The convergence performance of PSO is unstable. GSK converges better than MPPE in the optimization of f_2 but falls into the local optimal solution early in the optimization of f_{14}. Therefore, MPPE's comprehensive convergence performance in the optimization of these functions is better than the other six algorithms.

Table 1: Parameter settings of seven algorithms.

Algorithm	Parameter settings
MPPE	$N_p = 100$, groups = 5
PPE	$N_p = 100$, $c = 0.2$
PSO	$N_p = 100$, $c_1 = 2$, $c_2 = 2$, inertia constant from 0.8 to 0.2
SCA	$N_p = 100$
WOA	$N_p = 100$
GWO	$N_p = 100$
GSK	$N_p = 100$, $k_f = 0.5$, $k_f = 0.9$, $p = 0.1$, $K = 10$

4.2. Compared to Parallel Algorithms. To further verify the overall performance of MPPE, three parallel algorithms are selected to compare with this algorithm. These algorithms are PPSO [50], PWOA [51], and MMSCA [52]. The specific parameter settings are shown in Table 5. These algorithms were also compared on 28 functions of CEC 2013 and the experimental results of MPPE remain unchanged. Table 6 shows the optimization results of the four algorithms. MPPE can achieve the best results in the optimization of 20 functions, with multimode functions accounting for a large proportion. The results of PWOA and MMSCA are worse than MPPE in most functions, but PPSO can achieve better optimization results in 7 functions. Though the intergroup communication strategy of PPSO can avoid premature convergence to a certain extent, it still has the drawback of decreased accuracy when solving complex problems. Therefore, its overall optimization performance is worse than MPPE.

Table 7 shows the statistics of optimization results of four algorithms on three types of functions. MPPE ranks first with the most wins compared to the other three parallel algorithms. Especially in the optimization results of multimode and combination problems, MPPE occupies a significant advantage.

Table 8 shows the results of Wilcoxon signed rank-sum test between MPPE and the other three algorithms. In these sets of statistical results, the advantages of MPPE for the other three parallel algorithms are not significant. Therefore, we can at least infer that in this benchmark suite test, there is no obvious gap in the optimization results of MPPE compared to most algorithms under the parameter settings of this experiment.

Figure 6 shows convergence performance of the four parallel algorithms. We select three multimode and combination functions, respectively, for analysis. From the comparison of convergence speed, both MMSCA and PWOA are significantly slower than MPPE. PPSO and MPPE are similar in the early stage, but PPSO has a slower

Table 2: Optimization results for MPPE, PPE, PSO, SCA, WOA, GWO, and GSK over CEC 2013.

Function	MPPE	PPE	PSO	SCA	WOA	GWO	GSK
f_1	**−1.40E + 03**	**−1.40E + 03**	**−1.40E + 03**	1.13E + 04	**−1.40E + 03**	−6.66E + 02	**−1.40E + 03**
f_2	2.61E + 06	4.51E + 06	1.84E + 07	1.47E + 08	3.84E + 07	1.66E + 07	**2.57E + 05**
f_3	3.62E + 08	7.31E + 08	1.32E + 08	3.86E + 10	1.75E + 10	4.38E + 09	**−1.20E + 03**
f_4	**−2.06E + 02**	3.36E + 03	6.75E + 04	3.45E + 04	5.27E + 04	2.77E + 04	6.43E + 02
f_5	**−1.00E + 03**	**−1.00E + 03**	**−1.00E + 03**	1.45E + 03	−8.91E + 02	−2.95E + 02	**−1.00E + 03**
f_6	−8.33E + 02	−8.33E + 02	−8.20E + 02	−9.17E + 01	−7.74E + 02	−7.64E + 02	**−8.85E + 02**
f_7	−7.20E + 02	−6.95E + 02	−7.57E + 02	−6.07E + 02	−1.61E + 02	−7.49E + 02	**−8.00E + 02**
f_8	**−6.79E + 02**	**−6.79E + 02**	**−6.79E + 02**	**−6.79E + 02**	**−6.79E + 02**	**−6.79E + 02**	**−6.79E + 02**
f_9	−5.69E + 02	−5.69E + 02	−5.78E + 02	−5.60E + 02	−5.62E + 02	**−5.81E + 02**	**−5.63E + 02**
f_{10}	**−5.00E + 02**	−4.99E + 02	−4.99E + 02	1.21E + 03	−4.13E + 02	−2.65E + 02	**−5.00E + 02**
f_{11}	**−3.87E + 02**	−3.52E + 02	−3.77E + 02	−4.18E + 01	1.04E + 02	−3.08E + 02	−2.66E + 02
f_{12}	**−2.76E + 02**	−4.46E + 00	−2.20E + 02	8.75E + 01	2.10E + 02	−1.77E + 02	−1.30E + 02
f_{13}	1.29E + 02	1.79E + 02	1.72E + 02	1.91E + 02	2.96E + 02	−1.98E + 01	**−2.88E + 01**
f_{14}	**9.02E + 02**	1.15E + 03	2.59E + 03	7.06E + 03	5.07E + 03	2.93E + 03	5.93E + 03
f_{15}	**3.46E + 03**	4.22E + 03	7.13E + 03	7.63E + 03	5.61E + 03	3.80E + 03	7.38E + 03
f_{16}	**2.01E + 02**	**2.01E + 02**	2,02E + 02	2.03E + 02	2.02E + 02	2.03E + 02	2.02E + 02
f_{17}	**3.38E + 02**	3.79E + 02	3.62E + 02	7.84E + 02	9.25E + 02	4.54E + 02	4.74E + 02
f_{18}	6.23E + 02	6.74E + 02	6.32E + 02	9.04E + 02	9.73E + 02	6.45E + 02	**5.80E + 02**
f_{19}	**5.03E + 02**	5.07E + 02	**5.03E + 02**	3.95E + 03	5.67E + 02	2.03E + 03	5.13E + 02
f_{20}	6.15E + 02	6.15E + 02	6.15E + 02	6.14E + 02	6.15E + 02	**6.12E + 02**	6.13E + 02
f_{21}	1.09E + 03	1.09E + 03	**1.00E + 03**	2.61E + 03	1.09E + 03	1.54E + 03	**1.00E + 03**
f_{22}	**2.24E + 03**	2.48E + 03	5.05E + 03	8.58E + 03	7.12E + 03	3.91E + 03	5.04E + 03
f_{23}	5.67E + 03	5.69E + 03	5.77E + 03	8.83E + 03	7.62E + 03	**5.52E + 03**	8.02E + 03
f_{24}	1.28E + 03	1.28E + 03	1.29E + 03	1.32E + 03	1.31E + 03	1.25E + 03	**1.20E + 03**
f_{25}	1.42E + 03	1.44E + 03	1.41E + 03	1.43E + 03	1.42E + 03	**1.37E + 03**	1.41E + 03
f_{26}	**1.40E + 03**	1.47E + 03	1.48E + 03	1.41E + 03	1.53E + 03	1.53E + 03	**1.40E + 03**
f_{27}	2.42E + 03	2.28E + 03	2.29E + 03	2.70E + 03	2.63E + 03	**2.09E + 03**	**1.67E + 03**
f_{28}	2.52E + 03	4.69E + 03	3.47E + 03	4.04E + 03	5.41E + 03	2.63E + 03	**1.70E + 03**

Bold values represent the best results relative to other comparison methods.

Table 3: Statistics of optimization results of seven algorithms on three types of functions.

Algorithm	Single-mode	Multimode	Combination	Wins	Ranking
GSK	4	7	5	16	1
MPPE	3	9	2	14	2
GWO	0	3	3	6	3
PSO	2	2	1	5	4
PPE	2	2	0	4	5
WOA	1	1	0	2	6
SCA	0	1	0	1	7

Table 4: The results of Wilcoxon signed rank-sum test between MPPE and other six algorithms.

Algorithm	R^+	R^-	p value	+	≈	−	Sig.
MPPE versus PPE	340	21	0.63	18	9	1	≈
MPPE versus PSO	254	124	0.59	14	7	7	≈
MPPE versus SCA	400	0	0.035	24	3	0	+
MPPE versus WOA	391	0	0.33	23	5	0	≈
MPPE versus GWO	307	93	0.46	17	3	8	≈
MPPE versus GSK	172	206	0.81	9	7	12	≈

convergence rate in the late period. We can infer from this that MPPE is more likely to get rid of the local optimal solution and continue to seek the global minimum in the late period. Therefore, its overall convergence performance is better than the other three parallel algorithms.

4.3. Analysis of Algorithm Complexity. The maximum time complexity of PPE is described as $O(p \cdot n \cdot I)$, where p is the total number of stick insect populations, n is the dimension of the problem, and I is the maximum number of iterations. MPPE adopted a parallel method on the basis of PPE, and each

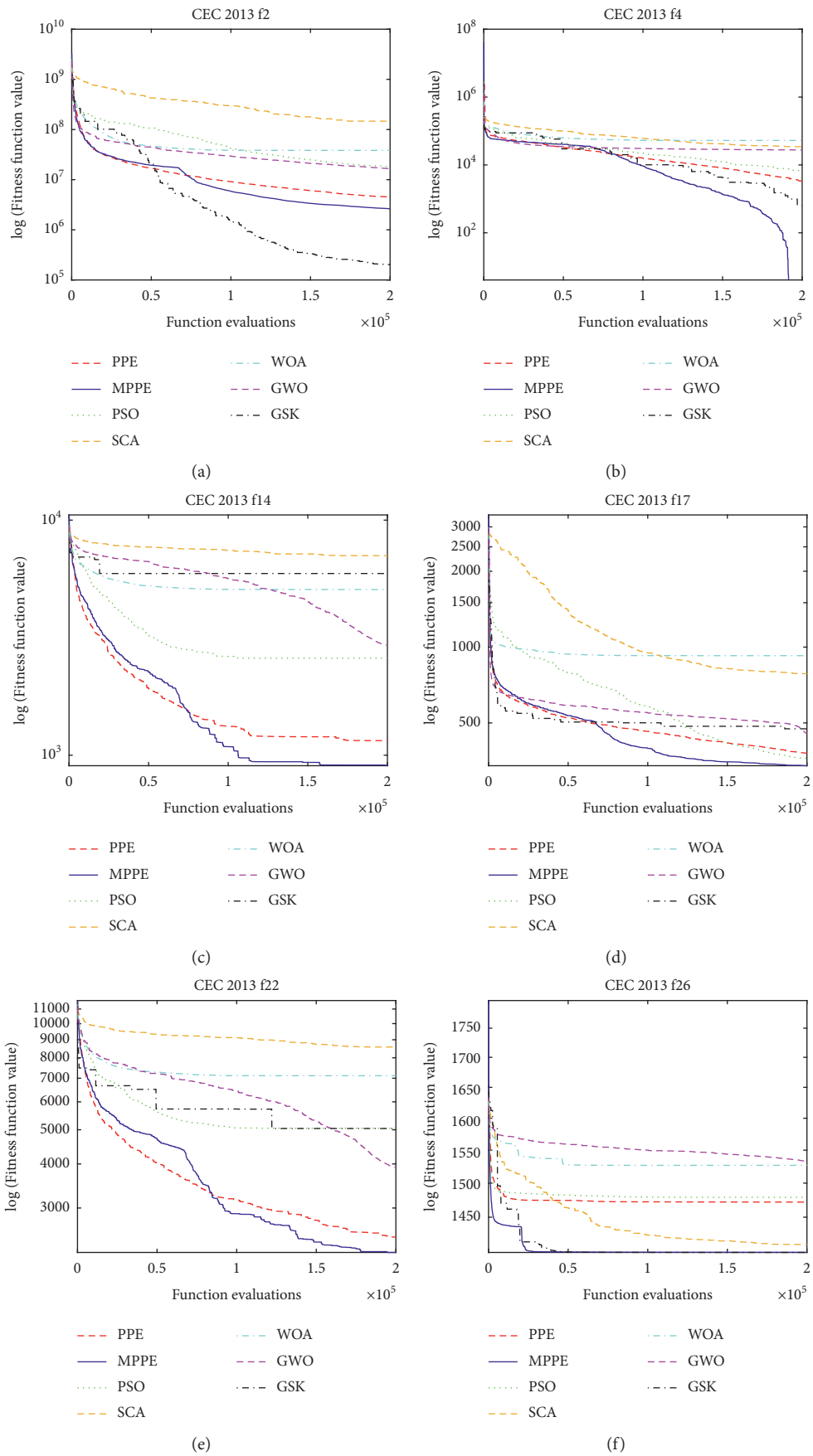


Figure 5: Convergence performance of seven algorithms. (a) f_2, (b) f_4, (c) f_{14}, (d) f_{17}, (e) f_{22}, and (f) f_{26}.

Table 5: Parameter settings of four parallel algorithms.

Algorithm	Parameter settings
MPPE	$N_p = 100$, groups = 5
PPSO	$N_p = 100$, groups = 5, $c_1 = 2$, $c_2 = 2$, inertia constant from 0.8 to 0.2,
PWOA	$N_p = 100$, groups = 5
MMSCA	$N_p = 100$, groups = 5

Table 6: Optimization results for MPPE, PPSO, PWOA, and MMSCA over CEC 2013.

Function	MPPE	PPSO	PWOA	MMSCA
f_1	**−1.40E + 03**	**−1.40E + 03**	−1.39E + 03	6.77E + 03
f_2	2.61E + 06	**8.09E + 05**	3.71E + 07	7.42E + 07
f_3	**3.62E + 08**	6.52E + 08	1.31E + 10	2.02E + 10
f_4	−2.06E + 02	**−7.43E + 02**	4.75E + 04	2.81E + 04
f_5	**−1.00E + 03**	**−1.00E + 03**	−8.88E + 02	6.56E + 02
f_6	−8.33E + 02	**−8.43E + 02**	−7.71E + 02	−4.33E + 02
f_7	**−7.20E + 02**	−7.08E + 02	−5.93E + 02	−6.68E + 02
f_8	**−6.79E + 02**	**−6.79E + 02**	**−6.79E + 02**	**−6.79E + 02**
f_9	−5.69E + 02	**−5.72E + 02**	−5.67E + 02	−5.62E + 02
f_{10}	**−5.00E + 02**	**−5.00E + 02**	−3.41E + 02	7.27E + 02
f_{11}	**−3.87E + 02**	−1.01E + 02	−3.99E + 01	−6.81E + 01
f_{12}	**−2.76E + 02**	−4.21E + 01	1.90E + 02	3.62E + 01
f_{13}	**1.29E + 02**	1.49E + 02	2.30E + 02	1.51E + 02
f_{14}	**9.02E + 02**	3.39E + 03	3.78E + 03	6.74E + 03
f_{15}	**3.46E + 03**	4.22E + 03	5.69E + 03	7.21E + 03
f_{16}	**2.01E + 02**	**2.01E + 02**	2.02E + 02	2.02E + 02
f_{17}	**3.38E + 02**	4.81E + 02	8.25E + 02	7.39E + 02
f_{18}	6.23E + 02	**5.61E + 02**	7.88E + 02	8.55E + 02
f_{19}	**5.03E + 02**	5.06E + 02	5.35E + 02	3.22E + 03
f_{20}	6.15E + 02	6.15E + 02	6.15E + 02	**6.14E + 02**
f_{21}	1.09E + 03	**1.03E + 03**	1.18E + 03	2.33E + 03
f_{22}	**2.24E + 03**	5.61E + 03	5.12E + 03	7.56E + 03
f_{23}	**5.67E + 03**	6.22E + 03	6.53E + 03	8.58E + 03
f_{24}	**1.28E + 03**	1.29E + 03	1.30E + 03	1.31E + 03
f_{25}	**1.42E + 03**	1.43E + 03	**1.42E + 03**	**1.42E + 03**
f_{26}	**1.40E + 03**	1.51E + 03	1.48E + 03	1.41E + 03
f_{27}	2.42E + 03	**2.40E + 03**	2.66E + 03	2.62E + 03
f_{28}	**2.52E + 03**	3.86E + 03	5.09E + 03	3.74E + 03

Bold values represent the best results relative to other comparison methods.

Table 7: Statistics of optimization results of four algorithms on three types of functions.

Algorithm	Single-mode	multimode	Combination	Wins	Ranking
MPPE	3	11	6	20	1
PPSO	4	6	2	12	2
MMSCA	0	2	1	3	3
PWOA	0	1	1	2	4

Table 8: The results of Wilcoxon signed rank-sum test between MPPE and the other three algorithms.

Algorithm	R^+	R^-	p value	+	=	−	Dec.
MPPE versus PPSO	245	125	0.87	13	8	7	≈
MPPE versus PWOA	391	0	0.39	23	5	0	≈
MPPE versus MMSCA	396	0	0.067	24	4	0	≈

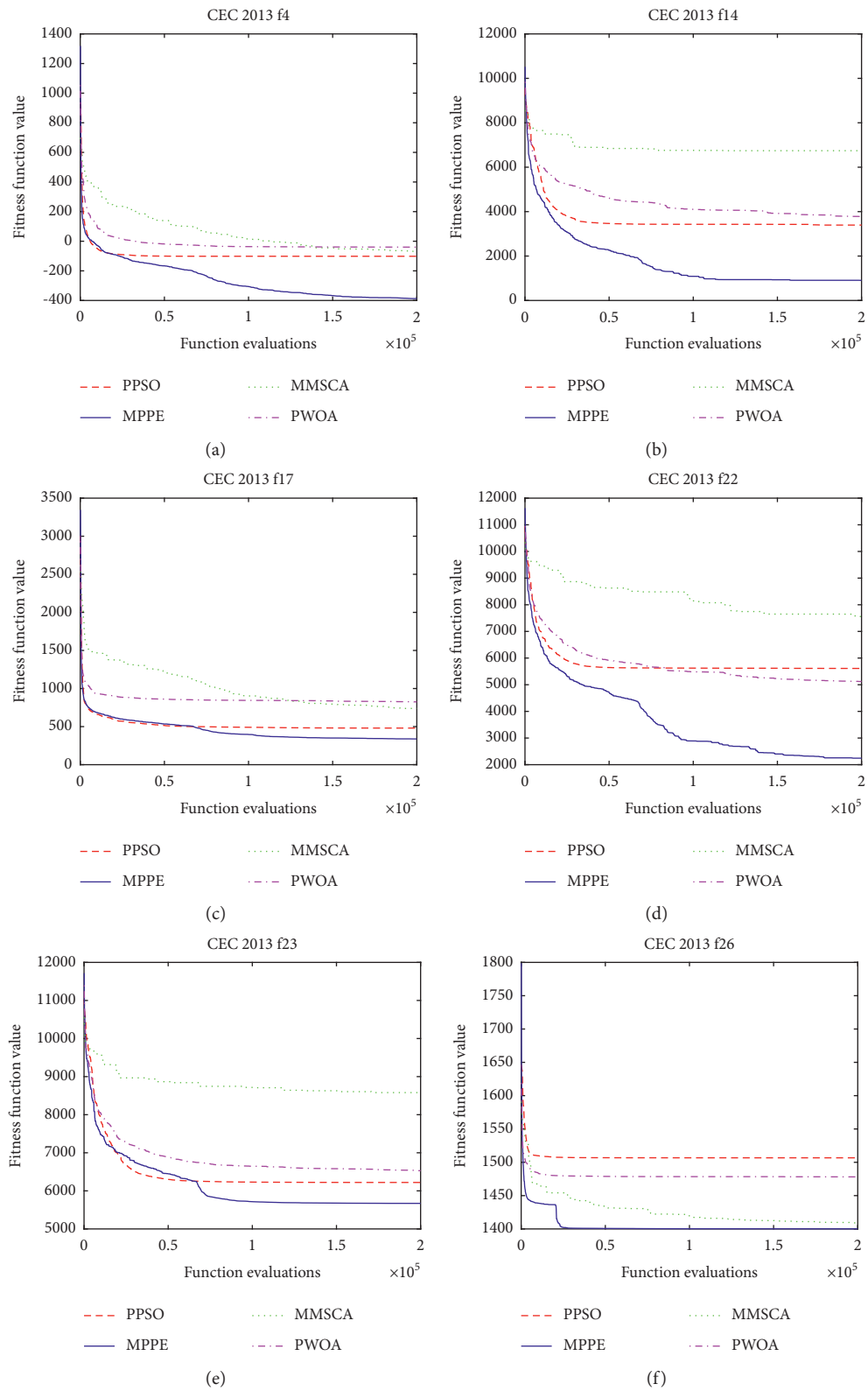


Figure 6: Convergence performance of four parallel algorithms. (a) f_4, (b) f_{14}, (c) f_{17}, (d) f_{22}, (e) f_{23}, and (f) f_{26}.

Table 9: Passenger flow in different time periods.

Period of time	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_9	S_{10}
6:00–9:00	401	101	317	339	106	11	10	9	20	12
9:00–12:00	410	65	107	154	117	86	151	107	33	8
12:00–16:00	177	32	50	28	98	100	138	126	10	16
16:00–19:00	144	72	154	254	177	138	162	250	90	23
19:00–21:00	70	62	35	48	47	9	41	33	11	8

Table 10: Comparison of results in $w_1 = 0.7$ and $w_2 = 0.3$.

Algorithm	Average result	Best result	Worst result	Scheduling scheme
SCA	2920.64	2893.12	2995.92	(2.3 2.7 5.5 2.8 3.5)
MCSCA	2944.07	2878.63	2995.92	(2.1 3.5 7.4 5.1 3.2)
PSO	2968.77	2895.50	3180.31	(3.1 5.5 1.6 3.0 3.3)
PPE	2868.47	2866.52	2873.99	(2.4 3.1 5.6 3.3 5.0)
MPPE	**2866.75**	**2865.77**	**2868.36**	(2.5 3.1 5.6 3.3 4.6)

Bold values represent the best results relative to other comparison methods.

Table 11: Comparison of results in $w_1 = 0.5$ and $w_2 = 0.5$.

Algorithm	Average result	Best result	Worst result	Scheduling scheme
SCA	3238.80	3176.37	3288.67	(2.9 4.3 4.2 3.1 3.2)
MCSCA	3291.97	3184.37	3296.17	(2.0 3.6 5.0 3.7 3.3)
PSO	3905.68	3151.46	5526.75	(8.2 8.2 7.9 1.5 9.6)
PPE	3139.95	3138.58	3141.18	(2.5 3.0 4.6 3.2 4.0)
MPPE	**3138.57**	**3137.83**	**3139.24**	(1.6 2.0 3.8 2.1 3.2)

Bold values represent the best results relative to other comparison methods.

Table 12: Comparison of results in $w_1 = 0.3$ and $w_2 = 0.7$.

Algorithm	Average result	Best result	Worst result	Scheduling scheme
SCA	3102.84	2912.69	3382.02	(2.2 3.3 5.0 3.6 3.2)
MCSCA	3129.38	2937.53	3382.02	(1.4 1.8 2.5 2.6 3.2)
PSO	4269.70	2925.66	6325.52	(8.6 9.5 7.8 3.7 3.6)
PPE	2884.01	2883.21	2885.38	(3.0 3.3 4.4 3.4 3.9)
MPPE	**2882.87**	**2882.41**	**2883.75**	(2.2 2.5 3.5 2.7 3.2)

Bold values represent the best results relative to other comparison methods.

iteration is accompanied by the simultaneous evolution of each group of populations. So, the maximum time complexity of MPPE can be described as $O((p_1 \cdot n + p_2 \cdot n + \cdots + p_g \cdot n) \cdot I)$, where p_i is the population size of the ith group and g is the number of groups. Since the number of populations in each group is equal to p/g, the maximum complexity of MPPE is described as $O(p \cdot n \cdot I)$.

Although the amount of calculation added by the intergroup communication strategy does not exceed the maximum time complexity, the actual running time of MPPE has increased considerably compared to PPE.

5. MPPE Applicating in IoT Electric Bus Scheduling

In this optimization algorithm applied to the bus scheduling experiment for urban waterlogging situation, the bus operating time is set to 6 am–9 pm, the total number of kilometers one way to 10 km, and the number of stations to 10; each site is named S_1 to S_{10}. The data simulation of passenger flow is shown in Table 9.

In this experiment, MPPE was compared with PPE, PSO, SCA, and MCSCA [53]. In the search process, all algorithms find their current optimal solutions through each iteration. The optimal solution is the minimum value of the objective function of IoT electric bus scheduling. The position corresponding to the best solution is the time interval of bus departure in each time period. Therefore, these algorithms are all feasible in solving this problem, but their results are usually different.

In the parameter settings, both Q_1 and Q_2 are set to 1. The maximum number of iterations for each algorithm is set to 100. The population size of all algorithms is set to 30. This is because in solving practical problems, it is necessary to set a small number of iterations and populations to improve efficiency. The value of the bus departure time interval in each time period is accurate to one decimal place, and their range is controlled in 0–10. All the algorithms are run 10 times in this experiment. Finally, we calculate the average result, worst result, and best result of these algorithms. To make the experimental results more convincing, we set the weights w_1 and w_2 to three different groups of values. They are as follows: (1) $w_1 = 0.7$ and $w_2 = 0.3$, (2) $w_1 = 0.5$ and $w_2 = 0.5$, and (3) $w_1 = 0.3$ and $w_2 = 0.7$.

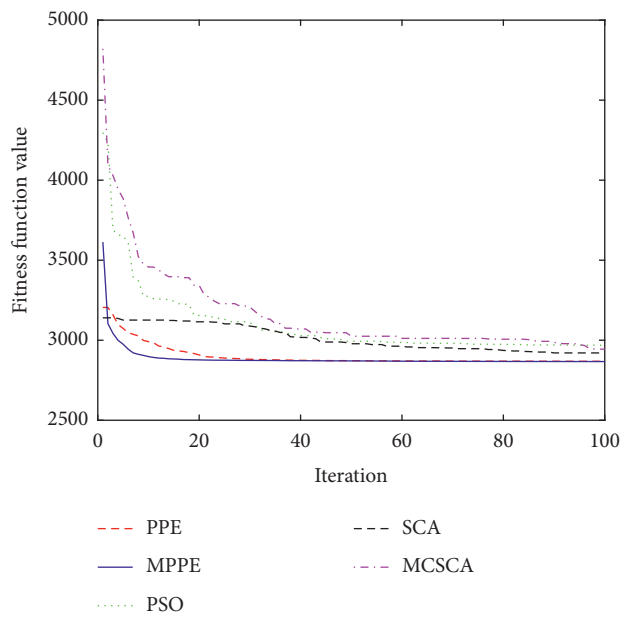


Figure 7: Convergence performances in $w_1 = 0.7$ and $w_2 = 0.3$.

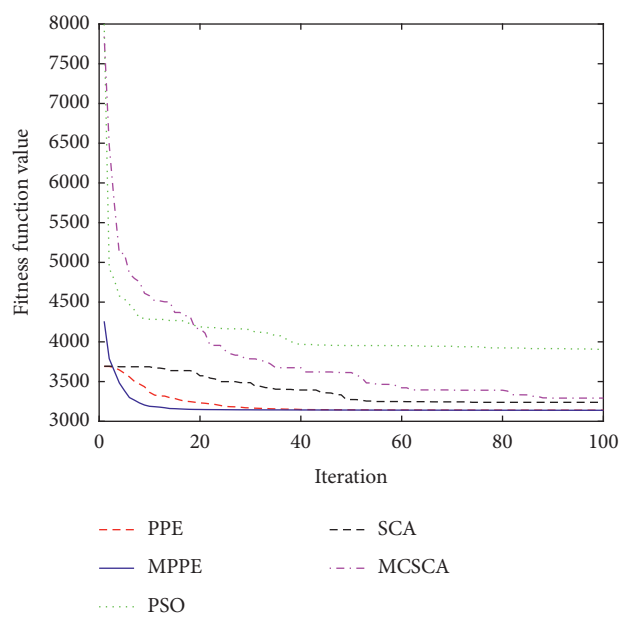


Figure 8: Convergence performances in $w_1 = 0.5$ and $w_2 = 0.5$.

Tables 10 to 12 show the comparison of fitness values and solutions of these algorithms under different weights. Figures 7 to 9 show the convergence performance of these algorithms.

The above three sets of experiments show that the MPPE can perform better in different weight selection, and its average results are better than the other three algorithms in all three cases. From the point of view of the optimal value, although the three results of MPPE are slightly smaller than PPE, MPPE also has better stability. Compared with MPPE, PSO, MCSCA, and SCA all have a big gap. In terms of convergence performance, MPPE has a faster convergence rate in the early stage. Compared to PPE, MPPE almost find the optimal value before 20 iterations, and PPE tends to end convergence in the middle of iterations. Although PSO also has a fast convergence rate in the early stage, it does not perform well in the later stage and tends to fall into local optimum.

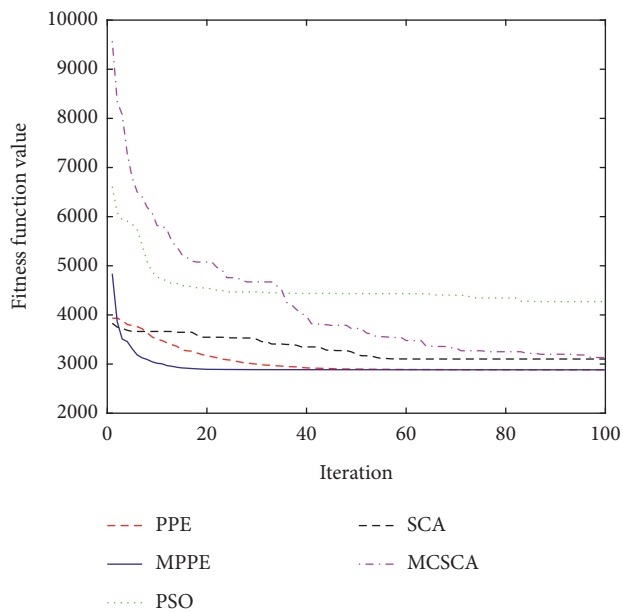


Figure 9: Convergence performances in $w_1 = 0.3$ and $w_2 = 0.7$.

Table 13: Ranking of average optimization results of MPPE under different weight combinations.

Weight combinations	Average result	Scheduling scheme	Ranking
$w_1 = 0.1$, $w_2 = 0.9$	1861.91	(5.1 6.0 9.6 6.7 8.6)	1
$w_1 = 0.9$, $w_2 = 0.1$	1890.99	(2.7 2.9 3.4 2.9 3.2)	2
$w_1 = 0.8$, $w_2 = 0.2$	2494.98	(3.3 3.9 7.3 4.2 6.7)	3
$w_1 = 0.2$, $w_2 = 0.8$	2518.76	(2.5 2.7 3.5 3.8 3.2)	4
$w_1 = 0.7$, $w_2 = 0.3$	2866.75	(2.5 3.1 5.6 3.3 4.6)	5
$w_1 = 0.3$, $w_2 = 0.7$	2882.87	(2.2 2.5 3.5 2.7 3.2)	6
$w_1 = 0.6$, $w_2 = 0.4$	3070.65	(1.9 2.4 4.6 2.6 3.8)	7
$w_1 = 0.4$, $w_2 = 0.6$	3079.02	(2.1 2.4 3.8 2.6 3.2)	8
$w_1 = 0.5$, $w_2 = 0.5$	3138.57	(1.6 2.0 3.8 2.1 3.2)	9

In summary, the MPPE algorithm can effectively solve the IoT electric bus rainy days scheduling problem. This is mainly reflected in the fast convergence speed and good optimization accuracy. The result of the solution is also relatively stable, and it is not prone to large deviations. From the perspective of people's experience, the scheduling scheme optimized by this algorithm is also reasonable.

In order to analyze the relationship between the weight setting and the optimization result, we only use the MPPE algorithm to solve this problem and test all possible weight combinations. Table 13 lists the rankings of the optimization results of all weight combinations and their corresponding scheduling schemes. The experimental results show that the theoretically best combination of weights is $w_1 = 0.1$ and $w_2 = 0.9$, because they achieve the best average results. However, decision-makers should make corresponding adjustments to the setting of weights based on actual conditions.

6. Conclusions

According to the limitations of the PPE algorithm, this research made improvements based on multiple groups and multiple strategies. The experiments of the benchmark CEC 2013 proved that our proposed MPPE algorithm has good convergence speed and optimization accuracy. In addition, we apply this algorithm to the IoT electric bus scheduling problem and achieve a smaller loss value compared to PPE, PSO, SCA, and MCSCA algorithm.

However, the MPPE algorithm still has shortcomings, which is mainly reflected in its long actual running time. In future research, we will consider this shortcoming and make further improvements. At the same time, some representative novel metaheuristic algorithms, such as MBO, SMA, GSK, and HHO, can also be used to solve optimization problems similar to bus scheduling. In the next step, we will consider combining these novel algorithms and expanding this type of optimization problem.

Data Availability

Relevant data used in this study are included in this article.

Conflicts of Interest

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

Acknowledgments

The authors are grateful for the financial support received from the National Key Project of China (no. 2020AAA0109300), the Open Research Project of Shanghai Key Laboratory of Information Security Integrated Management Technology (no. AGK2019004), and the Research on Intelligent Scheduling of Urban Waterlogging Emergency Facilities Based on Multisource Information (no. 21511103704).

References

[1] M. Abdel-Basset, L. Abdel-Fatah, and A. K. Sangaiah, "Metaheuristic algorithms: a comprehensive review," *Computational Intelligence for Multimedia Big Data on the Cloud with Engineering Applications*, Elsevier, Amsterdam, Netherlands, 2018.

[2] M. A. Al-Betar, M. A. Awadallah, I. Abu Doush, A. I. Hammouri, M. Mafarja, and Z. A. A. Alyasseri, "Island flower pollination algorithm for global optimization," *The Journal of Supercomputing*, vol. 75, no. 8, pp. 5280–5323, 2019.

[3] B. H. Abed-alguni, N. A. Alawad, M. Barhoush, and R. Hammad, "Exploratory cuckoo search for solving single-objective optimization problems," *Soft Computing*, vol. 25, pp. 1–14, 2021.

[4] S. Zhou, L. Xing, X. Zheng, N. Du, L. Wang, and Q. Zhang, "A self-adaptive differential evolution algorithm for scheduling a single batch-processing machine with arbitrary job sizes and release times," *IEEE Transactions on Cybernetics*, vol. 51, no. 3, pp. 1430–1442, 2021.

[5] B. H. Abed-alguni and N. A. Alawad, "Distributed Grey Wolf Optimizer for scheduling of workflow applications in cloud environments," *Applied Soft Computing*, vol. 102, Article ID 107113, 2021.

[6] S. Tiachacht, S. Khatir, C. Le Thanh, R. V. Rao, S. Mirjalili, and M. A. Wahab, "Inverse problem for dynamic structural health monitoring based on slime mould algorithm," *Engineering with Computers*, pp. 1–24, 2021.

[7] F. Zhao, X. He, and L. Wang, "A two-stage cooperative evolutionary algorithm with problem-specific knowledge for energy-efficient scheduling of no-wait flow-shop problem," *IEEE Transactions on Cybernetics*, vol. 51, no. 11, pp. 5291–5303, 2021.

[8] F. Zhao, L. Zhao, L. Wang, and H. Song, "An ensemble discrete differential evolution for the distributed blocking flowshop scheduling with minimizing makespan criterion," *Expert Systems with Applications*, vol. 160, Article ID 113678, 2020.

[9] F. Zhao, R. Ma, and L. Wang, "A self-learning discrete jaya algorithm for multiobjective energy-efficient distributed No-idle flow-shop scheduling problem in heterogeneous factory system," *IEEE Transactions on Cybernetics*, pp. 1–12, 2021.

[10] A. W. Mohamed, A. A. Hadi, and A. K. Mohamed, "Gaining-sharing knowledge based algorithm for solving optimization problems: a novel nature-inspired algorithm," *International Journal of Machine Learning and Cybernetics*, vol. 11, no. 7, pp. 1501–1529, 2020.

[11] R. Moghdani and K. Salimifard, "Volleyball premier league algorithm," *Applied Soft Computing*, vol. 64, pp. 161–185, 2018.

[12] M. A. Awadallah, M. A. Al-Betar, A. L. A. Bolaji, I. A. Doush, A. I. Hammouri, and M. Mafarja, "Island artificial bee colony for global optimization," *Soft Computing*, vol. 24, pp. 1–27, 2020.

[13] Y. Feng, S. Deb, G.-G. Wang, and A. H. Alavi, "Monarch butterfly optimization: a comprehensive review," *Expert Systems with Applications*, vol. 168, Article ID 114418, 2021.

[14] G.-G. Wang, "Moth search algorithm: a bio-inspired metaheuristic algorithm for global optimization problems," *Memetic Computing*, vol. 10, no. 2, pp. 151–164, 2018.

[15] A. A. Heidari, S. Mirjalili, H. Faris, I. Aljarah, M. Mafarja, and H. Chen, "Harris hawks optimization: algorithm and applications," *Future Generation Computer Systems*, vol. 97, pp. 849–872, 2019.

[16] A. Kaveh and A. Dadras, "A novel meta-heuristic optimization algorithm: thermal exchange optimization," *Advances in Engineering Software*, vol. 110, pp. 69–84, 2017.

[17] A. Tabari and A. Ahmad, "A new optimization method: Electro-Search algorithm," *Computers and Chemical Engineering*, vol. 103, pp. 1–11, 2017.

[18] A. F. Nematollahi, A. Rahiminejad, and B. Vahidi, "A novel physical based meta-heuristic optimization method known as Lightning attachment procedure optimization," *Applied Soft Computing*, vol. 59, pp. 596–621, 2017.

[19] P. C. Song, S. C. Chu, J. S. Pan, and H. Yang, "Phasmatodea population evolution algorithm and its application in length-changeable incremental extreme learning machine," in *Proceedings of the 2020 2nd International Conference on Industrial Artificial Intelligence (IAI)*, pp. 1–5, IEEE, Shenyang, China, October 2020.

[20] P. C. Song, S. C. Chu, J. S. Pan, and H. Yang, "Simplified Phasmatodea population evolution algorithm for optimization," *Complex and Intelligent Systems*, pp. 1–19, 2021.

[21] X. Yuan, T. Zhang, X. Dai, and L. Wu, "Master-slave model-based parallel chaos optimization algorithm for parameter identification problems," *Nonlinear Dynamics*, vol. 83, no. 3, pp. 1727–1741, 2016.

[22] A. Lančinskas, P. M. Ortigosa, and J. Žilinskas, "Parallel optimization algorithm for competitive facility location," *Mathematical Modelling and Analysis*, vol. 20, no. 5, pp. 619–640, 2015.

[23] S. Lalwani, H. Sharma, S. C. Satapathy, K. Deep, and J. C. Bansal, "A survey on parallel particle swarm optimization algorithms," *Arabian Journal for Science and Engineering*, vol. 44, no. 4, pp. 2899–2923, 2019.

[24] K. Shen, T. De Pessemier, L. Martens, and W. Joseph, "A parallel genetic algorithm for multi-objective flexible flow-shop scheduling in pasta manufacturing," *Computers and Industrial Engineering*, vol. 161, Article ID 107659, 2021.

[25] Z. Yang, B. Yu, and C. Cheng, "A Parallel ant colony algorithm for bus network optimization," *Computer-Aided Civil and Infrastructure Engineering*, vol. 22, no. 1, pp. 44–55, 2007.

[26] F. Yan, Y. Hu, J. Jia, Z. Ai, and X. Liu, "Interactive WEBVR visualization for online fire evacuation training," *Multimedia Tools and Applications*, vol. 79, no. 4, pp. 31541–31565, 2020.

[27] F.-t. Yan, Y.-h. Hu, J.-y. Jia, Q.-h. Guo, H.-h. Zhu, and Z.-g. Pan, "RFES: a real-time fire evacuation system for mobile

web3d," *Frontiers of Information Technology and Electronic Engineering*, vol. 20, no. 8, pp. 1061–1074, 2019.

[28] J. Zhang, S. Lin, H. Liu, Y. Chen, M. Zhu, and Y. Xu, "A small-population based parallel differential evolution algorithm for short-term hydrothermal scheduling problem considering power flow constraints," *Energy*, vol. 123, pp. 538–554, 2017.

[29] S. Lin and J. Wang, "Carbon emission reduction effect of transportation structure adjustment in China: an approach on multi-objective optimization model," *Environmental Science and Pollution Research*, vol. 29, pp. 1–18, 2021.

[30] S. Chen, "Highway transportation optimization control system based on OD forecast information," *International Journal of System Assurance Engineering and Management*, vol. 12, pp. 1–9, 2021.

[31] C.-H. Chen, "An arrival time prediction method for bus system," *IEEE Internet of Things Journal*, vol. 5, no. 5, pp. 4231-4232, 2018.

[32] F. T. Yan, J. Y. Jia, Y. H. Hu, Q. H. Guo, and H. H. Zhu, "Smart fire evacuation service based on Internet of Things computing for Web3D," *Journal of Internet Technology*, vol. 20, no. 2, pp. 521–532, 2019.

[33] Y. Bie, M. Hao, and M. Guo, "Optimal electric bus scheduling based on the combination of all-stop and short-turning strategies," *Sustainability*, vol. 13, no. 4, p. 1827, 2021.

[34] Y. Alwesabi, Y. Wang, R. Avalos, and Z. Liu, "Electric bus scheduling under single depot dynamic wireless charging infrastructure planning," *Energy*, vol. 213, Article ID 118855, 2020.

[35] W.-L. Liu, Y.-J. Gong, W.-N. Chen, Z. Liu, H. Wang, and J. Zhang, "Coordinated charging scheduling of electric vehicles: a mixed-variable differential evolution approach," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 12, pp. 5094–5109, 2020.

[36] Z. Yang, S. Zhao, and Q. Zhao, "Research on bus scheduling based on artificial immune algorithm," in *Proceedings of the 2008 4th International Conference on Wireless Communications, Networking and Mobile Computing*, pp. 1–4, IEEE, Dalian, China, October 2008.

[37] C. Wang, C. Guo, and X. Zuo, "Solving multi-depot electric vehicle scheduling problem by column generation and genetic algorithm," *Applied Soft Computing*, vol. 112, Article ID 107774, 2021.

[38] A. Xiao, "A self-adaptive quantum genetic algorithm for network flow vehicle scheduling problem," *Journal of Computer and Communications*, vol. 9, no. 7, pp. 43–54, 2021.

[39] Y. Zhou, Q. Luo, and J. Liu, "Glowworm swarm optimization for dispatching system of public transit vehicles," *Neural Processing Letters*, vol. 40, no. 1, pp. 25–33, 2014.

[40] Z. Fu, P. Hu, W. Li, J.-S. Pan, and S.-C. Chu, "Parallel equilibrium optimizer algorithm and its application in capacitated vehicle routing problem," *Intelligent Automation and Soft Computing*, vol. 27, no. 1, pp. 233–247, 2021.

[41] X.-S. Yang, "Flower pollination algorithm for global optimization," in *Proceedings of the International Conference on Unconventional Computing and Natural Computation*, pp. 240–249, Berlin, Heidelberg, September 2012.

[42] X.-S. Yang, M. Karamanoglu, and X. He, "Flower pollination algorithm: a novel approach for multiobjective optimization," *Engineering Optimization*, vol. 46, no. 9, pp. 1222–1237, 2014.

[43] T. Adithiyaa, D. Chandramohan, and T. Sathish, "Flower pollination algorithm for the optimization of stair casting parameter for the preparation of AMC," *Materials Today: Proceedings*, vol. 21, pp. 882–886, 2020.

[44] A. W. Mohamed, A. A. Hadi, and K. M. Jambi, "Novel mutation strategy for enhancing SHADE and LSHADE algorithms for global numerical optimization," *Swarm and Evolutionary Computation*, vol. 50, Article ID 100455, 2019.

[45] J. Derrac, S. García, D. Molina, and F. Herrera, "A practical tutorial on the use of nonparametric statistical tests as a methodology for comparing evolutionary and swarm intelligence algorithms," *Swarm and Evolutionary Computation*, vol. 1, no. 1, pp. 3–18, 2011.

[46] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proceedings of the ICNN'95-International Conference on Neural Networks*, pp. 1942–1948, IEEE, Perth, WA, Australia, November 1995.

[47] S. Mirjalili, "SCA: a sine cosine algorithm for solving optimization problems," *Knowledge-Based Systems*, vol. 96, pp. 120–133, 2016.

[48] S. Mirjalili and A. Lewis, "The whale optimization algorithm," *Advances in Engineering Software*, vol. 95, pp. 51–67, 2016.

[49] S. Mirjalili, S. M. Mirjalili, and A. Lewis, "Grey wolf optimizer," *Advances in Engineering Software*, vol. 69, pp. 46–61, 2014.

[50] S. C. Chu, J. F. Roddick, and J. S. Pan, "A parallel particle swarm optimization algorithm with communication strategies," *Journal of Information Science and Engineering*, vol. 21, no. 4, p. 9, 2005.

[51] Q.-w. Chai, S.-C. Chu, J.-S. Pan, P. Hu, and W.-m. Zheng, "A parallel WOA with two communication strategies applied in DV-Hop localization method," *EURASIP Journal on Wireless Communications and Networking*, vol. 2020, no. 4, 2020.

[52] Q. Yang, S.-C. Chu, J.-S. Pan, and C.-M. Chen, "Sine cosine algorithm with multigroup and multistrategy for solving CVRP," *Mathematical Problems in Engineering*, vol. 2020, no. 12, pp. 1–10, 2020.

[53] M. Zhu, S.-C. Chu, Q. Yang, W. Li, and J.-S. Pan, "Compact sine cosine algorithm with multigroup and multistrategy for dispatching system of public transit vehicles," *Journal of Advanced Transportation*, vol. 2021, no. 2, pp. 1–16, 2021.

Hindawi
Wireless Communications and Mobile Computing
Volume 2022, Article ID 8641096, 11 pages
https://doi.org/10.1155/2022/8641096

Research Article

An Intelligent Vision-Based Method of Worker Identification for Industrial Internet of Things (IoT)

Shuai Wang,[1,2] Xiaoyu Li,[1] Wei Chen,[1,3] Weiqiang Fan,[1] and Zijian Tian[1]

[1]*School of Mechanical Electronic & Information Engineering, China University of Mining and Technology (Beijing), Beijing 100083, China*
[2]*Department of Inner Mongolia, Administration of Coal Mine Safety Ordos Division, Ordos 017000, China*
[3]*School of Computer Science & Technology, China University of Mining and Technology, Xuzhou 221116, China*

Correspondence should be addressed to Wei Chen; chenwdavior@163.com

Received 22 October 2021; Revised 20 December 2021; Accepted 3 January 2022; Published 27 January 2022

Academic Editor: Chi-Hua Chen

With the rapid development of Internet of things (IoT) and computer vision (CV), the application of combining the IoT platform and CV technology to monitor the worker safety has attracted more and more attention in the field of industrial information. Worker identification is a prerequisite for safety management in industrial production, and safety helmet can not only protect worker's head from accidental injuries but also help to identify the work types of workers through different colors. Therefore, this study proposes an intelligent method for worker identification based on moving personnel detection and helmet color characteristics. First, the motion objects that contain personnel and nonpersonnel are detected by the Gaussian mixture model (GMM) and extracted to generate the region of interest (RoI) images. Then, the multiple-scale histogram of oriented gradient (MHOG) features of the RoI images are extracted, and the personnel images are identified by the support vector machine (SVM). Third, the workers' head images are obtained by the OpenPose model and personnel mask, and the GoogLeNet-based transfer learning network is established to extract the head images features and realize worker identification. This method is tested on our dataset, and the average accuracy of worker identification for multiple helmet color combinations reaches 99.43%, which is robust to workers' angle, scale, and occlusion.

1. Introduction

Safe production is the top priority of modern industrial development. With the advancement of Industry 4.0 era, it is an inevitable trend to use intelligent vision-based technology to identify worker information, and then feed it back to the data center through Internet of things (IoT) for analysis and processing. This not only automatically improves the efficiency of human work but also strengthens the guarantee of workers' life safety. In recent years, the vision-based technology of worker safety monitoring has shown significant advantages in industrial production and has played an important role in promoting the development of industrial IoT. Despite this, the industrial sectors report thousands of casualties every year, and a large proportion of the accidents are caused by falling objects [1]. In order to protect workers from being struck by falling objects, wearing helmet is the most convenient and effective measure of safety protection [2]. Therefore, to strictly require workers to wear helmets, a series of rules and regulations have been formulated by relevant security departments, especially for construction sites, coal mines, oil fields, factories, and other places.

From the above introduction, we know that personnel of any identity must wear safety helmets as long as they enter the production site. So, the helmet not only represents the existence of worker in a specific area but also reflects the different work types and worker identities in a certain industry. For example, in the power industry, white helmet represents the leader, blue helmet represents the management personnel, yellow helmet represents the construction worker, and red helmet represents the outsider. In the construction industry, red helmet represents the leader,

yellow helmet represents the ordinary worker, blue helmet represents the technician, and white helmet represents the manager or safety supervisor. In the coal mine industry, red helmet represents the technician or leader, yellow helmet represents the security worker, and black helmet represents the production worker. It can be seen that for each industry, the worker identities can be directly identified through their helmet colors, which is of great significance to the safety monitoring management and personnel scheduling of the enterprise.

To enhance workers' self-protection consciousness and ensure their personal safety, intelligent identification of safety helmet wearing is crucial for improving the safety management quantity of factories and supervisors [3]. At the same time, compared with traditional labor-intensive and time-intensive management methods, intelligent vision-based identification is helpful to promote the automation of industrial production. In the past years, some researchers have developed related technologies to identify safety helmets. Since the human face is the most prominent part near the helmet and the application of face recognition has become more and more mature recently, some scholars studied the discrimination algorithm about helmet wearing through face positioning. For instance, Chen et al. [4] introduced a safety monitoring system based on multiattribute face recognition to meet the safety detection of intelligent factory. Zhang et al. [5] determined whether the head wears a helmet by locating the upper area of the face and extracting its features of color and shape. Although the face-based methods have strong robustness and high accuracy for helmet detection, they cannot identify the helmet when the worker's back facing the camera or his face is obscured. Currently, with the rapid development of the deep learning (DL) and computer vision (CV), algorithms that directly rely on the convolutional neural network (CNN) for helmet recognition and detection are emerging. Huang et al. [6] used the YOLOv3 model to replace manual monitoring of site safety regulations to detect safety helmet. Zhou et al. [7] proposed a helmet detection algorithm based on the attention mechanism (AT-YOLO) for objects with small scales and obstructions. Wang et al. [8] presented a novel SHW detection model on the basis of improved YOLOv3 to heighten the capability of target detection on the construction site. Li et al. [9] developed a method based on the SSD-MobileNet algorithm for real-time detection of a safety helmet at the construction site. Wang et al. [10] introduced an approach to train and evaluate 8 DL detectors based on YOLO architectures, and samples include 4 colors of helmets, person, and vest. Chen and Demachi [11] provided a novel solution to identify improper use of personal protective equipment (PPE) by combining the DL detector and individual detector that uses geometric relationships analysis. Xiong and Tang [12] presented an extensible pose-guided anchoring framework for detecting proper use of PPE. Wu et al. [13] designed a one-stage system based on CNN to automatically monitor whether construction personnel are wearing hardhats and identify the corresponding colors. Fang et al. [14] used the high precision, high speed, and widely applicable Faster R-CNN method to detect construction workers' NHU in different construction site conditions. Long et al. [15] presented a DL approach for accurate safety helmets wearing detection by employing a single shot multibox detector. It can be seen from the above research that the integration of CV and DL technology has achieved remarkable results in the industrial field in recent years and has been maturely used in the application of helmet detection and recognition. However, for places with poor lighting and environmental conditions, such as underground coal mines, workers are not clear and sample collection is difficult. Especially for small sample problems, the poor fitting effect of the CNN-based training model leads to unsatisfactory detection results. In addition, for the image played by the video frame, the CNN method has high requirements on the detection time and equipment performance, so it has a higher application cost.

In order to avoid the shortcomings of the discussed methods, Mneymneh et al. [16] created an integrated framework that can automatically and efficiently detect any noncompliance with safety rules, which is achieved by isolating construction workers from the captured scene and then detecting the hardhat in the identified region of interest (RoI). Shine and Jiji [17] designed an automated system to identify motorcyclists without a helmet from traffic surveillance videos, which uses a two-stage classifier to extract motorcycles from surveillance videos, and the detected motorcycles are further fed to a helmet identification stage. Yogameena et al. [18] also presented a system for automatically detecting the motorcyclists with and without safety helmets, which segment the foreground object and detect the motorcycle, and then identify the motorcycle with and without helmet. Wu and Zhao [19] put forward an intelligent vision-based approach that used motion detection and pedestrian detection as the foundation technique to identify safety helmet. This kind of stage-based methods are not easily affected by the factors such as scale changes, sample size, and the posture of helmet carrier, so it has better practicability and higher identification efficiency compared with the methods previously discussed.

It is worth noting that although the approach in [19] is a cost-effective helmet identification algorithm, the usage of conventional histogram of the oriented gradient (HOG) descriptor will not only increase the number of redundant calculations but also cannot express the overall features of the personnel comprehensive. In addition, the layer-by-layer classification efficiency of hierarchical support vector machines (H-SVM) is relatively low. In view of the advantages of stage-based methods and the deficiencies of [19], this study proposes an algorithm for worker identification based on moving personnel detection and helmet characteristics. This can not only detect whether a worker wears a safety helmet on his head but also identify different worker identities through the helmet colors, and so, it has important research significance for the worker safety management in industrial production sites.

The major contributions of this study can be described as follows:

(1) Compared with popular pure CNN-based methods, this research provides a simple and easy to implement scheme for worker identification, which can be applied to a variety of application scenarios and small sample datasets. In addition, it is more effective and robust for multiscale workers, angle changes, and occlusions.

(2) We have designed a multiscale histogram of oriented gradient (MHOG) descriptor to extract features of RoI images, which can reduce the calculation complexity of single-scale HOG descriptor while extracting comprehensive information of RoI images

(3) We have provided a method of locating the worker's head by the OpenPose model, which is simple and effective, and is not affected by worker's posture

(4) We have proposed a method of head images identification based on the transfer learning network, which can identify different head samples at one time with a small number of training samples.

The rest of this study is organized as follows: Section 2 presents the overall architecture of the worker identification algorithm. Section 3 describes the methodology of personnel identification. Section 4 describes the methodology of helmet identification. Section 5 shows and discusses the experimental results. Section 6 concludes the current work and direction for future work.

2. Framework of the Worker Identification

This section presents the overall architecture of the proposed algorithm, as shown in Figure 1. The architecture is mainly composed of three parts: motion extraction, personnel identification, and helmet identification.

2.1. Motion Extraction. Extraction of foreground objects is necessary for many advanced image analysis tasks.

The application scenario of this study aims at the static background video image, and the motion object is the foreground object we are interested in this section. The motion detection is to extract objects that have changed spatial positions in a video. Over the years, scholars have proposed many different motion detection methods, mainly including the frame difference method, optical flow method, background subtraction, feature matching method, KNN, and variants of these methods. Among them, the background difference method is easy to implement, has fast detection speed and high accuracy, and can provide more comprehensive features of the motion area. The Gaussian mixture model (GMM) is the representative of the background subtraction method, which can get relatively ideal detection result. The process of motion extraction is shown in Figure 2; the GMM is first used to detect the motion objects in the video frames. Then, the morphological operations are applied to the resulting motion mask to eliminate noise. Finally, to further optimize the detection result of motion objects, blob analysis is adopted to detect groups of connected pixels that are likely to correspond to moving objects. So far, the motion mask has been extracted. In order to convenient the subsequent processing of the motion image, the detected motion pixel coordinates are used as the benchmark to draw rectangular region on the original frames, and the RoI images of motion extraction are obtained with the size of 256 pixels × 128 pixels.

2.2. Personnel Identification. Since not all the motion objects extracted by GMM are personnel, the task in this section aims to separate personnel objects from nonpersonnel objects, which can be treated as the classic issue of pedestrian detection. Considering that there may exist problems of poor lighting conditions and uneven illumination in the industrial environment, to more accurately identify personnel, the preprocessing operation of brightness [20] is performed on the RoI images before personnel identification.

HOG feature [21] is the most widely used pedestrian feature descriptor; it has achieved a detection success rate of nearly 100% on the MIT pedestrian database and has reached a detection rate of approximately 90% on the INRIA pedestrian database that includes changes in perspective, illumination, and background. Detailed and comprehensive features are more helpful to distinguish personnel from nonpersonnel; this study designs multiple block scales [22] to extract MHOG vectors that contain global features and local features. After that, the SVM classifier [23] is trained through MHOG vectors to classify the RoI images, and only the positive sample images are reserved.

2.3. Helmet Identification. The helmet identification is the core content of worker identification in this article, and the wearing of helmets directly reflects the workers' identities and their safety status. The focus in this section is to extract and identify the workers' head images, so as to realize the worker identification by helmet features. In order to facilitate the extraction and analysis of helmet features, the first step is to extract the workers' head images using the OpenPose model [24] and personnel mask and then resize the head images with the size of 64 pixels × 64 pixels. Furthermore, a transfer learning network based on GoogLeNet [25] is established to extract the head image features and identify whether workers wear helmets and the helmet colors.

3. Methodology of Personnel Identification

3.1. Multiscale Histogram of Oriented Gradient (MHOG). For the original HOG, when a block with the step size of one cell scans the image, it will cause any adjacent blocks to overlap each other, which means that redundant information will be brought into the final HOG feature vectors. However, if there is no overlap between the scan windows, the gradient histogram information on the spatial structure between adjacent blocks cannot be obtained, so the HOG features of the RoI image will be incomplete. For this reason, this section will introduce the MHOG descriptors to solve the problem. The core idea of MHOG is to decompose the

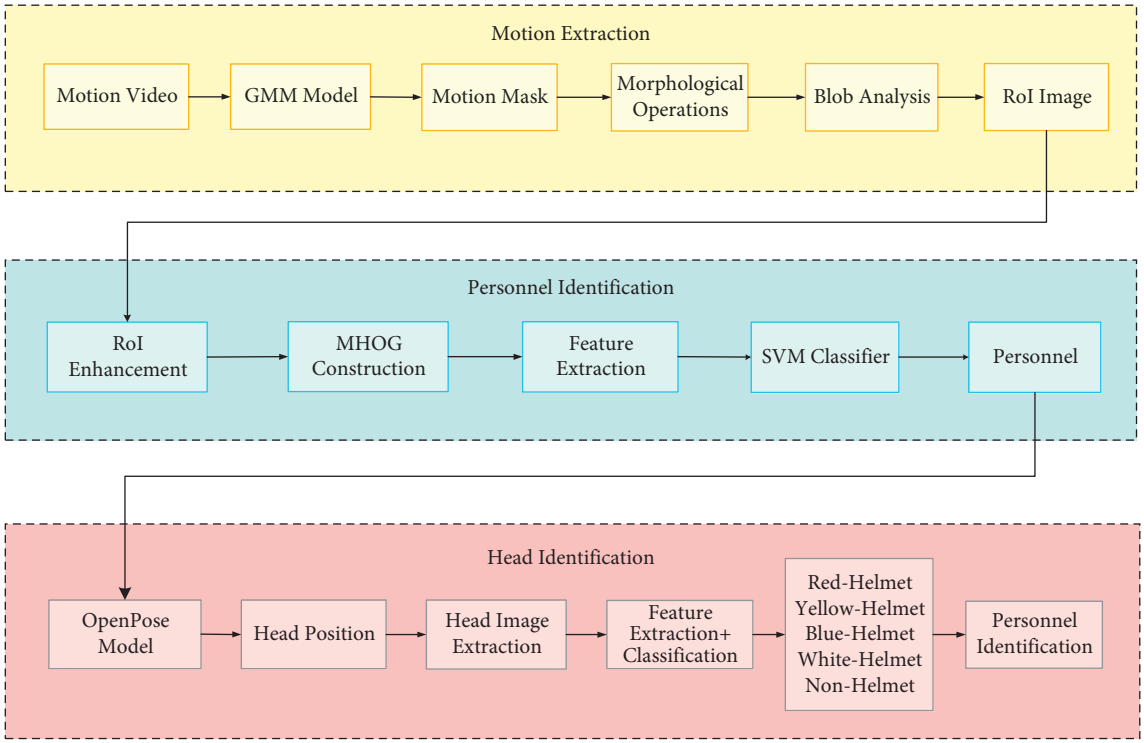


Figure 1: The architecture of the worker identification algorithm.

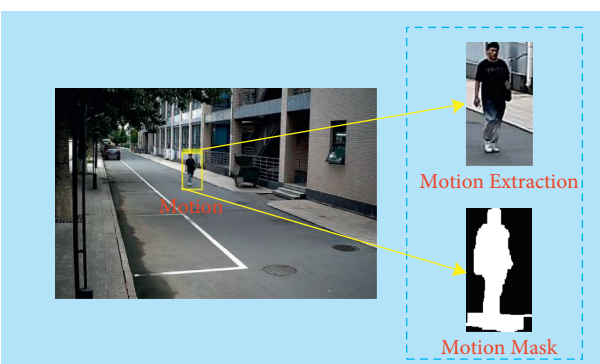


Figure 2: Motion extraction using the GMM model.

image in the form of a variety of different block scales, the large-scale HOG represents the global overall features of the object, and the small-scale HOG represents the local detail features of the object; then, the HOG feature vectors at each scale are connected to generate the MHOG feature descriptor.

The theoretical idea of MHOG is shown in Figure 3. The image is divided into blocks of different scales, and the cell size is a quarter of the block, that is, its length and width are all halved on the block. It can be seen that for an image with the size of 128 pixels × 64 pixels, if the block size of scale 1 is set as its own size of 128 pixels × 64 pixels, the cell number is 4 and the cell size is 64 pixels × 32 pixels. We specify that the proportional relationship between two adjacent block scales is that the length and width of the next block are half of the previous block, so the block size of scale 2 is 64 pixels × 32 pixels, whose cell number is 16, cell size is 32 pixels × 16 pixels; the block size of scale 3 is 32 pixels × 16 pixels, cell number is 64, cell size is 16 pixels × 8 pixels, and so on. Since each cell of HOG produces a 9-bin histogram, the image at scale 1 generates 36 feature vectors, which is defined as F_i^{Scale1} ($i = 1, 2, \ldots, 36$); the image at scale 2 generates 144 feature vectors, which is defined as F_j^{Scale2} ($j = 1, 2, \ldots, 144$); the image at scale 3 generates 576 feature vectors, which is defined as F_k^{Scale3} ($k = 1, 2, \ldots, 576$); and the image at scale 4 generates 2304 feature vectors, which is defined as F_l^{Scale4} ($l = 1, 2, \ldots, 2304$).

We know that the cell size of scale 3 and scale 4 is 16 pixels × 8 pixels and 8 pixels × 4 pixels, respectively, and 8 pixels × 8 pixels cell size of HOG is between scale 3 and scale 4. Obviously, MHOG feature dimensions at scale 3 and even scale 4 are all smaller than 3780 of HOG; moreover, the total dimensions number (3060) of MHOG at the four scales is also less than 3780 of HOG. Therefore, we take the connection result of the feature vector at four scales as the feature vector F_m^{MHOG} of MHOG, which can be expressed as

$$F_m^{\text{MHOG}} = \left[F_i^{\text{scale1}}, F_j^{\text{scale2}}, F_k^{\text{scale3}}, F_l^{\text{scale4}}, \cdots\right], \tag{1}$$

where m is the highest scale level of 4, and i, j, k, and l are the feature dimensions at each scale, respectively.

3.2. Support Vector Machine (SVM). As a representative algorithm in the machine learning field, SVM has developed rapidly since it was proposed and has achieved excellent results in many application scenarios. SVM is a two-class classifier model, whose task in worker identification is to separate F_m^{MHOG} of positive samples (personnel images) from the negative samples (nonpersonnel images). The basic idea of SVM learning is to solve the separation hyperplane that

can correctly divide the training dataset and has the largest geometric interval. As shown in Figure 4, w is the normal to the hyperplane, $\|w\|$ is the Euclidean norm, $|b|/\|w\|$ is the perpendicular distance from the hyperplane to the origin, and $w \cdot x + b = 0$ is the separating hyperplane. For a linearly separable dataset, there are infinitely many such hyperplanes, but the separating hyperplane with the largest geometric interval is the only one. The solution method of SVM is as follows:

Given, a training dataset on the feature space $T = \{(vn, \ln)\}$, where vn denotes the number of MHOG features, and $\ln \in \{1, -1\}$ denotes the two class labels of personnel and nonpersonnel. First step aims to construct and solve the convex quadratic programming problem:

$$\min_{\alpha} \frac{1}{2} \sum_{a=1}^{n} \sum_{b=1}^{n} \alpha_a \alpha_b l_a l_b (v_a \cdot v_b) - \sum_{a=1}^{n} \alpha_a, \tag{2}$$

$$\text{s.t.} \sum_{a=1}^{n} \alpha_a l_a = 0. \tag{3}$$

From this, the optimal solution $\alpha^* = (\alpha_1^*, \alpha_2^*, \ldots, \alpha_n^*)^T$ can be obtained. Second step aims to calculate w^* and b^* from the following formula:

$$b^* = l_b - \sum_{a=1}^{n} \alpha_a^* l_a (va - vb). \tag{4}$$

The third step is to achieve the solution of the separating hyperplane as follows:

$$w^* \cdot x + b^* = 0. \tag{5}$$

The classification decision function is

$$f(x) = \text{sign}(w^* \cdot x + b^*). \tag{6}$$

4. Methodology of Helmet Identification

4.1. Head Image Extraction. The OpenPose model [24] is a human posture recognition model developed by Carnegie Mellon University (CMU) based on CNNs and supervised learning and can realize posture estimation of human movements, facial expressions, and limb movements. It is suitable for single and multiple people and has excellent robustness. In order to accurately locate the worker's head, OpenPose is used to estimate the joint points of the worker body, and the joint point related to the head is extracted. In general, each worker has a set of joint points, but when two workers are severely occluded, sometimes only the joint points of the worker who is not occluded can be obtained. Therefore, it is necessary to perform detailed analysis on the worker images with one set of joint points. The schematic diagram of the human body with 18 joint points by OpenPose is shown in Figure 5, the right side of the figure lists the name of the human body organ corresponding to the number of each joint point.

For the worker image with one set of joint points, consider that different angles and postures of the worker may cause symmetrical joints (such as the shoulders, wrists, and knees) to be hidden to varying degrees, and the absence of front face makes head joint points undetectable. Hence, as shown in Figure 6(a), only the position of worker neck (no.1 joint point) is used as a reference, mark this point as $p1$ and make a horizontal straight line through it, locate the highest point $p2$ of the worker mask in the vertical direction passing through the no. 1 joint point, and make a horizontal straight line through point $p2$; these two horizontal straight lines form a horizontal band-shaped area. Then, search for the two outermost points $p3$ and $p4$ on the left and right sides of the worker mask in the upper half of the horizontal strip area and draw two straight lines in the vertical direction through these two points to form a vertical area. Next, analyze the changing characteristics of the worker mask boundary in the upper half of the common area where the two bands intersect. If there is only one highest point on the mask boundary and the highest point in the horizontal direction only falls but does not rise along both sides of the point $p2$, the image is regarded as a true single-worker image. As shown in Figure 6(b), the worker's head image is extracted from the blue dotted box formed by points of $p1$, $p2$, $p3$, and $p4$ shown in Figure 6(a). If the highest point in the horizontal direction first falls and then rises along one side of point $p2$, the image is regarded as a double-worker occlusion image. As shown in Figure 7(a), the lowest point that falls in the middle is marked as $p0$, and the highest point that rises later is marked as $p2$'. Draw a virtual straight line through point $p2$' in the vertical direction and take the line segment $p1$'$p2$' on this straight line so that the length of $p1$'$p2$' is equal to $p1p2$; thereby, the position of $p1$' point is obtained. In addition, take the midpoint $p5$ of points $p2$ and $p2$' in the horizontal direction and move $p5$ to the horizontal line where point $p4$ is located. As shown in Figure 7(b), the head image of the worker who is not blocked is extracted from the blue dotted box formed by points $p1$, $p2$, $p4$, and $p5$ shown in Figure 7(a), and the head image of the worker who is blocked is extracted from the red box frame formed by points $p1$', $p2$', $p3$, and $p0$ shown in Figure 7(a).

For the worker image with two sets of joint points, it also contains two situations: one situation is that the bodies of the two workers are blocking each other, but their heads are independent of each other, and the other situation is that the bodies and heads of the two workers are both blocking each other. First, select one of the no. 1 joint points $p1$, and use the method of true single-worker image shown in Figure 6 to locate the other three points $p2$, $p3$, and $p4$. If the change characteristics of the highest point $p2$ consist with the true single-worker image, the method shown in Figure 6 is used to extract the two head images, respectively. If the change characteristics of the highest point $p2$ consist with the double-worker occlusion image, the method shown in Figure 6 is used to locate the two sets of points $p1$, $p2$, $p4$ and $p1$', $p2$', $p3$, and the method shown in Figure 7 is used to determine points of $p0$ and $p5$. Finally, the two head images are extracted by the two sets of points $p1$, $p2$, $p4$, $p5$ and $p1$', $p2$', $p3$, $p0$ separately.

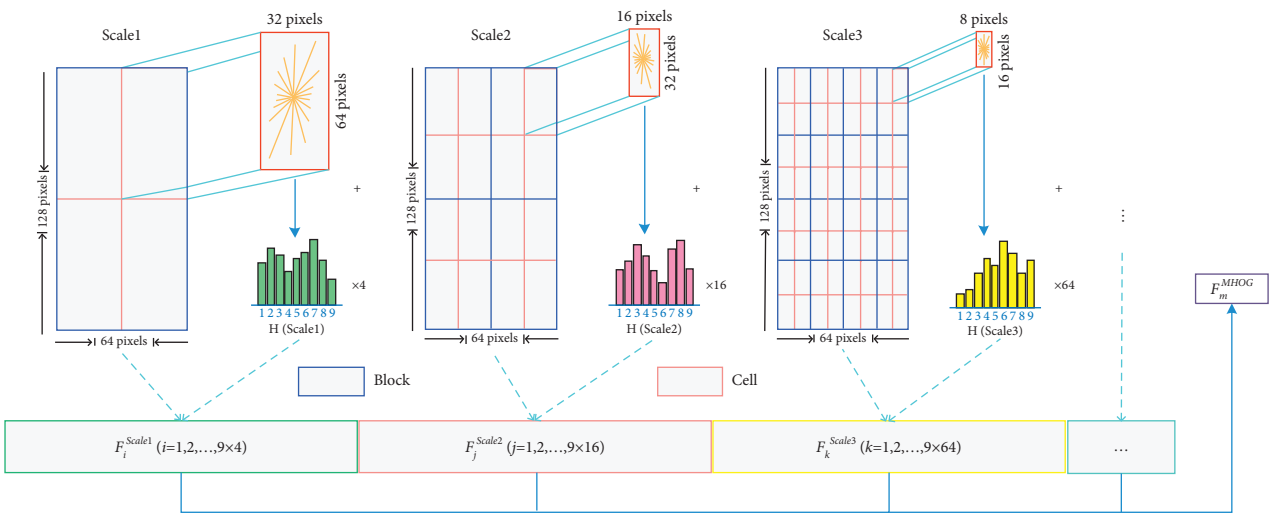


Figure 3: The extraction procedure of MHOG features.

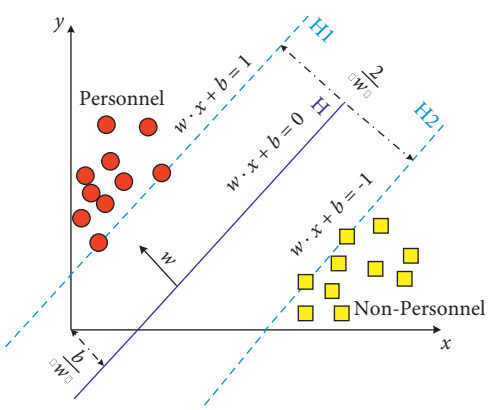


Figure 4: Support vector machine.

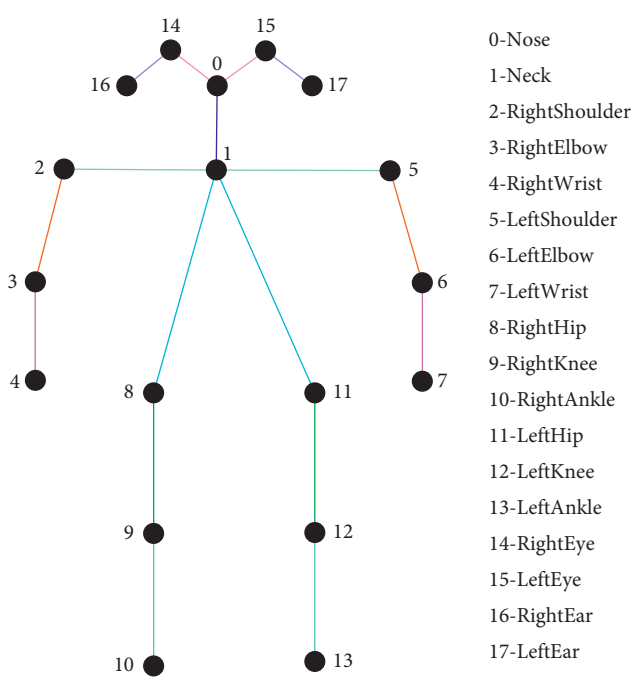


Figure 5: Structure diagram of human body joint point.

4.2. Head Image Identification. In this section, the problem of head image identification is achieved by the CNN network, which inputs the head images and outputs their labels together with classification probabilities. Specifically, the strategy of transfer learning is used to classify and identify the workers' head images by fine tuning the pretrained GoogLeNet. GoogLeNet has been trained on over a million images and can classify images into 1000 object categories, so it has learned rich feature representations for a wide range of images. Fine-tuning a network with transfer learning is usually much faster and easier than training a network with randomly initialized weights from scratch, which can help to quickly transfer learned features to the task of head images identification using a smaller number of training images.

The process of transfer learning based on GoogLeNet is shown in Figure 8. First, according to the requirements of the GoogLeNet input layer, the head images are uniformly adjusted to the size of $224 \times 224 \times 3$, and the training dataset and the testing dataset are allocated by a certain ratio. Then, the final two layers, "loss3-classifier" and "output," in GoogLeNet should be replaced with new layers adapted to the head image dataset. The original fully connected layer is replaced with a new fully connected layer that has number of outputs equal to the number of classes. In addition, to make learning faster in the new layers than transferred layers, the parameter values of WeightLearnRateFactor and BiasLearnRateFactor of the fully connected layer are increased. Next, the appropriate values of training option should be set to train the network. For transfer learning, the features from the early layers of the pretrained network (transfer layer weights) are retained. By setting the initial learning rate to a small value, the learning speed in the transfer layers is slowed down. When training the transfer learning network, one epoch is a complete training cycle for the entire training dataset, so there is no need to set a too large epoch and minibatch size. Finally, the trained network is used to predict and classify the testing dataset, the network will

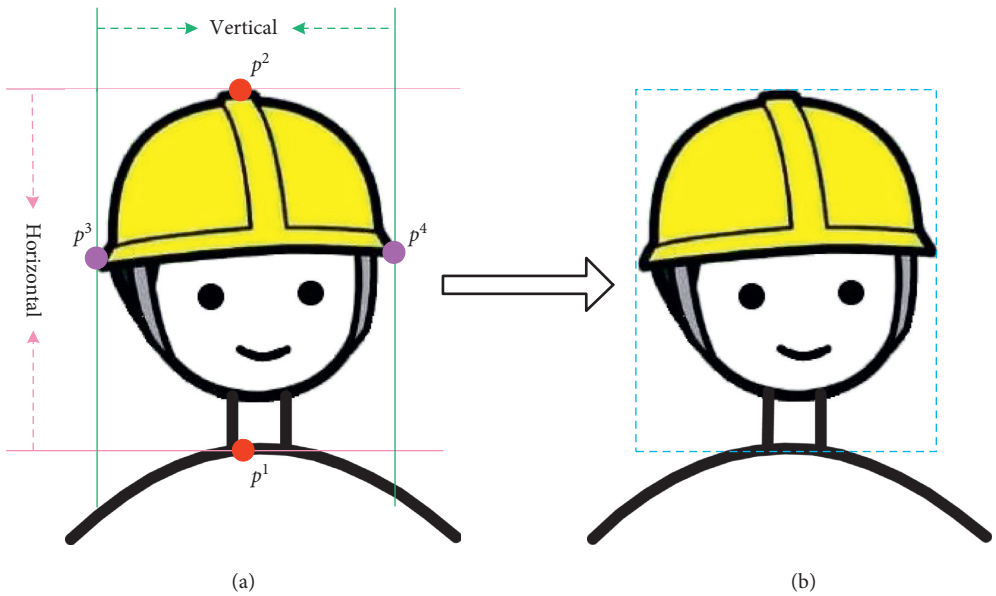


Figure 6: Head image extraction of one set of joint points. (a) Positioning of key points on the head. (b) Head extraction template.

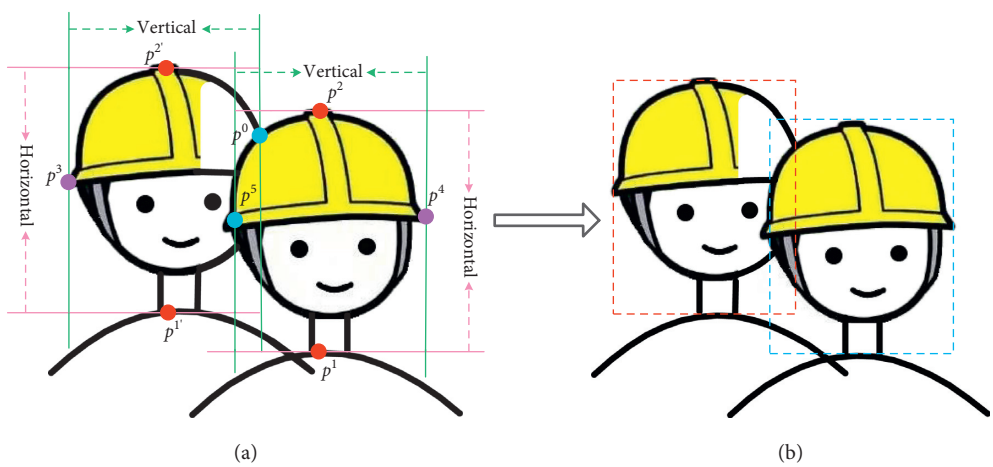


Figure 7: Head image extraction of two sets of joint points. (a) Positioning of key points on the head. (b) Head extraction template.

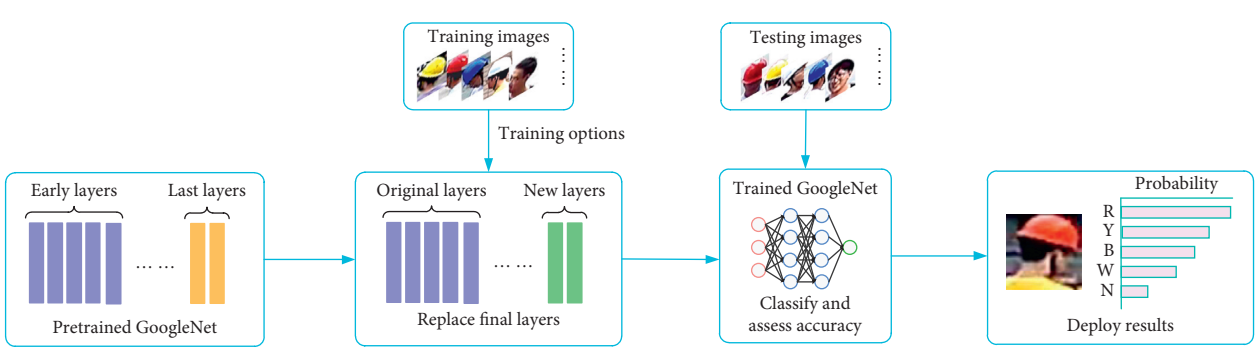


Figure 8: Flowchart of transfer learning.

output a classification probability value corresponding to each label for each test sample, and the label corresponding to the maximum probability value is the identification result of the test sample.

5. Experimental Results and Discussion

5.1. Personnel Identification Experiment. The dataset of personnel identification is self-collected motion videos with the size of 1920 pixels × 1080 pixel, which contains the aspects of different light intensity, distance, angle, and occlusion. Figure 9 shows a small part of the motion objects, in which the personnel images are regarded as the positive samples, and the nonpersonnel images are regarded as the negative samples. We selected a total of 2365 samples from the dataset for experiment, of which there are 1830 positive samples and 535 negative samples.

Table 1 provides the detailed quantity distribution of personnel images in the four aspects of illumination, distance, occlusion, and angle. The illumination conditions include two types of light and dark, the distance includes three types of far, near, and medium, the occlusion includes three types of light, heavy, and no occlude, and the angle includes three types of front, back, and side.

The accuracy of worker identification will directly affect the accuracy of helmet identification. In order to obtain better worker identification results, the data samples are set to three distribution ratios for experimental comparison, namely, 60% training and 40% testing, 70% training and 30% testing, and 80% training and 20% testing. Meanwhile, to improve the performance of the classifier, the automatic hyperparameter optimization method is adopted to train the SVM.

In this study, three conventional indicators [26] of precision, recall, and accuracy are used to evaluate the performance of the proposed algorithm, whose calculation formulas are as follows.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \tag{7}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \tag{8}$$

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}}, \tag{9}$$

where TP represents the positive sample that is correctly identified, TN represents the negative sample that is correctly identified, FP represents the positive sample that is incorrectly identified, and FN represents the negative sample that is incorrectly identified.

Table 2 provides the results of personnel identification with the method of MHOG + SVM; it can be seen that the three indicators that corresponding to MHOG (6, 4) perform best, and the recall of 99.73% means that the positive samples can basically be correctly identified. At the same time, in order to better illustrate the performance of MHOG, Table 3 provides the identification results of the HOG + SVM method in the (6, 4) allocation mode for comparison. Obviously, the precision and accuracy of the MHOG (6, 4) mode and MHOG (7, 3) mode are better than that of HOG + SVM, and recall of the MHOG (6, 4) is equal to that of the HOG (6, 4). This result proves that MHOG has obvious advantages over HOG, and on the basis of HOG, MHOG can effectively reduce the amount of calculation caused by feature dimensions.

In Table 2, the relatively low precision is due to the large FP, and large FP will inevitably reduce the accuracy. But it is worth noting that in Section 4.1, if the OpenPose does not detect any joint point information in the personnel images, the incorrectly classified positive samples can be further corrected into the true negative samples. Therefor, the FP samples of MHOG (6, 4) can be filtered out through the process of worker head extraction. Table 4 provides the identification result of MHOG (6, 4) after the optimization of Section 4.1. It can be seen that the values of precision and accuracy have been effectively improved on the basis of Table 2, and all the indicators prove that the final identification result is excellent.

5.2. Helmet Identification Experiment. All the head samples used in this experiment are obtained from Section 4.1. Figure 10 shows some samples that participated in the experiment, which contains five classes of head images, namely, red helmet, yellow helmet, blue helmet, white helmet, and no helmet.

Since the experiment purpose is to classify and identify any class of the head images, the experimental results should not be affected by the combination types of head images. To verify this view, Table 5 provides all combinations of sample types, which can be summarized as one type of sample, two types of samples, three types of samples, four types of samples, and all five types of samples. In the table, N represents not wearing a helmet, Y represents a yellow helmet, R represents a red helmet, B represents a blue helmet, and W represents a white helmet.

In addition, the experiment selected 65% of the total number of samples as training samples and the remaining 35% as testing samples. The number distribution of various classes of head images in the training dataset and testing dataset is given in Table 6.

The parameter values of the transfer learning network are given in Table 7.

For the 31 types of the helmet combinations in Table 5, except for the 5 single types, the remaining 26 multiple types are identified, respectively, and the identification results are given in Table 8. In the table, the column data corresponding to Y, R, B, W, and N represent the number of head samples correctly identified, and the last column is the identification accuracy of the proposed method. The identification accuracy is defined as the number ratio of correctly identified samples to the total samples.

It can be seen from Table 8 that 7 of the 10 types for the two colors combination reach 100%, only 2 of the 10 types for the three colors combination reach 100%, and all the 6 types of four colors combination and five colors combination dose not reach 100%. The mean accuracy of the two

(a) (b)

Figure 9: The data samples for personnel identification. (a) Positive sample. (b) Negative sample.

Table 1: Number distribution of the personnel images.

Sample	Illumination		Distance			Occlude			Angle		
	Bright	Dark	Far	Medium	Near	Light	Heavy	No	Front	Back	Side
Number	915	915	785	628	417	190	97	1543	506	587	737

Table 2: Identification results of MHOG + SVM.

Method	Precision (%)	Recall (%)	Accuracy (%)
MHOG (6, 4)	87.53	99.73	88.79
MHOG (7, 3)	87.52	99.64	88.72
MHOG (8, 2)	86.87	99.45	87.95

Table 3: Identification results of HOG + SVM.

Method	Precision (%)	Recall (%)	Accuracy (%)
HOG (6, 4)	87.43	99.73	88.69

Table 4: Identification results of MHOG + SVM after optimization.

Method	Precision (%)	Recall (%)	Accuracy (%)
MHOG (6, 4)	99.45	99.73	99.37

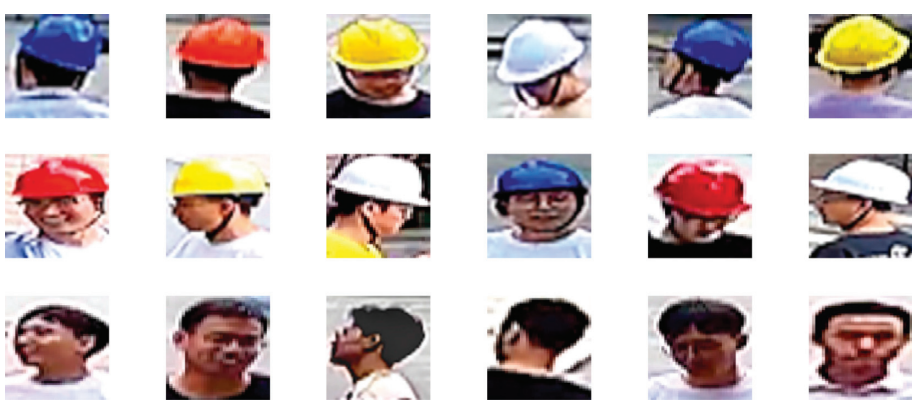


Figure 10: Data samples of workers' heads.

Table 5: Identification of types of helmets.

One	Two	Three	Four	Five
	RN	YRN		
	YN	YBN		
Y	BN	YWN	YRBN	
R	WN	BWN	YRWB	
B	RY	BRN	YBWN	YRBWN
W	RW	RWN	RWYN	
N	RB	YRB	RBWN	
	YB	YBW		
	YW	RBW		
	BW	YRW		

Table 6: Number distribution of workers' head samples.

Sample	Red helmet	Yellow helmet	Blue helmet	White helmet	No helmet
Training	496	205	174	199	157
Testing	267	110	94	107	85

Table 7: The parameter settings of transfer learning training.

Parameter	Value
MiniBatchSize	32
MaxEpochs	6
InitialLearnRate	0.0001

Table 8: Values of helmet identification accuracy.

Classes	Head sample					
	Y	R	B	W	N	Accuracy (%)
RN	×	267	×	×	85	100
YN	110	×	×	×	85	100
BN	×	×	92	×	84	98.32
WN	×	×	×	105	85	98.96
RY	109	266	×	×	×	99.47
RW	×	267	×	107	×	100
RB	×	267	94	×	×	100
YB	110	×	94	×	×	100
YW	110	×	×	107	×	100
BW	×	×	94	107	×	100
YRN	110	265	×	×	84	99.35
YBN	110	×	93	×	84	99.31
YWN	109	×	×	105	84	98.68
BWN	×	×	92	106	84	98.60
BRN	×	266	92	×	85	99.33
RWN	×	267	×	106	84	99.56
YRB	109	267	93	×	×	99.58
YBW	110	×	93	106	×	99.36
RBW	×	267	94	107	×	100
YRW	110	267	×	107	×	100
YRBN	110	266	93	×	85	99.64
YRWB	108	267	93	107	×	99.48
YBWN	109	×	91	106	84	98.48
RWYN	110	267	×	105	84	99.47
RBWN	×	266	91	106	83	98.73
YRBWN	108	265	90	106	84	98.49

colors is 99.68%, the three colors is 99.38%, and the four colors is 99.16%. Although the identification accuracy of the proposed method shows a downward trend with the increase of helmet types, its overall average accuracy is as high as 99.42%, and the identification accuracy is basically stable with various test schemes. This shows that the algorithm has strong robustness and adaptability and can well identify head images of different colors, angles, and postures.

In addition, the method in [19] and the proposed method are the same type of the helmet identification algorithm, so [19] is used as the comparison algorithm. For the head images identification, the tree structure of H-SVM is used in [19] to classify the samples layer by layer, and each layer classifies one class of samples. The test results of the H-SVM classifier on our dataset show that for the first-level classification, B-N + R + Y + W has the highest classification accuracy of 97.87% for the blue helmets among the five combinations. For the second-level classification, R-N + Y + W has the highest classification accuracy of 95.88% for the red helmets among the four combinations. For the third-level of classification, N-Y + W has the highest classification accuracy of 97.65% for the no-helmet samples among the three combinations. For the last level of classification, the YW has the classification accuracy of 98.18% and 97.20%, respectively. The average classification accuracy of H-SVM for the five layers is 97.36%.

The comparison shows that the average identification accuracy of the H-SVM method is 2.06% lower than that of the proposed method. More importantly, the H-SVM method requires hierarchical processing, and each level can only identify one type of head image, while the proposed method can identify multiple types of helmet colors at one time. In summary, the proposed method is not only higher than the H-SVM method in the identification accuracy but also far ahead of [19] in identification efficiency.

6. Conclusions

In this study, an intelligent vision-based method for worker identification is proposed. This method is implemented through three parts: motion extraction based on the GMM model, personnel identification based on MHOG and SVM, and helmet identification based on the OpenPose and transfer learning CNN.

The proposed method can be successfully used to identify the worker types in a variety of industrial scenarios, such as construction site, factory workshop, power construction site, and interior decoration. In addition, the method can be applied to small-scale datasets, and the identification accuracy cannot be affected by helmet occlusion, worker size, and different lighting conditions. The testing results on our self-collected dataset show that the accuracy values vary little under different types of identification situations, and the mean value of the accuracy reaches 99.43%.

It is worth considering that the visual features of the self-collected dataset in this study are relatively ideal, so the identification results are relatively excellent. Our next work will capture some worker videos in actual industrial environments and conduct in-depth research on the worker images to promote the application of vision-based worker identification.

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 there are no conflicts of interest.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (52074305 and 51874300), the National Natural Science Foundation of China and Shanxi Provincial People's Government Jointly Funded Project of China for Coal Base and Low Carbon (U1510115) and the Open Research Fund of Key Laboratory of Wireless Sensor Network and Communication, Shanghai Institute of Microsystem and Information Technology, and Chinese Academy of Sciences (20190902 and 20190913).

References

[1] W. Fang, L. Ding, H. Luo, and P. E. D. Love, "Falls from heights: a computer vision-based approach for safety harness detection," *Automation in Construction*, vol. 91, no. 7, pp. 53–61, 2018.

[2] H. Li, X. Li, X. Luo, and J. Siebert, "Investigation of the causality patterns of non-helmet use behavior of construction workers," *Automation in Construction*, vol. 80, no. 8, pp. 95–103, 2017.

[3] J. Shen, X. Xiong, Y. Li, W. He, P. Li, and X. Zheng, "Detecting safety helmet wearing on construction sites with bounding-box regression and deep transfer learning," *Computer-Aided Civil and Infrastructure Engineering*, vol. 36, no. 2, pp. 180–196, 2021.

[4] X. Chen, W. Chen, and P. Xu, "Attribute-based Face recognition and application in safety detection of intelligent factory," in *Proceedings of the 2018 International Conference on Computing & Artificial Intelligence (ICCAI)*, pp. 76–80, Chengdu, China, March 2018.

[5] G. Zhang, L. Lv, and L. Dan, "The method for recognizing recognition helmet based on color and shape," in *Proceedings of the 2017 5th International Conference on Machinery, Materials and Computing Technology (ICMMCT)*, pp. 1219–1223, Beijing, China, March 2017.

[6] L. Huang, Q. Fu, and M. He, "Detection algorithm of safety helmet wearing based on Deep learning," *Concurrency and Computation-Practice & Experience*, vol. 33, no. 13, Article ID e6234, 2021.

[7] Q. Zhou, J. Qin, X. Xiang, Y. Tan, and N. Xiong, "Algorithm of helmet wearing detection based on AT-YOLO deep mode," *Computers, Materials & Continua*, vol. 69, no. 1, pp. 159–174, 2021.

[8] H. Wang, Z. Hu, and Y. Guo, "A real-time safety helmet wearing detection approach based on CSYOLOv3," *Applied Sciences-Basel*, vol. 10, no. 19, p. 6732, 2020.

[9] Y. Li, H. Wei, and Z. Han, "Deep learning-based safety helmet detection in engineering management based on convolutional neural networks," *Advances in Civil Engineering*, vol. 2020, Article ID 9703560, 13 pages, 2020.

[10] Z. Wang, Y. Wu, and L. Yang, "Fast personal protective equipment detection for real construction sites using deep learning approaches," *Sensors*, vol. 21, no. 10, p. 3478, 2021.

[11] S. Chen and K. Demachi, "Towards on-site hazards identification of improper use of personal protective equipment using deep learning-based geometric relationships and hierarchical scene graph," *Automation in Construction*, vol. 125, Article ID 103619, 2021.

[12] R. Xiong and P. Tang, "Pose guided anchoring for detecting proper use of personal protective equipment," *Automation in Construction*, vol. 130, Article ID 103828, 2021.

[13] J. Wu, N. Cai, and W. Chen, "Automatic detection of hardhats worn by construction personnel: a deep learning approach and benchmark dataset," *Automation in Construction*, vol. 106, Article ID 102894, 2019.

[14] Q. Fang, H. Li, X. Luo et al., "Detecting non-hardhat-use by a deep learning method from far-field surveillance videos," *Automation in Construction*, vol. 85, pp. 1–9, 2018.

[15] X. Long, W. Cui, and Z. Zheng, "Safety helmet wearing detection based on deep learning," in *Proceedings of the 2019 IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC)*, Chengdu, China, March 2019.

[16] B. Mneymneh, M. Abbas, and H. Khoury, "Vision-based framework for intelligent monitoring of hardhat wearing on construction sites," *Journal of Computing in Civil Engineering*, vol. 33, no. 2, Article ID 04018066, 2019.

[17] L. Shine and C. Jiji, "Automated detection of helmet on motorcyclists from traffic surveillance videos: a comparative analysis using hand-crafted features and CNN," *Multimedia Tools and Applications*, vol. 79, no. 19-20, pp. 14179–14199, 2020.

[18] B. Yogameena, K. Menaka, and S. Saravana Perumaal, "Deep learning-based helmet wear analysis of a motorcycle rider for intelligent surveillance system," *IET Intelligent Transport Systems*, vol. 13, no. 7, pp. 1190–1198, 2019.

[19] H. Wu and J. Zhao, "An intelligent vision-based approach for helmet identification for work safety," *Computers in Industry*, vol. 100, pp. 267–277, 2018.

[20] D. Xuan, G. Wang, and Y. Pang, "Fast efficient algorithm for enhancement of low lighting video," in *Proceedings of the 2011 IEEE International Conference on Multimedia and Expo (ICME)*, July 2011.

[21] N. Dalal and B. Triggs, "Histograms of oriented gradients for Human detection," in *Proceedings of the 2005 IEEE Computer Society Confrrence on Computer Vision and Pattern Recognition (CVPR)*, pp. 886–893, San Diego, CA, USA, June 2005.

[22] X. Tian, H. Bao, and C. Xu, "Improved HOG algorithm of pedestrian detection," *Computer Science*, vol. 41, no. 9, pp. 320–324, 2014.

[23] C. Saunders, M. Stitson, and J. Weston, "Support vector machine," *Computer Science*.vol. 1, no. 4, pp. 1–28, 2002.

[24] Z. Cao, G. Hidalgo, T. Simon, S.-E. Wei, and Y. Sheikh, "OpenPose: realtime multi-person 2D pose estimation using Part Affinity fields," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 1, pp. 172–186, 2021.

[25] C. Szegedy, W. Liu, and Y. Jia, "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.

[26] P. David Martin Ward, "Evaluation: From precision, Recall and F-factor to ROC, informedness, markedness and correlation," *Journal of Machine Learning Technologies*, vol. 2, no. 1, pp. 2229–3981, 2011.

Hindawi
Wireless Communications and Mobile Computing
Volume 2022, Article ID 3622559, 11 pages
https://doi.org/10.1155/2022/3622559

Research Article

Deep Learning with Multisource Data Fusion in Electricity Internet of Things for Electricity Price Forecast

Ke Xie, Yiwang Luo, Wenjing Li, Zhipeng Chen, Nan Zhang, and Cai Liu

State Grid Information & Telecommunication Group Co. Ltd, Beijing 102211, China

Correspondence should be addressed to Cai Liu; noky@foxmail.com

Received 19 October 2021; Accepted 14 December 2021; Published 24 January 2022

Academic Editor: Chi-Hua Chen

More and more IoT (Internet of Thing) devices have been connected to our lives in recent years, making life more convenient. Many countries are also making use of Internet of Thing technology to carry out intelligent electricity network reform. One of the reform goals is balancing the supply and demand of electricity, which has become a top priority. Balancing electricity supply and demand through real-time electricity prices has become an effective way. However, using traditional machine learning models for real-time electricity price prediction requires complex feature engineering, and the results are not satisfactory. Also, the mainstream fusion methods use data-level fusion, which will put very high pressure on communication bandwidth and computer resources. In this paper, an LSTM- (long short-term memory-) based decision level fusion of multisource data is proposed and applied for real-time electricity price prediction on actual electricity price datasets. The method solves the difficulties of traditional machine learning models in dealing with complex nonlinear problems. It achieves local asynchronous processing of multisource data through decision-level fusion, reducing the requirement for bandwidth resources and providing perfect results in real-time electricity price prediction. The experimental results show that the prediction accuracy of the decision fusion prediction model based on LSTM is higher than that of the linear regression algorithm.

1. Introduction

With the development and progress of science and technology, more and more Internet of Things devices are connected to our life. It has provided us with more and more help in intelligent grid [1], intelligent transportation [2], smart home [3], and public safety [4], making life more convenient. Internet of Things devices collect all kinds of sensor information through various information sensors, radiofrequency identification technology, global positioning system, infrared sensors, laser scanners, and other devices and technologies and then connect all kinds of objects. Through various network links, they realize the ubiquitous connection between things and people and the intelligent perception of things and processes in identification and management [5].

In recent years, many countries have carried out intelligent electricity network reform with the help of Internet of Things technology. With the reform of the smart grid, the electricity Internet of Things has also introduced a variety of latest technologies, such as cloud computing [6], artificial intelligence [7], and big data [8], to realize information perception and processing in all links of the electricity Internet of Things [9]. The development of electricity Internet of Things is bound to be accompanied by the access of large-scale terminal equipment, producing a large amount of collected data [10]. Electricity Internet of Things realizes real-time sensing and dynamic control of electricity grid with the help of many sensing devices and various heterogeneous communication networks [11]. As shown in Figure 1, the Internet of Things in electricity system (IOTIES) is a three-tier structure composed of perception, network, and application layers. The perception layer, located at the bottom of the electricity Internet of Things, is the core of items [12]. The transmission layer is located in the second layer of the electricity Internet of Things. As the link between the perception and application layers, it is a center of information processing. Its primary function is to transmit information safely and reliably from the perception layer to the application layer—the application layer, as the top layer of the three-tier structure of the electricity Internet of Things. The

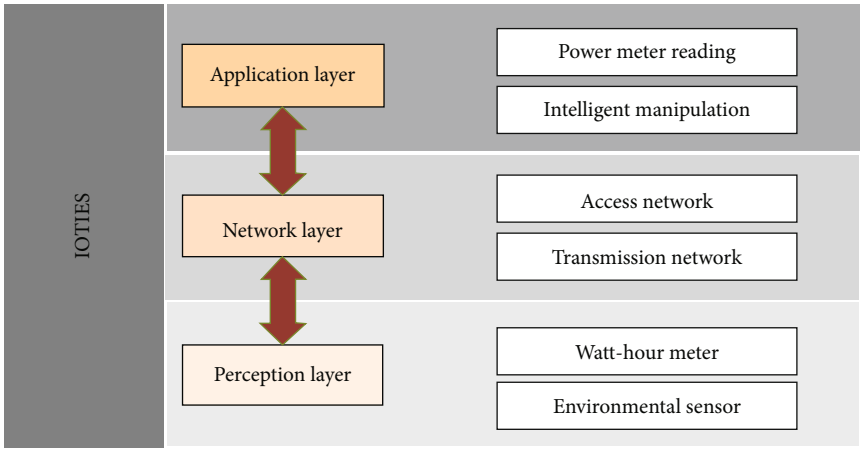


Figure 1: Architecture of Internet of Things in electricity systems.

application layer can calculate, process, and mine the data collected by the perception layer to realize real-time control, accurate management, and scientific decision-making of the physical world [13].

In promoting the development of the electricity Internet of Things, due to the access of large-scale terminal equipment and the collection of massive data in the electricity Internet of Things, some scholars also apply the appropriate data collection methods, cloud communication, and cloud computing. Data acquisition mostly depends on various sensors, such as sensors existing in user equipment, smart meters, and electricity stations. Leikanger et al. [14] proposed a new method to upload sensor data directly to the online cloud through NFC, which solves the disadvantage that early data needs to be submitted to the mobile phone before forwarding. To improve the security of the electricity information acquisition systems and reduce the risk of attack damage, Li et al. [15] proposes an electricity information acquisition system based on blockchain technology. After acquiring multisource data, cloud computing is introduced. A key benefit of connecting edge and cloud computing is achieving high throughput under high concurrent access, mobility support, real-time processing guarantee, and data persistence [16].

In parallel with the vigorous development of Internet of Things for electricity and the market-oriented reform of electricity around the world, coupled with the frequent occurrence of electricity outages due to energy scarcity worldwide, more and more scholars are focusing on predicting real-time electricity prices. Currently, electricity prices in most regions are fixed nondynamic in real-time, thus creating imbalances in the areas and times of electricity consumption. Traditional pricing methods have led to waste of energy and problems affecting grid stability. Balancing electricity consumption by adjusting prices in real time is a more effective way to help electricity companies coordinate the delivery of electricity by accurately forecasting real-time electricity prices. Accurate real-time tariff forecasting can help electricity companies to coordinate the delivery of electricity. It can also be used to cope with electricity shortages or excesses during specific periods, ensuring the safety [17] and stability [18] of the electricity system while achieving the national goal of carbon neutrality [19].

Currently, there are two main types of approaches to real-time electricity price forecasting: traditional machine learning models and neural network models. A part of the research has used traditional machine learning models that consider mostly historical charge sequences, climate, time, and other factors [20–22]. In terms of conventional machine learning models for predicting electricity prices, Azmira et al. [23] proposed a hybrid least squares support vector machine (LSSVM) and bacterial foraging optimization algorithm (BFOA) combined with a multistage optimization approach of LSSVM-BFOA, which can improve the accuracy and efficiency of prediction and provide an ultraconcise solution for electricity forecasting studies. Ding and Ge [24] proposed a new adaptive Kalman filter-based day-ahead electricity price forecasting. Under the condition that the state transfer matrix and the observed noise statistical characteristics of the forecasting model are unknown, the unknown parameters of the forecasting model are estimated based on the electricity market clearing tariff data to ensure that its unit capacity can participate in the market bidding and achieve the goal of maximizing its profit. The results show promising results on the Pennsylvania-New Jersey-Maryland (PJM) electricity market dataset.

In contrast, using traditional machine learning models involves feature engineering, which is a tedious process. Therefore, some scholars have applied the neural network approach to real-time electricity price prediction to get rid of the tediousness of feature engineering. Zou et al. [25] proposed a deep learning model based on stacked autoencoders for electricity price prediction to solve the problem that artificial neural networks are slow to train when the input data is extensive and easily fall into local optimum. The results showed that it could effectively solve the problem that neural networks are difficult to train. Li et al. [26] improved the genetic algorithm- (GA-) based BP neural network prediction algorithm. The traditional BP neural network tends to introduce error signals into local minima. The genetic algorithm can solve this problem by optimizing the weights and thresholds of the BP neural network. The improved

algorithm is used to predict electricity prices, and the results show good prediction results. However, the LSTM algorithm was attempted to be applied to time series problems because of the difficulty of modeling future situations using time-series features in traditional neural networks [27]. The LSTM algorithm can avoid gradient disappearance and also can solve long-term problems.

The research on information fusion methods can be divided into three levels: data-level fusion, feature-level fusion, and decision-level fusion. In data-level fusion, the tiniest information is lost in the fusion process. The rapid development of Internet of Things data collection technology brings severe challenges for all kinds of data collection systems. Shah [28] proposed a deep learning model called tensor deep learning (TDL), which is a proposed higher-order backpropagation algorithm based on the traditional backpropagation algorithm that extends the data from linear space to multiple linear areas and trains the parameters of the proposed model. The experimental results show that the proposed model not only performs well in heterogeneous data fusion. However, data-level fusion has very significant limitations, is highly influenced by the environment, and has very high requirements on communication bandwidth and computer resources.

In contrast, feature-level fusion provides objective information compression of data and solves the difficulties of data-level fusion. Gad et al. [29] proposed a multialgorithm feature-level fusion scheme for secure use of Internet of Things based on iris authentication for the current problem of how sensitive data can be safely used in Internet of Things through online banking, and the results show that the accuracy of verification of the correct client in sensitive applications is significantly improved. In feature-level fusion, it can handle the data coming from heterogeneous sensors. Still, the fused feature vectors are generally of high dimensionality, which will make the later pattern classification more difficult. Data from different sensors can be preprocessed locally first, feature extraction, and pattern classification in decision-level fusion. Decision-level fusion is less sensor-dependent. The participating sensors can be homogeneous or heterogeneous and can process asynchronous information. Decision-level fusion is currently widely used in medical research. Hypoxia in daily life is challenging to recognize in the short term. Still, if the situation is difficult, it may lead to decreased physical function or complete incapacity, so for the hypoxia problem that often occurs in people who work at high altitudes for long periods, Acharya et al. [30] proposed a parallel-based decision-level real-time hypoxia monitoring system, based on blood oxygen saturation and dysfunction at different measurements to build a model, tested on a real dataset, and showed outstanding results. It has also demonstrated its advantages in diagnosing depression. Zhang et al. [31] proposed a decision-level fusion method based on deep forest multimodal data, where a random forest regression model (RFR) was trained separately using heterogeneous data to obtain high-level feature vectors, and the cosine similarity of the two vectors was used as a complementary metric for the modal data. Experimental results on the DAIC-WOC dataset show that this fusion method significantly outperforms other methods and can better identify depressed patients.

The main contributions of this paper are listed as follows:

(1) A distributed decision-level fusion method based on power Internet of Things is proposed to predict real-time electricity prices

(2) At present, electricity data is multisource and heterogeneous. Therefore, a distributed data acquisition method of edge cloud is proposed. It can effectively obtain information and use the edge server for data processing, reduce the communication cost, and improve timeliness and prediction accuracy

(3) The electricity price of different regions is predicted through deep learning. The fusion is realized at the decision-making level to complete multiple regions' unified real-time electricity price prediction

The rest of the paper is organized as follows. Section 2 briefly introduces the architecture, multisource data fusion framework, and system flow chart of electricity Internet of Things. Section 3 describes the LSTM algorithm, decision-level fusion, and experimental evaluation index in detail. The feasibility and validity of the model and algorithm are verified by error comparison in Section 4. Section 5 presents the conclusions and future research.

2. Deep Learning with Multisource Data Decision-Level Fusion

Before the emergence of the Internet of Things, communication infrastructure resources and power system infrastructure resources were not integrated, resulting in low information levels and low infrastructure utilization efficiency. In the development of the Internet of Things, these problems are gradually solved, especially the proposed multisource data fusion architecture. The introduction of multisource data fusion systems under the Internet of Things includes multisource data acquisition topology and multisource data decision-level fusion machine learning model.

2.1. Multisource Data Collection Topology. Figure 2 shows a topology of an edge collecting data and uploading it to the cloud, presenting a tree structure. This topology is mainly composed of a decision center, a central cloud, and an edge cloud, where the edge cloud is connected to multiple wireless transmitters through which different sensor data are acquired. These sensor data sources include user equipment data, hydroelectric electricity plant data, and thermal electricity plant.

In Figure 2, three regions perform edge data acquisition. They rely on various sensors to obtain data, including sensors from hydroelectric electricity plants, thermal electricity plants, user equipment, and environmental collection devices. The electricity usage and environmental information are transmitted to the edge server through wireless transmission devices. Then, the edge server performs edge

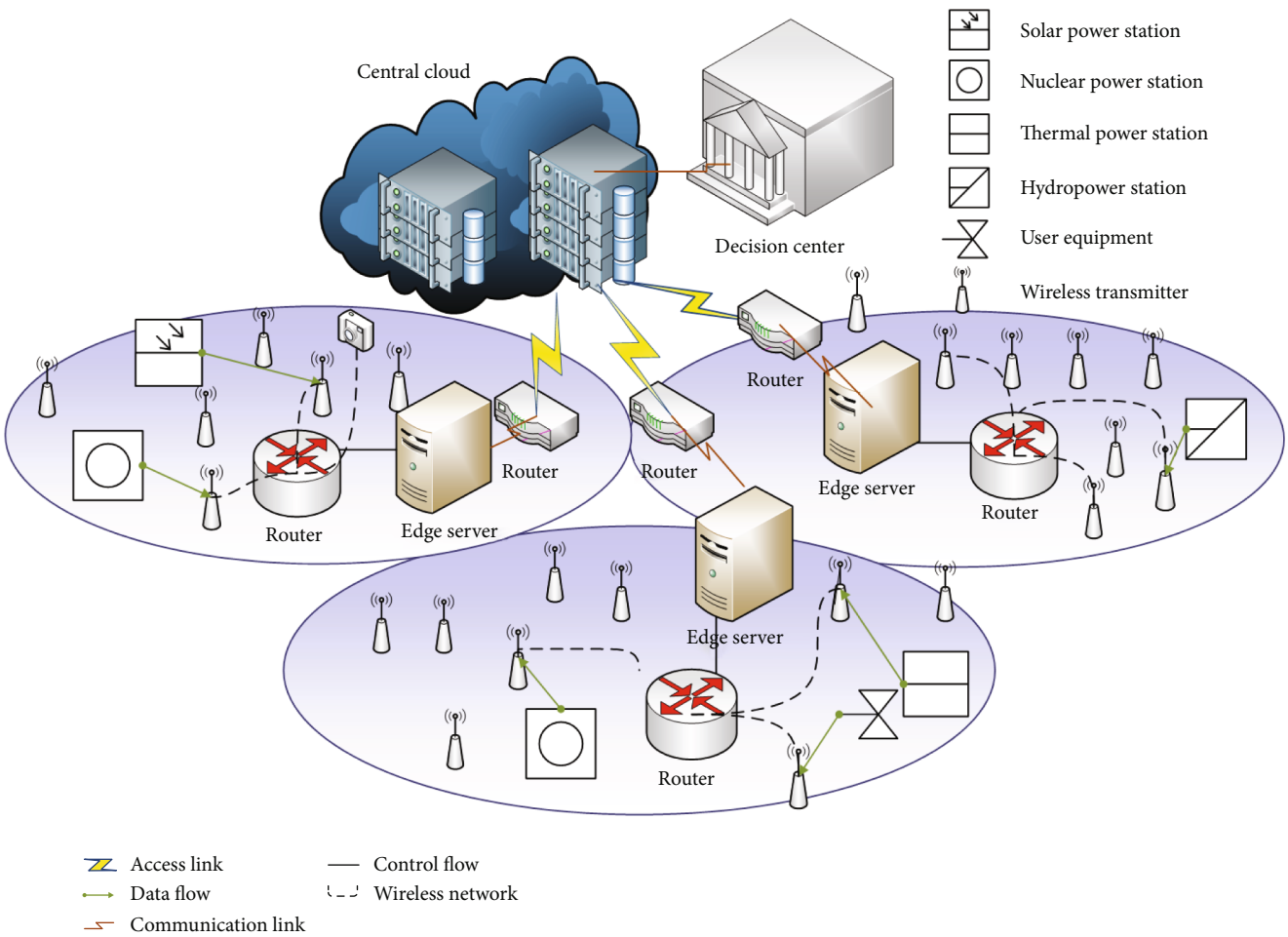


Figure 2: The architecture of electricity Internet of Things.

calculations to obtain the predicted real-time electricity prices for each region. The expected electricity value for each area is then transmitted to the central cloud through a router. The central cloud performs prediction through a decision-level fusion machine learning model to obtain the final real-time electricity price prediction value.

The communication between the central and edge clouds is transmitted in both directions through a wired network. And the edge cloud and each sensor are sent through the wireless network. This design speeds up the transmission and also makes the cost of acquiring various sensor data more favorable.

2.2. Overall Framework Diagram. First, each city in the same province collects historical electricity price information as the original dataset and obtains the input dataset in the Dataset Input module—Data_{city} (city $= A, B, \cdots, N$) through data preprocessing. Then, each city uses Data_{city} to train the LSTM model locally and uses the trained LSTM model to predict the future electricity price of each city–Predict_{city} (city $= A,B,\ldots,N$). Then, the electricity price information predicted by each city is uploaded to the center, and the final forecast result of the future electricity price of the province—PRICE—is obtained through a decision-level fusion algorithm.

After obtaining the prediction results in the first predict module, each city uploads the predicted electricity price information to the cloud center. According to the power load information of each city, the weight function of the decision-making level fusion is formed. After determining the fusion weight, the decision-level fusion calculation is performed on the future electricity price predicted by each city, and the unified forecast result of the future electricity price of the province—PRICE—is obtained. As shown in Figure 3.

2.3. Flow Chart. Figure 4 shows the flowchart of the LSTM-based decision-level fusion of multisource data for the electricity price prediction model. The first step of data preprocessing is to determine whether the information is correct or not, and if there are anomalies, remove the noisy data and fill in the vacant data, followed by data normalization, output the processed data, and divide it into a training set, validation set, and test set. The preprocessing of data can effectively avoid harmful data to the model training. The second step is to train the LSTM model with the training set,

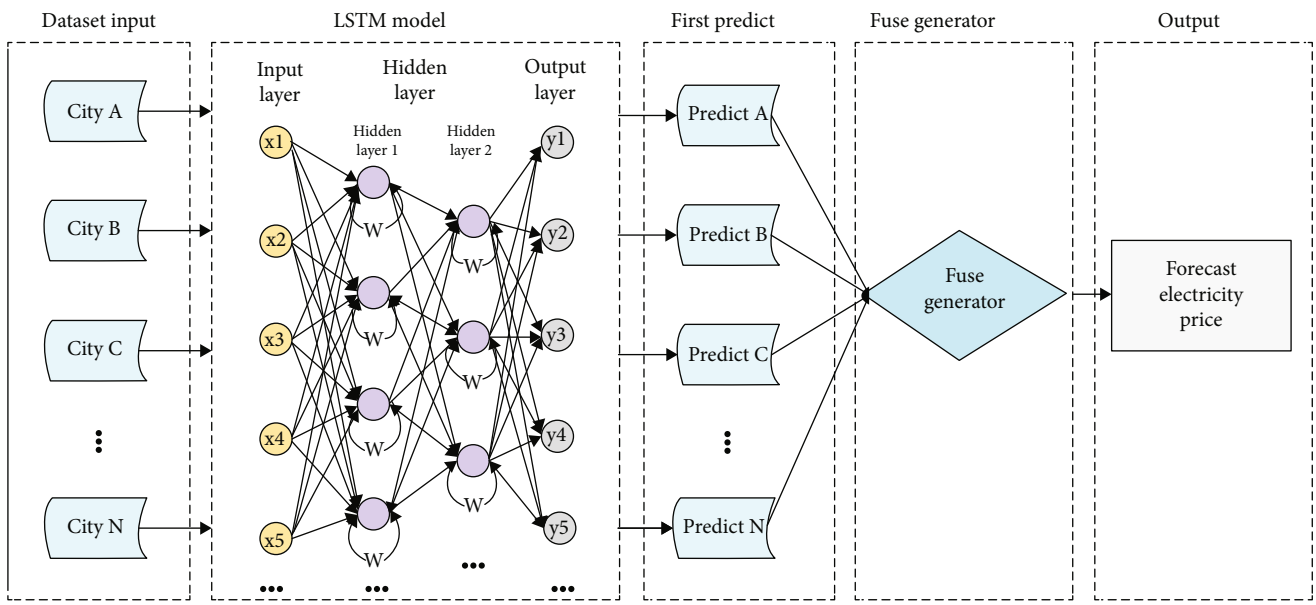


Figure 3: Overall framework diagram of multisource data decision-level fusion based on LSTM.

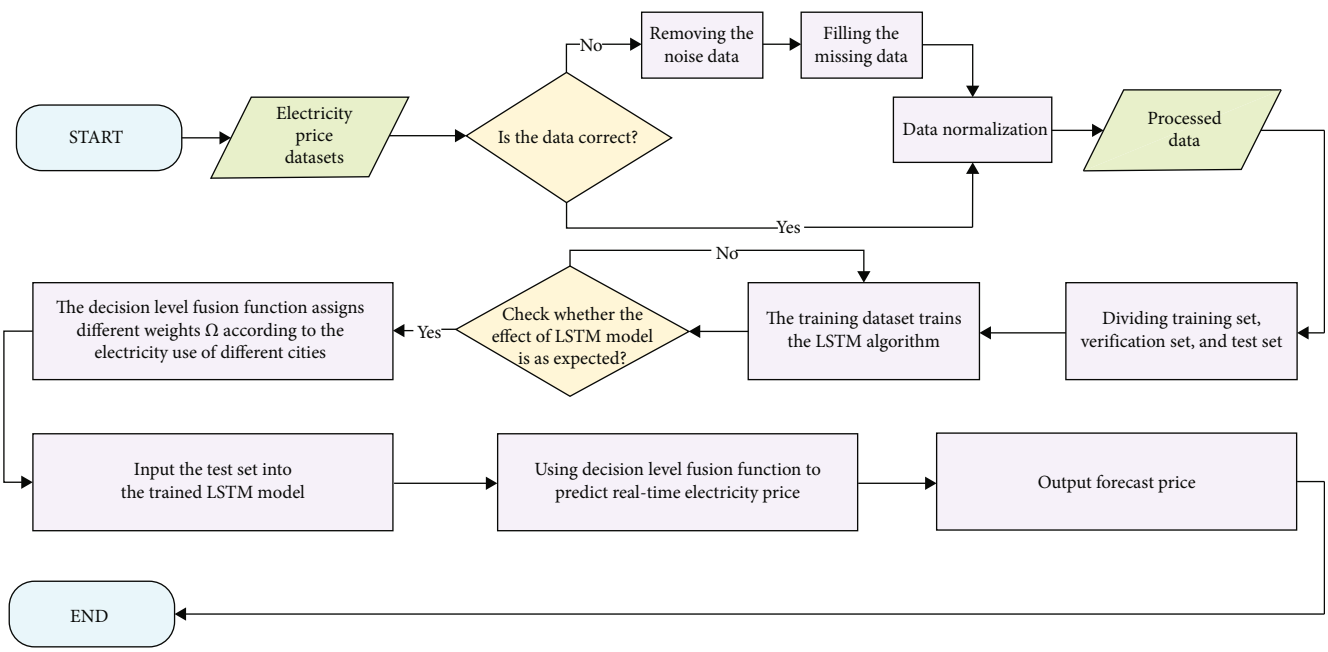


Figure 4: Flow chart of electricity price forecasting based on the LSTM model.

validate the LSTM model with the validation set, retrain it if it does not meet the expectation, and proceed to the next step if it does. The third step is to construct a decision-level fusion algorithm, assigning weights based on the percentage of electricity data allocated to each region and building a decision-level fusion algorithm based on these weights. The fourth step performs electricity price prediction on the test set data, using the decision-level fusion algorithm set up in the third step, and inputting the test set into the trained LSTM algorithm model for the final electricity price prediction.

3. The Specific Implement

This experiment uses the Australian electricity load and electricity price dataset. The original dataset is set to T, and the total data volume is $\mathrm{SUM} = 81649$. The dataset contains eight attributes, including date, hour, humidity, wet bulb temperature, dry bulb temperature, dew ooint temperature, electricity price, and electricity load. The data from 2006 to 2009 is used as the training set, and the data from 2010 is used as the test set.

Limited by the confidentiality of electricity price data, this experiment uniformly sampled the original dataset T. It divided it into three datasets of the same amount, representing the electricity price data of the three cities A, B, and C in the same province.

3.1. Data Preprocessing. The multisource data in training set P_{train} has different dimensions and units, resulting in other effects on model training. Normalize the data, and map the data of different sizes to the interval [0,1], which can avoid errors caused by the difference in dimensions. Since the attributes of the training set P_{train} are all numerical, this article uses the MinMaxScaler function provided in the python third-party library Sklearn to normalize the eight attributes of different dimensions in P_{train}. Assuming that the attribute set X contains n elements, the normalization formula is shown in formula (1):

$$x_i' = \frac{x_i - \min(x)}{\max(x) - \min(x)}, i = 1 \cdots n, \tag{1}$$

where x_i indicate an element in the attribute set X, $x_i \in X$. The min and max functions return the minimum and maximum values in the set X, respectively.

Due to the abnormal noise data in the "electricity price" attribute in the training set, this paper uses third-party libraries pandas and NumPy for data processing to improve the generalization ability of the training model. It fills in the abnormal data with electricity prices less than 0 and greater than 200 in the training set. The training set P'_{train} is preprocessed by normalization and denoising, and P'_{train} is obtained.

3.2. Deep Learning. Because of the characteristics of time series in electricity price data, this paper uses the Keras framework based on deep learning library TensorFlow 2.5 to construct an LSTM model as a local prediction model for each city.

RNN is a standard deep learning method of time series forecasting, but there is a phenomenon that the gradient disappears or explodes, which leads to short-term memory problems. The hidden layer information of the RNN at this moment only comes from the current input and the hidden layer information of the previous moment, and there is no memory function. When the time series is long, RNN will not learn the information in the early moments.

LSTM is a variant of RNN. The network structure of LSTM is much more complicated than that of RNN. Based on RNN, three gate structures of "forgotten gate," "input gate," and "output gate" are added to determine the preservation or forgetting of the information. As information enters the model, the cell in LSTM The information will be judged, the information that meets the rules will be left, and the noncompliant information will be forgotten. Based on this principle, LSTM can effectively solve the problem of long-term memory in RNN.

The "forgetting gate measures the importance of past memory," and the information of the previous moment and the current moment is input into the nonlinear activation function Sigmoid to obtain $f^t \in [0, 1]$, which is used to determine whether the past information should be saved or forgotten ($f^t = 1$ means all is saved, $f^t = 0$ means all is forgotten). The formula for the "forgotten door" is shown in

$$f^t = \sigma\left(W_f \cdot \left[h^{t-1}, x^t\right] + b_f\right), \tag{2}$$

where we call W_f is the weight matrix of the forgetting gate and call b_f is the bias of the forgetting gate. h^{t-1} represents the output of the unit state at the previous time, and x^t represents the input at the current time.

Then use the "input gate" to calculate the importance of the input information and get $i^t \in [0, 1]$ through the Sigmoid function, which is used to determine whether the data is compressed to between -1 and +1 by the tan$h$ function needs to be updated. The formula for "input gate" is shown in

$$i^t = \sigma\left(W_i \cdot \left[h^{t-1}, x^t\right] + b_i\right), \tag{3}$$

where we call W_i as the weight matrix of the input gate and call b_i as the bias of the input gate. h^{t-1} represents the output of the unit state at the previous time, and x^t represents the input at the current time.

Finally, the updated memory's importance to the next hidden layer is calculated through the "output gate" to determine the information to be input to the next hidden layer. The formula for "output gate" is shown in

$$o^t = \sigma\left(W_o \cdot \left[h^{t-1}, x^t\right] + b_o\right), \tag{4}$$

where we call W_o is the weight matrix of the output gate and call b_o is the bias of the output gate. h^{t-1} represents the output of the united state at the previous time, and it means the input at the current time.

In this paper, the mean square error (MSE) is selected as the loss function, the nonlinear activation function is elu, the Adam optimizer is used, and the learning rate is set to $lr = 0.01$ to construct an LSTM model with a 5-layer network structure. At the same time, set the number of nodes in the input layer to 128, the number of nodes in the three hidden layers to (32, 16, 8), and the number of nodes in the output layer to 1. The formula for MSE is shown in

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n} (y\hat{}_i - y_i)^2. \tag{5}$$

We input the training set P'_{train} into the LSTM model, set the number of iterations epoch = 100, and use the early stopping method to prevent the model from overfitting. We set the minimum amount of change min_delta = 0, and the number of tolerance periods patience = 15; that is, when the minimum difference of the monitored variable is less than 0 for more than 15 times, the iteration is stopped in advance, and the training of the model is ended.

3.3. Decision-Level Fusion. In cities A, B, and C, the local electricity price dataset is used to train the LSTM model,

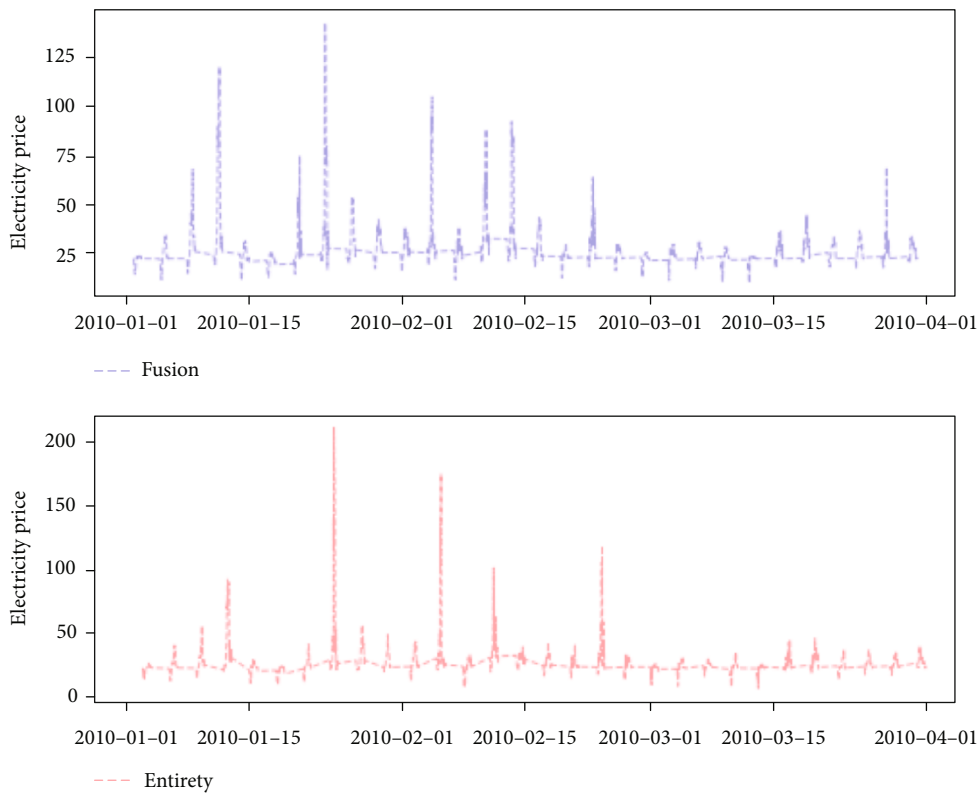


Figure 5: Season1: January to March.

and then, the trained LSTM model is used to predict the future electricity price of the city—$\text{Predict}_{\text{city}}$ (city = A, B, C). Finally, cities A, B, and C upload the predicted future electricity price information to the cloud center.

Due to the difference of urban population and economic foundation, the total electricity consumption of different cities varies greatly. The greater the total electricity consumption, the greater the impact of the prediction results uploaded by the city on the decision-making level fusion calculation. We use the Softmax function to convert the average electricity load into the corresponding impact weight W'_{city}. The calculation formula is shown in

$$W'_{\text{city}} = \text{Softmax}\left(L_{\text{city}}\right), \tag{6}$$

where L_{city} indicates the average electricity load of each city.

Finally, the unified electricity price Electricity_Price of the province is obtained using the decision-making level fusion algorithm. The calculation formula is shown in

$$\text{Electricit}_{\text{Price}} = \sum\left(W'_{\text{city}} \cdot \text{Predict}_{\text{city}}\right), \tag{7}$$

where $\text{Predict}_{\text{city}}$ indicates the forecast electricity price uploaded by each city.

4. Experiment and Result Analysis

This experiment uses the Australian electricity load and electricity price dataset. Among them, the data from 2006 to 2009 is the training set, and the data from 2010 is the test set. The dataset contains a total of 8 attributes, namely, date, hour, humidity, wet bulb temperature, dry bulb temperature, dew point temperature, electricity price, and electricity load.

Limited by the confidentiality of electricity price data, this experiment is uniformly sampled from the original dataset T and divided into 3 equal datasets, representing the electricity price data of the three cities A, B, and C in the same province.

4.1. Decision-Level Fusion Experiment and Comparison. In cities A, B, and C, the local electricity price dataset is used to train the LSTM model, and then, the trained LSTM model is used to predict the future electricity price of the city. Finally, cities A, B, and C upload the future electricity price information to the cloud center.

To verify the feasibility of decision-level fusion, we upload the undivided raw dataset T to the cloud center. Then, we train the LSTM model according to the training set simultaneously as the cities A, B, and C. Then, we use the trained LSTM model to predict the province's future overall electricity price—entirety.

The decision-level fusion is performed on the electricity price forecast results uploaded by cities A, B, and C, and the fusion result obtained is fusion.

Comparing fusion with entirety, considering the periodicity of electricity price fluctuations, the experimental results are divided into four groups: Season 1, Season 2, Season 3, and Season 4 based on quarters, as shown in Figures 5–8.

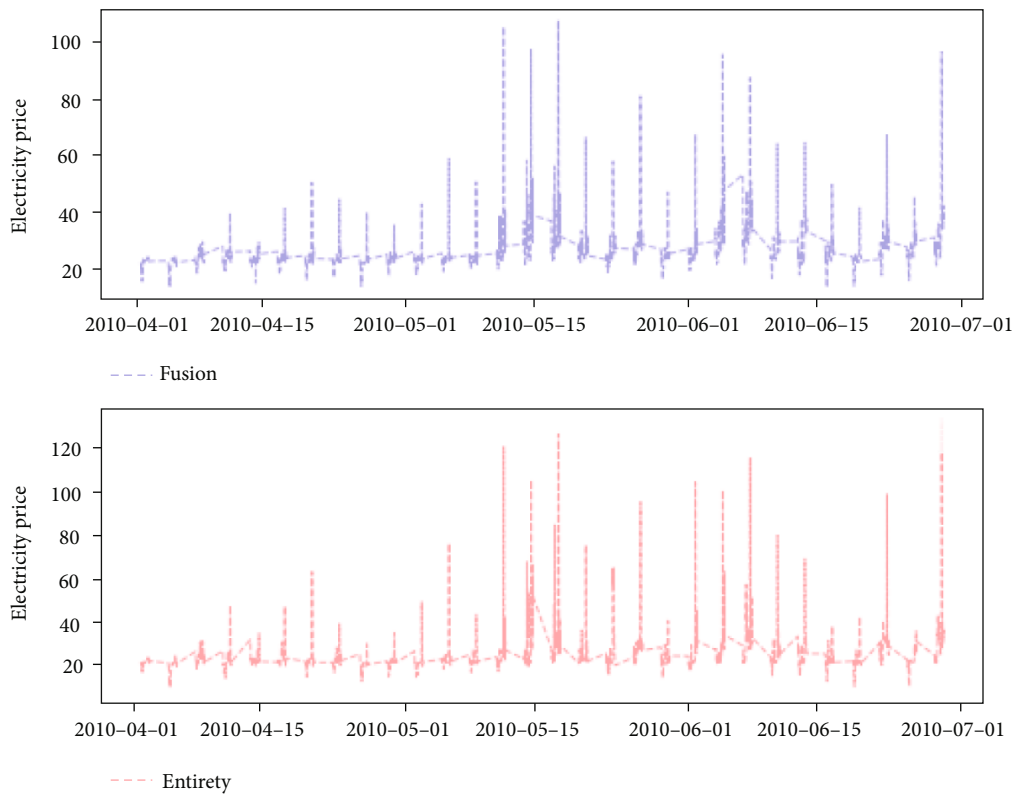


Figure 6: Season2: April to June.

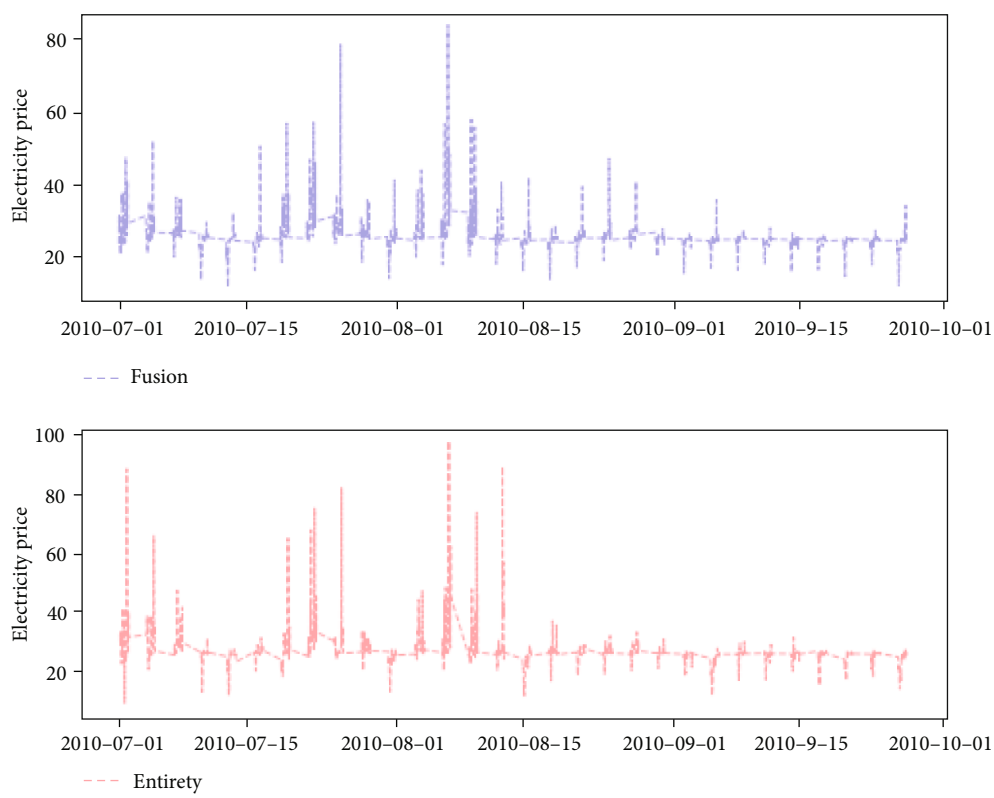


Figure 7: Season3: July to September.

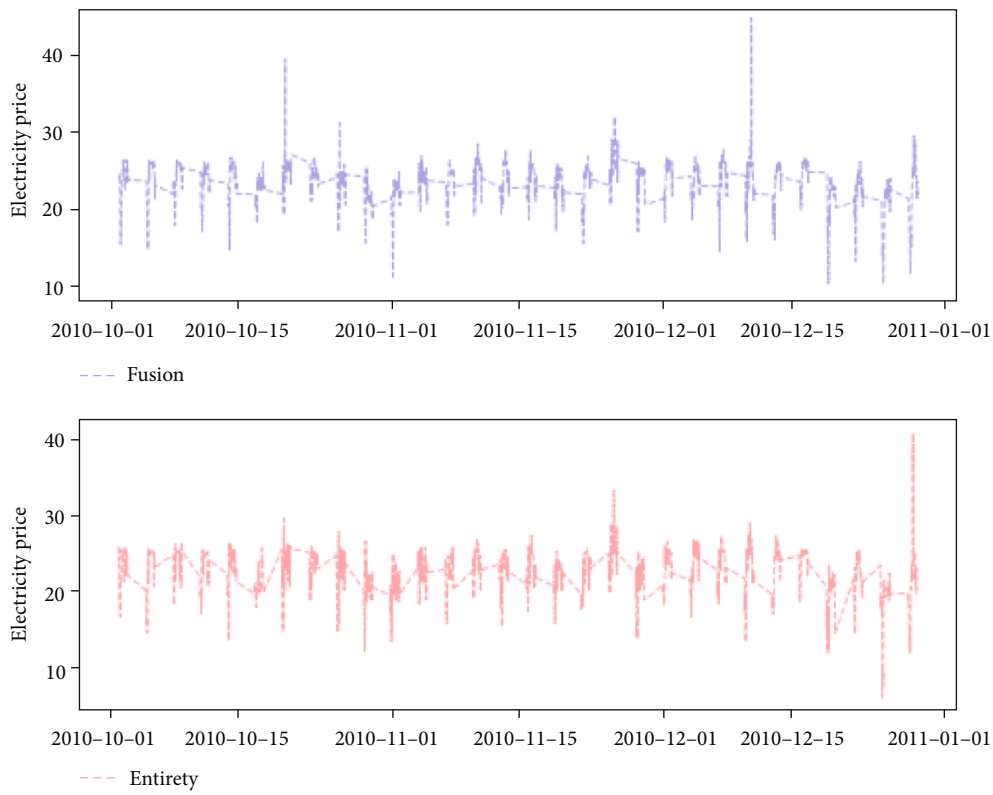


Figure 8: Season4: October to December.

Season 1 compares the decision-level fusion results—fusion and entirety from January to March, as shown in Figure 5. Among them, the fusion result in March is the closest to entirety. Due to the influence of subjective factors such as human control, the electricity prices on January 11 and 20 and February 17 have extreme fluctuations, causing specific errors, making Season 1's fusion error (me) much higher than other quarters. However, in different periods of the month, the decision-level fusion results can approach entirety well. They can perform well in additional months when individual electricity prices fluctuate sharply, such as January 23 and February 4.

Season 2 compares the decision-level fusion results—fusion and entirety from April to June, as shown in Figure 6. As can be seen from the figure, the decision-level fusion result—fusion has always maintained the same trend of change as entirety. Although there is a specific error when individual electricity prices fluctuate sharply, in other periods of Season 2, the fusion result can guarantee a higher accuracy rate.

Season 3 compares the decision-level fusion results—fusion and entirety from July to September, as shown in Figure 7. Although electricity price fluctuations have a particular impact on the accuracy of decision-level fusion, the decision-level fusion result—fusion still maintains the same trend as entirety, and more accurate fusion results can be obtained at most of the moments when electricity price fluctuations are flat.

Season 4 is the comparison between the decision-level fusion results—fusion and entirety from October to December, as shown in Figure 8. Due to the relatively stable fluctuations in electricity prices from October to December, the accuracy of the decision-level fusion results in Season 4 is significantly higher than that of other quarters. In Season4, the fusion result always maintains a high accuracy rate, and the average error is as low as 2.67.

Combining Figures 5–8, it can be seen that in the case of flat electricity price fluctuations, the decision-level fusion result (fusion) can maintain the same trend as Entirety and a high accuracy rate. And in extreme cases where electricity prices fluctuate sharply, the error of decision-level fusion is also within an acceptable range.

Through the above analysis, it can be concluded that while the decision-level fusion algorithm has a high accuracy rate, it also ensures that the error in the event of severe electricity price fluctuations is within a reasonable range. While reducing the communication burden, this solution also ensures that the error is controllable. Therefore, the decision-level fusion algorithm has high feasibility.

5. Conclusions

This paper presents LSTM-based decision-level fusion of multiple sources of data for real-time electricity price prediction. Experiments are conducted on an Australian electricity dataset, which is split into multiple copies, and the electricity consumption of different datasets assigns the weights to achieve the prediction of real-time electricity prices finally. The experimental simulation results show that the machine learning model based on the decision-level fusion of the

LSTM algorithm significantly outperforms other machine learning models using the same dataset, such as the model of linear regression and XGboost, in terms of prediction performance. The decision-level fusion machine learning model based on the LSTM algorithm for multisource data provides reasonable accuracy for predicting the electricity market prices in multiple regions. In the future, we will improve the model for decision-level fusion to maintain or improve the prediction accuracy while speeding up the real-time prediction and using multiple models to integrate the forecast to avoid the drawbacks of a single model.

Data Availability

We have not used specific data from other sources for the simulations of the results. The datasets in this paper can be downloaded with the website: https://download.csdn.net/download/loading_123/11074748

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Acknowledgments

This work is supported by the funding of the National Key R&D Program of China (2020YFB0905900).

References

[1] A. Alhariry, S. Brown, D. Eshenbaugh, N. Whitt, and A. F. Browne, "A survey of sensing methodologies in smart grids," in *SoutheastCon 2021*, Atlanta, GA, USA, 2021.

[2] S. S. Hlaing, M. M. Tin, M. M. Khin, P. P. Wai, and G. R. Sinha, "Big traffic data analytics for smart urban intelligent traffic system using machine learning techniques," in *2020 IEEE 9th Global Conference on Consumer Electronics (GCCE)*, pp. 299-300, Kobe, Japan, 2020.

[3] T. Chaurasia and P. K. Jain, "Enhanced smart home automation system based on Internet of Things," in *2019 Third International conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*, pp. 709–713, Kobe, Japan, 2019.

[4] C. Chatzigeorgiou, M. Feidakis, D. G. Kogias, and C. Z. Patrikakis, "Increasing safety and security in public places using IoT devices," in *2020 IEEE 6th World Forum on Internet of Things (WF-IoT)*, New Orleans, LA, USA, 2020.

[5] L. Chen, J. Xinghong, and D. Gang, "On the technical characteristics and broad application of internet of things," *Scientific Consultation*, vol. 9, pp. 86–86, 2011.

[6] H. Tianfield, "Towards edge-cloud computing," in *2018 IEEE International Conference on Big Data (Big Data)*, pp. 4883–4885, Seattle, WA, USA, 2018.

[7] D. Choi and K. Lee, "An artificial intelligence approach to financial fraud detection under IoT environment: a survey and implementation," *Security and Communication Networks*, vol. 2018, no. 4, Article ID 5483472, 2018.

[8] M. R. Bashir and A. Q. Gill, "Towards an IoT big data analytics framework: smart buildings systems," in *2016 IEEE 18th International Conference on High Performance Computing and Communications; IEEE 14th International Conference on Smart City; IEEE 2nd International Conference on Data Science and Systems (HPCC/SmartCity/DSS)*, pp. 1325–1332, Sydney, NSW, Australia, 2016.

[9] Z. Feng, Z. Hui, and D. Yinglong, "Development ideas of key technologies of ubiquitous power internet of things intelligent perception," *Proceedings of The Chinese Society for Electrical Engineering*, vol. 40, no. 1, pp. 70–82, 2020.

[10] T. Li, W. Liu, Z. Zeng, and N. N. Xiong, "DRLR: a deep reinforcement learning based recruitment scheme for massive data collections in 6G-based IoT networks," *IEEE Internet of Things Journal*, 2021.

[11] X. Sui, H. Zhang, and Y. Lv, "Coverage performance analysis of grid distribution in heterogeneous network," in *2017 IEEE 17th International Conference on Communication Technology (ICCT)*, pp. 1424–1428, Chengdu, China, 2017.

[12] S. N. Swamy, D. Jadhav, and N. Kulkarni, "Security threats in the application layer in IOT applications," in *2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*, pp. 477–480, Palladam, India, 2017.

[13] J. Yao, Y. Zhang, F. Liu, and Y. C. Liu, "Object detection based on decision level fusion," in *2019 Chinese Automation Congress (CAC)*, Hangzhou, China, 2019.

[14] T. Leikanger, C. Schuss, and J. Häkkinen, "Near field communication as sensor to cloud service interface," in *2017 IEEE Sensors*, pp. 1–3, Glasgow, UK, 2017.

[15] M. Li, W. Chen, L. Chen, and Y. Sun, "Preliminary applications of blockchain technology in electricity information acquisition system," *2020 IEEE 9th Joint International Information Technology and Artificial Intelligence Conference (ITAIC)*, vol. 9, pp. 679–683, 2020.

[16] C. Esposito, A. Castiglione, F. Pop, and K. R. Choo, "Challenges of connecting edge and cloud computing: a security and forensic perspective," *IEEE Cloud Computing*, vol. 4, no. 2, pp. 13–17, 2017.

[17] I. Swapna, "Cyber security for smart grid, cryptography, and privacy," *International Journal of Digital Multimedia Broadcasting*, vol. 2011, Article ID 372020, 2011.

[18] K. Ren, H. Li, S. Li, and H. Dong, "Voltage stability analysis of front-end speed controlled wind turbine integrated into regional power grid based on bifurcation theory," *Complexity*, vol. 2020, Article ID 8816334, 11 pages, 2020.

[19] W. Xu, Y. Lin, Y. Zhou, and Z. Xu, "Impact analysis of regional grid interconnection on transient stability of Guangdong power grid," *Automation of Electric Power Systems*, vol. 37, no. 21, pp. 34–38, 2013.

[20] J. J. Guo and P. B. Luh, "Selecting input factors for clusters of Gaussian radial basis function networks to improve market clearing price prediction," *IEEE Transactions On Power Systems Pwrs*, vol. 18, no. 2, pp. 665–672, 2003.

[21] S. K. Aggarwal, L. M. Saini, and A. Kumar, "Electricity price forecasting in deregulated markets: a review and evaluation," *International Journal of Electrical Power and Energy Systems*, vol. 31, no. 1, pp. 13–22, 2009.

[22] "Improving market clearing price prediction by using a committee machine of neural networks," *IEEE Transactions on Power Systems*, vol. 19, no. 4, pp. 1867–1876, 2004.

[23] W. A. R. Intan Azmira, A. Ahmad, I. Z. Abidin, K. S. Yap, M. N. M. Nasir, and W. R. Upkli, "Electricity price prediction with support vector machine and bacterial foraging optimization algorithm for day-ahead model," in *2020 IEEE Student*

Conference on Research and Development (SCOReD), pp. 159–164, Batu Pahat, Malaysia, 2020.

[24] L. Ding and Q. Ge, "Electricity market clearing price forecast based on adaptive Kalman filter," in *2018 International Conference on Control, Automation and Information Sciences (ICCAIS)*, pp. 417–421, Hangzhou, China, 2018.

[25] Y. Zou, M. Tu, X. Teng, R. Cao, and W. Xie, "Electricity price forecast based on stacked autoencoder in spot market environment," in *2019 9th International Conference on Power and Energy Systems (ICPES)*, Perth, WA, Australia, 2019.

[26] Y. Li, R. Liang, Z. Li, J. Gao, Y. Wang, and T. Wu, "Research on electricity price forecasting method based on genetic algorithm and neural network in power market," in *2018 2nd IEEE Conference on Energy Internet and Energy System Integration (EI2)*, Beijing, China, 2018.

[27] F. Hamami and I. A. Dahlan, "Univariate time series data forecasting of air pollution using LSTM neural network," in *2020 International Conference on Advancement in Data Science, E-learning and Information Systems (ICADEIS)*, Lombok, Indonesia, 2020.

[28] S. C. Shah, "Recent advances in mobile grid and cloud computing," *Intelligent automation and soft computing*, vol. 24, no. 2, pp. 285–298, 2018.

[29] R. Gad, A. A. Abd El-Latif, S. Elseuofi, H. M. Ibrahim, M. Elmezain, and W. Said, "IoT security based on iris verification using multi-algorithm feature level fusion scheme," in *2019 2nd International Conference on Computer Applications & Information Security (ICCAIS)*, pp. 1–6, Lombok, Indonesia, 2019.

[30] S. Acharya, A. Rajasekar, B. S. Shender, L. Hrebien, and M. Kam, "Real-time hypoxia prediction using decision fusion," *IEEE Journal of Biomedical and Health Informatics*, vol. 21, no. 3, pp. 696–707, 2017.

[31] Y. Zhang, Y. Nie, B. Sun, H. E. Jun, and B. Yang, "A method of depression evaluation based on decision level fusion of deep forest with multimodality data," *Journal of Beijing Normal University (Natural Science)*, 2018.

Hindawi
Wireless Communications and Mobile Computing
Volume 2022, Article ID 4401653, 13 pages
https://doi.org/10.1155/2022/4401653

Research Article

Cooperative Multiagent Attentional Communication for Large-Scale Task Space

Qijie Zou,[1] **Youkun Hu**,[1] **Dewei Yi**,[2] **Bing Gao**,[1] **and Jing Qin**[1]

[1]*Department of Information Engineering Faculty, Dalian University of China, 116622, China*
[2]*Department of Computing Science, University of Aberdeen, Aberdeen AB24 3UE, UK*

Correspondence should be addressed to Youkun Hu; huyoukun163@163.com

Received 8 September 2021; Revised 20 November 2021; Accepted 9 December 2021; Published 24 January 2022

Academic Editor: Chi-Hua Chen

With the rapid development of mobile robots, they have begun to be widely used in industrial manufacturing, logistics scheduling, intelligent medical, and other fields. For large-scale task space, the communication between multiagents is the key to affect cooperation productivity, and agents can coordinate more effectively with the help of dynamic communication. However, the traditional communication mechanism uses simple message aggregation and broadcast and, in some cases, lacks the distinction of the importance of information. Multiagent deep reinforcement learning (MDRL) is valid to solve the problem of informational coordination strategies. However, how different messages affect each agent's decision-making process remains a challenging task for large-scale task. To solve this problem, we propose IMANet (Import Message Attention Network). It divides the decision-making process into two substages: communication and action, where communication is considered to be part of the environment. First, an attention mechanism based on query vectors is introduced. The correlation between the query vector agent's own information and the current state information of other agents is estimated, and then, the results are used to distinguish the importance of information from other agents. Second, the LSTM network is used as the unit controller for each agent, and individual rewards are used to guide the agent training after communication. Finally, IMANet is evaluated on tasks on challenging multi-agent platforms, Predator and Prey (PP), and traffic junction. The results show that IMANet can improve the efficiency of learning and training, especially when applied to large-scale task space, with a success rate 12% higher than CommNet in baseline experiments.

1. Introduction

Multiagent system is very practical in distributed control, remote scheduling, and modeling analysis [1]. Compared with a single agent, it can complete tasks more effectively and has better robustness, reliability, and scalability. Communication is the basis for maintaining the efficiency and organization of multiagent systems [2]. In the DEC-POMDP (Decentralized Partially Observable Markov Decision Process) environment, through communication [3], agents can exchange their observations to better discover the current global state and understand the actions and intentions of other agents. However, traditional predefined communication protocols and broadcast messages cannot allow multi-agent to effectively "learn to communicate" in a large-scale task. Reinforcement learning (RL) is mainly to study how agents choose actions by perceiving local and global states and constantly interact with the dynamic environment in order to find optimal policy that maximize cumulative rewards [4, 5].

The reinforcement learning problem in multiagent scenarios is more complex than in single-agent scenarios. Because the agent interacts with the environment at the same time and treats other agents as part of the environment, the agent usually faces the dimensionality disaster (grows exponentially with the number of agents; so the multiagent system dimension is very large and computationally complex) [6], credit assignment (identify the impact of the agent's behavior on this global return), and other issues [7]. As the number of agents increases, the task space is extended to large scale and cooperative information is augmented. Deep learning is an efficient representation learning that can discover the key information in the original message [8]. The main reason is that the neural

network can process the input high-dimensional data and extract useful expressions [9, 10]. The main advantage of deep reinforcement learning (DRL) is that it can extend RL to high-dimensional state and action spaces [11].

In recent years, multiagent deep reinforcement learning (MDRL) [12], a combination of deep learning (DL) and multiagent reinforcement learning (MARL), has been successful in many complex environments, such as StarCraft [13] and particle environment [14]. However, these tasks either assume that the environment is completely observable or there is a lack of communication between agents.

In this paper, we will investigate how to use Deep MARL's approach for effective communication learning in a partially observable distributed environment.

In the centralized training paradigm, as the number of agents increases, the linear growth of the input dimension and the exponential growth of the output space make, the centralized training method using the traditional central controller makes the algorithm not easily scalable to large-scale tasks, the convergence of the algorithm becomes poor or even unable to converge, and the problem of poor scalability. In the communication process, broadcast communication is a common setting for the study of "learning communications" between agents, but this does not allow selective attention to the observations and actions of other agents, does not provide useful information to agents in the decision-making process, and leads to unstable learning processes. These problems are caused by the inability of traditional reinforcement learning methods to learn cooperative strategies through effective communication in DEC-POMDP conditions [15]. To this end, we propose the IMANet method for multiagent deep reinforcement learning.

The main contributions of this paper are as follows:

(1) We use a query vector-based attention mechanism in the IMANet communication structure to identify the messages contained in more favorable specific agents. In this algorithm, the local observations of each agent are encoded and attention is directed to different agents according to the magnitude of the attention weights, generating dynamically changing communication vectors to coordinate policies. The problem of inability to judge and distinguish the importance of messages is solved to make the learning process more efficient and stable

(2) Using a single individual LSTM network as a controller for each agent [16], individual observations and communication vectors from the agents themselves are processed. Our independent control model selectively outputs important information, thus alleviating the problems associated with dimensional explosion, which makes it possible for agents to learn coordination strategies in large-scale spaces

2. Notation and Background

2.1. Technical Background. Dec-POMDP is a multiagent extension of partially observable Markov decision process. First, our approach requires the introduction of the necessary notation, and then, three common reinforcement learning framework structures are introduced.

2.1.1. Decentralized Partially Observable Markov Decision Processes (Dec-POMDPs). We consider a fully cooperative multiagent setting that can be formulated as DEC-POMDP [17]. It is formally defined as a tuple $\langle N, S, A, T, R, O, Z, \gamma\rangle$, where N is the number of agents, S is the state space; $A = A_1 \times A_2 \cdots \times A_N$ is the action space of all agents and A_j is the set of local action a_j that agent j can take; and $T : S \times A \times S \longrightarrow [0, 1]$ is the state transition probability. $R = (r_1, r_2, \cdots, r_N)$ is the set of rewards, where $r^j : S \times A \times S \longrightarrow \mathbb{R}$ is the reward function of an agent j; $O = [O_1, \cdots, O_N]$ is the set of joint observation o; $Z : S \times A \longrightarrow O$ is observation function; and $\gamma \in [0, 1]$ is the discount factor. The set of action policies is $\pi = \{\pi_1, \pi_2, \cdots, \pi_N\}$; each policy is represented by a neural network and its parameter set $\theta = \{\theta_1, \theta_2, \cdots \theta_N\}$.

A policy $\pi(a_j^t \mid h_j^t; \theta^j)$ is the probability of taking action a_j^t when encountering history h_j^t under the policy parameters θ^j. The agent's action a_j^t depends on the encountered history h_j^t. We generally omit the parameter θ^j in policy $\pi(a_j^t \mid h_j^t; \theta^j)$ for brevity and denote the policy as $\pi(a_j^t \mid h_j^t)$. We use h_j^t to denote the history of individual observations o_j, individual rewards r_j, and individual actions a_j encountered by the agent j following policy π at the time t and defined as $h_j^t = \{s_j^0, o_j^0, r_j^0, a_j^0, \cdots, r_j^{t-1}, a_j^{t-1}, s_j^{t-1}\}$. Sequence $h_j^0 = \{s_j^0\}$ denotes the history at time $t = 0$ and contains only the star state s_j^0 of the agent j. The return of an agent j that interacts with the environment to produce a history h_j^t is written as $R(h_j^t) = \sum_{t=0}^{T} \gamma^t r_j^t$.

2.2. Multiagent Reinforcement Learning Architecture. The following are the three architectures of multiagent reinforcement learning:

(1) Decentralization: without a central controller, the agent makes independent decisions based on its own policy network.

(2) Fully centralized: the central controller makes decisions for all agents.

(3) Centralized training and decentralized execution: the central controller is used only by the training process. Each agent makes a decision on its own policy network by its own observations.

2.2.1. Decentralization. All agents are independent individuals, and they do not communicate with each other. As shown in Figure 1, each agent independently interacts with the environment to obtain individual observation o^i and individual rewards r^i. Each agent deploys its own policy network and trains its own policy network independently, exactly the same as the reinforcement learning of a single agent. After the training, each agent utilizes its own policy network to make a decision, and the observed o^i is the input

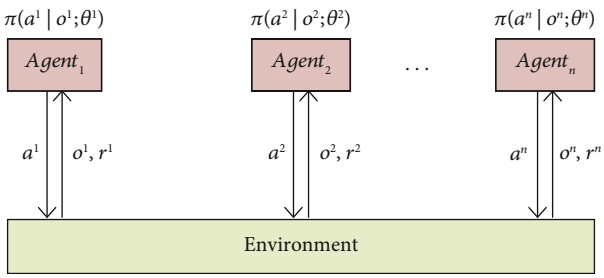


Figure 1: Decentralized architecture.

of the policy network and the probability distribution $\pi(a^i \mid o^i;\theta^i)$ is the output. To produce a discrete action a^i, we sample from this distribution: $a^i \sim \pi(\cdot \mid o^i;\theta^i)$, and then execute the action a^i. Regardless of training or execution, there is no communication between agents. The essence of such decentralization is single-agent reinforcement learning, rather than true multiagent reinforcement learning. Although this decentralized structure can well deal with the problems caused by the growth of the number of agents, the effect of single-agent reinforcement learning for multiagent reinforcement learning does not usually work. The single-agent algorithm assumes that these functions are stable, and in a multiagent scenario, it will face the problem of environmental instability. The reason is that the influence between them should not be ignored.

2.2.2. Fully Centralized. There are n agents interacting with the environment, and each agent will change the environment, thereby affecting other agents. As shown in Figure 2, there is no policy network on the agent, so the agent cannot make decisions by itself, and all have to be commanded by the central controller.

During training, the agent reports its observation o^i and reward r^i to the central government. The policy network is in the center, and the center transmits the decision a^i to the agent i. The agent performs actions in accordance with the instructions of the center and interacts with the environment to do it in the center. The central controller uses all observation o, reward r, and action a to train the policy network. Even after the training is completed, the central controller is needed when it comes to decision-making.

N policy networks are trained on the central controller, and their network structures are the same, but the parameters may be different. Use θ^i to denote the parameters of the i-th agent. The input of the policy network is all the observations $o^1 \sim o^n$ of the agent. The i-th policy network determines the action a^i of the agent i, and the decision can only be made by the central government. This is because the policy network needs to use the observations of all agents. An agent only knows its own observation o^i, it does not have enough information to make a decision, so the policy network cannot be deployed on the agent and can only stay on the central controller. In the execution, all agents report their observations to the central government; then, the central government determine each agent what to do. The center transmits $a^1 \sim a^n$ to the corresponding agent.

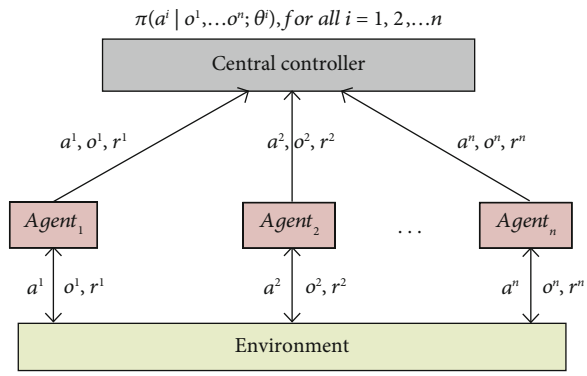


Figure 2: Fully centralized architecture.

The advantage of centralization is to know the overall information, which can help all agents make better decisions. But centralization also has disadvantages. The execution speed is slow. The agent itself has no decision-making power, and all decisions are made by the central government. The agent reports all its observations to the central government, and the central government collects global information before making decisions. The biggest problem faced by this centralized structure is the huge input and output space. With the increase in the number of agents, the input dimension increases linearly, and the space exponential of the output joint policies increases.

2.2.3. Centralized Training Decentralized Execution (CTDE). In the traditional CTDE architecture, each agent has its own policy network and the team has a central controller during training, which helps the agent train the policy network. After training, there is no need for a central controller, and each agent interacts with the environment independently. The agent has own policy network to act based on your own local observations. There are many different models of CTDE, which are popular nowadays (such as MAAC proposed by Sha et al. [18] and COMA proposed by Foerster et al. [19].).

3. Related Work

Recently, in the direct communication method of embedding communication channels in deep neural networks, the use of specific communication channels can selectively complete the information exchange between various agents [20]. For learning communication protocols, it has been proven to be very effective [21–23]. Under normal circumstances, the continuous transmission between agents through the network forms a communication channel, which makes the agents consider local information and global information at the same time during the learning process. The protocol can be optimized at the same time as the network is optimized.

Learning an effective multiagent communication protocol is mainly divided into the following two aspects.

3.1. Broadcast Partial Observations of Each Agent to All Agents. Sukhbaatar et al. proposed the CommNet algorithm.

CommNet is just a single network designed for all agents [24]. It solves the Dec-POMDP problem. It follows the centralized training and decentralized execution method. This single-network communication channel cannot be easily extended to a large-scale agent environment. Due to continuous communication, the controller can learn through back-propagation. However, the information transmission method adopted by this algorithm is to broadcast communication content to all agents, which will also cause waste of bandwidth resources. Aiming at the problem of channel occupation when broadcasting a message to each agent within the communication range, Kim et al. proposed SchedNet [25], a learning method for multiagent deep reinforcement learning. This method introduces the CSMA protocol, which is a contention-type access medium. When the agent sends a message to the channel, it monitors whether the channel is occupied at this time and stops sending if it is occupied. When a conflict occurs, it will stop sending the message, wait for a suitable time, and send it randomly. This method also alleviates the message loss problem caused by the agent through broadcast communication. The Message-Dropout MADDPG uses the message dropout technology in multiagent reinforcement learning in order to allocate communication resources reasonably [26]. This method uses a centralized training decentralized execution framework under fully or partially observable conditions, discards the received message with a certain probability in the training phase, and compensates by multiplying the weight of the discarded block unit by the correction probability this influence. This method is also robust against communication errors.

3.2. Selective and Targeted Communication through the Use of Attention Mechanism. As it is difficult for multiagents to distinguish between valuable information and shared information, Jiang and Lu proposed the ATOC based on the actor-critic framework [27]. In this algorithm, the local observation value of each agent is coded. The attention unit is used to determine which agents to communicate with (screening agents for information sharing), and a two-way LSTM network is used as the communication channel between communication groups. Agents in this communication group exchange information with each other in the communication channel. The attention module is an RNN network and will face the problem of vanishing gradients. Without a centralized controller, decentralized agents cannot learn effective and decentralized cooperation policies in a complex environment. Geng et al. proposed a new attention-based communication neural network (CommAttn) [28]. CommAttn can use the display communication method to automatically learn and explore the cooperation policies in the problem, by modeling the interaction between agents and introducing the attention mechanism. Calculate the relevance of the received message, and determine whether communication between agents is required, so that the agent can decide who to communicate with. The learned communication model of the system is more able to adapt to the dynamically changing environment. Facebook AI Research proposed a collaborative multiagent deep reinforcement learning method, namely, TarMAC [29]. Allow targeted communication between agents. The purpose is similar to ATOC's attentional mechanism. It determines not only who to send the message to but also the part of the observation that is most relevant to the goals of other multiagents. This targeted communication behavior is achieved through a signature-based attention mechanism: Together with the message, the sender broadcasts a key that encodes the attributes of the multiagent targeted by the message and is used by the receiver to measure the relevance of the message. To cope with how to select important messages and how to process important messages efficiently, Mao et al. proposed a DRL method called Double Attentional Actor-Critic Message Processor (DAACMP) [30], which embeds a first class attention mechanism in the actor part, where the importance of a message is positively related to the distance between two agents; the Actor Attention is able to pay attention to the messages of nearby agents adaptively. Embedding the second type of attention mechanism in the Critic part, where the joint action policies of teammates are modeled using Q-values and similar joint actions are grouped and processed instead of processing the action policy of individual agent, the Critic Attention has a more sophisticated ability to process all important messages.

In the above method, using the method of multiagent deep reinforcement learning with direct communication, there will be an obvious information transmission process between the agents, and the communication objects can be selected to reduce the problem of the explosion of the joint action space or by optimizing the communication content. This method allows the agent to pay attention to more important messages when sending information, reduce the bandwidth occupation in the communication channel, and improve calculation efficiency.

The IMANet method in this article can be seen as an extension of the CommNet. The CommNet algorithm is to chain the messages together and broadcast them; IMANet adds an attention vector to learn the importance weight of each message from other agents, obtain their weighted sum, and use it to perform operations. You can selectively conduct partial interactions, and determine which multiagents provide shared information that can improve performance and make the training process more stable. And because CommNet uses a central controller for centralized training, the scalability is poor, so we have an independent controller.

4. IMANet Method

4.1. Basic Description. We first introduce the necessary assumptions for our approach and then describe in detail our multiagent communication architecture.

Hypothesis 1. The decision-making process is divided into two subphases: communication subphase and action subphase.

Hypothesis 2. During training, the current content of the agent's communication is related to the agent's own encounter history and the encounter histories of other agents (i.e., the hidden layer states in the next section).

Hypothesis 3. Each agent's policy depends only on its own hidden layer state.

4.2. IMANet Architecture. We propose a new deep MARL framework with targeted communication, called IMANet, whose overall architecture is depicted in Figure 3. IMANet consists of the following three levels of control structure.

NC (NO Communication) is an independent control that uses LSTM network structure, where each agent is controlled by an independent LSTM. IMANet without communication is exactly NC. For the j-th agent, its policy is defined as

$$s_j^{t+1} = \mathrm{LSTM}\left(e\left(o_j^t\right), h_j^t, s_j^t\right), \tag{1}$$

$$h_j^{t+1} = \mathrm{LSTM}\left(e\left(o_j^t\right), h_j^t, s_j^t\right), \tag{2}$$

$$a_j^t = \pi\left(h_j^t\right), \tag{3}$$

where o_j^t is the observation of the agent j at a time t, $e(.)$ is the encoder function parameterized by the fully connected neural network, and π is the action policy of the agent. In addition, h_j^t and s_j^t, respectively, represent the hidden layer state and memory cell state of the agent j at a time t. Use the same LSTM model for all agents and share their parameters. All agents share a unit, which has a higher utilization rate of samples; shared parameters reduce model complexity.

IMANet extends this independent controller model NC; a vector c_t containing communication is introduced, which allows agents to obtain local information observed by other agents through exchange to observe the global state of the system. As in Hypothesis 1, the IMANet model divides the decision-making into two subphases: communication and action, as shown in Figure 4. Before choosing an action, use the method based on the attention mechanism to decide which important information to pay attention to.

The policy for the j-th agent in the IMANet network is defined as

$$s_j^{t+1} = \mathrm{LSTM}\left(e\left(o_j^t\right), c_j^t, h_j^t, s_j^t\right), \tag{4}$$

$$h_j^{t+1} = \mathrm{LSTM}\left(e\left(o_j^t\right), c_j^t, h_j^t, s_j^t\right), \tag{5}$$

$$a_j^t = \pi\left(h_j^t\right). \tag{6}$$

According to Figure 3(c), the candidate memory cells $\tilde{s}_j^{t+1}$ and the gate value functions Γ_u, Γ_f, and Γ_o are defined as

$$\tilde{s}_j^{t+1} = \tanh\left(\tilde{w}_s\left[h_j^t, e\left(o_j^t\right), c_j^t\right] + b_{\tilde{s}}\right), \tag{7}$$

$$\Gamma_u = \delta\left(w_u\left[h_j^t, e\left(o_j^t\right), c_j^t\right] + b_u\right), \tag{8}$$

$$\Gamma_f = \delta\left(w_f\left[h_j^t, e\left(o_j^t\right), c_j^t\right] + b_f\right), \tag{9}$$

$$\Gamma_o = \delta\left(w_o\left[h_j^t, e\left(o_j^t\right), c_j^t\right] + b_o\right). \tag{10}$$

Among them, since the same LSTM model is used for each agent, that is, the update gate Γ_u, which determines what information we want to store in the cell state, has parameters w_u and b_u. The forget gate Γ_f, which determines what information we want to discard from the cell state, has parameters w_f and b_f. The next tanh layer creates a candidate vector $\tilde{s}_j^{t+1}$, which will be added to the cell state. In the next step, we will combine these two vectors to create the update value, which includes parameters $\tilde{w}_s$ and $b_{\tilde{s}}$. Finally, the output gate Γ_o, where we need to decide what we want to output, will be based on our cell state, which includes parameters w_o and b_o. That is, all the parameters are in the four w and b. δ represents the sigmoid function, which makes the gate value very close to 0 or 1. At each time step, Through an activation function tanh, the current input $e(o_j^t)$, c_j^t, and the h_j^t passed down from the previous state are spliced and trained to obtain $\tilde{s}_j^{t+1}$. Here, tanh is used because $\tilde{s}_j^{t+1}$ is used as input data instead of a gate signal. The three gate values above the update gate Γ_u, the forget gate Γ_f, and the output gate Γ_o allow the values flowing through the network to be adjusted.

The update gate and the forget gate are used to update the value of the state s_j^{t+1}, the state value is defined as

$$s_j^{t+1} = \Gamma_u * \tilde{s}_j^{t+1} + \Gamma_f * s_j^t. \tag{11}$$

Specifically, Γ_f is used as a forgetting gate to control which information of the previous state s_j^t should be retained and which should be forgotten. As an update gate Γ_u, select and memorize the inputs coding observations $e(o_j^t)$ and communication vectors c_j^t at this stage, and record more important contents. Adding the results of the above two, it means that part of the information of the current state h_j^t is deleted and some information of the new input $\tilde{s}_j^{t+1}$ is added to obtain the next state s_j^{t+1}.

The output gate is used to update the hidden layer state function h_j^{t+1}, the hidden layer state is defined as

$$h_j^{t+1} = \Gamma_0 * \tanh s_j^{t+1}. \tag{12}$$

The s_j^{t+1} obtained in the previous stage is scaled by an activation function tanh and controlled by the output gate Γ_o. This stage will determine which states will be used as the output of the current hidden layer state h_j^{t+1}.

4.2.1. Training. We use the reinforcement learning method based on Policy gradient to train the action policy. IMANet uses independent controllers to train different agents for guidance, executed in a decentralised manner, as each agent

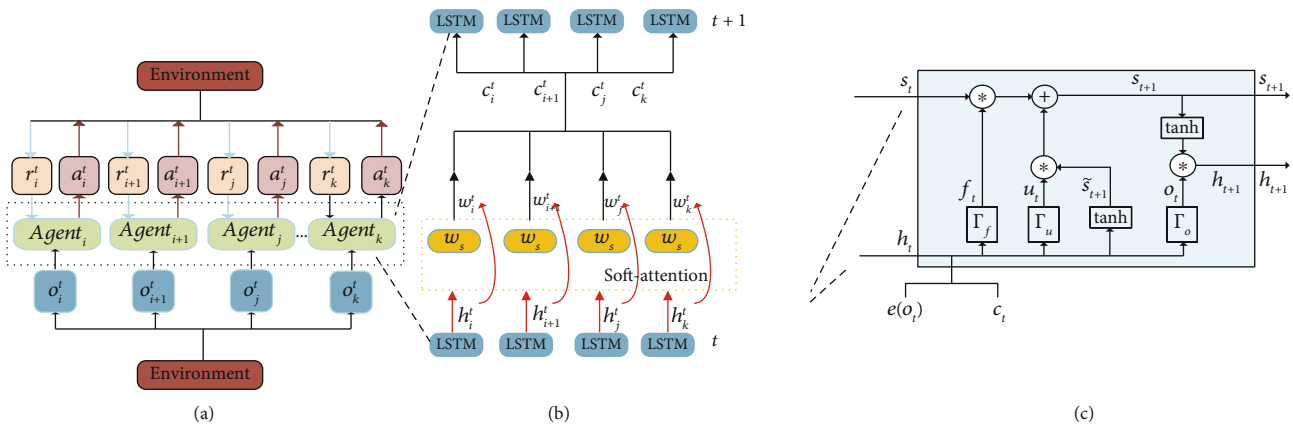


Figure 3: An overview of our IMANet model. (a) The optimization of the individual rewards of each agent based on the observation results. (b) The production of communication vectors for each agent in a single communication step. (c) The module view of the LSTM unit.

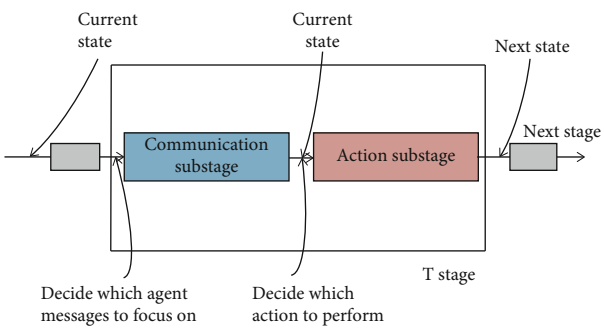


Figure 4: Decision substage.

only needs its local state vector and the weighted sum of incoming messages to act, as shown in Figure 5.

Each controller trains only one agent, and the system has n agents. As shown in Figure 3(a), the agent j interacts with the environment at time t to obtain individual observations o_j^t, individual actions a_j^t, and individual rewards r_j^t from the environment. For training, the agent j needs an independent controller. As shown in Figure 5, for agent j at time t, the attention unit performs a purposeful fusion based on the received hidden layer state $h_i^t \sim h_n^t (i \neq j)$ of other agent to generate a communication vector c_j^t. The hidden layer state h_j^t of agent j and the communication vector c_j^t containing the encounter histories of other agents are sent to the independent controller of agent j. Through a communication mechanism, the different agents can exchange information about their observations, actions, and intentions to stabilize their training process. After training, the agent j makes its own decisions based on its own hidden layer state (encounter history) h_j^t. And the policy network outputs a probability distribution $\pi(a_j^t \mid h_j^t)$.

The overall performance measure of the policy π is denoted as $J(\pi)$. It is defined as

$$J(\pi) = E\left\{R\left(h_j^t\right)\right\} = \int p\left(h_j^t\right) R_j^t dh_j^t, \tag{13}$$

where $p(h_j^t)$ is the probability of the existence of each sequence h_j^t given the parameter θ^j of the policyπ.

To optimize the policy π, we want to update the parameter θ^j along the gradient of J to an optimal; it is defined as

$$\theta_j^{t+1} = \theta_j^t + \alpha \nabla_\theta J(\pi), \tag{14}$$

where α is the learning rate, the policy are updated by ascent with the following gradient, and the policy gradient is defined as

$$\nabla_\theta J(\pi) = \nabla_\theta \int_{h_j^t} p\left(h_j^t\right) R\left(h_j^t\right) dh_j^t = \int_{h_j^t} \nabla_\theta p\left(h_j^t\right) R\left(h_j^t\right) dh_j^t. \tag{15}$$

4.2.2. Communication. Establishing effective collaboration policies requires targeted communication, that is, the ability to send specific messages to agents. We use the attention mechanism based on query vector in the communication structure to identify more beneficial specific agent messages and realize the fusion of messages, leading to specific communication links according to the size of the attention weight. The attention model based on query vector is shown in Figure 6.

As shown in Figures 3(b) and 6, in the attention model, the communication vector C_j^t is defined as

$$C_j^t = \sum_{i \neq j} w_i^t h_i^t, \tag{16}$$

$$e_i^t = \text{Score}\left(h_j^t, h_i^t\right), \tag{17}$$

$$w_i^t = \frac{\exp\left(e_i^t\right)}{\sum_{n=1}^{T} \exp\left(e_n^t\right)}. \tag{18}$$

In the communication process, we try to understand the content of the communication received by agent j from other agents. C_j^t is the communication vector of the agent j

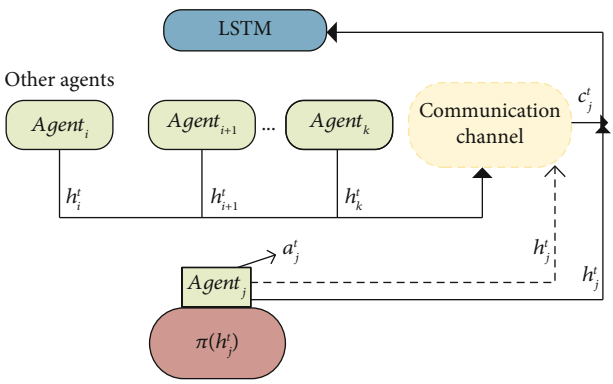


Figure 5: IMANet controller model.

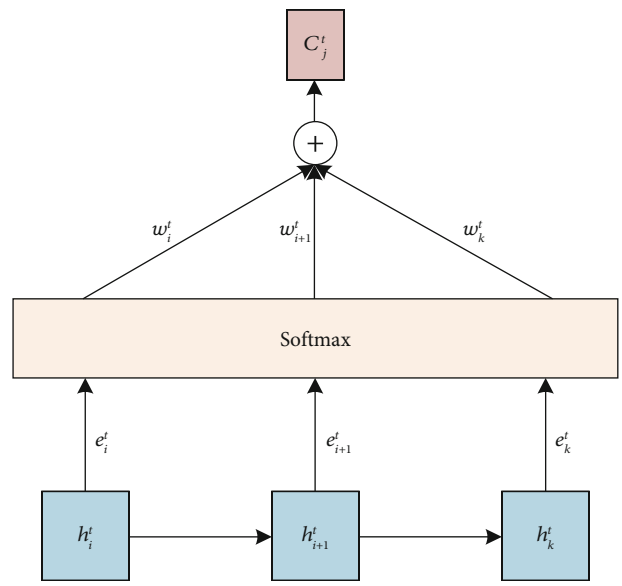


Figure 6: Attention model for query vector.

at time t. The score function ($\text{Score}(h_j^t, h_i^t)$) is used to calculate the hidden layer state h_j^t at the current moment as the query vector and do the inner product operation with the hidden layer vectors $h_i^t \sim h_n^t (i \neq j)$ passed by other agents at the current moment, respectively, to get the weighting coefficient w_i^t of the similarity size and the size of the coefficient reflects the importance of the content at the same time. The attention mechanism module is a simple LSTM network. The communication vector at the current moment can be obtained through the weighted addition of the hidden layer. At this time, the agent focuses on more important information. Take the agent's own local observation and state coding as input, and use the communication vector generated by the query vector attention mechanism, that is, the state information observed by other agents as additional input, and the fused state of the hidden layer is output to guide the cooperation policies π.

5. Experiment

5.1. Experimental Setup

5.1.1. Experiment Environment Settings. We evaluated IMANet, CommNet, and NC in the traffic junction task [24] and Predator and Prey task [31]. CommNet is a communication method that uses broadcast. The detailed experimental environment will be described in detail in the following subsections. The experimental hardware environment uses Intel(R) Core(TM) i7-7700 CPU+GeForce GTX 1650+16GB; the software environment for the experiments uses PYTorch+Gym [32]; the agent updates the policy according to their respective reward functions. Using the RMSProp approach, the configuration of learning rate hyperparameters is done automatically by the algorithm, and RMSProp can be targeted to provide different learning rates for each parameter; this method was proposed by Geoff Hinton [33], thus improving the problem of fading learning rate.

We use a learning rate of 0.001, set the hidden size to 128 units, do 10 weight updates every round, and conduct 1000 rounds of experiments in a Predator and Prey task and traffic junction task. Use LSTM to realize the attention unit.

5.1.2. Scenarios

(1) Predator and Prey. In the grid map of different sizes, there are n agents as predators to capture a static prey, and each predator has the same size of perception area. When the prey is caught, a positive reward is given. Until the end of this round, that is, other agents have reached the position of the prey, as shown in Figure 7.

Figure 7 shows the red circle representing the predator and the red arrow representing the direction in which the predator is moving. Each predator can take five actions: up, down, left, right, and stop. The green circle represents fixed prey. The yellow 3×3 square represents the perception range of the agent.

(2) Traffic Junction. In the task of traffic junction, the total number of cars is fixed avoiding collisions while passing through the intersection. The cars randomly follow one of the possible routes to reach the grid destination and are again added back to the environment with a different route assignment. We set up two one-way roads on a 7×7 grid, as shown in Figure 8(a), and the four connected junctions of two-way roads in 15×15 grid, as shown in Figure 8(b).

Figure 8 shows different colored circles representing cars controlled by agents, while different colored dashed lines represent possible routes. The cars take a probability from entering the arrivals point, and the car can take two actions: braking and forward. When the car reaches the target position on the edge of the grid, it will be removed. One or more road intersections exist in different maps, so the diverse routes are optional. It is considered a collision that two cars occupy the same position at the same time, and the agents will be penalized but will not affect the simulation. The identification of each car, its current location, and the assigned route are encoded in a one-hot binary vector set; each car

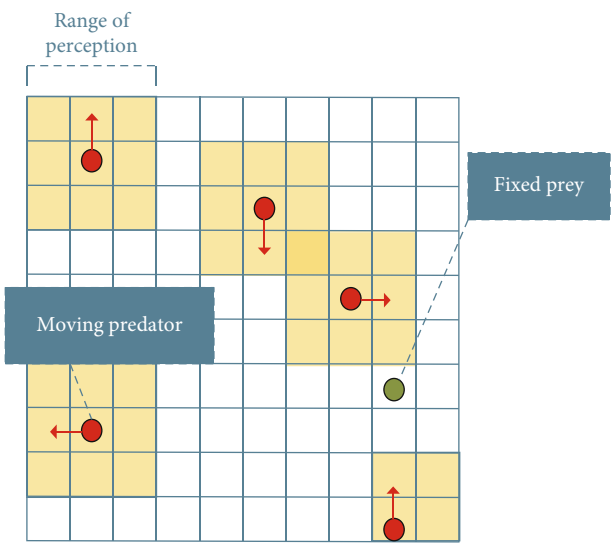


Figure 7: Predator-Prey environments' visualizations.

is finitely perceived as a grid of area (2×2) and can communicate with other cars under certain conditions.

5.1.3. Metrics. We will evaluate our method and baseline algorithm in terms of the following indicators.

In the Predator and Prey task:

(1) Maximum allowable steps taken $T_{\max}$:

$$T_{\max} = kT_{\text{finish}}, \tag{19}$$

where T_{finish} is the average time period from the start of the movement of the first predator to the location of the last predator to reach the prey. Select k values according to different map sizes and number of agents

(2) Success rate β:

$$\beta = \frac{n_{\text{success}}}{n}, \tag{20}$$

where β is the ratio of the number of successful events to the number of test events. An event is considered successful if all predators capture prey before the maximum allowed step taken $T_{\max}$

(3) Agent density ρ_{agent}:

$$\rho_{\text{agent}} = \frac{n_{\text{agent}}}{W \times H}, \tag{21}$$

where ρ_{agent} denotes the sparsity of the environment, n_{agent} is the number of agents in the map, and $W \times H$ is the size of the map

(4) Average step taken $T_{\text{steps-taken}}$:

$$T_{\text{steps-taken}} = \frac{T_{\text{total}}}{n_{\text{agent}}}, \tag{22}$$

where $T_{\text{steps-taken}}$ denotes the average step taken of each agent to reach the goal, T_{total} is the number of steps to complete the epoch, and n_{agent} is the sum of the number of agents

(5) Average reward r_{average}:

$$r_{\text{average}} = \frac{1}{n_{\text{agent}}} \sum_{i=1}^{n} r_i, \tag{23}$$

where r_{average} denotes the average of rewards obtained by each agent at each epoch, r_i is the individual reward of agent i, and n_{agent} denotes the number of agents

In the Predator and Prey environment, the reward function for agent i is set as follows:

$$r_i(t) = (1 - \alpha_i) * r_{\text{penalty}} + \alpha_i * n_t * r_{\text{success}}, \tag{24}$$

where α_i indicates whether the agent i has captured the prey and it is set to 1 or 0.1 indicating that the agent i has captured the prey at a time t; otherwise, it is 0. r_{penalty} represents the penalty value at a time step, equal to -0.05, which can encourage the agent to actively explore the environment. n_t represents the number of agents that captured the prey at a time t. r_{success} represents the reward value for catching the prey, which is equal to 0.05.

In the traffic junction task:

(1) Maximum allowable steps taken $\tilde{T}_{\max}$:

$$\tilde{T}_{\max} = k\tilde{T}_{\text{finish}}, \tag{25}$$

where $\tilde{T}_{\text{finish}}$ is the average time period from the start of the movement of the first car to the arrival of the last one at the target position at the edge of the grid. Selecting k values according to different size of tasks

(2) Success rate $\tilde{\beta}$:

$$\tilde{\beta} = \frac{\tilde{n}_{\text{success}}}{n}, \tag{26}$$

where $\tilde{\beta}$ is the ratio of the number of successful events to the number of test events. The simulation ends after the Maximum allowable steps taken $\tilde{T}_{\max}$. No collision is classified as a success, and if one or more collisions occur, it is classified as a failure

In the traffic junction environment, the reward function for agent i is set as follows:

$$\tilde{r}_i(t) = r_{\text{coll}} d_i^t + r_{\text{time}} \tau_i, \tag{27}$$

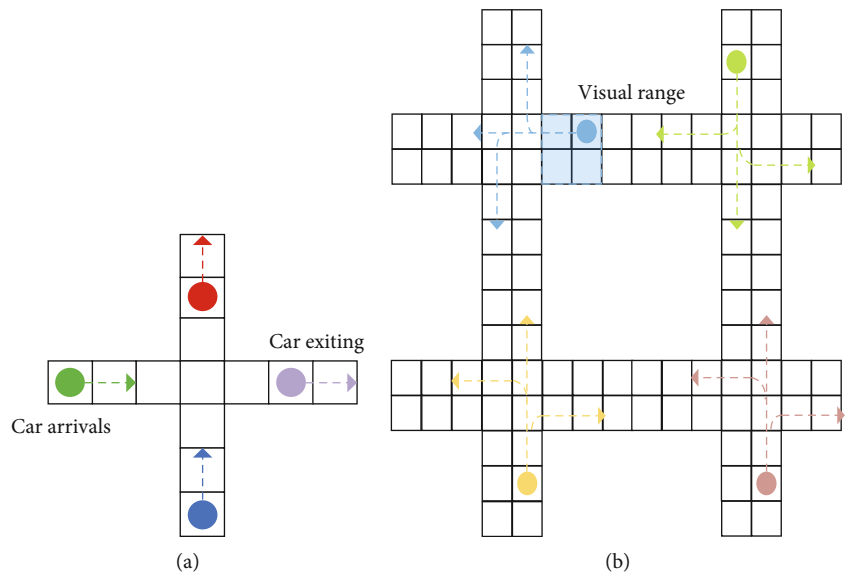


Figure 8: Easy and difficult versions of traffic junction task: (a) 7×7 grid of traffic junction task; (b) 15×15 grid of traffic junction task.

Table 1: Success rate in three density levels of Predator-Prey environment in cooperative setting.

	Predatory-Prey (success rate)		
Agent density (n_{agent})	10×10, 6 agents, sparse	10×10, 8 agents, normal	10×10, 10 agents, crowded
NC	$83.3 \pm 0.9\%$	$83.4 \pm 0.8\%$	$81.3 \pm 0.5\%$
IMANet	$90.3 \pm 0.8\%$	$86.8 \pm 0.8\%$	$82.7 \pm 0.8\%$

where r_{coll} denotes the penalty incurred when two cars collide, $r_{\text{coll}} = -10$. d_i^t is the number of times car i collides at time t, but the collision does not affect the car's route. The actions the car can take at each time step are braking and moving forward. τ_i is the number of steps that car i has elapsed from the time it starts to the moment t, $r_{\text{time}}\tau_i = -0.01\tau_i$.

5.2. Ablation Experiment. IMANet is attentional communication model, we conduct ablation experiments on the structure of the attention unit, and NC is a simplified version of IMANet without communication. NC has to train an independent policy network for each agent. In order to verify the influence of independent controllers in IMANet on the increase in the number of agents and the advantages of communication. We increase the number of agents on a map of the same size and the density of agents increases accordingly. We can test this by performing it in three different Predator and Prey environments with sparse, normal and crowded agent densities ρ_{agent}. With the success rate β and average reward r_{average}, we can evaluate the performance of the IMANet and NC. The success rate β is shown in Table 1, and the average reward value r_{average} is shown in Figure 9.

5.2.1. Scene Setting. There are sparse scene task with $n_{\text{agent}} = 6$ and $\rho_{\text{agent}} = 0.06$, normal scene task with $n_{\text{agent}} = 8$ and $\rho_{\text{agent}} = 0.08$, and crowded scene task with $n_{\text{agent}} = 10$ and $\rho_{\text{agent}} = 0.1$. Set the maximum allowable steps taken $T_{\max} = 40$.

As shown in Table 1, NC has success rates β of $83.3 \pm 0.9\%$ and $81.3 \pm 0.5\%$ on Sparse and Crowded, respectively. In the same size map environment, as the number of agents increases, the independent controller structure we use to guide each agent training effectively improves the dimensional explosion problem. As the number of agents increases on the same size map, independent controllers can help IMANet models scale to large teams of agents.

In Figures 9(a) and 9(b), the average reward value r_{average} of NC under the three task situations of sparse, normal, and crowded is close to 4.5. However, the average reward value of IMANet is close to 7.4, and the performance of IMANet is better than that of NC in all three density levels. This shows that attentional communication can improve the collaboration between agents in the cooperative task scenario.

5.3. Baseline. We conducted experiments comparing IMANet with NC and CommNet in predator and prey tasks and traffic junction task: (1) no communication, (2) communication but broadcast communication, and (3) targeted communication; set up task scenarios of different size; the benefits of communication and attention increase with the complexity of the task space. The comparison of IMANet with baseline work is shown in Table 2.

In the Predator and Prey environment, we set up three grid worlds of different sizes and different numbers of

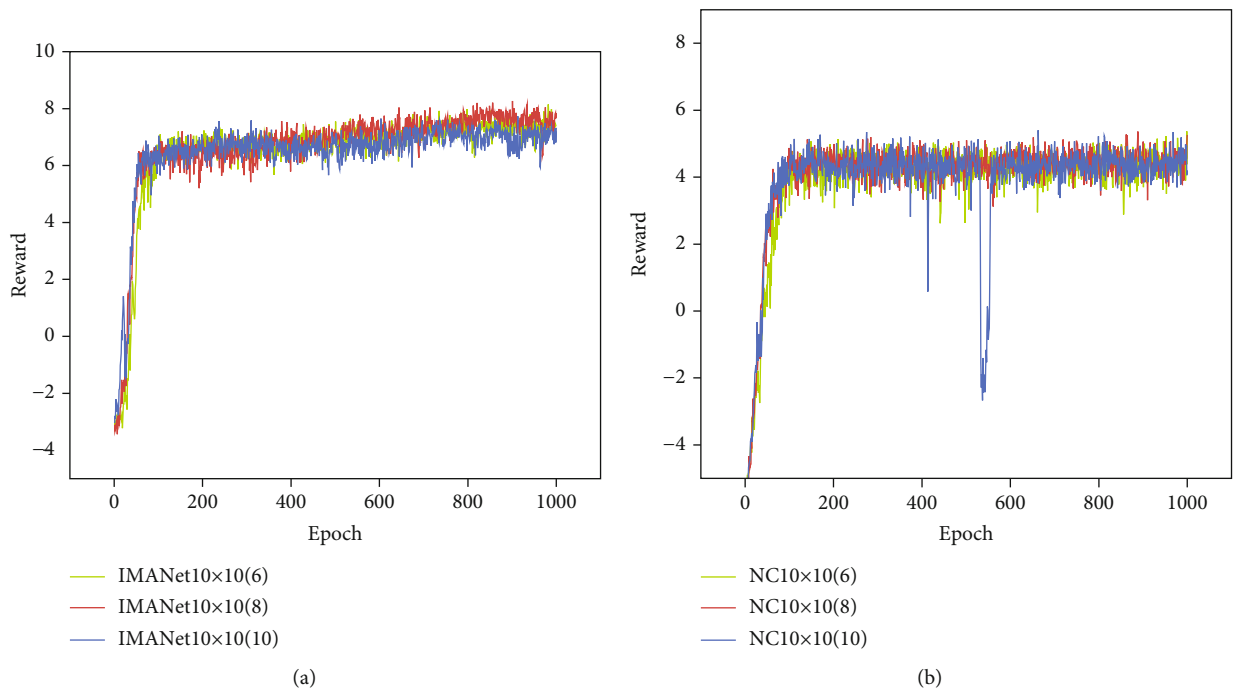


Figure 9: The average rewards (r_{average}) against the change of agent density setup: (a) average rewards (r_{average}) of IMANet; (b) average rewards (r_{average}) of NC.

Table 2: Comparison of IMANet's work with NC and CommNet's multiagent collaborative communication.

	Communication category	Execution
CommNet	Broadcast communication	Decentralized
NC	No communication	Decentralized
IMANet	Targeted communication	Decentralized

cooperative predators, and we evaluated IMANet against CommNet and NC on small, medium and large scales. The observation horizon of the agents is effective, thus emphasizing the importance of communication.

5.3.1. Scene Setting. The small task has 3 agents on a 4×4 grid with $T_{\max} = 20$, the medium task has 5 agents on a 8×8 grid with $T_{\max} = 40$, and the large task has 8 agents on a 15×15 grid with $T_{\max} = 75$.

We compare the convergence effect of the average step taken $T_{\text{steps-taken}}$ in 1000 epochs and the success rate β of IMANet, CommNet, and NC. Figure 10 shows the fast and slow convergence of the step taken as well as the smoothness, and Tables 3 and 4 report and evaluate the success rate and average step taken of the three algorithms at different scales, respectively.

In the traffic junction task, we can test what communication methods the agent performs during the task to avoid collisions while going through the intersection. At any time, the maximum number $N_{\max}$ of car is set at the intersection, each time a new car is added to the environment with a probability p_{arrive} from either of the different directions. We study IMANet with CommNet and NC in both easy and difficult traffic junction environments.

5.3.2. Scene Setting. The easy version is a connecting junctions consisting of two 7×7 one-way roads with $N_{\max} = 4$, $p_{\text{arrive}} = 0.25$, and $\tilde{T}_{\max} = 20$. The difficult version is a four-connection intersection consisting of four 15×15 two-way roads with $N_{\max} = 14$, $p_{\text{arrive}} = 0.05$, and $\tilde{T}_{\max} = 80$.

In order to analyze the impact of target communication on car through intersections in traffic scenarios, we use success rate $\tilde{\beta}$ evaluation of IMANet with CommNet and NC; Tables 5 reports and evaluates the success rate.

5.4. Analysis. AS shown in Table 3, on small, CommNet, NC, and IMANet get close to 99.9%, a communication baseline NC has success rates of$43.9 \pm 6.9\%$ on large-scale, a broadcast communication baseline CommNet has success rates of$49.0 \pm 0.4\%$on large-scale, and a targeted communication IMANet has success rates of$67.3 \pm 6\%$ on large-scale, which is 12% absolute improvement over CommNet.

As shown in Table 5, in the traffic junction experiment, there can be no good performance without communication. The baseline NC proves this. NC has success rates of $74.0 \pm 0.9\%$and $50.3 \pm 0.7\%$ on easy and difficult, respectively. On easy, both CommNet and IMANet get close to 93%. IMANet has success rates of $68.3 \pm 7.8\%$ on difficult, which is 10% absolute improvement over CommNet.

Figure 10 shows the average step taken convergence efficiency on cooperation task in Predator and Prey environment. As shown in Figures 10(a) and 10(b), on small scale, the average step taken convergence efficiency of IMANet, CommNet and NC are all close to 95%, on medium-scale,

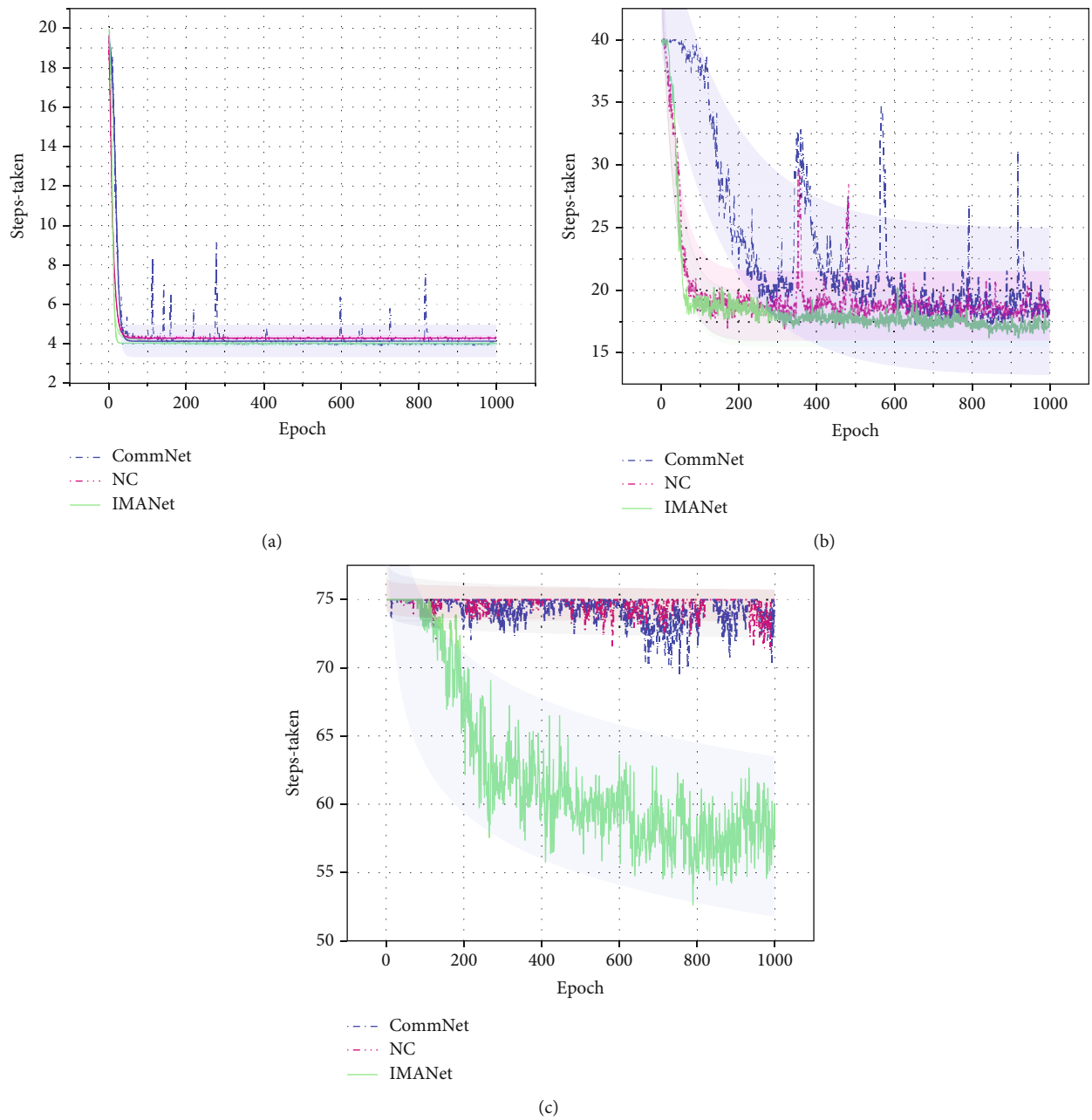


Figure 10: Average steps taken ($T_{\text{steps-taken}}$) to complete an episode of Predator-Prey environment. (a) Average step taken ($T_{\text{steps-taken}}$) at small scale. (b) Average step taken ($T_{\text{steps-taken}}$) at medium scale. (c) Average step taken ($T_{\text{steps-taken}}$) at large scale.

Table 3: Success rates of cooperative tasks with three different settings in a Predatory-Prey environment.

	Predatory-Prey (success rate)		
Map size (n_{agent})	4×4, 3 agents, small	8×8, 5 agents, medium	15×15, 8 agents, large
CommNet	$99.0 \pm 0.9\%$	$86.6 \pm 10\%$	$49.0 \pm 0.4\%$
NC	$98.0 \pm 0.6\%$	$92.6 \pm 0.5\%$	$43.9 \pm 6.9\%$
IMANet	$99.9 \pm 0.1\%$	$99.0 \pm 0.9\%$	$67.3 \pm 6\%$

the average step convergence efficiency obtained by CommNet is 71%, and the average step efficiency obtained by NC and IMANet is close to 92%. As shown in Figure 10(c), on large scale, the average step taken convergence efficiency of IMANet is 62%, which is much higher than the CommNet and the IC. As shown in Table 4, IMANet agents take 58.3 steps to capture prey on average vs. 73.9 for NC vs. 74.7 for CommNet; IMANet largely outperforms all the baselines.

Table 4: Average number of steps taken to complete the episode in three different environment size settings.

Map size (n_{agent})	Predatory-Prey (Avg. steps) 4×4, 3 agents, small	8×8, 5 agents, medium	15×15, 8 agents, large
CommNet	4.1 ± 0.01	18.1 ± 0.13	74.7 ± 0.04
NC	4.0 ± 0.00	18.7 ± 0.04	73.9 ± 0.06
IMANet	4.0 ± 0.00	17.6 ± 0.03	58.3 ± 0.09

Table 5: Success rate on various difficulty levels.

Level	Traffic junction (success rate) Easy	Difficult
CommNet	$93.2 \pm 2.5\%$	$53.4 \pm 3.7\%$
NC	$74.0 \pm 0.9\%$	$50.3 \pm 0.7\%$
IMANet	$93.0 \pm 5.2\%$	$68.3 \pm 7.8\%$

The above experiments verify that our method can be effectively extended to large-scale agent teams.

5.4.1. NC vs. IMANet. This communication-free approach has the advantage that it can cope well with the increase or decrease of agents, but it cannot observe the information of other agents and cannot coordinate the actions of agents. IMANet outperforms NC; we can see communication indeed helps.

5.4.2. CommNet vs. IMANet. However, CommNet also has communication. Why does it perform so much worse in a large-scale task space? CommNet uses centralized training and decentralized execution, where the policies of multiagents form a large network. In this structure, all hidden layer vectors are stacked though broadcast communication to generate a communication vector. Input the joint observation value of the agent, and the output is the action of all the agents. The action generation of this structure is based on a joint policies. As the number of agents increases, the number of observations increases linearly and the space of joint actions increases exponentially. Therefore, as shown in Figure 10(b), the average step taken learning training process for the baseline experiment appears to be unstable. Our IMANet uses an independent LSTM network as a controller to guide the generation of action strategies for the agents. The agent uses an attention unit to weight and aggregate the hidden layer vectors passed by each agent, and since each agent pays more attention to information that is more similar to its own observations, the agent has continuity in its observations and so does the generated communication messages, which makes the training process smoother. This decentralized execution allows the agents to generate appropriate actions based on their own hidden layer states. This independent network structure allows each agent to update its policy according to its own reward function. It can effectively deal with the problems caused by the increase in dimensionality. Through the above experiments, we can conclude that our content-based optimization method IMANet has better scalability and stability in the cooperative agent setting compared to the traditional broadcast communication method. It can better handle the problems that arise with the increase of task space.

6. Conclusions

We propose IMANet to learn the communication between agents in a fully cooperative multiagent task. In IMANet, we embed attention units based on query vectors. The attention units can compress the state values of the hidden layer more efficiently by assigning different attention sizes, and since the observations of the agents have continuity, the generated communication information also has continuity, which makes the training process smoother. Also this independent network structure allows each agent to update its policy according to its own reward function. This decentralized execution allows the agents to generate more beneficial value actions based on their own hidden layer states. The algorithm performs significantly better than the other two algorithms in experiments with Predator and Prey and traffic junction. Moreover, the agent remained significantly more scalable in a larger task space. Attention-based communication can indeed help the agent to perform tasks more effectively in the MARL environment.

Data Availability

The data used to support the findings of this study are available from the authors upon request.

Conflicts of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the Basic Scientific Research Project of Education Department of Liaoning Province: Research on Large-scale Cooperative Learning Methods of Multi-agent in Uncertain Environment (No. LJKZ1180), and the National Natural Science Foundation of China: Research on the stability of multisurface high-speed unmanned boat formation and the method of cooperative collision avoidance in complex sea conditions (No. 61673084).

References

[1] R. Olfati-Saber, J. A. Fax, and R. M. Murray, "Consensus and cooperation in networked multi-agent systems," *Proceedings of the IEEE*, vol. 95, no. 1, pp. 215–233, 2007.

[2] C. V. Goldman and S. Zilberstein, "Decentralized control of cooperative systems: categorization and complexity analysis," *Journal of Artificial Intelligence Research*, vol. 22, pp. 143–174, 2004.

[3] C. V. Goldman and S. Zilberstein, "Optimizing information exchange in cooperative multi-agent systems," in *Proceedings of the second international joint conference on Autonomous*

agents and multiagent systems, pp. 137–144, Sydney, Australia, 2003.

[4] R. S. Sutton, "Policy gradient method for reinforcement learning with function approximation," in *Proceedings of the 1998 IEEE International Conference on Robotics & Automation*, Leuven,Belgium, 2000.

[5] Y. Hu, W. Wang, H. Jia et al., "Learning to utilize shaping rewards: a new approach of reward shaping," 2020, https://arxiv.org/abs/2011.02669.

[6] L. Busoniu, R. Babuska, and B. D. Schutter, "A comprehensive survey of multiagent reinforcement learning," *IEEE Transactions on Systems*, vol. 38, no. 2, pp. 156–172, 2008.

[7] A. K. Agogino and K. Tumer, "Unifying temporal and structural credit assignment problems," *IEEE Computer Society*, vol. 4, pp. 980–987, 2004.

[8] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems*, pp. 1097–1105, 2012.

[9] R. Shen, Y. Zheng, J. Hao et al., "Generating behavior-diverse game AIs with evolutionary multi-objective deep reinforcement learning," in *Proceedings of the Twenty-Ninth International Conference on International Joint Conferences on Artificial Intelligence*, pp. 3371–3377, Yokohama , Japan, 2020.

[10] Y. Yang, J. Li, and L. Peng, "Multi-robot path planning based on a deep reinforcement learning DQN algorithm," *CAAI Transactions on Intelligence Technology*, vol. 5, no. 3, pp. 177–183, 2020.

[11] V. François-Lavet, P. Henderson, R. Islam, M. G. Bellemare, and J. Pineau, "An introduction to deep reinforcement learning," *Foundations and Trends in Machine Learning*, vol. 11, no. 3-4, 2018.

[12] T. T. Nguyen, N. D. Nguyen, and S. Nahavandi, "Deep reinforcement learning for multiagent systems: a review of challenges, solutions, and applications," *IEEE transactions on Cybernetics*, vol. 50, no. 9, pp. 3826–3839, 2020.

[13] P. Peng, Y. Wen, Y. Yang et al., "Multiagent bidirectionally-coordinated nets: emergence of human-level coordination in learning to play," 2017, https://arxiv.org/abs/1703.10069.

[14] R. Lowe, Y. Wu, A. Tamar, J. Harb, P. Abbeel, and I. Mordatch, "Multi-agent actor-critic for mixed cooperative-competitive environments," pp. 6379–6390, 2017, https://arxiv.org/abs/1706.02275.

[15] K. Zhang, Z. Yang, and T. Baar, *Multi-Agent Reinforcement Learning: A Selective Overview of Theories and Algorithms*, Handbook of Reinforcement Learning and Control, 2021.

[16] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[17] D. S. Bernstein, R. Givan, N. Immerman, and S. Zilberstein, "The complexity of decentralized control of Markov decision processes," *Mathematics of Operations Research*, vol. 27, no. 4, pp. 819–840, 2002.

[18] S. Iqbal and F. Sha, "Actor-attention-critic for multi-agent reinforcement learning," in *International Conference on Machine Learning*, Stockholm ,Sweden, 2018.

[19] J. Foerster, G. Farquhar, T. Afouras, N. Nardelli, and S. Whiteson, "Counterfactual multi-agent policy gradients," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 32, San Francisco, USA, 2017no. 1.

[20] L. Haitao, *Research on several communication technologies in multi-agent robot system*, Harbin Institute of Technology, 2007.

[21] H. Mao, W. Liu, J. Hao et al., "Neighborhood cognition consistent multi-agent reinforcement learning," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34no. 5, pp. 7219–7226, Hilton New York Midtown, New York, USA, 2020.

[22] R. Wang, X. He, R. Yu, W. Qiu, B. An, and Z. Rabinovich, "Learning efficient multi-agent communication: an information bottleneck approach," in *InInternational Conference on Machine Learning*, pp. 9908–9918, Vienna, AUSTRIA, 2020 Nov 21.

[23] H. Mao, Z. Gong, Z. Zhang, Z. Xiao, and Y. Ni, "Learning multi-agent communication under limited-bandwidth restriction for internet packet routing," 2019, https://arxiv.org/abs/1903.05561.

[24] S. Sukhbaatar, A. Szlam, and R. Fergus, "Learning multiagent communication with backpropagation," *Advances in neural information processing systems*, vol. 29, pp. 2244–2252, 2016.

[25] D. Kim, S. Moon, D. Hostallero et al., "Learning to schedule communication in multi-agent reinforcement learning," 2019, https://arxiv.org/abs/1902.01554.

[26] W. Kim, M. Cho, and Y. Sung, "Message-dropout: an efficient training method for multi-agent deep reinforcement learning," *Proceedings of the AAAI conference on artificial intelligence*, vol. 33, pp. 6079–6086, 2019.

[27] J. Jiang and Z. Lu, "Learning attentional communication for multi-agent cooperation," 2018, https://arxiv.org/abs/1805.07733.

[28] M. Geng, K. Xu, X. Zhou, B. Ding, H. Wang, and L. Zhang, "Learning to cooperate via an attention-based communication neural network in decentralized multi-robot exploration," *Entropy*, vol. 21, no. 3, p. 294, 2019.

[29] A. Das, T. Gervet, J. Romoff et al., "Tarmac: targeted multi-agent communication," *International Conference on Machine Learning*, vol. 97, pp. 1538–1546, 2018.

[30] H. Mao, Z. Zhang, Z. Xiao, Z. Gong, and Y. Ni, "Learning multi-agent communication with double attentional deep reinforcement learning," *Autonomous Agents and Multi-Agent Systems*, vol. 34, no. 1, 2020.

[31] P. Stone and M. Veloso, "Multiagent systems: a survey from a machine learning perspective," *Autonomous Robots*, vol. 8, no. 3, pp. 345–383, 2000.

[32] G. Brockman, V. Cheung, L. Pettersson, and J. Schneider, "OpenAI Gym," 2016, https://arxiv.org/abs/1606.01540.

[33] G. Hinton, N. Srivastava, and K. Swer-sky, "Neural networks for machine learning lecture 6a overview of mini-batch gradient descent," vol. 14, 2012.

Hindawi
Wireless Communications and Mobile Computing
Volume 2023, Article ID 9850836, 1 page
https://doi.org/10.1155/2023/9850836

Retraction

Retracted: A System for Trusted Recovery of Data Based on Blockchain and Coding Techniques

Wireless Communications and Mobile Computing

Received 10 October 2023; Accepted 10 October 2023; Published 11 October 2023

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. Chen, Y. Yan, S. Guo, Y. Ren, and F. Qi, "A System for Trusted Recovery of Data Based on Blockchain and Coding Techniques," *Wireless Communications and Mobile Computing*, vol. 2022, Article ID 8390241, 12 pages, 2022.

Hindawi
Wireless Communications and Mobile Computing
Volume 2022, Article ID 8390241, 12 pages
https://doi.org/10.1155/2022/8390241

Research Article

A System for Trusted Recovery of Data Based on Blockchain and Coding Techniques

Jinqian Chen,[1] **Yong Yan**,[2] **Shaoyong Guo**,[1] **Yinlin Ren**,[1] **and Feng Qi**[1]

[1]*State Key Laboratory of Networking & Switching Technology, Beijing University of Posts and Telecommunications, Beijing 100876, China*
[2]*Electric Power Research Institute, Zhejiang Electric Power Corporation, Hangzhou 310007, China*

Correspondence should be addressed to Feng Qi; qifeng@bupt.edu.cn

Received 9 August 2021; Revised 26 October 2021; Accepted 3 December 2021; Published 17 January 2022

Academic Editor: Chi-Hua Chen

With the continuous development of information technology, the Internet of Things has also been widely used. At the same time, in the power Internet of Things environment, reliable data is essential for data use and accurate analysis. Data security has become a key factor in ensuring the stable operation of the power grid. However, the power Internet of Things devices is extremely vulnerable to network attacks, leading to data tampering and deletion. Resisting tampering, preventing data loss, and reliably restoring data have become difficult to ensure data security. In order to solve this problem, this paper proposes a trusted data recovery system based on blockchain and coding technology. Data nodes of the power Internet of Things encode key data and back them up to the blockchain network through a data processing server located on the edge. The data processing server performs real-time detection of the data integrity of the data nodes. When the data is tampered with or deleted, the data processing server promptly obtains the corresponding data encoding blocks from the blockchain network, decodes them, and sends them to the data node to complete the data recovery task. According to the test result, the data backup speed of this system is increased by 15.3%, and the data recovery speed is increased by 19.8% compared with the traditional scheme. It has good security and real-time performance. Meanwhile, it reduces the network and storage resource overhead in the data backup and recovery process.

1. Introduction

With the rapid development of network information technology, network attacks have become more and more common. Network hackers can easily use Trojan horses, worms, and local vulnerabilities to attack devices on the network. In the power Internet of Things, attackers usually maliciously delete or tamper with the data (such as collected and summarized data, firewall configuration information, virus database data, and log file) stored in the terminal to interfere with the stable operation of the power Internet of Things. For example, attackers can tamper with the firewall configuration of the device to further implement a distributed denial of service (DDOS) attack. IoT administrators can use data backup and recovery technology to reconstruct the firewall. Then, the IoT environment can be repaired. So, data backup and recovery technology is an important method to ensure the stable operation of the power Internet of Things [1]. Currently, mainstream data recovery solutions mainly include two types: (1) centralized data backup and recovery solutions. For example, all data nodes in the Internet of Things back up their own data to a central server (usually, it is a cloud service) according to the rule. When the data of a node is lost, it downloads the corresponding data from the centralized server to complete the recovery; (2) distributed data backup and recovery scheme. For example, in a distributed storage system, data is usually stored redundantly. When the node data is lost or tampered with, the IoT node can use the P2P network to download the required data from the remaining surviving nodes to complete data recovery [2, 3]. However, both of these schemes have certain shortcomings. The first scheme has the

advantages of low cost and easy management, but it also brings low security and reliability [4]. For example, a centralized storage system may be attacked by network hackers, resulting in data stored in the centralized server being tampered with or stolen. At the same time, studies have found that data recovery in a centralized manner is uneconomical [5] because it has to solve the problems of large storage data, network congestion, complex asynchronous processing, and low efficiency [6, 7]. In the second scheme, the local data node is vulnerable to hijacking or worm infection. Malicious data may be sent to the data nodes during the recovery process, such as the Stuxnet worm. Therefore, the second solution is difficult to solve the problem of untrusted data sources in the data recovery process. It is difficult to meet the security and reliability requirements of the power Internet of Things data recovery system.

The blockchain is essentially a distributed database. Blockchain has the characteristics of decentralization and can ensure that the stored data is difficult to tamper with. At the same time, the blockchain's oracle mechanism can ensure that reliable data is provided for the blockchain, and IoT data nodes can accurately and reliably obtain data on the blockchain through smart contracts. Therefore, blockchain technology provides a promising solution to the above problems. However, the storage scalability of the blockchain is poor, and each node needs a complete ledger in the storage system to maintain the decentralization and consistency of the ledger, which puts high storage requirements on each node server. This has caused the storage scalability of the blockchain to become a major bottleneck for the application of blockchain technology to data backup and recovery in the power Internet of Things. Coding technology can effectively reduce storage resource overhead, so coding technology has begun to be applied to improve the scalability of blockchain storage [8]. As a typical coding scheme, erasure coding can effectively solve the problem that the distributed storage system cannot complete data reconstruction due to partial loss of data slices during data transmission, improving data transmission efficiency and reducing network resource overhead. At the same time, the data coding block on the blockchain is publicly visible to the entire network. In order to prevent the coding matrix from being stolen by network attackers and cause IoT data leakage and to ensure the privacy of power IoT data, it is necessary to apply access control technology to protect the coding matrix.

In summary, the existing Internet of Things data backup and recovery methods have disadvantages such as high storage overhead and low data recovery efficiency. At the same time, these methods face the challenges of data leakage and untrusted data sources in data recovery. Therefore, this paper proposes a trusted recovery system for power Internet of Things data with high performance. The system can effectively ensure the privacy of power Internet of Things data while achieving decentralized and credible recovery. Our contributions are summarized below.

(i) A trusted data recovery system is implemented based on blockchain technology. Data is stored in the blockchain in the form of coded blocks. The data of the blockchain is verified based on the oracle mechanism to ensure the reliability of the data. We built the corresponding smart contract. The data processing server will automatically and accurately obtain the coding block for data recovery from the blockchain network

(ii) Based on the Jerasure coding library, an improved coding scheme is proposed. This scheme increases the data size of each matrix operation to 128 bytes and sets the word size to 64 bits. At the same time, preprocessing and concurrent computer mechanisms are added to read the data in advance. It is processed in the cache to reduce I/O operations and increase the encode and decode rate

(iii) Based on the CP-ABE algorithm, an attribute-based access control model is proposed to protect the encoding matrix with fine-grained granularity. System administrators formulate access control strategies, and the access control model will automatically distribute the coding matrix according to the access control strategy based on the subject attributes of the data nodes of the Internet of Things and the object attributes of the data

(iv) Through simulation tests, the system has better data backup and recovery efficiency than previous data recovery systems. The data backup rate has increased by 15.3%, and the data recovery rate has increased by 19.8%. The system can effectively guarantee the privacy of IoT data while reducing the storage and communication resource overhead during data backup and recovery

2. Related Work

With the continuous development of information technology, the Internet of Things has been widely used. In applying the Internet of Things and wireless sensors, the terminals in the Internet of Things will use the data stream for specific data processing and analysis. Due to unexpected power on of devices and network attacks, local data nodes may experience data tampering and loss. Reliable data is essential to data accurate analysis. The safe storage and trusted recovery of data become critical in ensuring the reliable operation of the Internet of Things.

In terms of data security storage, Tchernykh et al. [9] proposed a multicloud storage architecture called WA-MRC-RRNS in the Internet of Things environment, which combines a weighted access control scheme, threshold secret sharing, and remaining redundant data. The system also proposes a multinode fault detection/data recovery mechanism based on the redundant residual number system. Through simulation tests, this architecture can effectively ensure data security and improve the data credibility of the IoT infrastructure.

Xia et al. [10] proposed a highly robust, secure, and trust-oriented IoT data storage model (RoSES) based on

edge storage and completely partial reconstruction codes. The storage model can achieve data robustness, high security, and lightweight local computing. At the same time, the model supports a trust-oriented data access (TODA) strategy, which can realize the legal data access of uncertain access requesters under the premise of wide applicability, thereby realizing the secure sharing of IoT data.

Wang et al. [11] studied the data recovery problem based on QoS guarantee and system robustness in the information-centric Internet of Things system (IC-IoT). They proposed a data recovery algorithm based on rarity perception. Its core idea is to establish a rarity index to evaluate data copies and service demand distribution comprehensively. The algorithm will eliminate unnecessary copies and gradually restore the original data according to the rarity of the original data and the priority of recovery. Experimental and simulation results show that compared with the traditional direct data recovery method, the algorithm has better QoS performance, and the robustness has been significantly improved.

The Internet of Things (IoT) has become an emerging technology in the past decade, and the number and research prospects of smart devices and related technologies have grown rapidly. Due to IoT terminals' low processing power and storage capacity, the existing security or encryption technology is not suitable for protecting IoT data [12]. At the same time, the importance of IoT security will become more obvious and huge. However, the Internet of Things still has many security issues and is vulnerable to attacks from some potential factors [13]. Therefore, some researchers have used blockchain technology as a decentralized method to solve security and privacy issues in the Internet of Things, such as data management, access control, and data recovery.

Bae and Shin [14] proposed an automatic recovery mechanism for data systems based on the blockchain. This mechanism uploads the data copy summary to the blockchain and regularly checks and restores the data. Before the data recovery process, a comparison will be made. Blockchain copy summary value information and local redundant copy summary value information are checked to ensure that the copy is correct. This mechanism solves the problem that the local data copy cannot be restored correctly after the local data copy is tampered with. In smart cities, Mishra and others [15] have designed a data protection system for key document data of large organizations based on the blockchain, which includes inspection, location, and recovery. The malicious behaviour of data tampering is detected and recovered, but this system is only suitable for recovering document data, not real-time data. In important business areas, Zhang and Li [16] combined blockchain and smart contract technology to improve the existing backup and recovery technology and adopted role-based access control strategies to strictly audit the data backup and recovery process to prevent data from being compromised. In the field of the supply chain, Cha et al. [17] proposed a data management and recovery system that uses blockchain and key agent encryption. This system enhances data integrity, availability, and traceability while solving the failures, denial of service attacks, and undeniable problems. In order to reduce storage costs, Liu et al., N. Liang et al., and W. Liang et al. [18, 19, 20] used coding techniques to improve the storage efficiency of distributed storage systems. In the field of smart grid, blockchain technology is also beginning to be applied. Ferrag and Shu [21] proposed a novel deep learning and blockchain-based energy exchange framework for smart grids. This framework uses blockchain for facilitating the exchange of excess energy among neighboring nodes. It also uses a recurrent neural network to detect network attacks and fraudulent transaction in exchange.

Through research on data security storage and trusted recovery related literature, it is found that people usually do not need to back up their data in a central database. Data can be safely stored in distributed storage systems or different blockchain network nodes, and the blockchain can guarantee its authenticity. Moreover, it can prevent unauthorized access. However, some of the above studies are still in the discussion of theoretical concepts [15, 16], some apply blockchain to data recovery in specific fields [17], and some apply access control technology or key technology to data credible recovery to ensure data recovery security, and some just use coding technology to optimize storage efficiency [18, 19, 20]. In summary, there are few studies on improving the recovery efficiency while ensuring the credibility of the data recovery process.

At present, researchers have adopted blockchain technology to ensure the security and privacy of the data recovery process [17, 22]. However, the former blockchain-based data recovery methods use symmetric encryption to protect data privacy. Therefore, these methods have the disadvantage of low data recovery speed. These methods also ignore the reliability of data transmission. The method proposed in this paper uses code techniques to complete the storage of the backup data, which can effectively improve the reliability of data transmission. And this method uses an access control strategy to strictly protect the coding matrix. It can protect the privacy of data while ensuring the high efficiency of the system.

3. System Architecture

Figure 1 shows the architecture of the system. The system mainly includes four entities: blockchain network, local data node, data processing server, and access controller. The power Internet of Things sensors broadcast the raw data to a distributed storage system composed of multiple local nodes. According to the corresponding backup strategy, the local data sends the key data to the data processing server on the edge side. After receiving the data, the data processing server uses the corresponding encoding matrix to encode the key data to generate a data encoding block and finally upload it to the blockchain network to complete the data backup. When the data is lost, the local data node obtains the corresponding coding matrix from the access controller and then sends the data recovery request and the coding matrix to the data processing server. After receiving the data recovery request, the data processing server downloads the corresponding data encoding block from

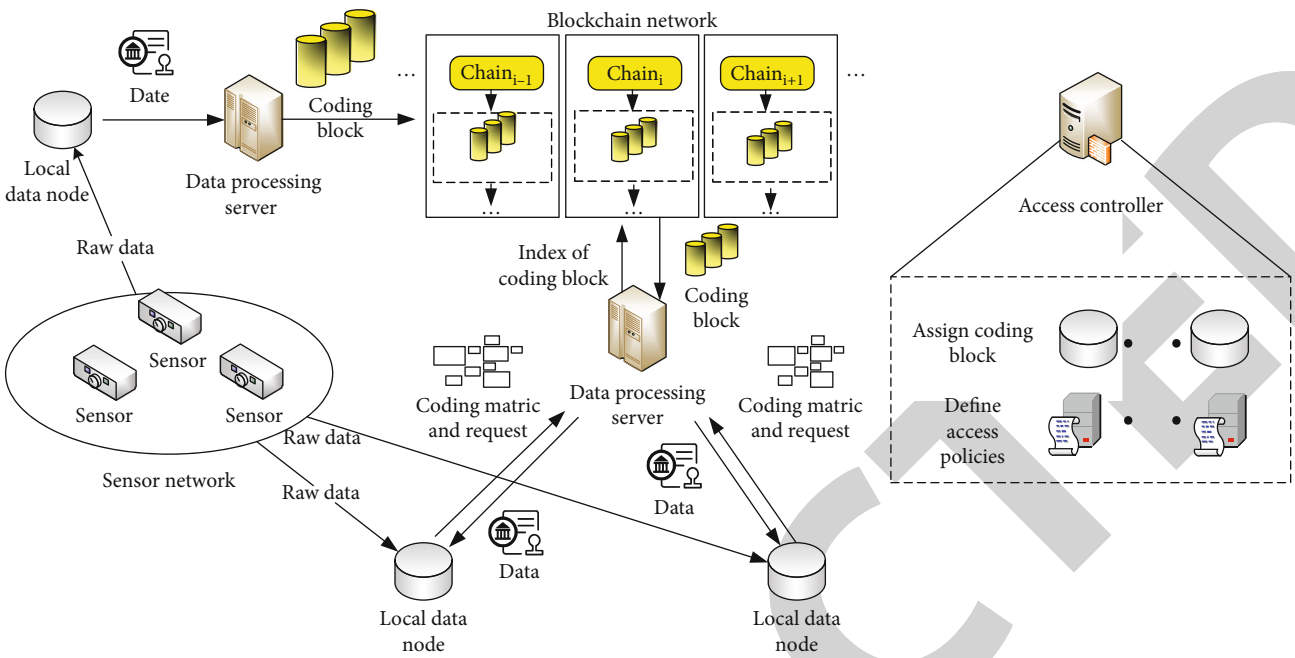


Figure 1: Architecture of system.

the blockchain network, uses the encoding matrix submitted by the local data node to complete the data decoding operation, and finally sends the data to the local data node to complete the data recovery process. The system applies coding and access control technology to data trusted recovery work to solve the problems of large data storage overhead, low transmission efficiency, and privacy leakage in the blockchain system.

3.1. Entities. As shown in Figure 1, the system includes four types of entities.

Local data nodes: the sensors in the power Internet of Things store the collected data in local data nodes, and multiple local data nodes of the same kind form a distributed storage system in full replication mode. The local data node can send data backup and recovery requests to the data processing server.

Data processing server: the data processing server is located on the edge of the local data node and is responsible for processing the data backup and recovery requests submitted by the local data node. Furthermore, it completes the data encoding and decoding operations. The data processing server is trusted for the power Internet of Things.

Access controller: the system uses an access controller to fine-grained control of the key data of the power Internet of Things. The access controller defines the data access strategy to protect the coding matrix and distributes the coding matrix to the data nodes. If the data node has sufficient permissions, it can obtain the corresponding data, and the access controller is centrally maintained and configured by the administrator in the power Internet of Things.

Blockchain network: the blockchain network is responsible for storing the coding blocks of the key data of the power Internet of Things. The local data node obtains the code block from the blockchain network through the data processing server to complete data recovery. The blockchain network is composed of multiple blockchains, and each blockchain is maintained by a single or multiple blockchain nodes.

3.2. Operations. This system provides two types of operation: data backup and data recovery.

Data backup: data backup needs to go through the following three steps:

(1) The local data node sends the data to be backed up and the coding matrix to the data processing server. The data processing server first preprocesses the data and divides it into blocks. Assuming that the data size is M, divide it into k original data blocks of fixed size, denoted as $(F_i)_{i=1,2,\cdots,n}$, and the size of each data block is M/k

(2) The data processing server uses the erasure coding scheme to encode the original data block to obtain $k+m$ coded data blocks (m is the number of check blocks), denoted as $(P_i)_{i=1,2,\cdots,k+m}$. Each coding block is a linear combination of original data blocks. After the encoding is completed, the data processing server sends the used encoding matrix E, data identification, and attributes to the access controller

(3) The data processing server sends $k+m$ code blocks to the blockchain network. The blockchain network stores the coded data blocks according to the corresponding rules. There are $k+m$ blockchains in the system, and each node stores n coded data blocks

When $k=2$, $m=2$, and $n=1$, the data backup process is shown in Figure 2.

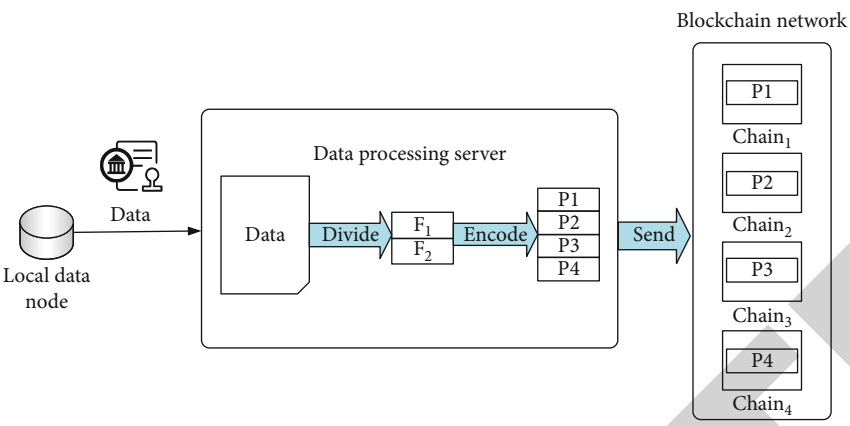


Figure 2: Data backup.

Data recovery: data recovery needs to go through the following three steps.

(1) The local data node sends its digital signature and identification of the data to be restored to the access controller. The controller assigns the coding matrix *EM* to the local data node according to the access control strategy

(2) The local data node sends the data recovery request and the coding matrix *EM* to the data processing server, and the data processing server downloads the required data coding block from the node in the blockchain network according to the data identification (usually the node with a smaller load).

(3) The data processing server decodes the downloaded data code blocks to obtain k original blocks. The server combines these blocks to obtain the original data and returns it to the local data node

When $k = 2$, $m = 2$, and $n = 1$, the data recovery process is shown in Figure 3.

3.3. Design Goals. We list the design goal of this system as follows:

(1) Integrity of data: the data recovery system should ensure that the data was submitted by an authenticated data node and has not been tampered with

(2) Data access control: the encoding matrix should be enforced with fine-grained access control policies so that only authorized data nodes can obtain raw data

(3) Data recovery efficiency: critical data should be backed up and restored efficiently through blockchain and edge computing nodes to adapt to large industrial systems

4. Proposed System

4.1. Data Node Registration Module. The system uses an attribute-based access control mechanism, and each data node has a unique ID as its identifiable identifier. The access controller formulates a registration strategy. Each data node in the system can apply to the access controller for registration and obtain subject attributes through its own MAC address and identifier *ID*. If the data node is verified, the access controller will generate the registration transaction and sign the hash value and timestamp of the registration transaction. Finally, the access controller will pack the registration transaction, signature information, and timestamp together and put it into its transaction pool. In this workflow, all parties communicate through the TCP protocol. The workflow chart is shown in Figure 4.

4.2. Data Encoding and Decoding Module. Erasure code is a forward error correction technology that originated in the field of communications. Erasure codes have low redundancy and high accuracy. We use erasure codes in a storage system, first set up an encoding matrix $E.M.$ based on certain rules, and use the encoding matrix and data slices to do matrix multiplication operations to obtain a set of encoded data blocks, as shown in Figure 5 below. The coding matrix here is an 8-row 5-column matrix composed of a fifth-order unit matrix and three rows. Through calculation, the coded data obtained has a total of 8 rows; the first five rows are data slices, and the last three rows satisfy

$$\begin{aligned} C_1 &= B_{11} * D_1 + B_{12} * D_2 + B_{13} * D_3 + B_{14} * D_4 + B_{15} * D_5, \\ C_2 &= B_{21} * D_1 + B_{22} * D_2 + B_{23} * D_3 + B_{24} * D_4 + B_{25} * D_5, \\ C_3 &= B_{31} * D_1 + B_{32} * D_2 + B_{33} * D_3 + B_{34} * D_4 + n_{35} * D_5. \end{aligned} \tag{1}$$

As shown in Figure 6, if a node needs to reconstruct the original data, it needs to randomly select k different coding blocks from the data coding block set. The node obtains the inverse matrix corresponding to the coding matrix and multiplies the inverse matrix with the taken coding block, and then, the original data D is completely reconstructed.

In this system, the Vandermonde matrix is selected as the coding matrix. When calculating the coding matrix,

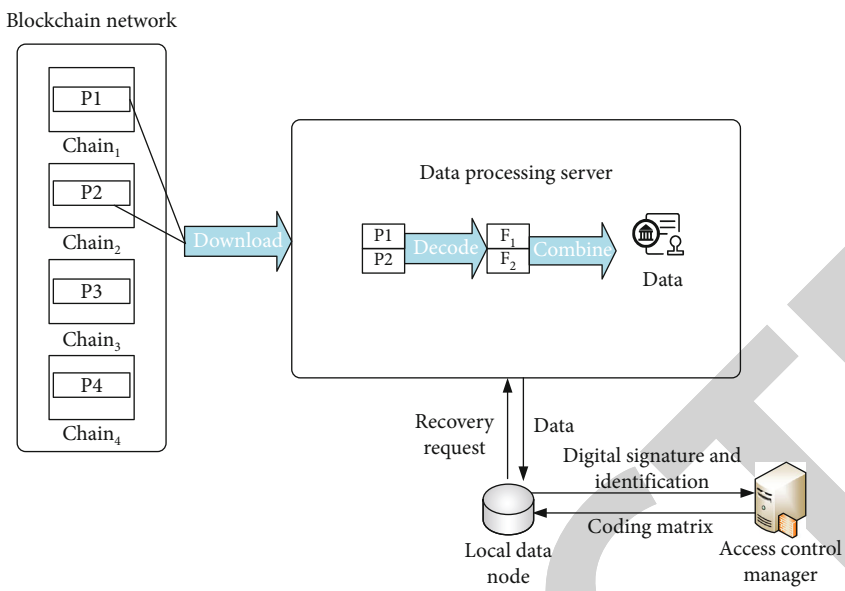


Figure 3: Data recovery.

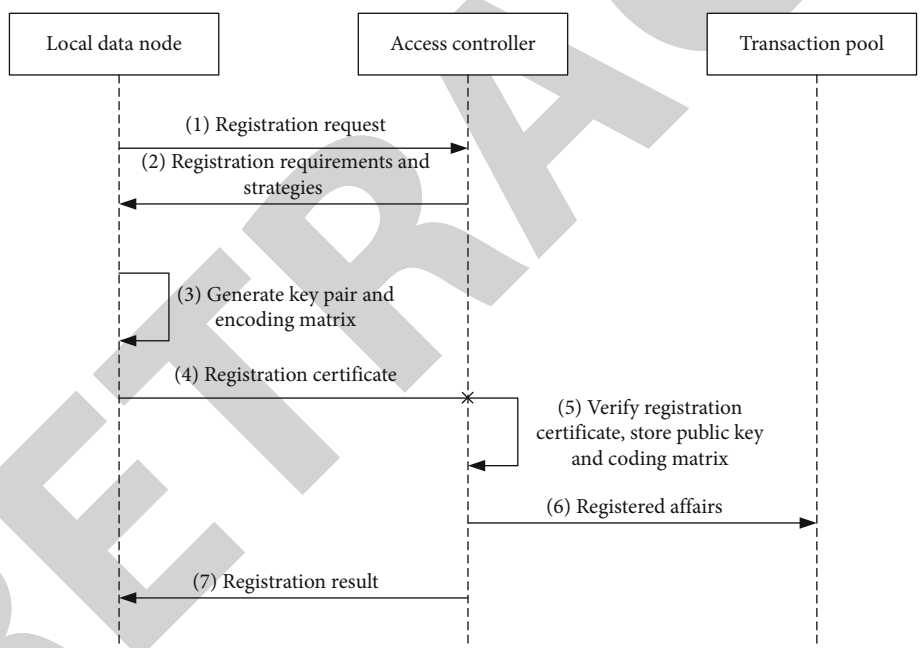


Figure 4: Data node registration flowchart.

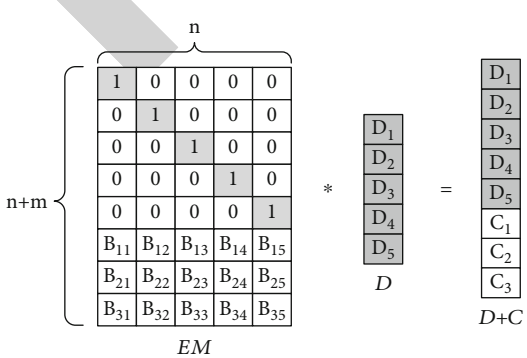


Figure 5: Process of data coding.

different matrix factors can be generated according to the different identification IDs of each data node.

4.3. Data Backup Module. Blockchain can be divided into three categories: public blockchain, private blockchain, and consortium blockchain. The private blockchain has the characteristics of fast transaction speed, and its data storage access efficiency is very close to conventional databases, which meets the real-time requirements of the system. The system designs a data blockchain network based on private blockchain. Single or multiple nodes maintain each private blockchain in the blockchain network to achieve safe and reliable data recovery. The model is shown in Figure 7.

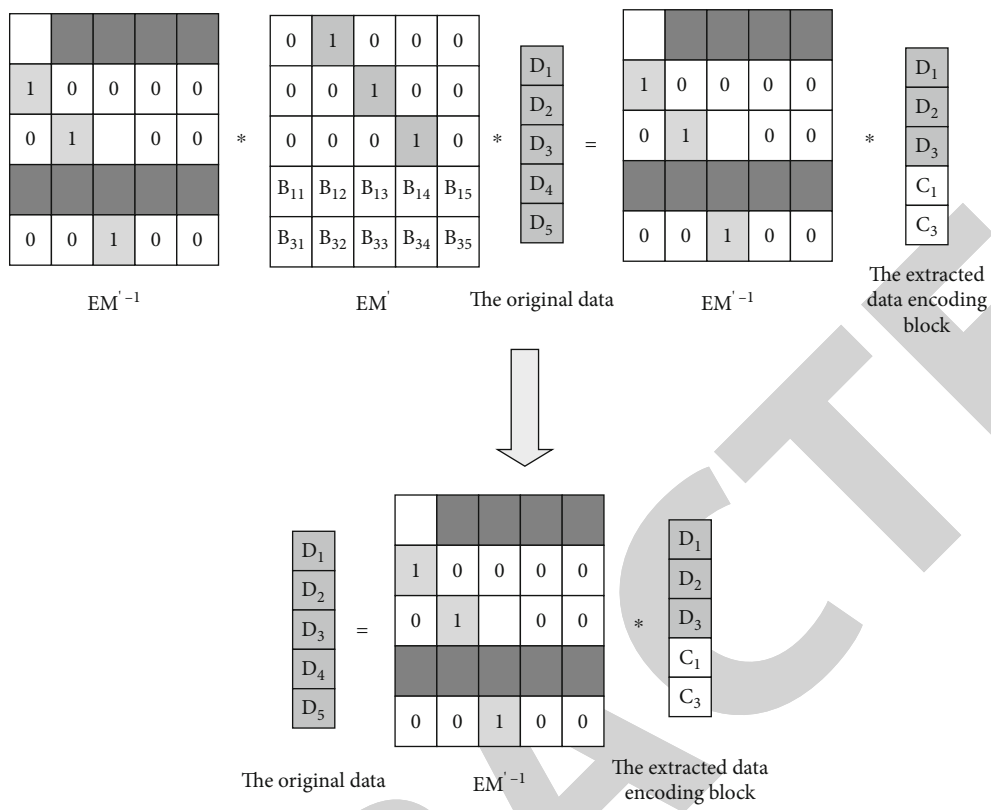


Figure 6: Process of data decoding.

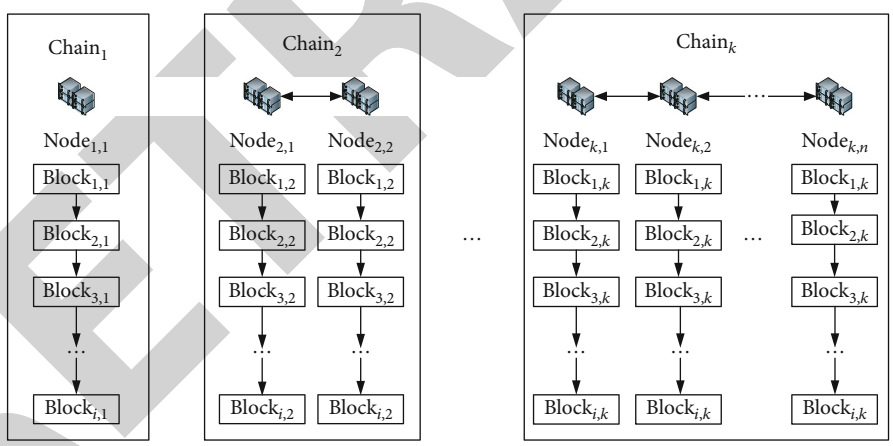


Figure 7: Data blockchain network model.

Chain_k in the figure represents the k_{th} private chain in the blockchain network. $\text{Node}_{k,n}$ represents the n nodes in the k_{th} private chain in the blockchain network. $\text{Block}_{i,k}$ represents the k_{th} coding block of the i_{th} data.

The data backup process is as follows: (1) the local data node uses the private key to sign the data B and the encoding matrix, then sends them to the data processing server; (2) the data processing server checks the signature of data B. When the signature is complete, it uses the coding matrix of the data node to encode the data to obtain a set of coded blocks E; (3) the data processing server sends the set of coded data blocks E to multiple private blockchains in the blockchain network; (4) after receiving the corresponding coded blocks, the nodes in the blockchain network generate consensus results according to the hybrid consensus mechanism and feedback the consensus results to the data processing server; and (5) the data processing server adds the relevant information of the data to the data directory in the access controller after receiving the successful result of the consensus.

4.4. Data Recovery Module. Attribute-based access control (ABAC), using the attributes of the subject and object as the basis for the basic judgment of authority, can realize

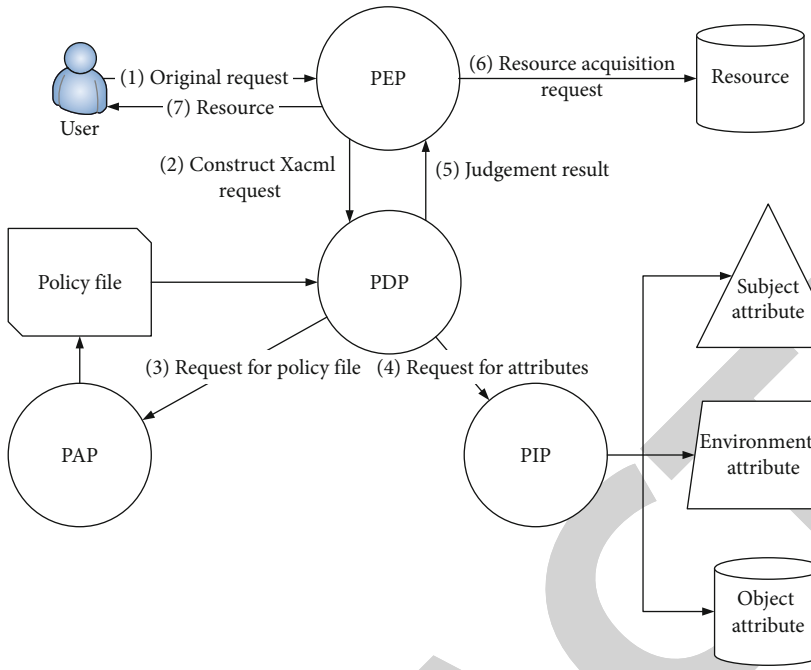


Figure 8: Authorization judgment process. (1) The user sends the original request to apply for access to the resource. (2) After receiving the original request, the policy enforcement point (PEP) constructs a request in xacml format and sends it to the policy deployment point(PDP). (3) PDP requests to obtain the policy file in the policy administration point (PAP) according to the xacml request; (4) After the PDP obtains the policy file, it sends a request to the policy information point (PIP) to obtain the required attribute values (theme attribute, environment attribute, and resource attribute) in the policy file. (5) The PDP makes the judgment result (permit, deny, uncertain, and not applicable) according to the policy file and returns it to the PEP. (6) If the result is permit, the PEP sends an access request to the resource. (7) After the PEP obtains the resource, it returns the resource to the user.

many-to-many access control. It can well separate policy management and authority judgment, has higher flexibility, and can better support fine-grained access control of large-scale information systems. The introduction of environmental attributes makes ABAC support dynamic access control. The authorization judgment process of the attribute-based access control model is shown in Figure 8.

At the same time, the data processing server will perform real-time detection of the data integrity of the local data node to prevent data loss or tampering.

Based on this control and recovery model, the process of data recovery is shown in Figure 9.

5. Simulation

5.1. Simulation Environment. We use Ethereum and the Jerasure library to build a system prototype and the Edge X Foundry framework to deploy the data server on the edge of the power Internet of Things data node. We use the Raspberry Pi 4b as a data node in the power Internet of Things, a desktop-level terminal as a data processing server. We use the Ethereum client (geth) to build a private blockchain network. The consensus mechanism of this private blockchain is PoA. Each private blockchain contains two nodes. Each node in the blockchain has a separate account. The blockchain is deployed in a server cluster with ten identical servers. The hardware configuration is shown in the Table 1. We tested the data encoding and decoding speed, data backup and recovery speed, recovery success rate, and system security under different system scales and encoding matrix parameters.

5.2. Speed of Encoding and Decoding. Assuming that the original data to be backed up by each data node is 100 MB, the original data is an arbitrary binary file. This paper proposes an improved encoding scheme based on the Jerasure encoding library. The scheme increases the data size for each encoding and decoding to 128 bytes and sets the word size to 64 bits. At the same time, it increases the preprocessing and concurrent computer mechanism to save the data in advance. The server reads data into the cache for processing to reduce I/O operations and increase the codec speed. We set the coding matrix as $(m+k)\times k$-order Vandermonde matrix, where $k = 4$, and test the encoding and decoding speed at different m (m is the number of check blocks, and its initial value is 2). As shown in Figure 10, when $m = 2$, there is the maximum encoding and decoding speed, the encoding speed is about 222 MB/s, and the decoding speed is about 196 MB/S. As m increases, the data encoding and decoding speed gradually decreases. Finally, the data encoding speed converges to 54 MB/s, and the decoding speed converges to 42 MB/s.

We tested the speed of the encoding scheme based on the Intel E.C. library in the same environment. The data size is also 100 MB. The result is shown in Figure 10. When $m = 2$, there is the maximum encoding and decoding speed. The encoding speed is about 192 MB/s, and the decoding

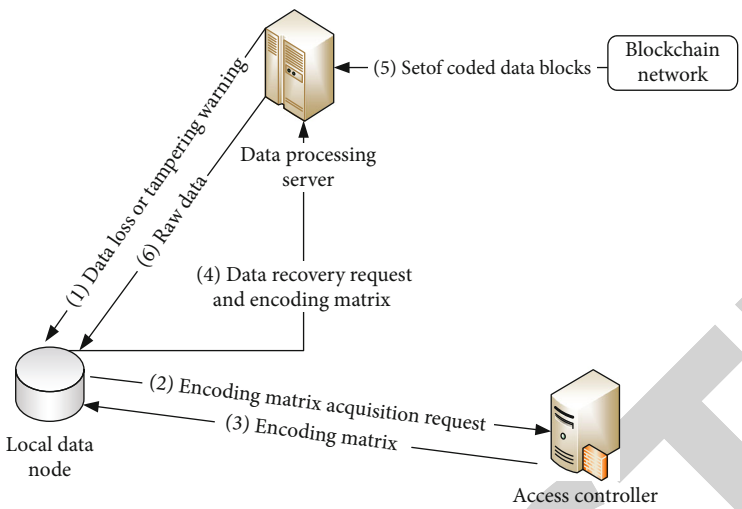


Figure 9: Data recovery. (1) After the data processing server detects that the data of the local data node is lost or tampered with, it sends an alarm message to the data node. (2) The local data node obtains the data identification from the data direction and requests the corresponding coding matrix according to the alarm information. (3) The access controller automatically distributes the coding matrix according to the access control strategy. (4) The local data node sends the data recovery request and the coding matrix to the data processing server. (5) The data processing server obtains the required set of coded data blocks from the blockchain network. (6) The data processing server decodes the coding blocks, then gets the original data and sends it to the data node.

Table 1: System hardware configuration information table.

Platform	OS	CPU	Memory
Power IoT data node	Raspbian 10	BCM2711b0	4 GB
Data processing server	Ubuntu 16.04	AMD Ryzen 7 5800H	16 GB
Server	Centos 7.4	Intel Xeon Gold 5118	64 GB

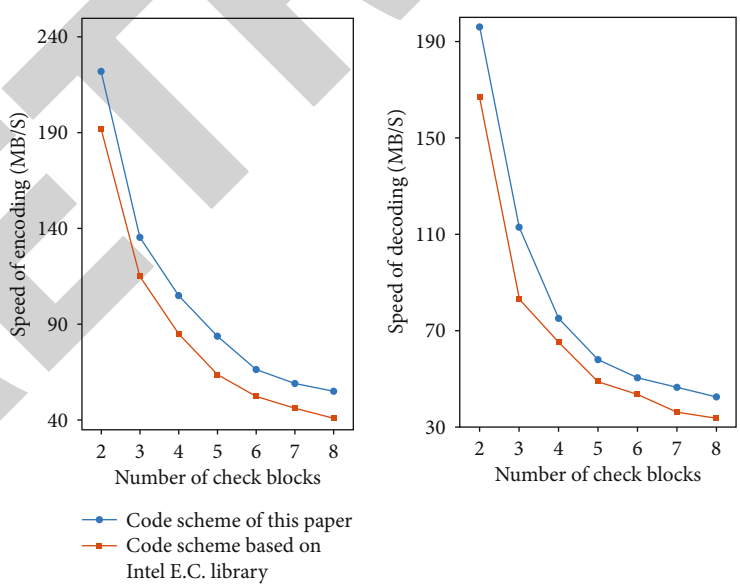


Figure 10: Data encoding and decoding speed.

speed is about 167 MB/S. As m increases, the data encoding and decoding speed gradually decreases. Finally, the data encoding speed converges to 40 MB/s, and the decoding speed converges to 33 MB/s.

We can compare the experimental results and draw the following conclusion. When the number of check blocks is less than 4, the data encoding and decoding speed of the encoding scheme proposed by this paper is greatly improved

compared with the Intel EC encoding scheme. So the encoding scheme proposed by the system is more advantageous.

5.3. Speed of Data Backup and Recovery. This article builds a blockchain network on a server cluster. The blockchain network contains five private chains, and two blockchain nodes maintain each private chain. The coding matrix is a 5×4 -order Vandermonde matrix, and each blockchain node stores one coding block of the data coding block set. Under the above conditions, the system was tested for performance comparison with the Data Protect (DP) data recovery system based on cloud computing proposed by Hewlett-Packard (HP company). DP (originally Omniback) is an automated backup and recovery software for single-server to enterprise environments, supporting disk storage or tape storage targets. It provides crossplatform, online backup of data for Microsoft Windows, Unix, and Linux operating systems. The working mode of DP is server-client. The user can install the client in the distributed database and set the relevant configuration to back up the data to the central server (usually, it is a cloud server). It is a centralized management data recovery system that can use symmetric encryption to prevent privacy leakage [23]. The test data set is sensor data randomly generated by the Raspberry Pi in a simulated environment, and the data size is 1-10 MB. The test network download speed is about 200 Mbps, and the upload speed is about 160 Mbps.

The speed of data backup is shown in Figure 11. From the results, the data backup speed of the system proposed in this paper increases as the data size increases and eventually stabilizes. When the data is small, the time required for data slicing, connection establishment, and service response is a fixed value, which will take up a large system overhead and cause the backup speed to slow down. As the total amount of data increases, edge computing technology can improve system data processing and transmission capabilities. At the same time, compared with the DP data recovery system, the system proposed in this paper has higher data backup efficiency.

Also, we use different sizes of recovery data to test the speed of data recovery. The results of the data recovery speed are shown in Figure 12 below. Compared with the DP data recovery system, the system proposed in this article also has a higher data recovery efficiency.

5.4. The Success Rate of Data Backup and Recovery. In the trusted data recovery method proposed by this system, the success rate of data recovery is related to the number of nodes in the system and the setting of the coding matrix. We set the coding matrix as $(m+k) \times k$-order Vandermonde matrix and assume that there are a total of x nodes participating in the system, and each node stores n coding blocks. In the case of $k = 4$ and $m = 1$, the failure probability of the blockchain node and the link failure rate are both 50%, and we use Matlab simulation software to test the influence of x and n on the success rate of data recovery, the success rate of data recovery = the successful number of recovery/the number of recovery requests. We test 100 times, and the test results are shown in Figure 13.

It can be seen from the figure that the success rate of data recovery increases as the number of nodes and the number of code blocks stored by nodes increase. When $n = 1$ and $c = r/n = 25\%$, the storage optimization efficiency is the highest. However, when the number of nodes in the system is small, the data recovery success rate will decrease. After the number of nodes reaches 14, the data recovery success rate will approach 100%. With the increase of n, the data

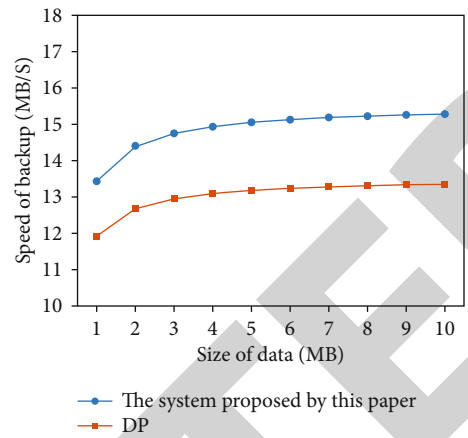


Figure 11: Data backup speed.

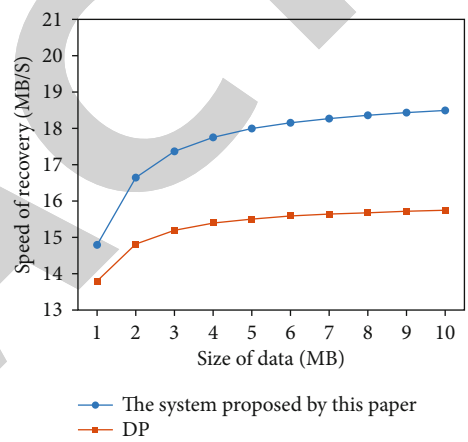


Figure 12: Data recovery speed.

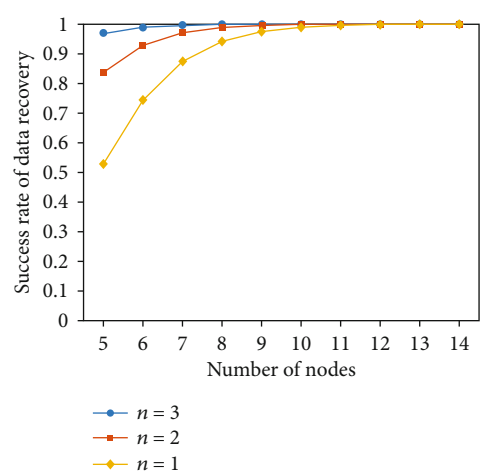


Figure 13: Data recovery success rate based on the coding scheme.

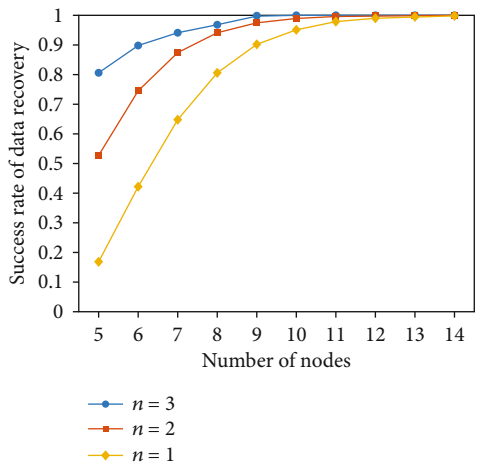


Figure 14: Data recovery success rate based on fragmentation scheme.

recovery success rate of only eight nodes can reach 100% in the end.

In the same environment, the data recovery system using the fragmentation scheme is tested consistently, and the test results are shown in Figure 14.

We compared the experimental results and drew the following conclusions. In the same environment, the data recovery system based on the coding scheme has a higher data recovery success rate than the data recovery system based on the fragmentation scheme under the same number of nodes. Therefore, the system has lower storage and network resource overhead and stronger robustness in data credible recovery.

5.5. Security. This system modifies CP-ABE and implements the access control based on attributes to protect the coding matrix. After testing, the malicious data nodes cannot obtain the encoding matrix without authorization. It is also difficult for the malicious data nodes to decode the encoding block to obtain the original data without the encoding matrix so that the system can prevent data leakage. The block data is also protected by the hash function within the block, which hashes the previous block of the verification chain and mechanically verifies the whole blockchain. The tampered block data cannot pass the verification, thus ensuring that the data in the system is trusted and complete. At the same time, in this system, the edge computing node and the central node of the blockchain will cooperate in supervising the blockchain and preventing the failure of a node. When the fault is detected, the block data will be repaired in time to ensure the security of the data and the continuity of the service.

6. Conclusion

This article proposes a data trusted recovery system in the power Internet of Things environment based on blockchain and coding technology. First, the system uses erasure coding technology to improve blockchain storage performance and data transmission efficiency, providing a basis for efficient and reliable data recovery. Then, the system uses edge computing technology to offload the coding and decoding tasks of local data nodes to the data processing server on the edge, which greatly increases the speed of data backup and recovery. At the same time, the system adopts an attribute-based access control model to protect the coding matrix in a fine-grained manner. It can ensure the privacy of power Internet of Things data and effectively prevent attackers from stealing key information collected by sensors to destroy the power Internet of Things. The system is superior to previous data recovery systems through simulation tests regarding storage performance, data recovery efficiency, and security, mainly reflected in the following aspects: the system uses the blockchain system to store the coding blocks of the key data of the power Internet of Things. Compared with ordinary distributed storage systems, the data stored on the blockchain has extremely strong integrity and is effective enough. It can prevent attackers from tampering with key data, ensure that data nodes obtain original key data after the data recovery, and ensure the stable operation of the power Internet of Things. The system supports dynamic backup of local data nodes and rapid recovery of key data and has good security and real-time performance while reducing the resource overhead in the data backup and recovery process. The system uses the data processing services to monitor the data integrity of the local data node in real time, avoiding the dependence of the data recovery system on the management personnel, so the system is universal, suitable, and easy to use.

However, this system still has some shortcomings: (1) to ensure the reliability of the data source, the system uses the blockchain to store data coding blocks, so there are certain shortcomings in the data access speed; (2) the system uses an attribute-based access strategy, which is more complicated in strategy formulation, has certain requirements for the professionalism of strategy formulation personnel; (3) in terms of coding technology, the system uses an improved RS code coding scheme. The improved coding scheme is only suitable for terminals equipped with a 64-bit operating system, so there may be insufficient compatibility; and (4) the system uses edge computing technology, but no further research has been done on task offloading optimization, which may cause the system to crash under heavy data recovery tasks.

Compared with traditional data trusted recovery system, the system proposed by this paper has many obvious advantages, especially in terms of security, data storage performance, and data recovery efficiency. The system can efficiently help local data nodes complete data backup and recovery to prevent the leakage of key data of the power Internet of Things. With the development of blockchain technology, coding technology, and the power Internet of Things, data credible recovery technology based on blockchain and coding technology will surely receive more and more attention. This advanced data credible recovery system also has high application value. In the future, we will research the optimization of the blockchain consensus mechanism to improve the speed of blockchain data access.

Hindawi
Wireless Communications and Mobile Computing
Volume 2023, Article ID 9782397, 1 page
https://doi.org/10.1155/2023/9782397

Retraction

Retracted: Effects of Attentional Withdrawal on Chinese Disabled Biathletes' Shooting Performance

Wireless Communications and Mobile Computing

Received 11 July 2023; Accepted 11 July 2023; Published 12 July 2023

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.

In addition, our investigation has also shown that one or more of the following human-subject reporting requirements has not been met in this article: ethical approval by an Institutional Review Board (IRB) committee or equivalent, patient/participant consent to participate, and/or agreement to publish patient/ participant details (where relevant).

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. Li and S. Li, "Effects of Attentional Withdrawal on Chinese Disabled Biathletes' Shooting Performance," *Wireless Communications and Mobile Computing*, vol. 2021, Article ID 7264264, 9 pages, 2021.

Hindawi
Wireless Communications and Mobile Computing
Volume 2021, Article ID 7264264, 9 pages
https://doi.org/10.1155/2021/7264264

Research Article

Effects of Attentional Withdrawal on Chinese Disabled Biathletes' Shooting Performance

Jin Li[1] **and Sihua Li**[2]

[1]*Minjiang Teacher College, Fujian 350108, China*
[2]*Teaching and Research Section of Psychology and Education, Capital University of Physical Education and Sports, Beijing 100191, China*

Correspondence should be addressed to Sihua Li; lisihua168@126.com

Received 23 October 2021; Accepted 15 December 2021; Published 30 December 2021

Academic Editor: Chi-Hua Chen

Under two situations of static empty gun preview and simulated competition, the validity of self-suggestion language "preloading compaction," which suggests attentional withdrawal, was verified, respectively. *Methods*. The experiments in the two parts were designed in a single factor group. The participants (Chinese disabled biathletes) fired 20 times (4 groups in total, 5 times in each group) both in pretest and posttest, and they used self-suggestion before each group of firing in posttest. Optoelectronic shooting device based on wireless laser transceiver technology was selected to collect the shooting technical indicators, scores, and other information. The operation evaluation form reflected the subjective feelings of the participants, and the heart rate was measured by the chest band. *Results*. By the method of big data analysis, under the condition of static empty gun preview, the results showed that there was an extremely significant difference between shooting score and quality of triggering. All the marks on the operation evaluation form improved after the intervention. In the mock competition situation, shooting score has highly significant differences with quality of triggering and relative triggering value and significant difference with aiming ability. The marks of the operation evaluation form have been improved; the heart rate of firing all decreased in different degrees. *Conclusions*. The attentional withdrawal has positive effects on the shooting scores, technical indicators of optoelectronic shooting device, subjective feelings, and heart rate of disabled biathletes under the two conditions. TIRE (the timing of firing) is the most sensitive indicator and can be used as the focus of future training and scientific research.

1. Introduction

The disabled biathlon is a combination of cross-country skiing and air rifle shooting, which requires athletes to complete precise technical movements in high-intensity sports and quickly switch between dynamic and static. Therefore, it is characterized by contradiction and complexity and requires athletes to meet high requirements in all aspects [1]. Whenever one shot is missed in the competition (short-distance and medium-distance competitions), it will be punished by sliding 150 meters. The result will be changed by 1%-1.5% [2]. Therefore, it is key for the competition result to find the shooting technical characteristics and improve the shooting stability.

Attentional withdrawal is the core of shooting techniques, which has been widely recognized and applied in target shooting and other precision events. It refers to the shooter's process from aiming to firing, in which he withdraws all his attention into the shooting action and focuses on the relationship of front sight and iron sight and preloading of the trigger's index finger (or other trigger parts) [3, 4]. Good attentional withdrawal has clear requirements for shooting skills and shooter psychology. In fact, shooting training is both technical aspect and psychological aspect. As the researcher put forward the comprehensive model of psychological construction, psychological training skills should be purposefully and systematically integrated into the shooting technology, to achieve the state of combination of mentality and technology, in order to better help athletes improve their performance [5]. However, some biathletes often separate psychological training from technical training in the daily shooting training, which results in the athletes'

shooting skills cannot match with the characteristics of disabled biathlon. Besides, there is also a common phenomenon that they have good shooting performance in training but failed to play their normal level in competition.

Compared with the air rifle shooting and the disabled biathlon, they both need to accurately hit a stationary target 10 meters away. The difference is that the air rifle shooting can be completed within the specified time; athletes have enough time to complete a series of actions of attentional withdrawal, such as "sight is clear and target is blur," and then attentional withdrawal to the preloading index finger, so as to sense the balance of the body [6]. The results of the disabled biathlon competition are calculated by time. One round of shooting is average about 20 seconds. Athletes will greatly shorten the preparation and thinking time for shooting while ensuring a hit [7]. They cannot complete all the technical actions like air rifle shooting. In the process of attentional withdrawal, mounting the gun and aiming are certainly important, but a good preload can make up for the disadvantages caused by the unstable control and poor aiming of the gun, and also achieve good results [8]. The disabled biathlon athletes are shooting after high-intensity sliding. Muscle tension or fatigue and fast heart rate make it difficult for athletes to be stable, which increases the difficulty of gun controlling and aiming. In addition, the bull's eye of biathlon is larger than air rifle shooting, and the biathlon has no concept of ring number, which reduce the vision requirement of the enlarged aiming area. Therefore, the importance of preloading becomes more prominent. This study puts forward the hypothesis that the disabled biathlon athletes may not be able to follow the requirements and steps of attentional withdrawal in the target shooting, and they need to be done in a more concise way that can better match the characteristics of the event. That is, they can neglect the search for the relationship of front sight and iron sight and emphatically withdraw attention to the preloading.

However, the current evaluation of shooting technology is mainly based on whether the shooting hits or misses, the experience of coaches, and the subjective report of the athletes, which cannot provide detailed guidance for the technical movements at a deeper level and lacks objective and quantitative evaluation means [9]. The appearance of optoelectronic shooting device has changed this situation. As a commonly used shooting training tool at present [10], it has been tested and recognized in both target shooting [11] and biathlon [12]. Optoelectronic shooting device can convert shooting techniques such as control of the gun, aiming, and preloading into quantized technical indexes, which not only correspond to the definition of attentional withdrawal but also provide convenience for statistical analysis.

This study used the indicators provided by the optoelectronic shooting device, in conjunction with the athlete's subjective feelings and the heart rate of firing, to verify whether the withdraw of attention to the preloading trigger can effectively improve the athlete's shooting performance and enhance shooting technique both in static empty gun preview and simulation game two kinds of situations, so as to provide guidance and reference for the follow-up training and research and help biathletes achieve better performance in winter of 2022 Beijing Paralympics.

2. Research Subjects and Methods

2.1. Research Subjects. The participants of this study are all national athletes of the disabled biathlon in China. A total of 25 athletes (M = 23.32 years old, SD = 2.45 years old) participated in the static empty gun preview situation, including 18 males and 7 females, 16 for sitting position and 9 for standing position. Because the teams have to alternately participate in external training and experimental conditions, 9 athletes (M = 23.56 years old, SD = 2.95 years old) who have participated in static air gun preview are randomly selected as the research participants of simulated competition situation, among which 6 males and 3 females, 6 for sitting position and 3 for standing position.

2.2. Research Tools and Means

2.2.1. Self-Suggestion Instruction. The researchers conducted interviews with 17 Olympic champions of target shooting, and the results showed that self-suggestion is the main choice for athletes to conduct self-regulation, whether before or during the competition, and plays a role of stabilizing mentality and technology [13]. Therefore, self-suggestion is used as a way to prompt athletes to withdraw their attention. The content of self-suggestion should be concise and clear and focus on the core technical actions [14]. According to the hypothesis proposed above, "preloading compaction" should be formulated as a self-suggestion to prompt attentional withdrawal. "Preloading" can encourage athletes to withdraw their attention on the shooting action and specify where to withdraw their attention, so as to avoid confusion of independent variables of the experiment between preloading and aiming. "Compaction" is an action request that prompts the athletes to press the trigger to the most appropriate position in the shooting process, so that the muscles are evenly exerted, and the strength of the index finger (or the other trigger position) is consciously controlled [4].

2.2.2. Optoelectronic Shooting Device. Optoelectronic shooting device transmits and receives signals through the wireless laser optoelectronic transceiver of Noptel ST-2000 SPORT made in Finland. The signal irradiates to the specific reflecting target before the target and then reflects and recovers the transceiver. The optoelectronic signal is converted into electronic signal and transmitted to the notebook computer, and the supporting software will process and analyze the electrical signal on the base of big data analysis [15]. The two experiments of this study need to measure the abilities of the athletes' controlling gun, aiming, preloading, etc., in the process of attentional withdrawal. Therefore, indicators such as the average number of rings, controlling gun stability, aiming accuracy, and trigger pressure ability were selected (Table 1).

The deviation in horizontal (Dev X) and vertical (Dev Y) directions is the most powerful indicator for measuring muzzle stability when aiming, and the combination of the two indicators can also reflect the dispersion degree of shooting performance to a certain extent [16]. Trigger pressure control capability is an important part of shooting technology and also an important factor affecting shooting

Table 1: Detailed explanation of optoelectronic shooting device indicators.

Name	Abbreviations	Explanation	Note
Controlling gun stability	Dev X	Standard deviation of the aiming point at the last second in the X-axis direction.	The closer Dev X and Dev Y values are, the better.
	Dev Y	Standard deviation of the aiming point at the last second in the Y-axis direction.	
Aiming accuracy	COG	The simulated target redrawn by the system corresponds to the actual target; represents the accuracy of the aim.	The higher the value, the higher the accuracy.
Trigger control capability	RTV	The relative value of trigger pressure is calculated as the ratio of the distance moved by the gun 0.2 seconds before firing and 0.6 seconds before firing. If the distance is less than 1, it indicates that the trigger instant movement is less than the average value.	Indicates the cleanliness of the trigger.
	ATI	The absolute value of trigger pressure is calculated as the absolute distance of muzzle displacement 0.2 seconds before firing.	Low value, stable muzzle. Large value, obvious shaking.
	TIRE	The average falling point of the aiming point is closest to the target center, 0-0.2 seconds before firing TIRE = 3, 0.2-0.4 seconds TIRE = 2, 0.4-0.6 seconds TIRE = 1.	The higher the score, the more decisive the shot, and the better the timing.
The average number of ring	SCORE	The average number of rings on target (1.0-10.9).	

performance. It can be divided into three indicators to explain: RTV represents the accurate and subtle control power of the index finger from the preloading fire to the second fire when pressing the trigger. ATI represents whether pressing on the trigger makes the muzzle shake violently; TIRE refers to the ability of firing when the aiming point moves to the bull's eye rather than moving outward from the bull's eye, indicating the determinateness of hitting, namely, the timing of firing [17]. COG refers to whether the athletes can accurately aim at the bull's eye. SCORE is the actual number of rings displayed on the optoelectronic system after the athlete presses the trigger.

2.2.3. Operation Evaluation Form. According to the definition of attentional withdrawal and the meaning of self-suggestion words, the shooting techniques related to preloading are selected, and the "operation evaluation form" of the shooting suggestion words is compiled, which reflects the subjective evaluation of the athletes on the completion degree of the shooting techniques (Table 2). Likert's 7-point scoring method was adopted to answer a total of 13 questions, among which the first 6 questions were answered twice, respectively, before and after the self-suggestion intervention. The seventh question was the evaluation of the completion degree of attentional withdrawal action, so the answers were given after the intervention.

2.3. Research Methods. In the two experiments of this study, single factor intragroup experiment design was adopted. Paired sample t-test statistical method was used to analyze the pretest and posttest data of the technical indicators, operation evaluation form, heart rate, and other indicators of the optoelectronic shooting device. Before each experiment, the optoelectronic shooting device was calibrated and zeroed.

For static empty gun preview situation, researchers need to convey the information for athletes, including the concept of attentional withdrawal and the method of self-suggestion, etc., and empty gun preview can effectively help participants experience inner feelings and muscle changes [9]. So, it is the most appropriate that static empty gun prepare intervention situation was selected firstly. The experimental site was the Biathlon training hall of a certain ski team. The participants fired 20 times without guidance (simulated competition rules, 5 times in one round, 4 rounds in total, with short rest and debugging between each round). After all rounds were completed, the participants were asked to fill in the "operation evaluation form" and then complete the pretest part of the experiment. Combined with the results of the pretest, the concept, function, and the meaning and usage of self-suggestion were explained to the participants. Then, participants were asked to repeat how to understand the attentional withdrawal and how to complete the preloading compaction. When there is a "do not aim at deliberately," "focus on the preloading of the index finger," "grasp well the magnitude of the preloading," "fire decisively at a good time," and the same or similar key words, which showed the participants understood the self-suggestion language meaning and usage compiled by the researchers. In the posttest, participants were required to silently recite the self-suggestion words before each round of shooting, and they could be used again if they felt that their states were not good and needed to be adjusted. Fill in the "operation evaluation form" after all the assignments are completed.

Table 2: Title and evaluation method of operation evaluation form.

No.	The title	Degree of completion						
		Very low	Low	A little low	General	A little high	High	Very high
1	Withdraw your attention on your index finger (or other trigger parts)							
2	Index finger (or other trigger parts) preloading compaction							
3	The trigger force can be controlled when firing							
4	Be decisive when firing	1	2	3	4	5	6	7
5	The muzzle is stable after firing							
6	Shooting action consistency							
7	Complete the actions suggested by the self-suggestion							

For situation of simulated competition, biathlon for the disabled requires athletes to shoot under a certain load. Whether attentional withdrawal can really improve athletes' shooting performance is not enough to be proved by static experiments only. Therefore, it is necessary to conduct experiments under the situation of simulated competition [18]. In order to simulate the competition situation as much as possible and according to the actual situation of the site, the long-distance competition rules were adopted to unify the sliding standards of both group and gender. One round of shooting was conducted for every 3 km (2 laps of the ski team compound), and a total of 4 rounds were carried out. Subjects participating in the experiment departed at intervals of 1 minute. The Polar H10 heart rate chest band was used to record the heart rate data, and the researchers manually recorded heart rate at the moment of each firing. For each experiment, two to three subjects were scheduled to participate. The aim was to ensure that the same group of subjects performed the experiment on the same day. After completing 4 rounds of shooting, fill in "operation evaluation form." In the posttest, participants were required to say "preloading compaction" before each round of shooting, so as to avoid forgetting to use self-suggestion words after high-intensity sliding and to eliminate the interference caused by researchers' reminder to participants. The pre- and posttests were arranged in the adjacent two weeks, and the pre- and posttest time of each subject was the same, so as to ensure that the physical strength was sufficient.

The research participants of simulated competition situation had already carried out the intervention of attentional withdrawal under static condition, which may have certain influence on the experimental results of this stage. The objective condition is caused by the number of athletes and the actual condition. In this study, the intragroup experiment design was adopted, and the posttest was improved compared with the pretest, which proved that the intervention under this situation was effective. At the same time, in order to eliminate interference better, the athletes were told to conduct experiments according to the shooting rhythm in the actual competition during the pretest, and the participants were confirmed again after the end of the pretest. All the participants indicated that they had completed the experiment as required.

3. Research Results

3.1. Static Empty Gun Preview Situation

3.1.1. Difference between Pre- and Posttest of Optoelectronic Shooting Device Indicators and Shooting Performance. The technical indicators and shooting results of 25 participants were analyzed. As shown in Table 3, there is a highly significant difference in SCORE and indicator TIRE ($p \leq 0.001$) among the participants after "preloading compaction," and the mean value was improved. Compared with the pretest, in terms of controlling gun stability, the mean value and standard deviation of X and Y axis offset decreased. Although COG had not achieved statistical difference, the overall trend of change was the same as SCORE. The mean and standard deviation of ATI decreased slightly, while the mean value of RTV decreased and the standard deviation increased.

3.1.2. The Difference between the Pre- and Posttest of the Operation Evaluation Form. Statistical analysis was conducted on the "operation evaluation form" scores of 25 participants, as shown in Table 4. After the participants used self-suggestion to complete the experiment, the subjective feeling scores of each subject increased (this indicated an increase in positive feelings), and the standard deviation showed a downward trend (this indicated that positive feelings were more stable). Questions 1, 2, and 3 were very significantly different ($p \leq 0.001$), and questions 4 and 5 were significantly different ($p < 0.01$). The consistency score of shooting action improved slightly, but there was no statistical difference. The mean score and standard deviation of the actions indicated in the self-suggestion words reached 5.800 points and 0.500 points, which exceeded the mean value of the evaluation form, meaning that all the participants completed the intervention according to the experimental requirements.

3.2. Simulated Competition Situation

3.2.1. Difference between Pre- and Posttest of Optoelectronic Shooting Device Technical Indicators and Shooting Performance. The optoelectronic shooting device data of 9 participants were statistically analyzed. As shown in Table 5, pre- and posttest data of SCORE, COG, RTV, and TIRE are significantly increased under the condition of self-suggestion

Table 3: Paired sample t-test results of optoelectronic shooting device technical indicators and shooting performance in pre- and posttest.

Indicators	n	Pretest		Posttest		$t(99)$	p	95% CI		Cohen's d
		M	SD	M	SD			LL	UL	
SCORE	25	9.498	0.542	9.830	0.437	6.779	$\leq 0.001^{***}$	0.429	0.235	0.674
Dev X	25	0.558	0.288	0.545	0.194	0.437	0.663	0.046	0.072	0.053
Dev Y	25	0.667	1.042	0.488	0.296	1.722	0.088	0.027	0.384	0.234
COG	25	9.570	1.244	9.694	0.888	0.855	0.395	0.411	0.164	0.115
ATI	25	5.206	2.218	5.100	2.143	0.557	0.578	0.272	0.485	0.049
RTV	25	1.077	0.408	0.891	0.911	1.851	0.067	0.013	0.387	0.264
TIRE	25	1.657	0.337	2.301	0.357	12.404	$\leq 0.001^{***}$	0.747	0.541	1.855

* means $p < 0.05$; ** means $p < 0.01$; *** means $p \leq 0.001$, the same as below.

Table 4: Paired sample t-test results of "operation evaluation form" in pre- and posttest.

The title	Pretest		Posttest		$t(24)$	p	95% CI		Cohen's d
	M	SD	M	SD			LL	UL	
1	2.640	1.254	5.080	0.400	9.939	$\leq 0.001^{***}$	2.947	1.933	2.622
2	4.040	1.274	5.680	0.802	7.624	$\leq 0.001^{***}$	2.084	1.196	1.541
3	4.400	1.225	5.360	0.860	4.529	$\leq 0.001^{***}$	1.397	0.523	0.907
4	4.960	1.620	5.800	0.764	2.871	0.008^{**}	1.444	0.236	0.663
5	4.840	0.850	5.400	0.577	3.412	0.002^{**}	0.899	0.221	0.771
6	4.640	1.075	4.800	0.816	0.609	0.548	0.702	0.382	0.168
7	—	—	5.800	0.500	—	—	—	—	—

Table 5: Paired sample t-test results of optoelectronic shooting device technical indicators and shooting performance in pre- and posttest.

Indicators	n	Pretest		Posttest		$t(35)$	p	95% CI		Cohen's d
		M	SD	M	SD			LL	UL	
SCORE	9	9.378	0.596	9.922	0.253	5.446	$\leq 0.001^{***}$	0.747	0.341	1.188
Dev X	9	0.445	0.141	0.448	0.172	0.064	0.949	0.072	0.068	0.019
Dev Y	9	0.412	0.119	0.397	0.148	0.571	0.572	0.039	0.070	0.112
COG	9	9.444	1.371	10.071	0.336	2.622	0.013^{*}	1.111	0.141	0.628
ATI	9	4.289	1.725	4.242	1.571	0.151	0.881	0.595	0.691	0.028
RTV	9	1.129	0.248	0.716	0.165	8.230	$\leq 0.001^{***}$	0.311	0.515	1.961
TIRE	9	1.744	0.227	2.367	0.293	9.628	$\leq 0.001^{***}$	0.753	0.491	2.377

language indicating attentional withdrawal, showing significant ($p < 0.05$) or extremely significant difference ($p \leq 0.001$). The average offset of the X and Y axes changed very little and basically remained at the same level before and after intervention. The mean and standard deviation of ATI decreased only slightly in the posttest.

3.2.2. Pre- and Posttest Differences of the Operation Evaluation Table. The "operation evaluation form" scores of 9 participants were statistically analyzed. As shown in Table 6, the average scores on questions 1 to 5 were highly significantly different ($p \leq 0.001$) or significantly improved ($p < 0.01$). The consistency score of shooting action was consistent with the experimental results under static condition. Although the score was improved, there was no statistical difference. The performance of the participants in completing the suggested actions was also relatively good, higher than the average of 4 points, indicating that in the simulated competition, the participants still performed well in attentional withdrawal according to the experimental requirements. On the whole, the posttest scores of the form were higher than those under the static empty gun preview (Table 4).

3.2.3. Heart Rate Statistics at the Moment of Firing. After the manually recorded heart rate values at the moment of firing were input into the computer, the mean heart rate and standard deviation of all participants from the first to the fifth times were counted, and paired sample t-test was performed (Table 7). The average heart rate from the first to the fifth shot in the posttest was slightly lower than in the pretest, and the second, third, and fourth shot showed significant difference after intervention ($p < 0.01$).

Table 6: Paired t-test results of "operation evaluation form" in pre- and posttest.

The title	Pretest		Posttest		$t(8)$	p	95% CI		Cohen's d
	M	SD	M	SD			LL	UL	
1	3.000	0.866	5.330	0.707	9.899	≤0.001***	2.877	1.790	2.947
2	4.000	1.414	5.780	1.093	6.400	≤0.001***	2.418	1.137	1.409
3	4.330	1.225	5.780	0.667	3.833	0.005**	2.313	0.576	1.470
4	4.670	1.323	6.220	0.441	4.128	0.003**	2.424	0.687	1.572
5	4.440	1.014	5.670	1.118	8.315	≤0.001***	1.561	0.883	1.152
6	4.560	1.130	5.110	0.928	1.890	0.095	1.233	0.122	0.532
7	—	—	5.560	0.882	—	—	—	—	—

Table 7: Paired t-test results of heart rate of at the moment of firing.

Times	n	Pretest		Posttest		$t(8)$	p	95% CI		Cohen's d
		M	SD	M	SD			LL	UL	
1	9	151.580	4.023	150.528	3.470	1.811	0.108	0.289	2.400	0.280
2	9	145.580	4.690	143.667	4.142	3.645	0.007**	0.703	3.126	0.432
3	9	139.470	4.633	137.390	4.773	4.767	0.001**	1.076	3.091	0.442
4	9	134.861	4.727	132.694	4.501	3.813	0.005**	0.856	3.477	0.470
5	9	130.361	4.318	128.860	4.545	2.058	0.074	0.181	3.181	0.339

4. Analysis and Discussion

4.1. Analysis of the Influence of Attentional Withdrawal on Optoelectronic Shooting Device Indicators. In the static empty gun preview situation, TIRE represents the decisiveness of pressing the trigger, the posttest average point increased from 1.657 to 2.301, and the other two indicators RTV and ATI also showed a downward trend. It was proved that after the self-suggestion, when the large circle of the front sight just caught the small circle of the bull's eye, that is, when the best firing point appeared, the athletes could press the trigger more decisively and could control the pressure of the trigger, so as to keep the muzzle as stable as possible at the moment of firing. Such a situation can be explained from two aspects. On the one hand, after "preloading compaction," the index finger (or the other trigger parts) needs dozens of grams or even grams of force to trigger, so when a good firing time occurs, it only needs to gently touch the trigger [19]. On the other hand, after self-suggestion, the attention of the participants was better withdrawn into the shooting action of preloading, and the psychological orientation was more definite, and they did not hesitate for fear of poor aiming or bad shooting [20].

The values of Dev X and Dev Y represent the amplitude of the muzzle shake, and the smaller the value is, the more stable the gun is. Researchers conducted experiments on biathlon of healthy people and found that, since the five targets of biathlon were arranged in a row, the movement from one target to another was roughly "U" shaped, and the horizontal offset should be smaller than the vertical offset, which was more conducive to the stability of shooting [21]. This has a certain relationship with the shooting rhythm. In the process of "one breath, one shot" or "two breath, one shot," the muzzle will be lifted up gradually from the bottom. The vertical deviation is inevitable, while the horizontal deviation should also be reduced as far as possible. The horizontal offset value of 25 participants changed little in the posttest, and the vertical offset value was still higher than the horizontal offset value despite a significant decrease. The target of optoelectronic shooting device is not a row of 5 targets, but a simulated target to complete all the shooting, so the X-axis offset is relatively stable. However, after attention was withdrawn to pressure, the mean value and standard deviation of athletes' deviation in Y-axis decreased significantly, indicating that the stability of guns was improved and more stable after intervention.

COG reflects the ability of the subject to aim with the muzzle pointing to the center of the target. The shooting program will simulate a virtual target according to the aiming trajectory, reflecting the accuracy of aiming and corresponding to the actual shooting performance [22]. In the comparison of the pre- and posttest results, the shooting scores increased by 0.3 rings on average and presented a very significant difference ($p \leq 0.001$), which proved that the participants' shooting scores and stability could be effectively improved after attentional withdrawal through self-suggestion. Shooting performance can reflect the dispersion of shooting from the side. The more concentrated the dispersion, the better the consistency of shooting action and the more stable the performance [23].

In the simulated game situation, Dev X and Dev Y did not change significantly in this part of the experiment, and the muzzle shake amplitude and frequency were consistent when taking the gun, regardless of whether the participants used self-suggestion or not. Compared with the static situation, the X and Y axis offsets were closer and smaller. The similarity between the two values indicated that the muzzle

shake tends to be round, which was more coordinated and reasonable according to the action of the gun. Although offset than air rifle shooter (first-line players, Dev X = 0.25, Dev Y = 0.18; second-line players, Dev X = 0.4, Dev Y = 0.3) still have a certain gap [8], the disabled biathlon is also acceptable.

In three indicators of trigger control capability, two indicators of RTV and TIRE presented extremely significant difference ($p \leq 0.001$), showing that the method of self-suggestion prompted athlete's attentional withdrawal, also improved the preloading problem in the shooting of simulation competition. These made the shot more clean and more decisive. The change of ATI in pre- and posttest was not significant, which may be related to the source of the indicator. ATI calculated the absolute value of gun muzzle movement 0.2 seconds before firing. At this time, the movement of gun muzzle may be caused by inadequate control of finger strength or the involuntary shaking of body due to breathing, fatigue, and other conditions, which makes this indicator not sensitive in the experiment. This showed from the side, after the athlete withdraw attention to preloading, even if the muzzle or the body had a violent shaking, the score was also made up for by the preloading in place, smooth, and decisive shot.

The indicators of COG showed significant differences, and the shooting accuracy and precision of the participants after intervention were also significantly improved, and the average value of COG in the posttest reached 10 rings. Combined with the shooting performance, the actual shooting performance in the posttest also reached 9.9 rings. This indicated that the participants deliberately aimed less during the aiming process. When the bull's eye was aligned with the front sight, they did not hesitate to press the trigger, instead of pursuing the chance of the "two circles" to stay completely in place.

Overall, the two experiments of the study, the three indicators of "trigger control capability" were more sensitive than other indicators, especially TIRE. That is, the decisive change in firing is the most significant. On the one hand, it showed that withdraw attention to prepressure can indeed help athletes to mend their bad shooting movements and habits. No matter whether the main contradiction of athletes is focused on the point of hesitation in shooting, it can play a role of consolidation and strengthening. On the other hand, it is more in line with the characteristics of the disabled biathlon to complete a large number of shots in a short period of time. Other technical indicators did not change significantly before and after the intervention, which was inconsistent with the results of the study of optoelectronic shooting device application in biathlon in healthy people. The researchers tested 17 biathlon athletes and found that in addition to the quality of trigger (TIRE) and trigger relative value (RTV), vertical stability (Dev X) and aiming ability (COG) were also important factors in determining shooting performance [24]. The researchers tested 22 biathlon athletes after exhaustion and found that vertical stability (Dev X) was the focus to strengthen and improve [25]. Because this study only focuses on the preloading part of attentional withdrawal, athletes pay more attention to the pre-preloading trigger. Although the indicators of controlling gun stability and aiming ability were not significantly different, they also showed a positive trend. Moreover, according to the controlling gun, aiming, preloading, and firing, the series of actions are coherent and mutually influenced [26]. The gun is kept shaking as regularly as possible, which lays the foundation for aiming at the bull's eye. In one "breath" of the disabled biathlon, the best firing point may be only once. After finding the "point" according to their own shooting rhythm, they will press the trigger decisively. While the gun is being held and aimed, attention is being withdrawn into the index finger (or the other trigger parts), and the trigger is being prepressed into position, that is "compacting." It ensures that the shot is fired most smoothly when the opportunity comes. Therefore, in this study, by implying preloading, other indicators were also improved.

The shooting performance was also significantly improved in both situations. Although it was not the basis for judging the competition, the improved performance meant that the bullets spread smaller and were closer to the center of the target, and the missed caused by accident factors could be reduced in the case of a guaranteed hit [6]. However, in both static and simulated competition situations, the average shooting scores of the participants were kept above 8 rings. Strictly speaking, the participants hardly missed the target in the experiment. In the actual competition and previous studies, even excellent athletes will miss the target [27]. Researchers believed that such a situation happens for the following reasons. Firstly, live bullets were used to record matches, and the hit or miss dichotomy was applied as the statistical method in the previous studies. However, in this study, optoelectronic shooting devices were used for simulation, and the number of rings was applied as the unit for statistics. So there were inconsistencies in the presentation of results. Secondly, the real target and the simulated target are also different, which may have a certain impact on the athletes' performance in visual sense. Meanwhile, some unexpected situations that may occur in the real target, such as the steel target does not topple over, and the misfiring. Those can be excluded from the simulated target. Thirdly, there is a calibration problem with the optoelectronic shooting device. If the target center simulated by the system is deviated from the actual target, the final statistical result will be seriously affected. However, the COG and SCORE values in this study were very close, so the calibration deviation of the optoelectronic shooting device system can be excluded to some extent. Fourth, although the gliding intensity and rules of the competition were simulated, the participants' psychological preparation, psychological pressure, and environment were far from the formal competition. So they might still be in the state of training, and their shooting performance could be guaranteed naturally.

4.2. Analysis of the Influence of Attentional Withdrawal on Subjective Feelings. In the static empty gun preview, the average values had big differences between the pre- and posttest, and the standard deviations were further reduced. These meant that without instructions, the athletes cannot fully perform the technical action of attentional withdrawal even if they preview the empty gun, and they may forget or skip the preloading and directly fire. After the instruction of the

suggestive language, the attentional withdrawal of athletes was significantly improved, which was related to the score of question 7, indicating that the athletes had conscientiously completed the actions prompted by the self-suggestive language in the posttest experiment. The feeling of decisiveness in firing (question 4) has the highest score in the posttest, which can be corresponding to the significant difference of the "TIRE" indicator of the optoelectronic shooting device. Because of preloading compaction, the original hesitant athletes had a better improvement, and for other athletes also played a consolidated and strengthened role. The questions 2, 3, and 5 are about the evaluation of preloading. There were significant differences between the pre- and posttest, but the data of ATI and RTV indicators did not get the corresponding reflected. Probably after using self-suggestion, the technology was, in fact there was no obvious improvement, but athletes felt more confident and had more distinct perceptions of shooting techniques. There was only a small improvement in the score of question 6, and the athletes believed that there was no significant change in the consistency of shooting actions before and after the self-suggestion intervention. On the one hand, the experiment was conducted under the static situation. Even if the shooting action is not standard or correct, the athletes can maintain a high consistency. So the effect of intervention is not obvious. On the other hand, self-suggestion methods and language cannot be applied to every shot.

In simulated competition situation, the average value of question 7 was 5.560, which proved that the athletes can still maintain the use of self-suggestion after intense exercise. Questions 1 to 5 were consistent with the static evaluation and showed statistical differences. However, the difference lied in that the posttest score under the simulated competition situation was higher, and even one more than 6 points, indicating that attentional withdrawal under this situation had a more obvious subjective influence on the athletes. It reflected the practical application value of the self-suggestion of "preloading compaction."

In general, the scores of "operation evaluation form" were significantly improved in both parts of the study and corresponded to the changes of optoelectronic shooting device technical indicators, which proved that the participants not only improved their shooting skills in the process of attentional withdrawal but also believed that they could better play their skills psychologically. In this way, when encountering the state imbalance such as excessive tension and paying attention to others' achievements, a good withdrawal of attentional can help them effectively adjust. It can be seen that it is feasible and necessary to combine effective psychological training methods with shooting techniques. The self-suggestion of "preloading compaction" not only deepened the participants' understanding of attentional withdrawal but also established the correct movement imagery. It also helped biathletes to form a correct psychological orientation, so that in the shooting process, more attention was paid to the action itself rather than the performance [28].

Therefore, common psychological training methods can also be used, such as imagery method and abdominal breathing method. These training methods should match the actual problems of the disabled biathlon athletes.

4.3. Analysis of the Influence of Attentional Withdrawal on the Heart Rate at Firing Moment. Researchers believed that scientific and rational monitoring of cardiovascular function should be strengthened to improve the shooting training and competition results of Chinese biathletes [29]. And shooting accuracy was influenced by the cardiac cycle phase [30]. The main purpose of monitoring the heart rate in this study was to detect whether the exercise intensity of the participants met the requirements, so as to ensure the effectiveness of simulated competition situation.

In the event of disabled biathlon, the first and the final shots are more important than the other three. The first hit or not related to the follow-up shooting rhythm and the mentality of the athletes. Too fast heart rate will increase the body shaking and is not conducive to the performance of technical action. A slow heart rate requires more time; it may be out of step with the overall competition and may not have a better shooting performance. Therefore, it is crucial to lower the heart rate to an appropriate level before starting shooting [31].The final shot is the end of the round and is the key step between shooting and gliding. Researchers put forward that it is most appropriate for biathlon athletes to adjust the psychology of first shot to 140-150 times/min [32, 33]. The average heart rate of the 9 participants reached 152 beats/min at the first moment in the pretest but dropped to 151 beats/min after the intervention, which was more conducive to the stability of the first shot. The heart rate of the last four shots at the moment of firing was within the appropriate range and decreased to different degrees. The corresponding shooting performance and subjective feeling were improved, indicating that the participants can activate the parasympathetic nerve more quickly after using the self-suggestion language, so as to reduce the heart rate and improve the shooting stability and performance [34].

5. Conclusions

Under static empty gun preview and simulated competition conditions, attentional withdrawal has positive effects on shooting performance, optoelectronic shooting device technical indicators, subjective feelings, and heart rate of disabled biathlon athletes.

Various technical indicators of optoelectronic shooting device, the quality of press trigger (TIRE) is the most sensitive; athletes suggest themselves that "preloading compaction" can be more decisive and in line with the characteristics of the project and can be the focus of training and scientific research in the future.

Data Availability

The (SPSS Statistics Data Document) data used to support the findings of this study are available from the corresponding author upon request. The data of the test will be disclosed by the author in the near future.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Hindawi
Wireless Communications and Mobile Computing
Volume 2021, Article ID 1333250, 11 pages
https://doi.org/10.1155/2021/1333250

Research Article

Designing Compact Convolutional Filters for Lightweight Human Pose Estimation

Shili Niu,[1] Weihua Ou,[1] Shihua Feng,[1] Jianping Gou,[2] Fei Long,[3,4] Wenchuan Zhang,[1] and Wu Zeng[5]

[1]*School of Big Data and Computer Science, Guizhou Normal University, Guiyang 550025, China*
[2]*School of Computer Science and Communication Engineering, Jiangsu University, Zhenjiang 212000, China*
[3]*College of Artificial Intelligence and Electrical Engineering, Guizhou Institute of Technology, Guiyang 550003, China*
[4]*Special Key Laboratory of Artificial Intelligence and Intelligent Control of Guizhou Province, Guiyang 550003, China*
[5]*School of Electrical and Electronic Engineering, Wuhan Polytechnic University, Wuhan 430000, China*

Correspondence should be addressed to Weihua Ou; ouweihuahust@gmail.com

Received 19 August 2021; Revised 18 November 2021; Accepted 29 November 2021; Published 17 December 2021

Academic Editor: Chi-Hua Chen

Existing methods for human pose estimation usually use a large intermediate tensor, leading to a high computational load, which is detrimental to resource-limited devices. To solve this problem, we propose a low computational cost pose estimation network, MobilePoseNet, which includes encoder, decoder, and parallel nonmaximum suppression operation. Specifically, we design a lightweight upsampling block instead of transposing the convolution as the decoder and use the lightweight network as our downsampling part. Then, we choose the high-resolution features as the input for upsampling to reduce the number of model parameters. Finally, we propose a parallel OKS-NMS, which significantly outperforms the conventional NMS in terms of accuracy and speed. Experimental results on the benchmark datasets show that MobilePoseNet obtains almost comparable results to state-of-the-art methods with a low compilation load. Compared to SimpleBaseline, the parameter of MobilePoseNet is only 4%, while the estimation accuracy reaches 98%.

1. Introduction

Human pose estimation is also called human key point detection. Its main task is to detect the key points of human body (eyes, nose, shoulders, elbows, etc.) in a given RGB picture. Human pose estimation is one of the basic tasks of computer vision and has many practical applications, such as human-computer interaction [1], human tracking [2], and motion analysis [3]. In recent years, with the quick development of neural networks, human pose estimation based on deep neural networks [4–9] has gained a high accuracy. However, these works have focused only on improving the accuracy of pose estimation through the use of complex and computationally expensive models, while largely ignoring the issue of the cost of model inference. Many methods already require computational resources beyond the capabilities of many mobile and embedded devices. At the same time, information security is a growing concern for people, and it is important to deploy applications directly on edge devices for personal information protection, which leads to high requirements for the computational volume and complexity of human pose estimation models.

Many works have been proposed to solve this problem by building human pose estimation networks with small model size and low computing cost [7, 10, 11]. For example, there is a recent attempt [10] to construct pose estimation models with fewer parameters using quantitative methods, but the performance of the obtained model largely degraded. Also, some researchers [7] try to use knowledge distillation to reduce the parameters of the model, but the model training time and deployment time are increased. On the other hand, some works attempt [12] to find a lightweight pose

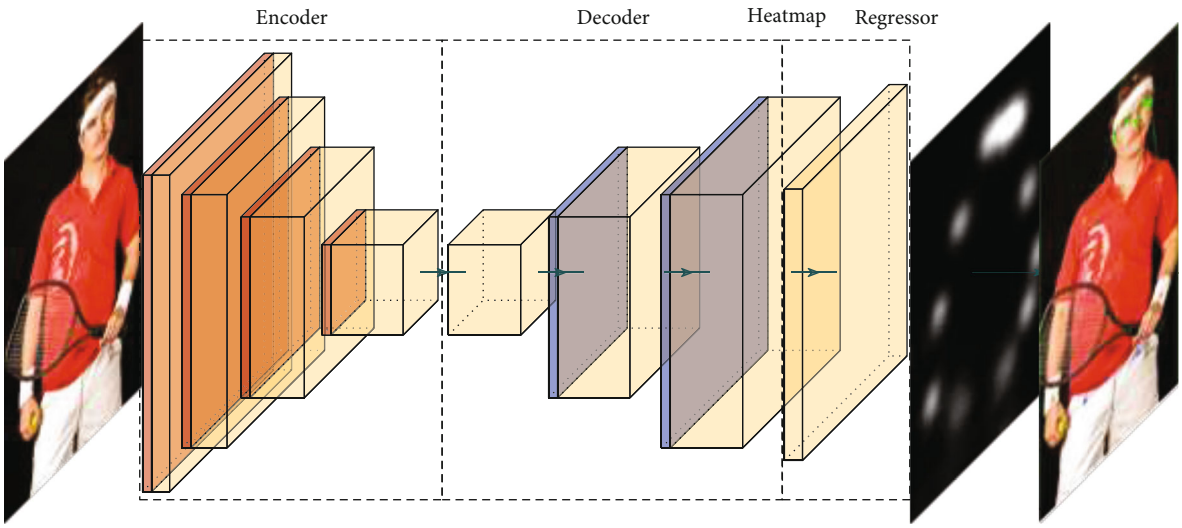


Figure 1: The architecture of the presented MobilePoseNet.

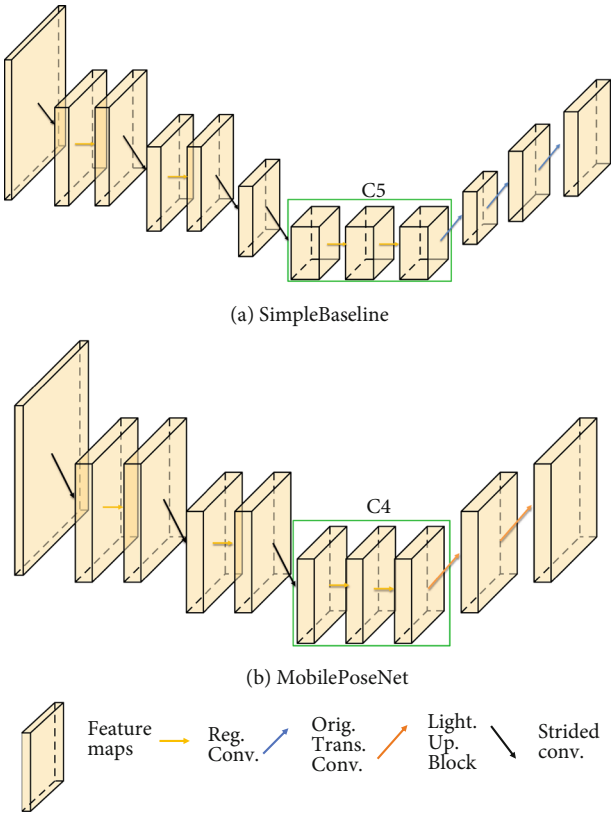


Figure 2: The differences between our proposed network and the SimpleBaseline structure. We can see from the figure that we choose C4 as the input for upsampling, and the implement time of our network is less compared to SimpleBaseline.

estimation model using neural structure search methods. However, the obtained model has a complex structure and slow inference speed. The key problem is how to balance the model accuracy and the inference efficiency.

To address this problem, in this paper, we propose a lightweight human pose estimation network specifically for mobile and resource-constrained environments by designing compact convolutional filters. As shown in Figure 1, our model contains three main parts: an encoder, a decoder, and a heat map regressor that estimates each key point. To keep the model lightweight, we use the first 13 layers of MobileNetV3 [13] as our encoder. Intuitively, high resolution is beneficial for human pose estimation, so we design the encoder with less downsampling, and the structure of the specific model and the comparison of SimpleBaseline can be seen in Figure 2. In the decoder part, inspired by the bottleneck block, we propose a lightweight upsampling module, whose concrete structure is shown in Figure 1. The detailed structure of the overall model is shown in Table 1. Finally, we also propose a parallel OKS-based NMS to further improve the operation speed of pose estimation. Experimental results show that our method can achieve 69.0 AP with only 1.5M model parameters and 1.23 GFLOP calculation amount under the condition of less cost. The contributions of the proposed method are summarized as follows:

(i) We design a lightweight upsampling block that integrates separable transpose convolution and channel-based attention. This is achieved by extensively examining the upsampling modules in existing state-of-the-art deep convolutional networks

(ii) We reduce the number of upsampling and use lightweight upsampling blocks to achieve a lightweight pose estimation network. In particular, we balance the accuracy of the model and the inference speed of the model, which is a key issue to be addressed in extending existing depth-pose estimation methods to practical applications

(iii) We propose a parallel OKS-NMS by combining Matrix-NMS [14] and OKS-NMS [15] to further improve the efficiency of the human pose estimation system

The rest of this paper is organized as follows. We briefly review the related work in the second section and followed

Table 1: Specification for MobilePoseNet. SE denotes whether there is a squeeze-and-excite in that block. NL denotes the type of nonlinearity used. Here, HS denotes h-swish and RE denotes relu. LPB is our proposed lightweight upsampling block. bneck is the bottleneck block in MobileNetV3. k is the number of key points.

Input channel	Input size	Operator	Exp size	#out	Attention	NL	s
3	256×192	Conv2d	—	16	—	HS	2
16	128×96	bneck, 3×3	16	16	—	RE	1
16	128×96	bneck, 3×3	64	24	—	RE	2
24	64×48	bneck, 3×3	72	24	—	RE	1
24	64×48	bneck, 5×5	72	40	SE	RE	2
40	32×24	bneck, 5×5	120	40	SE	RE	1
40	32×24	bneck, 5×5	120	40	SE	RE	1
40	32×24	bneck, 3×3	240	80	—	HS	2
80	16×12	bneck, 3×3	200	80	—	HS	1
80	16×12	bneck, 3×3	184	80	—	HS	1
80	16×12	bneck, 3×3	184	80	—	HS	1
80	16×12	bneck, 3×3	480	112	SE	HS	1
112	16×12	bneck, 3×3	672	112	SE	HS	1
112	16×12	bneck, 5×5	672	160	SE	HS	1
160	16×12	bneck, 5×5	960	160	SE	HS	1
160	16×12	bneck, 5×5	960	160	SE	HS	1
160	16×12	LPB, 4×4	320	160	SE	RE	2
160	32×24	LPB, 4×4	320	160	SE	RE	2
160	64×48	Conv2d 1×1	—	k	—	RE	1

by description of the proposed method. Then, we conduct experiments on the MSCOCO and MPII datasets and conclude this work.

2. Related Works

2.1. Human Posture Estimation. In recent years, deep learning-based pose estimation methods [16] have made great progress. Despite significant performance improvements, these prior works focused only on improving the accuracy of pose estimation by using complex networks and large tensors, while largely ignoring the cost issues of model inference. This state of affairs significantly limits their deploy ability in real-world applications, especially when the available computational budget is very limited.

In the literature, there are some recent works aimed at improving model efficiency. Bulat and Tzimiropoulos [10] designed a binary hourglass network using quantitative methods, but the restricted binary network has weak information representation and low accuracy of the model. Zhang et al. [7] proposed a new fast pose distillation (FPD) model learning strategy. A pretrained teacher network can be used to obtain a computationally fast and computationally inexpensive student network. However, it requires too much time to train. Yu et al. [17] proposed conditional channel weighting blocks and constructed the HR-Lite network, which achieves a great advantage in model accuracy and scale. However, the network structure is too complex, resulting in slow model inference. Zhang and Tang [11] proposed lightweight bottleneck block with depthwise convolution and attention mechanism, while the model size is still up to 2.7M parameters.

Compared with previous methods, comprehensively considering the accuracy of the model, the speed of inference, and the complexity of the model, we directly designed a model with simple structure and low complexity, which makes the model more practical and reliable in practical application scenarios.

2.2. Efficient Upsampling Module. Recent work [13, 18–20] has shown that deep convolutional neural networks have reached state-of-the-art performance. For advanced vision problems such as semantic segmentation [21], pose estimation [16], and object detection [22], existing approaches pass inputs through a network, usually consisting of high- to low-resolution subnetworks and a main network of raised resolutions. Many approaches have been designed to improve the resolution of the main network in different ways. For example, networks such as hourglass [6] reduce the input high-resolution features to low-resolution features and then use interpolation upsampling to scale the low-resolution features to the original input features, fuse the information with the previously input high-resolution features, and finally expect to generate fused high semantic and high resolution. Although it achieved very good results, the large tensor is used in the process of feature fusion. Zhou et al. [23, 24] constructed an attention-driven feature fusion upsampling network in an attempt to reduce the complexity of the model and reduce the use of large tensor using heterogeneous convolution. However, the network structure is complex and does not solve the problem of slow model inference fundamentally. SimpleBaseline [25] uses several transposed convolutional layers to generate high-resolution representations and achieves very good results. Although the model structure is simple, however, transposed convolution introduces a large number of parameters and computational effort, which is not friendly to small devices.

Therefore, we propose an efficient upsampling module that achieves a significant reduction in the number of

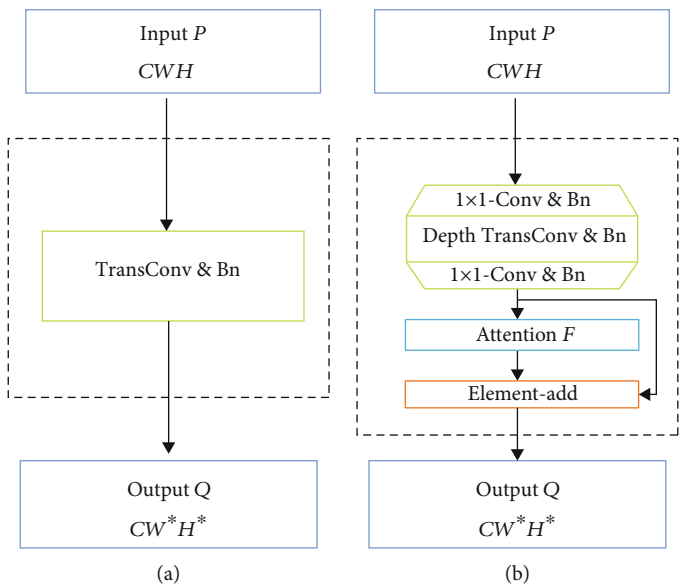


Figure 3: The comparison of two different upsampling methods. (a) The traditional transposed convolution, which has a large computational overhead. (b) The proposed lightweight upsampling block, which includes depthwise transposed convolution operation, pointwise convolution operation, and attention F. In (a), the features are amplified directly by transposed convolution. In (b), we first use 1×1 point convolution to expand the number of channels of the feature so that the number of channels goes from C to C^* and then use the depthwise transposed convolution to generate high-resolution feature maps. Finally, we use a 1×1 point convolution to change the number of channels to C and the attention mechanism to make feature map stronger.

parameters and computation during upsampling while ensuring the simplicity of the model structure and inference accuracy.

3. Proposed Method

In this section, we detail a simple and low computational cost human pose estimation network (MobilePoseNet), which designs a lightweight upsampling block (LPB) and directly uses high-resolution features to achieve high-resolution representation while maintaining lightweight features.

3.1. Lightweight Upsampling Block. Transposed convolution was first introduced into pose estimation by SimpleBaseline and achieved excellent performance. However, this operation brings a model with nearly a third of the parameters and calculations. Specifically, given the input of feature maps $C_{\text{in}} \times W_{\text{in}} \times H_{\text{in}}$ and the output of feature maps $C_{\text{out}} \times W_{\text{out}} \times H_{\text{out}}$, the amount of computation for conventional transposed convolution is

$$C_{\text{in}} \times C_{\text{out}} \times W_{\text{out}} \times H_{\text{out}} \times K \times K. \tag{1}$$

The number of parameters of traditional transpose convolution is

$$C_{\text{in}} \times C_{\text{out}} \times K \times K, \tag{2}$$

where K is the kernel size of traditional transposed convolution.

To reduce the burden of calculation and the number of parameters, while maintaining the effect of transpose convolution, we designed a lightweight upsampling block inspired by the intuition that the bottlenecks actually contain all the necessary information, as shown in Figure 3(b), which composed of three parts: depthwise transposed convolution, 1×1 point convolution, and attention module. Specifically, we first expand the low-latitude information to high-latitude information by 1×1 point convolution and use depth transpose convolution on each channel of the feature map for the spatial transformation. Finally, we use 1×1 point convolution to fuse the information between each channel and compress the high-latitude information to the original input latitude.

As shown in Figure 3(b), the computation of the lightweight upsampling block is the sum of the depth transpose convolution and the two point convolution computations:

$$\begin{aligned} C^* \times (&C_{\text{in}} \times W_{\text{in}} \times H_{\text{in}} \times 1 \times 1 + W_{\text{out}} \times H_{\text{out}} \\ &\times K \times K + C_{\text{out}} \times W_{\text{out}} \times H_{\text{out}} \times 1 \times 1). \end{aligned} \tag{3}$$

The number of parameters of the lightweight upsampling block is

$$C^* \times (K \times K + C_{\text{in}} + C_{\text{out}}), \tag{4}$$

where C^* is the number of channels for high-latitude features. Compared to the traditional transposed convolution, our method reduces the calculation amount to 83.2% and the number of parameters reduces to 74%.

Since LPB separates space operation and channel operation into two independent steps, the decoding effect of transpose convolution will be weakened. To solve this problem,

we enhance the feature responses through channel attention mechanism. Here, we directly use SENet [26] as our channel attention mechanism to dynamically adjust the weight of each channel, as shown in Figure 3(b). To sum up, we assumed the input feature map $X \in R^{C_{\text{in}} \times W_{\text{in}} \times H_{\text{in}}}$, the feature output through the LPB is $X' \in R^{C_{\text{out}} \times W_{\text{out}} \times H_{\text{out}}}$ as input to the channel attention mechanism.

The feature output through the channel attention mechanism is $X_{\text{att}} \in R^{C_{\text{out}} \times 1 \times 1}$. Then, the feature output of LPB and the feature output of the channel attention mechanism are multiplied and summed to obtain the final fusion information Y, i.e.,

$$Y = X' + X' X_{\text{att}}. \tag{5}$$

3.2. Lightweight Human Pose Estimation. Usually, the pipeline [5, 6, 27, 28] for poses estimation consists of three parts: the upsampling, the downsampling, and estimation of the heat map. In this work, we focus on the design of a lightweight upsampling and downsampling.

Different from SimpleBaseline which uses a ResNet backbone as the downsampling and three traditional deconvolutional layers as the upsampling, we use MobileNetV3 as our downsampling, which reduces the size of the parameters up to 96% and reduces the computation load up to 79%. For the upsampling, we replace each traditional deconvolution layer with a lightweight upsampling block. The details of the model are shown in Table 1.

As shown in Figure 2, different with SimpleBaseline, we use a higher resolution feature map as the input for upsampling. The rationale behind this that it is beneficial to maintain high-resolution representations before upsampling.

3.3. Parallel Pose NMS. In pose estimation, human body detectors inevitably generate redundant detection, and pose estimation also generates redundant poses. Therefore, nonmaximum suppression (NMS) is required to eliminate redundant postures.

Given pose P_i with m joints $\{<k_1^i, s_1^i>, <k_2^i, s_2^i>, \cdots, <k_m^i, s_m^i>\}$ where k_j^i and s_j^i are the location and confidence score of the jth joint, respectively. Corresponding detection boxes b^i with b_s^i confidence score. The general pose NMS is as follows: firstly, the pose with the highest confidences was chosen as the reference, and the poses similar to it were suppressed or discarded. This process is repeated for the rest of the pose set until only one pose is left.

However, the main problems for this process are sequential and cannot be implemented in parallel, resulting in slower speeds. Inspired by Matrix-NMS, we proposed parallel nonmaximum suppression considering following two key factors:

(1) The confidence of pose: the higher the confidence of pose, the lower the probability of joints being suppressed, i.e., if the pose P_i and P_j with confidence $(p_i > p_j)$, P_j will have a high probability of being suppressed

(2) Similarity between the pose and other poses: the lower the similarity between one pose and other poses, the lower the suppression ratio of the poses

For the pose confidence, we set the product of the average of the confidence of the key points and the confidence of the human detector as the final pose confidence below

$$p_i = \left(\frac{\sum_m s_m^i \delta\left(s_m^i, \text{threshold}\right)}{\sum_m \delta(s_m^i, \text{threshold})} \right) \cdot b_s^i, \tag{6}$$

where b_s^i is the confidence of the detection box and δ is defined as follows:

$$\delta(s, \text{threshold}) = \begin{cases} 1, s > \text{threshold}, \\ 0, s \leq \text{threshold}. \end{cases} \tag{7}$$

We consider the key point prediction to be true if s_m^i is bigger than threshold and otherwise to be false.

For the similarity between two poses, we use the object key point similarity (OKS) [29] to measure the pose distance function as follows:

$$f\left(P_i, P_j\right) = 1 - O_{i,j}, \tag{8}$$

where $O_{i,j}$ is given by

$$O_{i,j} = \frac{\sum_m \exp\left\{ -\left(k_m^i - k_m^j\right)^2 / \left(2 b_a^i b_a^j\right) \right\}}{\sum_m \delta(s_m^i, \text{threshold}) \bullet \delta\left(s_m^j, \text{threshold}\right)}, \tag{9}$$

where b_a^i is the area of the detection box.

We define a new decay factor for pose NMS. For $f(P_i, P_j) = 1 - O_{i,j}$, we can get a new decay_j, where

$$\text{decay}_j = \min_{\forall p_i > p_j} \frac{f\left(P_i, P_j\right)}{f(:, P_i)}, \tag{10}$$

where $f(:, P_i) = \min_{\forall p_l > p_i} f(p_l, p_i)$.

Finally, we get a new pose confidence $p_j \longleftarrow \text{decay}_j \cdot p_j$. For usage, we just need threshold and selecting top-k scoring masks as the final predictions.

Like Matrix-NMS, all the operations in pose NMS could be implemented in one shot without recurrence. We first get a N pose confidence and then compute a $N \times N$ pairwise OKS matrix for the N pose sorted descending by pose confidence score. The decay factors of each pose can be obtained by looking up the table of the OKS matrix. Finally, the pose scores are updated by the decay factors. For usage, we just need threshold and select top-k pose scoring as the final predictions. The whole procedure is summarized in Algorithm 1.

Input: the area of the detection boxer $b_a = \{b_a^i\}$, the confidence of the detection boxer $b_s = \{b_s^i\}$, the location of the key point $K = \{k_j^i\}$, the confidence of the key point $S = \{s_j^i\}$, and parameter threshold. Here, i is the i-th person, $i \in \{1, 2, 3, \cdots, n\}$, j is the j-th key point, and $j \in \{1, 2, 3, \cdots, m\}$.
Output: the confidence of the key point $p = \{p^i | i = 1, 2, 3, \cdots n\}$.
Initialize threshold = 0.1
Calculate p by equation (6) and parameter b_s, S, threshold
Sort b_a, b_s, and K in descending order by $p = \{p^1, p^2, \cdots p^n\}$
Calculate B_+ using $b_s \times (b_s)^T$
Calculate K_+ using $(K + K^T)^2 - 4 \times (K + K^T)$
Calculate OKS matrix O by equation (9) and parameter K_+, B_+, threshold
Update $O = \{O_{i,j} = 1 | i \geq j, i, j = 1, 2, 3, \cdots n\}$
Set $o_m = \{\max_j (O_{:,j}) | j = 1, 2, 3, \cdots n\}$
Set O_m by repeating $o_m n$ times
Calculate decay matrix D using $(I - O)/(I - O_m)$
Set decay $o_d = \{\min (D_{:,j}) | j = 1, 2, 3, \cdots n\}$
Update the confidence of the key point p by $o_d \odot p$

Algorithm 1: Parallel pose NMS.

Table 2: Comparisons of results on the MSCOCO validation set.

Method	Backbone	Input	#Params	GFLOPs	AP	AP^{50}	AP^{75}	AP^{M}	AP^{L}	AR
8-stage hourglass [6]	Hourglass	256×192	25.6M	26.2	66.9	—	—	—	—	—
CPN [31]	ResNet-50	256×192	27.0M	6.2	68.4	—	—	—	—	—
SimpleBaseline [25]	ResNet-50	256×192	34.0M	8.9	70.4	88.6	78.3	67.1	77.2	76.3
HRNet-W32 [5]	ResNet-50	256×192	28.5M	12.4	**73.4**	**89.5**	**80.7**	**70.2**	**80.1**	**79.8**
DARK [32]	HRNetV1-W48	128×96	63.6M	3.6	71.9	89.1	79.6	69.2	78	77.9
MobileNetV2 [19]	MobileNetV2	256×192	9.6M	1.48	64.6	87.4	72.3	61.1	71.2	70.7
MobileNetV2 1×	MobileNetV2	384×288	9.6M	3.33	67.3	87.9	74.3	62.8	74.7	72.9
ShuffleNetV2 [33]	ShuffleNetV2	256×192	7.6M	1.28	59.9	85.4	66.3	56.6	66.2	66.4
ShuffleNetV2 1×	ShuffleNetV2	384×288	7.6M	2.87	63.6	86.5	70.5	59.5	70.7	69.7
Small HRNet	HRNet-W16	256×192	1.3M	0.54	55.2	83.7	62.4	52.3	61	62.1
Small HRNet	HRNet-W16	384×288	1.3M	1.21	56	83.8	63	52.4	62.6	62.6
Lite-HRNet	Lite-HRNet-18	256×192	1.1M	0.20	64.8	86.7	73	62.1	70.5	71.2
Lite-HRNet	Lite-HRNet-18	384×288	1.1M	0.45	67.6	87.8	75	64.5	73.7	73.7
MobilePoseNet	MobilNetV3	256×192	1.5M	0.55	66.2	87.3	74.2	63.1	72.5	72.4
MobilePoseNet	MobilNetV3	384×288	1.5M	1.23	69	88.2	75.9	65.5	75.5	74.9

4. Experiments

We conduct experiments on the MSCOCO and MPII datasets to evaluate the performance of our method in multiperson pose estimation.

4.1. Datasets

(i) The MSCOCO dataset contains over 200K images, 250K human body instances, and 17 key points. We trained our model on the MSCOCO train2017 dataset, including 57K images and 150K person instances and evaluated our approach on val2017 and test-dev2017, which contained 5000 images and 20K images, respectively

(ii) The MPII Human Pose dataset contains about 25K images of more than 40,000 people with annotated human joints, which are taken from a wide range of real-world activities with full-body pose annotations

We selected the object key point similarity (OKS) as an evaluation metric for the MSCOCO dataset. The standard metric [30], the PCK (probability of correct key point normalized by head) score, was used to evaluate the MPII dataset.

4.2. Implement Details. In MSCOCO, we extend the human detection box into a fixed aspect ratio with 4 : 3, and crop the box from the image with fixed size, 256×192 or 384×288. In MPII, the input size is cropped to 256×256 for fair comparison with other methods. In addition, the same data

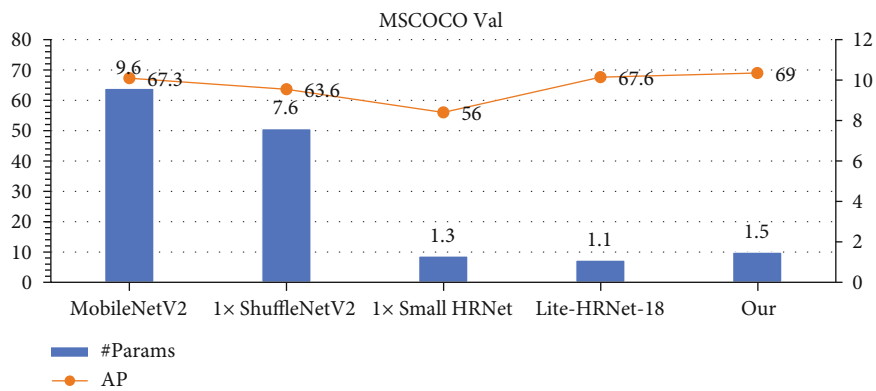


Figure 4: The complexity and accuracy comparison of MSCOCO val for 384×288 input size.

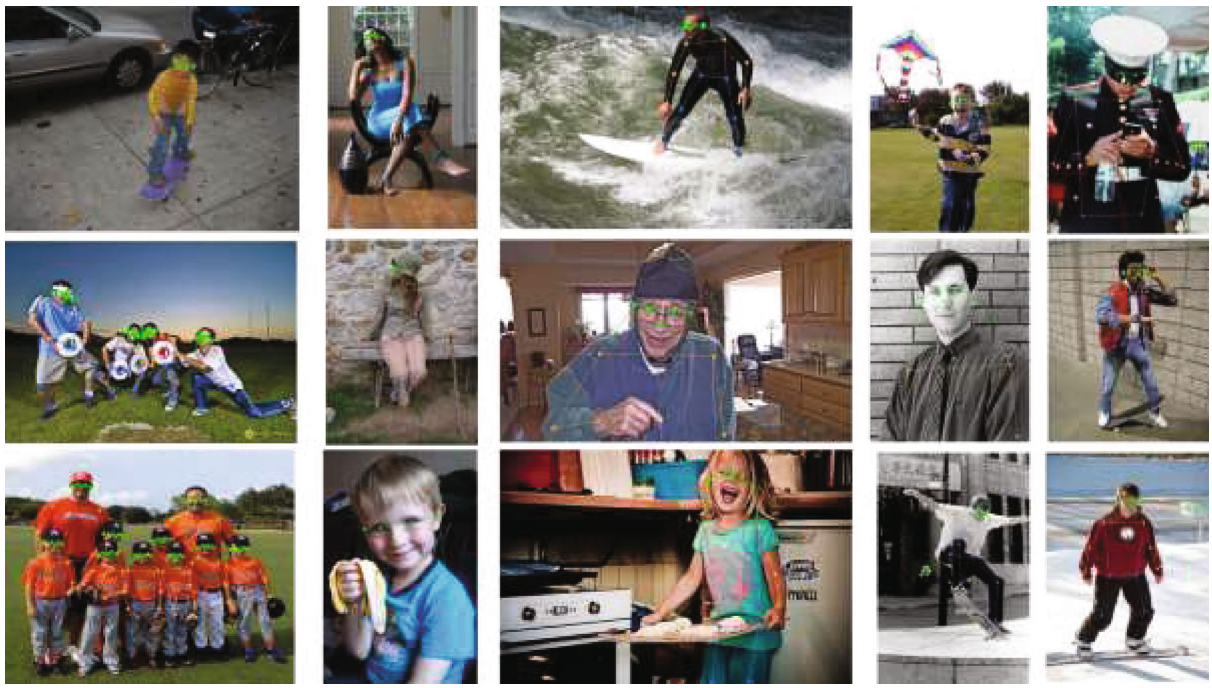


Figure 5: Visualization results of human pose estimation based on MobilePoseNet for MSCOCO validation set.

augmentation and the training strategy are utilized for both datasets. The data augmentation includes random rotation ([-45, 45]), random scale ([0.65, 1.35]), and flipping. In MSCOCO, half body data augmentation is also involved.

We all use the Adam optimizer with initial learning rate $1e-3$. The model was trained on a single Nvidia TITAN RTX GPU with a minibatch size 32 and stop at 210 epochs.

4.3. Experimental Results

4.3.1. Results on MSCOCO Dataset. From the results, as shown in Table 2, we can see that our method has a significant advantage in terms of model size and complexity with comparable accuracy. For input size 256×192, our method achieved comparable accuracy with less than 6% the parameters with respect to hourglass network. Compared with MobileNetV2 and ShuffleNetV2, our method obtained better accuracy with low complexity. For the small network HRNet-W16 and Lite-HRNet-18, our model is also better in terms of accuracy although the model size is slightly large. For the input 382×288, we can also derive the same conclusion.

Figure 4 illustrates the comparison of accuracy and complexity of small networks. Figure 5 shows the visualization results of our method in MSCOCO. It can be seen that our model achieved better balance between complexity and accuracy and can estimate the accurate joints under different complex scenes.

Table 3 lists the mAP, input size, Params, and GFLOP values of compared methods and our method on the MSCOCO dataset.

4.3.2. Results on MPII Human Pose Dataset. Table 4 reports the results of our network and other lightweight networks on MPII val data. Compared with MobileNetV2, MobileNetV3,

Table 3: Comparisons of results on MSCOCO test-dev2017 set. #Params and flops are calculated for the pose estimation network, and those for human detection are not included.

Method	Backbone	Input	#Params	GFLOPs	AP	AP^{50}	AP^{75}	AP^{M}	AP^{L}	AR
Bottom-up: key point detection and grouping										
OpenPose [34]	—	—	—	—	61.8	84.9	67.5	57.1	68.2	66.5
Associative embedding [35]	—	—	—	—	65.5	86.6	72.3	60.6	72.6	70.2
PersonLab [4]	—	—	—	—	68.7	89	75.4	64.1	75.5	75.4
MultiPoseNet [36]	—	—	—	—	69.6	86.3	76.6	65.0	76.3	73.5
HigherHRNet [37]	HRNet-w32	512×512	28.6M	47.9	66.4	87.5	72.8	61.2	74.2	—
Top-down: human detection and single-person key point detection										
Large network										
Mask-RCNN [22]	ResNet-50-FPN	—	—	—	63.1	87.3	68.7	57.8	71.4	—
G-RMI [15]	ResNet-101	353×257	42.6M	57	64.9	85.5	71.3	62.3	70.0	69.7
IPR [27]	ResNet-101	256×256	45.0M	11	67.8	88.2	74.8	63.9	74.0	—
RMPE [38]	PyraNet [39]	320×256	28.1M	26.7	72.3	89.2	79.1	68.0	78.6	—
CPN [28]	—	384×288	—	—	72.1	91.4	80.0	68.7	77.2	78.5
SimpleBaseline [25]	ResNet-152	384×288	68.6M	35.6	73.7	91.9	81.1	70.3	80.0	79.0
Small network										
MobileNetV2 [19]	MobileNetV2	384×288	9.8M	3.33	66.8	90.0	74.0	62.6	73.3	72.3
ShuffleNetV2 [33]	ShuffleNetV2	384×288	7.6M	2.87	62.9	88.5	69.4	58.9	69.3	68.9
Small HRNet [17]	HRNet-W16	384×288	1.3M	1.21	55.2	85.8	61.4	51.7	61.2	61.5
Lite-HRNet [17]	Lite-HRNet-18	384×288	1.1M	0.45	66.9	89.4	74.4	64.0	72.2	72.6
MobilePoseNet	MobileNetv3 [13]	256×192	1.5M	0.55	64.8	88.8	72.4	61.9	70.2	70.7
MobilePoseNet	MobileNetv3	384×288	1.5M	1.23	67.4	89.4	74.2	64.1	73.3	73.3

Table 4: Comparisons on the MPII val set. The GFLOPs is computed with the input size 256×256.

Model	#Params	GFLOPs	PCKh
MobileNetV2 1×	9.6M	1.97	85.4
MobileNetV3 1×	8.7M	1.82	84.3
ShuffleNetV2 1×	7.6M	1.70	82.8
Small HRNet-W16	1.3M	0.72	80.2
Lite-HRNet-18	**1.1**M	**0.27**	86.1
MobilePoseNet	1.5M	0.74	**87.3**

ShuffleNetV2, and SmallHRNet-W16, our model achieves better accuracy with lower number of parameters and calculation weights. Compared to Lite-HRNet-30, our model achieves 87.3 PCKh@0.5 in terms of the number of parameters with 0.3M less than Lite-HRNet-30. Compared to MobileNetV2, MobileNetV3, ShuffleNetV2, and Small HRNet-W16, our model improved by 1.9%, 3.0%, 4.5%, and 7.1%, respectively. Figure 6 illustrates the comparison of accuracy and complexity.

4.4. Inference Speed. FLOPs and Param are only the properties that measure the size and complexity of the model. In this section, we study the actual inference speed of the human pose estimation network by inference items per second (Inference Items Per Second). The speed is tested on devices with GPU and without GPU, respectively, with a batch size of 32 and full precision (fp32). We use the Nvidia TITAN TRX as the GPU device and the Intel Core I9-10900k device without GPU as the non-GPU device. To better reflect the running speed of the model, all methods are tested on the MSCOCO validation set. We use the same person detector provided by the SimpleBaseline validation set. In the tests without GPU, a thread was used for evaluation. As can be seen in Table 5, thanks to the simple structure of our model, our actual inference is 3 times faster than the less computationally intensive Lite-HRNet on the GPU speed test. In the GPU-free speed test, our method is faster than a large network like HRNet. Also, our model has a significant advantage in complexity and computational power compared to other models, which means easier deployment to embedded devices.

5. Ablation Study

We study the effect of each component of our approach on the validation set of MSCOCO.

5.1. Deconvolution Blocks. In this section, we analyzed the impact of reducing the number of upsampling and using different upsampling blocks in terms of accuracy with resolution 384×288. From Table 6, it can be seen that the number of parameters and the computation of our model

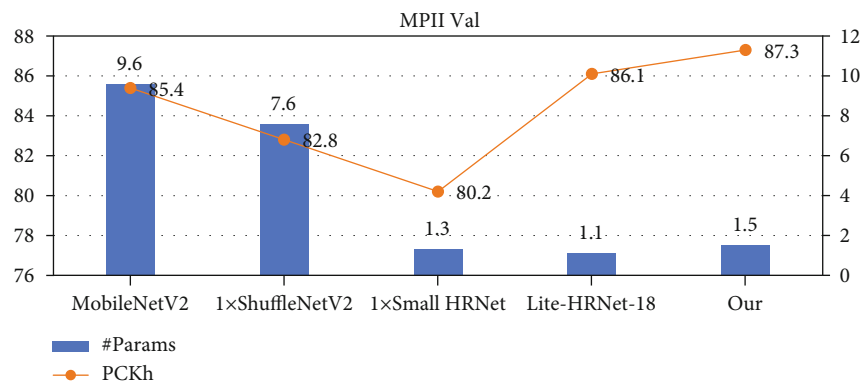


Figure 6: The complexity and accuracy comparison of MPII val sets for 256 × 256 input size.

Table 5: Inference speed comparisons on the MSCOCO validation set. Speed* refers to the result on non-GPU device. Speed refers to the result on GPU device. Bold values are the optimal results.

Method	Backbone	#Params	GFLOPs	Input size	AP	Speed*	Speed
HRNet	HRNetV1-W32	28.5M	7.1	256 × 192	74.4	7.5	19.2
HRNet	HRNetV1-W32	28.5M	16	384 × 288	**75.8**	4	18.8
SimpleBaseline	ResNet-50	34.0M	8.9	256 × 192	70.4	8.1	**273.1**
NLite-HRNet-18	HRNet-W16	0.7M	0.19	256 × 192	62.8	11	18.9
WNLite-HRNet-18	HRNet-W16	1.3M	0.3	256 × 192	66	12	18.6
ShuffleNetV2 1×	ShuffleNetV2	7.6M	1.28	256 × 192	59.9	**17**	71.3
ShuffleNetV2 1×	ShuffleNetV2	7.6M	2.87	384 × 288	63.6	10	64.1
MobileNetV2 1×	MobileNetV2	9.6M	1.48	256 × 192	64.6	6.8	83.1
MobileNetV2 1×	MobileNetV2	9.6M	3.33	384 × 288	67.3	4.5	73.1
Lite-HRNet	Lite-HRNet-18	1.1M	0.2	256 × 192	64.8	12	17.4
Lite-HRNet	Lite-HRNet-18	1.1M	0.45	384 × 288	67.6	7.1	16.3
MobilePoseNet	MobileNetV3	1.5M	0.55	256 × 192	66.2	7.8	54.8
MobilePoseNet	MobileNetV3	1.5M	1.23	384 × 288	69.0	5.1	50.8

Table 6: Ablation experiments on reduced downsampling with the use of lightweight upsampling blocks, on the MSCOCO val dataset. V1 denote the model that uses C5 as the input for upsampling, using the first 16 layers of MobileNetV3 as the downsampling and three layers of deconvolution as the upsampling part. V2 denote the model that uses C4 as the input for upsampling, using the first 13 layers of MobileNetV3 as the downsampling part, then uses three layers of 5 × 5 bottleneck with a stride of 1, and finally uses two layers of the same deconvolution as V1 as the upsampling part.

Model	Input size	FLOPs	#Params	AP	AP^{50}	AP^{75}	AP^{M}	AP^{L}	AR
V1	256 × 192	604M	2.5M	65.22	87.05	73.05	62.2	71.47	71.45
V2	256 × 192	684M	2.1M	66.23	87.23	74.19	63.21	72.48	72.38
V1	384 × 288	1.33G	2.5M	68.44	87.71	75.41	64.89	75.01	74.47
V2	384 × 288	1.5G	2.1M	68.97	87.72	75.47	65.32	**75.7**	74.85
Ours	256 × 192	**557**M	1.5M	66.23	87.38	74.25	63.13	72.52	72.4
Ours	384 × 288	1.23G	1.5M	**69.03**	**87.72**	**75.95**	**65.52**	75.55	**74.98**

are reduced compared to other models, while the precision has indeed been improved.

5.2. OKS-Based Nonmaximum Suppression. We compared the proposed OKS-based nonmaximum suppression and other OKS-based nonmaximum suppression methods on the accuracy and speed with the same pose estimator. As shown in Table 7, we can find that our proposed OKS-based nonextreme suppression has significant advantages in terms of accuracy and speed.

Table 7: Performance (AP) and speed (ms) of the MSCOCO validation set for different pose NMS. Matrix-OKS-NMS outperforms both Hard- and Soft-OKS-NMS methods in terms of speed and accuracy. Input size is 384×288.

Method	AP	AP^{50}	AR	Time (ms)
Hard-OKS-NMS	67.7	**87.1**	91.4	6.14
Soft-OKS-NMS	67.7	86.7	**91.5**	6.8
Parallel-OKS-NMS	**67.8**	86.7	**91.5**	**6.0**

6. Conclusion

In this paper, we propose a lightweight pose estimation network, which can achieve an AP score of 69.0 on the MSCOCO val set with only 1.5M parameters and 1.23 GFLOPs. However, we found that our model has some gaps compared to high-performance algorithms, mainly because we are missing the fusion of multiscale information. Designing complex networks and introducing the fusion of multiscale information will increase the inference speed of the model. In future work, we will redesign the backbone network for human pose estimation by introducing multiscale information to balance accuracy and speed.

Data Availability

The datasets used in this paper are the public datasets MSCOCO and MPII.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Nos. 61962010 and 61976107), the Excellent Young Scientific and Technological Talent of Guizhou Province ([2019]-5670), the Natural Science Foundation of Guizhou Province (Grant No. [2017]5726-32), the National Natural Science Foundation (No. 61863006), and the Basic Research Project (Key Project) of Guizhou Province ([2019]-1416).

References

[1] J. Shotton, A. Fitzgibbon, M. Cook et al., "Real-time human pose recognition in parts from single depth images," in *Computer Vision and Pattern Recognition 2011*, pp. 1297–1304, Colorado Springs, USA, 2011.

[2] N.-G. Cho, A. L. Yuille, and S.-W. Lee, "Adaptive occlusion state estimation for human pose tracking under self- occlusions," *Pattern Recognition*, vol. 46, no. 3, pp. 649–661, 2013.

[3] G. Cheron, I. Laptev, and C. Schmid, "P-CNN: pose-based CNN features for action recognition," in *International Conference on Computer Vision*, pp. 3218–3226, Santiago, Chile, 2015.

[4] G. Papandreou, T. Zhu, L.-C. Chen, S. Gidaris, J. Tompson, and K. Murphy, "PersonLab: person pose estimation and instance segmentation with a bottom-up, part-based, geometric embedding model," in *European Conference on Computer Vision*, pp. 282–299, Munich, Germany, 2018.

[5] K. Sun, B. Xiao, D. Liu, and J. Wang, "Deep high-resolution representation learning for human pose estimation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5693–5703, Long Beach, USA, 2019.

[6] A. Newell, K. Yang, and J. Deng, "Stacked hourglass networks for human pose estimation," in *European Conference on Computer Vision*, pp. 483–499, Glasgow, United Kingdom, 2016.

[7] F. Zhang, X. Zhu, and M. Ye, "Fast human pose estimation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3517–3526, Long Beach, USA, 2019.

[8] Y. Chen, Y. Tian, and M. He, "Monocular human pose estimation: a survey of deep learning-based methods," *computer vision and image understanding*, vol. 192, article 102897, 2020.

[9] W. Li, Z. Wang, B. Yin et al., "Rethinking on multi-stage networks for human pose estimation," 2019, https://arxiv.org/abs/1901.00148.

[10] A. Bulat and G. Tzimiropoulos, "Binarized convolutional landmark localizers for human pose estimation and face alignment with limited resources," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pp. 3726–3734, Venice, Italy, 2017.

[11] Z. Zhang, J. Tang, and G. Wu, "Simple and lightweight human pose estimation.," https://arxiv.org/abs/1911.10346.

[12] M. Ding, X. Lian, L. Yang et al., "HR-NAS: searching efficient high-resolution neural architectures with lightweight transformers," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2982–2992, 2021.

[13] A. Howard, R. Pang, H. Adam et al., "Searching for MobileNetV3," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 1314–1324, Seoul, Korea (south), 2019.

[14] X. Wang, R. Zhang, T. Kong, L. Li, and C. Shen, "SOLOv2: dynamic and fast instance segmentation," pp. 17721–17732, 2020, https://arxiv.org/abs/2003.10152.

[15] G. Papandreou, T. Zhu, N. Kanazawa et al., "Towards accurate multi-person pose estimation in the wild," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3711–3719, Hawaii, USA, 2017.

[16] A. Toshev and C. Szegedy, "DeepPose: human pose estimation via deep neural networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1653–1660, Columbus, USA, 2014.

[17] C. Yu, B. Xiao, C. Gao et al., "Lite-HRNet: a lightweight high-resolution network," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 10440–10450, 2021.

[18] 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 (CVPR),*, pp. 770–778, Las Vegas, USA, 2016.

[19] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: inverted residuals and linear bottlenecks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4510–4520, Salt Lake City, USA, 2018.

[20] W. Lu, R. Yu, S. Wang, C. Wang, P. Jian, and H. Huang, "Sentence semantic matching based on 3D CNN for human–robot language interaction," *ACM Transactions on Internet Technology (TOIT)*, vol. 21, no. 4, pp. 1–24, 2021.

[21] Q. Zhou, X. Wu, S. Zhang, B. Kang, Z. Ge, and L. Jan Latecki, "Contextual ensemble network for semantic segmentation," *Pattern Recognition*, vol. 122, article 108290, 2022.

[22] K. He, G. Gkioxari, P. Dollar, and R. Girshick, "Mask R-CNN," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pp. 2980–2988, Venice,Italy, 2017.

[23] Q. Zhou, Y. Wang, Y. Fan et al., "AGLNet: towards real-time semantic segmentation of self-driving images via attention-guided lightweight network," *applied soft computing*, vol. 96, p. 106682, 2020.

[24] Q. Zhou, Y. Wang, J. Liu, X. Jin, and L. J. Latecki, "An open-source project for real-time image semantic segmentation," *SCIENCE CHINA Information Sciences*, vol. 62, no. 12, article 227101, 2019.

[25] 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. 472–487, Munich,-Germany, 2018.

[26] J. Hu, L. Shen, S. Albanie, G. Sun, and E. Wu, "Squeeze-and-excitation networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2011–2023, Salt Lake City, USA, 2018.

[27] X. Sun, B. Xiao, F. Wei, S. Liang, and Y. Wei, "Integral human pose regression," in *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 536–553, Munich,Germany, 2018.

[28] Y. Chen, Z. Wang, Y. Peng, Z. Zhang, G. Yu, and J. Sun, "Cascaded pyramid network for multi-person pose estimation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 7103–7112, Salt Lake City, USA, 2018.

[29] T.-Y. Lin, M. Maire, S. J. Belongie et al., "Microsoft COCO: common objects in context," in *European Conference on Computer Vision*, pp. 740–755, Zurich, Switzerland, 2014.

[30] M. Andriluka, L. Pishchulin, P. Gehler, and B. Schiele, "2D human pose estimation: new benchmark and state of the art analysis," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3686–3693, Columbus, USA, 2014.

[31] S. Huang, M. Gong, and D. Tao, "A coarse-fine network for keypoint localization," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pp. 3047–3056, Venice, Italy, 2017.

[32] F. Zhang, X. Zhu, H. Dai, M. Ye, and C. Zhu, "Distribution-aware coordinate representation for human pose estimation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 7093–7102, Seattle, USA, 2020.

[33] N. Ma, X. Zhang, H.-T. Zheng, and J. Sun, "ShuffleNet V2: practical guidelines for efficient CNN architecture design," in *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 122–138, Munich,Germany, 2018.

[34] 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 (CVPR)*, pp. 1302–1310, Hawaii, USA, 2017.

[35] A. Newell, Z. Huang, and J. Deng, "Associative embedding: end-to-end learning for joint detection and grouping," *Advances in neural information processing systems*, vol. 30, pp. 2278–2288, 2017.

[36] M. Kocabas, S. Karagoz, and E. Akbas, "MultiPoseNet: fast multi-person pose estimation using pose residual network," in *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 437–453, Munich,Germany, 2018.

[37] B. Cheng, B. Xiao, J. Wang, H. Shi, T. S. Huang, and L. Zhang, "HigherHRNet: scale-aware representation learning for bottom-up human pose estimation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5386–5395, Seattle, USA, 2020.

[38] H.-S. Fang, S. Xie, Y.-W. Tai, and C. Lu, "RMPE: regional multi-person pose estimation," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pp. 2353–2362, Venice,Italy, 2017.

[39] W. Yang, S. Li, W. Ouyang, H. Li, and X. Wang, "Learning feature pyramids for human pose estimation," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pp. 1290–1299, Venice,Italy, 2017.

Hindawi
Wireless Communications and Mobile Computing
Volume 2021, Article ID 7386280, 11 pages
https://doi.org/10.1155/2021/7386280

Research Article

Group-Based Atrous Convolution Stereo Matching Network

Qijie Zou,[1] **Jing Yu**,[1] **Hui Fang**,[2] **Jing Qin**,[1] **Jie Zhang**,[1] **and Shengkai Liu**[1]

[1]*Department of Information Engineering Faculty, Dalian University of China, 116622, China*
[2]*School of Computer Science, Loughborough University, LE113TU, UK*

Correspondence should be addressed to Jing Yu; dl_yujing163@163.com

Received 21 August 2021; Revised 11 October 2021; Accepted 2 November 2021; Published 28 November 2021

Academic Editor: Chi-Hua Chen

Stereo matching is the key technology in stereo vision. Given a pair of rectified images, stereo matching determines correspondences between the pair images and estimate depth by obtaining disparity between corresponding pixels. The current work has shown that depth estimation from a stereo pair of images can be formulated as a supervised learning task with an end-to-end frame based on convolutional neural networks (CNNs). However, 3D CNN puts a great burden on memory storage and computation, which further leads to the significantly increased computation time. To alleviate this issue, atrous convolution was proposed to reduce the number of convolutional operations via a relatively sparse receptive field. However, this sparse receptive field makes it difficult to find reliable corresponding points in fuzzy areas, e.g., occluded areas and untextured areas, owing to the loss of rich contextual information. To address this problem, we propose the Group-based Atrous Convolution Spatial Pyramid Pooling (GASPP) to robustly segment objects at multiple scales with affordable computing resources. The main feature of the GASPP module is to set convolutional layers with continuous dilation rate in each group, so that it can reduce the impact of holes introduced by atrous convolution on network performance. Moreover, we introduce a tailored cascade cost volume in a pyramid form to reduce memory, so as to meet real-time performance. The group-based atrous convolution stereo matching network is evaluated on the street scene benchmark KITTI 2015 and Scene Flow and achieves state-of-the-art performance.

1. Introduction

Some complex advanced visual tasks depend on depth perception [1], such as robot control and navigation [2], three-dimensional measurement [3], unmanned aerial vehicles (UAVs), virtual reality, and microoperating system parameter detection, showing the significance of distance information acquisition for vision works. Stereo matching is one of visual tasks, which computes the disparity of each pixel when given a pair of rectified images. Extensive work has been proposed for the task, including conventional methods and newly deep learning (DL) methods. Nevertheless, in practical application, the size of the target object is diverse. For larger objects (such as indoor walls and tables and outdoor sky and ground), we may ignore fine details; for smaller objects (such as pedestrians and vehicles), it will be a lack of global information. These make it difficult in dealing with ill-posed regions which still need to be calculated accurately, such as weakly textured regions, occluded areas, and reflective surfaces. Moreover, CNN-based algorithms are computationally expensive. These deficiencies make the research based on vision develop continuously.

Stereo matching estimates depth by matching pixels from a rectified image pair captured by two cameras, in which the goal is to obtain distance and contextual information from disparity quickly and accurately. The performance depends on robustly segmenting objects at multiple scales. Features provide contextual information for the stereo matching process to compute disparity.

Conventional stereo matching can be expressed as a multi-level optimization problem, which generally includes four steps: matching cost calculation, cost aggregation, disparity calculation, and disparity optimization [4]. The method achieves a trade-off between the computational complexity and the quality of the results obtained. However, the performance of the traditional stereo matching method is heavily limited by the

handcrafted features, such as graph cut [5], belief propagation [6], BM (Block Matching), and DP (Dynamic Programming) [7], which are adopted by cost functions, with poor robustness [8–11].

At present, stereo matching algorithms have become a deep learning task resorting to the development of CNN. CNN is introduced to replace one or more components in the legacy stereo pipeline. Some CNN-based algorithms have shown their stronger feature extraction ability than traditional methods. MC-CNN [12] first utilizes a convolutional neural network to learn how to match corresponding points in the matching cost computation process. End-to-end disparity estimation network as one of the CNN-based algorithms, which integrates all steps in the stereo matching pipeline for concatenating optimization, produces dense disparity maps from stereo images directly. Stereo matching networks with end-to-end approach are able to generate highly accurate depth estimation from stereo image pairs. However, they require huge memory and computation consumption. Meanwhile, it is difficult to infer reliable correspondences in ill-posed with limited receptive field and lack of contextual information. Atrous Spatial Pyramid Pooling (ASPP) explores an incoming convolutional feature layer with filters at multiple sampling rates and effective fields-of-views, thus capturing objects as well as rich image context at multiple scales [13].

In particular, we consider that the challenge of stereo matching is how to reduce memory usage while ensuring that the contextual information is fully utilized. Therefore, in this work, we design a network which could make use of rich contextual information as well as minimize GPU memory occupation. Specifically, we propose a small dilation atrous convolution group and a light tailored cascade cost volume, which boost the accuracy and performance by extracting features with more contextual information and lessen consumed memory and time.

Our main contributions are listed as follows:

(1) A group-based atrous convolution pyramid module is presented in this paper. The module ensures multiscale context information captured from various receptive fields when reducing the network size significantly

(2) The tailored cascade cost volume is constructed by changing the output channels and utilizing pyramid construction of cascade structure leading to efficiency to calculate the disparity

It is worth noting that the above two properties of GASPP make the training error converge faster in matching, and the accuracy of disparity regression is higher. Compared with the classic CNN-based model, PSM-Net, the average time of each iteration of our model is shortened by about 30%. In addition, the model is less dependent on the batch size.

2. Background Knowledge

2.1. Stereo Matching Based on Deep Learning. Currently, stereo matching algorithms can be treated as a deep learning task. For convenience, we classify stereo matching based on deep learning algorithms into two categories: non-end-to-end learning algorithm and end-to-end algorithm. An end-to-end algorithm learns to map an image pair directly to a disparity map using a deep learning network.

CNN-based algorithms use a convolutional neural network for geometry learning by making full use of the contextual information in stereo images. It is crucial to make utmost use of environmental contextual information, which includes local and global features, for further preserving subtle details.

GC-Net [14] is presented as an end-to-end supervised learning model that expresses the correspondences between stereo image patches by constructing a cost volume. To fully explore global contextual information, multiscale 3D convolution and 3D deconvolution are used to cost volume regularization. PSM-Net [15] introduces the Spatial Pyramid Pooling (SPP) [16] module, which extends pixel-level features to regional-level features of different scales, combing global and local feature cues to form a reliable cost volume [15, 16]. In addition, to gather more contextual features, Song et al. [17] propose a multitask network named EdgeStereo, which is composed of backbone subnetworks and edge subnetworks, to integrate edge cues by embedding edge perception smoothness loss regularization to improve matching performance.

2.2. Atrous Convolution. Atrous convolution is a method to expand the receptive field, known as "dilated convolution" due to contain dilation filters [18]. The receptive field can be understood as the size of the receptive range of neurons in the network to the original image, and it also refers to the size of the pixel point on the original image of the output feature map of each layer. Another pooling method can also extend the receptive field through downsampling. Downsampling has to unavoidably downsample the resolution of the image. When performing a pooling operation each time, the spatial resolution will be sacrificed. Thus, downsampling causes the loss of information and affects the performance of the disparity regression. An atrous kernel can be dilated by inserting zeros into suitable positions. In other words, atrous convolution has the ability to implement a larger receptive field size without sacrificing spatial resolution. Compared with a traditional convolution operator, the feature map generated by atrous convolution has the same size of its input while representing features from a larger receptive field, which means that higher-level semantics can be encoded.

In contrast, ordinary convolution can obtain a larger receptive field size by choosing a large size convolution kernel, connecting several convolutional layers with small size or other methods. However, these approaches will increase the number of parameters and calculation cost, which cannot guarantee real-time processing performance [19]. To this end, atrous convolution is developed to gather feature information. It inserts "holes" (that is, zero) into the encoding part of the convolution kernel, which can obtain a larger receptive field size without increasing the number of kernel parameters. This approach ensures that it can gain more accurate predictions while keeping the same computation cost. However, the atrous convolution framework has an inherent problem, which is the gridding issue, which incurs information loss for feature extraction. The architecture is shown in Figure 1.

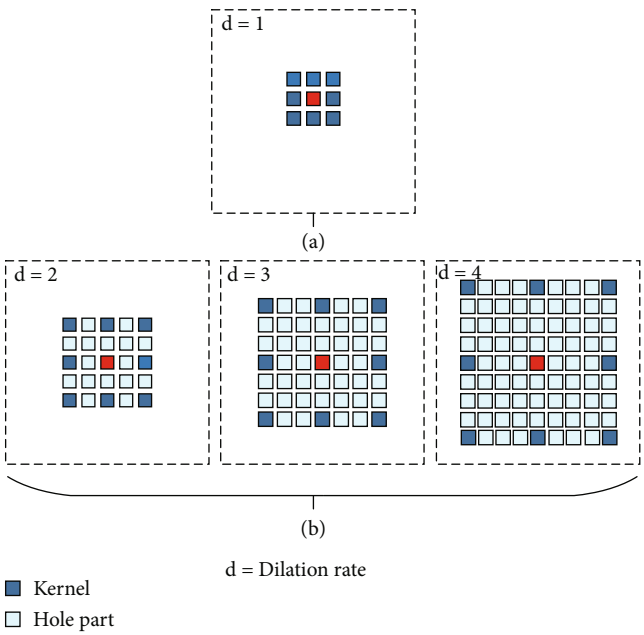


Figure 1: The comparison of ordinary convolution and hole convolution: (a) schematic diagram of ordinary convolution; (b) schematic diagram of atrous convolution.

The receptive field size of a convolution kernel can be formulated as the following equation:

$$R = (d - 1) \times (K - 1) + K, \tag{1}$$

where d represents the dilation rate and K represents the size of a kernel. R is the size of the receptive field.

2.3. Cost Volume. In stereo matching tasks, the cost volume performs matching cost calculation, whose purpose is to measure the correlation between the pixels to be matched and the candidate pixels [20–23]. Whether two pixels are homonymy points, the matching cost can be calculated by the matching cost function. The cost is smaller, inversely represents greater correlation, and also implies that the probability of being homonymy points is greater. Thus, cost volume is also equivalent to a space similarity measurement.

As shown in Figure 2, a standard cost volume usually consists of $W \times H \times D \times F$, which contains four dimensions, where $\mathrm{W} \times \mathrm{H}$ is regarded as the spatial resolution, D is the number of disparity planes, and F represents the number of channels of the feature map. A four-dimension cost volume retains feature dimensions and integrates it into the cost volume.

3. Related Work

3.1. Atrous Convolution-Based Methods. In the stereo matching task, contextual information refers to the relationship between an object and its surroundings or subregions [24], for instance, the relationship between a vehicle and its subareas (windshield, door). The size of the receptive field

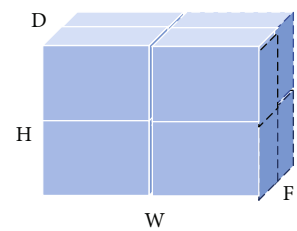


Figure 2: Standard cost volume.

indicates where the contextual information comes from. The atrous convolutional layer is typically used to solve problems related to semantic segmentation, which resolves the contradiction between resolution and the receptive field [25, 26]. In our paper, we apply it to an end-to-end stereo matching network for implementing multiscale information feature extraction.

Although atrous convolution solves the contradiction between feature map resolution and a larger receptive field [27–30], it has unicity when generating semantic information from the feature map. Specifically, all neurons in the feature map are generated by the atrous convolution with the same receptive field, which implies that only a single-scale feature is applied to complete the semantic information generation process. ASPP [13] uses atrous convolution that increases the dilation rate layer by layer to extract features and organizes the atrous convolutional layers in a parallel manner to obtain multiscale information. Multiscale information is a significant factor in segmentation of varying scales and fuzzy pixels; it requires gaining diverse ranges of contextual information to provide richer information for subsequent cost aggregation. However, when the dilation rate increases to a certain extent, the number of effective filtering parameters decreases gradually. In the extreme, when the sizes of the receptive field and the feature map are the same, the convolution does not capture the contextual information of the entire image but degenerates into a $1 * 1$ convolution, gradually losing its modeling ability.

Thus, to design a network architecture that encodes multiscale information while keeping a large receptive field sufficiently, Yang et al. [29] propose a method called DenseASPP that stacks atrous convolutional layers in parallel and in cascade. The cascade structure is mainly composed of multilevel atrous convolutions with a gradually increasing dilation rate. The parallel structure deals the same input feature with a sequential of atrous convolutional layers with various dilation rates, followed by connecting the results together and obtaining an output feature that is the input multiscale sampling feature. In DenseASPP, each atrous convolution makes full use of an atrous filter with a reasonable dilation rate. Through a series of atrous convolutions, the neurons in the later layers possess a larger receptive field without encountering the issue of kernel degradation. Extracting features by the cascade method implements a very large coverage and denser features, improving the recognition ability of the algorithm when the target ratio changes, so as to improve the robustness of matching.

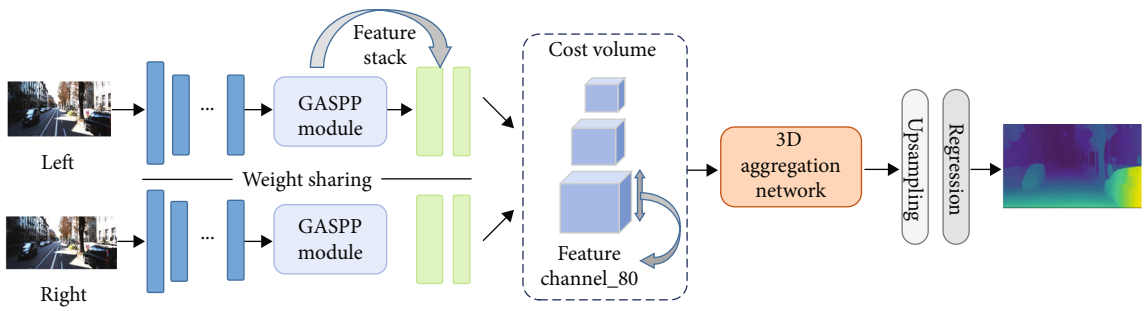


Figure 3: The architecture of proposed group-based atrous convolution stereo network.

3.2. Cost Volume Optimization. GC-Net [14] firstly uses unary features to form the cost volume to compute the stereo matching cost. This allows the network to learn incorporating context which can operate over features. But the large space in the cost volume also leads huge computation cost, which requires more active research effort to deal with.

Cascade cost volume [31] is constructed based on the idea of a feature pyramid to realize disparity estimation from coarse to fine level. A lower resolution disparity map is constructed by a smaller cost volume to complete the first estimate. In the following stage, it can reduce the disparity hypothesis range of the current scale and adjust the disparity map output by utilizing the previous stage, which can reduce the memory usage. GWCNet [32] adopts a group-related strategy, and group multichannel features are mapped with the disparity channel to form a group-based cost volume, which is comprised by concatenated volume and group-related volume. The concatenated volume is the same as that of PSM-Net but contains fewer channels, alleviating the need for parameters. Based on GC-Net [14], Lu et al. [33] propose the Sparse Cost Volume (SCV) to introduce the concept of step size (the distance that right feature moves) in the process of gathering the features of the image pair to form the cost volume to compress adjacent pixels to the center pixel to reduce redundant calculations. The Convolutional Attention Residual Network (CAR-Net) [34] combines residual network and attention mechanism to extract features, with simplifying network parameters while retaining the semantic information of the feature map.

Although methods based on cost volume remarkably boost the performance, they still rely on interpolation operation and downsampled cost volumes, which need high GPU storage. Our tailored cascade volume modifies feature channels to improve real-time performance and GPU memory efficiency.

4. Group-Based Atrous Convolution Stereo Network

4.1. Motivation. In semantic image segmentation, atrous convolutions can realize large receptive fields and meet the requirements of multiscale feature information through various dilation operations, which makes an effective method to deal with the challenging scale changing problem of objects. However, when the dilation rate increases, the spacing between the atrous convolution kernels also increases and prompts some local information loss, which can be understood as a gridding problem. Atrous convolution becomes more and more ineffective with increased dilation rate, and the number of parameters becomes higher. Consequently, these issues can influence the segmentation accuracy. Thus, we expect to design a method that enables large receptive fields utilizing small dilation convolutional layers and is more suitable for the stereo matching model. Based on the above ideas, we present group-based atrous convolutional layers to solve the contradiction between the receptive field and the gridding problem.

Moreover, the stereo matching model requires high real-time performance. At present, the deep learning model occupies a lot of memory and necessarily takes a long time to train; this obeys the real-time thought. Therefore, we introduce a tailored cascading cost volume. On the one hand, the structure is designed as a pyramid to match the previous GASPP. On the other hand, it enables reducing the dependence on batch size and the consumption of computer memory to enhance real-time performance.

4.2. Network Architecture. We present an end-to-end group-based atrous convolution stereo network. The architecture is shown in Figure 3. Our stereo matching model is composed of four parts: feature extraction, cost volume, cost aggregation, and disparity regression.

In the feature extraction part, a group-based atrous spatial pyramid module is introduced to generate features to reduce the loss of local information. The cost volume is tailored on the basis of the cascaded cost volume, and we adjust the number of channels to reduce the memory usage and accelerate computing speed, so as to cooperate with the group-based atrous pyramid module. Cost aggregation uses a three-dimensional stacked hourglass network; at last, we use soft-argmin function [13] to complete the disparity regression.

4.3. Group-based Atrous Convolution Spatial Pyramid Pooling Module. A group-based atrous pyramid structure is proposed for feature extraction. Feature extraction is the prerequisite first step for correct disparity estimation. In PSM-Net, the SPP module is used to extend pixel-level features to regional-level features with different scales, then utilizes global contextual information for stereo matching [14].

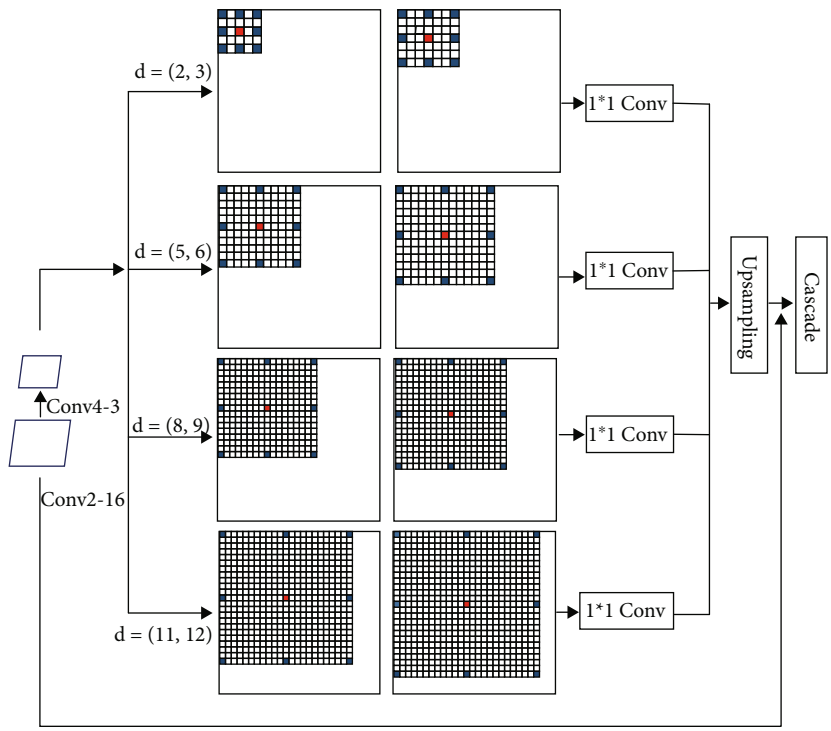


Figure 4: The structure of GASPP.

It is crucial to acquire contextual information for stereo matching, and the amount of contextual information is closely related to the size of its receptive field.

In the atrous convolution kernel, zero is filled between two pixels, so the receptive field only covers a gridding region, which implies that only nonzero positions are sampled and some adjacent information is lost. When the dilation rate increases, the gridding problem becomes more serious. For instance, ASPP uses an atrous convolutional layer with dilation rates of 6, 12, 18, and 24; the interval is 6, which can implement the purpose of extracting multiscale information. However, since the kernel of atrous convolution is not continuous, its calculation formula is similar to the gridding format as shown in Figure 1. As the dilation rate increases, the receptive field becomes larger and the distance between the kernels also increases. Then, there is less correlation between the obtained information, and more local information will be lost, which will affect the performance of the disparity regression later. DenseASPP obtains wider and denser feature information by cascading, but it still has a gridding problem.

To incorporate large context and compute feature maps more densely, whilst reducing the loss of local information and improving the accuracy of the disparity map, we propose the Group-based Atrous Convolution Spatial Pyramid Pooling (GASPP). The model is shown in Figure 4. The basic idea is to substitute an original atrous convolution with a large dilation rate with two continuous atrous convolutions which have a smaller dilation rate. The purpose is to use a smaller dilation rate but maintain the size of receptive field at the same time. In other words, the correlation between the information gets strengthened and the hole area smaller. Consequently, it improves the disparity regression performance in those ill-posed areas, e.g., occluded regions. We design four groups of atrous convolutional layers in the GASPP module and assign two atrous convolutional layers with continuous increased dilation rate to each group. The designed dilation rate between each group is gradually increased, forming{(2,3) (5,6) (8,9) (11,12)} four parallel branches, which can provide feature maps of spatial information with different scales, and the four branches complement each other. Ultimately, the output is accumulated to obtain a feature map that contains multiscale information, with wider and denser receptive fields.

To simplify notations, we use $G^n_{K,d}(x)$ to represent a group-based atrous convolution and consequently write GASPP y as the following equation:

$$y = \begin{cases} G^1_{3,2}(x) + G^1_{3,3}(x) + G^2_{3,5}(x) + G^2_{3,6}(x), \\ +G^3_{3,8}(x) + G^3_{3,9}(x) + G^4_{3,11}(x) + G^4_{3,12}, \end{cases} \tag{2}$$

where d is the dilation rate, K denotes the kernel size, and n is the group number. The receptive field size of GASPP can be formulated as the following equation:

$$\begin{cases} d_2 = d_1 + 1, \\ R_g = (2d_1 - 1) \times (K - 1) + 2K, \end{cases} \tag{3}$$

where d_1 and d_2 represent the size of a kernel of a group. Otherwise, R_g contributes to the size of the group-based atrous convolution receptive field.

In this manner, if only one atrous convolution with a dilation rate of 6 and kernel of 3 is used, the size of the receptive field is $5 * 2 + 3 = 13$, while if two atrous convolutions, one with a dilation rate of 3 and kernel of 3 and another with a dilation rate of 4 and kernel of 3, are used, the size of the receptive field is $(2 * 2 + 3) + (3 * 2 + 3) = 16$. Otherwise, if we chose two atrous convolutions, one with a dilation rate of 2 and kernel of 3 and another with a dilation rate of 3 and kernel of 3, the size of the receptive field is $(1 * 2 + 3) + (2 * 2 + 3) = 12$. Ultimately, we choose the $(2, 3)$ combination that has the most similar receptive field with dilation rate of 6. This same setting is applied to all other GASSP groups.

After feature extraction, we take some measures to make the cost volume be appropriate for the GASPP model. In the following section, we will show the formulation about the tailored cascade cost volume.

4.4. Tailored Cascade Cost Volume. Gu et al. [31] introduce the cascade cost volume. The algorithm is based on the idea of the feature pyramid, which processes the cost volume in stages by utilizing a gradually refined scale (forecast output from coarse to fine). The formation of the cost volume mainly goes through three steps: firstly, the space of the cost volume is determined, which is the disparity plane; secondly, what to do is to warp the features extracted from the stereo image pair to the disparity plane to construct feature volume; ultimately, the feature volume is fused to build the cost volume.

The cascaded cost volume is divided into two stages in the stereo matching task. For the processing of the feature volume, the initial number of channels is changed from 32 to 320 through a convolutional layer, which is consistent with channels for feature extraction. We utilize two 2D convolutional layers to tailor the cascade cost volume. One 2D convolutional layer is used to reduce the number of channels to 160, and another one is used to cut down the number of feature channels to 80 as shown in Figure 3. In this way, the characteristic channels are {1/2, 1/4}, respectively, and the number of final channels is fewer, which occupy less space during training process. This design is similar to a layered structure, which reduces layer by layer, losing less information and ensuring the performance and effect of the network.

4.5. Output Module and Loss Function. In the output module, we use two 3D convolutions to generate a 4D volume with 1 channel, followed by the upsampling operation and transforming it into a probability volume with a softmax function along with a disparity dimension. The probability of each pixel is calculated from the predicted cost through softmax operation $\sigma(\bullet)$. Due to the higher the cost, the probability of matching is lower, so we take prediction negative cost $-c_d$. At last, the predicted disparity $\bar{d}$ is obtained by calculating the weighted probability sum of each disparity d; the disparity regression function is defined as the following equation:

$$\bar{d} = \sum_{d=0}^{\max} d \times \sigma(-c_d), \tag{4}$$

where d and $\sigma(-c_d)$ denote a possible level and the corresponding probability. The final loss is given by the following equation:

$$L = \sum_{k=0}^{3} \lambda_k \bullet \mathrm{Smooth}_{L_1}\left(\bar{d}_k - \hat{d}\right), \tag{5}$$

where λ_k denotes the coefficient of the k_{th} disparity prediction and $\hat{d}$ denotes ground-truth disparity maps. The predicted disparity map in the last four output modules are represented by $\bar{d}_0$, $\bar{d}_1$, $\bar{d}_2$, and $\bar{d}_3$. The Smooth_{L_1} loss function is defined as the following equation:

$$\mathrm{Smooth}_{L_1}(x) = \begin{cases} 0.5x^2, & \text{if } |x| < 1, \\ |x| - 0.5, & \text{otherwise.} \end{cases} \tag{6}$$

5. Experiment

This section introduces the experimental settings and results. We evaluate the key components of our model on the Scene Flow [35] and KITTI 2015 [36] datasets. In addition, we also compare our method with advanced stereo matching methods on the KITTI benchmark.

5.1. Experiment Details

5.1.1. Experiment Environment Settings. The end-to-end network is implemented in Windows environment and runs under the PyTorch deep learning framework. In terms of hardware facilities, we apply NVIDIA 1070Ti GPU to train our model, and the batch size is set to 2.

For all datasets, we set the resolution of the training stereo image pair to 512×256, the RGB values of all images are normalized in [-1, 1], and the initial maximum disparity $D_{\max}$ is set to192. The model is trained with the Adam optimizer, and several optimization parameters are set to $\beta_1 = 0.9$ and $\beta_2 = 0.99$. In addition, the four output coefficients are adjusted, respectively, as

$$\lambda_0 = 0.5, \lambda_1 = 0.5, \lambda_2 = 0.7, \lambda_3 = 1.0. \tag{7}$$

Commonly, end-to-end stereo networks are pretrained from scratch on the Scene Flow dataset, then further to optimize the network on a smaller target dataset such as KITTI 2015. The Scene Flow dataset is more effective when moving to KITTI. Therefore, for the Scene Flow dataset, we train 15 epochs with a fixed learning rate of 0.001. For the KITTI 2015 dataset, we employ the pretrained model on the Scene Flow for further optimization. The KITTI optimization training is set for 300 epochs. The learning rate for the first 200 epochs is 0.001, and that of the other 100 epochs is 0.0001.

5.1.2. Datasets

(1) Scene Flow. The dataset is a large synthetic dataset, which includes 35454 pairs of stereo images for training and 4370

Table 1: Experimental evaluation of different components of our network.

Network model							>3px	EPE	Run time (s)
The dilation of GASPP				Channel of cost volume					
(2,3)	(5,6)	(8,9)	(11,12)	40	80	160			
✓	✓						4.21	0.87	1.63
	✓	✓			✓		3.40	0.81	1.36
✓	✓	✓			✓		3.02	0.73	1.23
	✓	✓	✓	✓			3.21	0.75	1.32
	✓	✓	✓			✓	3.18	0.72	1.28
✓	✓	✓	✓	✓			3.12	0.71	1.26
✓	✓	✓	✓			✓	2.92	0.69	1.19
✓	✓	✓	✓		✓		**2.81**	**0.64**	**1.15**

The bold data represents the optimal data result under this column.

pairs of stereo images for testing. It also provides dense and exhaustive ground truth disparity maps and camera parameter information for each pair, in which the resolution of all images is 960×540.

The Scene Flow subset comprise three scenes. Among them, FlyingThings3D is a scene with random-type objects including extensive floating objects with rich details, the Driving dataset is a street scene captured in the process of simulating car driving, and Monkaa is a scene that includes monkeys in a deep forest environment, which involves closer objects and means more areas with larger disparity values.

(2) KITTI 2015. This is a dataset collected from a real street scene, including 200 pairs of stereo images for training and 200 pairs of stereo images for testing. The dataset provides a sparse disparity map collected by LiDAR as the ground truth value.

For the Scene Flow dataset, we select End-Point-Error (EPE) as the evaluation metric, which is defined as the mean average disparity error in pixels, calculating the Euclidean distance between the prediction error of each pixel and true error, followed by taking the average value. The rate between error and the accuracy of matching is the inverse proportion, so when the error is smaller, the matching accuracy is inversely higher. The calculation method of EPE is defined as follows:

$$\mathrm{EPE} = \frac{1}{N} \sum_{m \in N} \sqrt{(d_m - d\wedge_m)^2}, \tag{8}$$

where N denotes the total number of pixels, d_m represents the ground truth at the m_{th} pixel, and $\widehat{d}_m$ represents the predicted disparity value at the m_{th} pixel.

For the KITTI 2015 dataset, we apply a 3 px error as the predicted error, which differs from the label by a threshold of three pixels. In other words, the 3 px error refers to the ratio of the number of pixels whose absolute value between the predicted disparity value and the ground truth value exceeds 3 to the entire image.

5.2. Ablation Experiment. In this section, to understand the impact of each model components on the final performance, we conducted a comprehensive ablation study by adding each component individually and showed how the components make an impact on the performance. The overall comparative experimental results of our network are shown in Tables 1 and 2. Moreover, we compare our proposed method with one of the most advanced stereo matching methods, PSM-Net, to evaluate our network jointly on the Scene Flow and KITTI 2015 datasets.

5.2.1. Ablation Experiment for GASPP Module. We first conduct ablation experiments on the GASPP structure. GASPP is a group of two consecutive convolutional layers with a continuous dilation rate, which consists of four groups. This part mainly compares the pooling module, GASPP module, original SPP module, ASPP module, and DenseASPP module.

Combining the experimental results in evaluation of different components of our network in Tables 1 and 3 indicated that the GASPP module improves the effect and performance of disparity estimation to a certain extent. The application of the group-based atrous convolutional layer reduces EPE from 0.62 to 0.57. The analysis shows that for the complex KITTI 2015 dataset, the SPP module loses more spatial information in the process of pooling and upsampling, resulting in slightly worse effect of the network, while GASPP can collect contextual information more closely and preserve the salient local information.

Inspired by DenseASPP, we propose the GASPP module based on the idea of grouping. Among them, the size of the dilation rate within the group and the interval of the dilation rate between each group need to be manually set. In order to achieve the optimization, the study conducted multigroup experiments on the dilation rate and interval between groups of the GASPP module, setting a total of 7 sets of parameters. The interval between the first three groups is 1, and the initial dilation rate is different. The last four groups are set with different numbers of interval, and the dilation rate of the initial group is the same for testing. The experimental results are shown in Table 2. It can be seen from the results in Table 2 that the best result can be obtained when the interval is 2.

Table 2: Evaluation of experimental effects of GASPP module with various dilation rates.

Serial number	Interval of dilation rate	Dilation group				>3px	EPE	Run time (s)
		Group 1	Group 2	Group 3	Group 4			
1	1	(2.3)	(4.5)	(6.7)	(8.9)	3.22	0.83	1.23
2	1	(3.4)	(5.6)	(7.8)	(9.10)	3.15	0.87	1.34
3	1	(5.6)	(7.8)	(9.10)	(10.11)	3.21	0.85	1.28
4	2	(2.3)	(5.6)	(8.9)	(11.12)	**2.91**	**0.68**	**1.15**
5	3	(2.3)	(6.7)	(10.11)	(14.15)	3.02	0.70	1.20
6	4	(2.3)	(7.8)	(13.14)	(18.19)	3.11	0.72	1.25
7	5	(2.3)	(8.9)	(14.15)	(20.21)	3.18	0.85	1.28

The bold data represents the optimal data result under this column.

Table 3: Experimental evaluation of GASPP module.

Pooling	EPE	>3px
SPP	0.62	2.62
ASPP	0.60	2.63
DenseASPP	0.59	**2.60**
GASPP	**0.57**	2.61

The bold data represents the optimal data result under this column.

5.2.2. Ablation Experiment for the Tailored Cost Volume. In this paper, we add the tailored cost volume to our model and explore to limit the computational cost by adding a 2D convolutional layer at the end to design the number of feature channels. We provide a comparison between the results of using the 2D convolution layers for the proposed tailored cascade cost volume and the original cascade cost volume in the first row in Table 1. In order to explore the effect of different feature channel numbers on the cascaded cost volume, we set up three groups of feature numbers for verification, which are 40, 80, and 160. The experimental results are shown in Figures 5 and 6.

Figure 5 shows the comparison of the number of network parameters and the run time between PSM-Net and the cascade cost volume of GASPP with different feature channels. The number of network parameters of PSM-Net, Cascade-40, and Cascade-160 decreases by 36.2%, 9.12%, and 3.21%, respectively. Figure 6 displays the correspondence between the number of iterations and the 3 px error. Cascade-80 of GASPP implements the minimum error and the fastest convergence speed at the same time.

The combined relationship is depicted in Figures 5 and 6; with the increase of the number of feature channels, the accuracy of stereo matching is improved. From Cascade-40 to Cascade-160, the number of feature channels has quadrupled, but the error is increased instead. Therefore, the increase of feature channels leads to expensive computational cost but insignificant performance improvement. From the results, one may conclude that the Cascade-80 is the most suitable channel for our network model.

5.2.3. Benchmark Performance on KITTI 2015 Dataset. In this section, we conduct a detailed study of Figure 7 disparity maps trained by PSM-Net and the cascade cost volume and contrasted them. We use a rectangular box to mark the areas where our network has different matching effects with PSM-

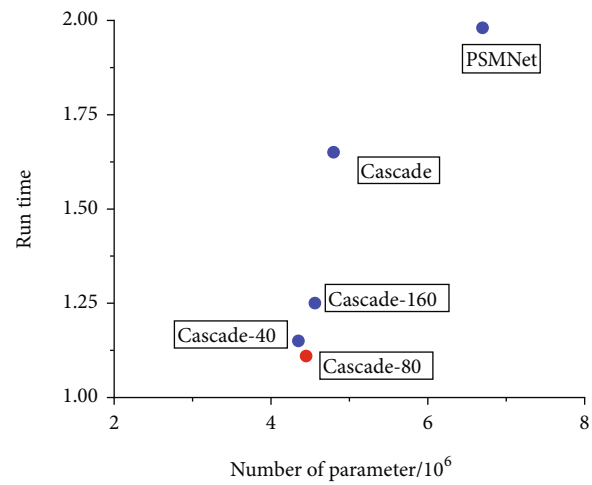


Figure 5: Correspondence between the number of network parameters and run time.

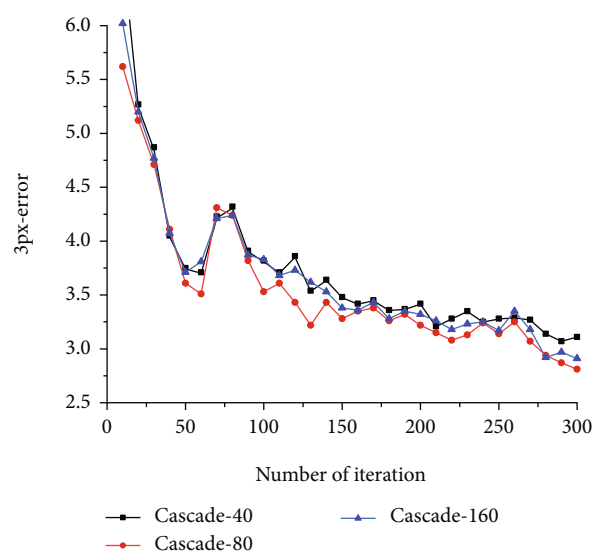


Figure 6: Correspondence between the number of iterations and 3 px error.

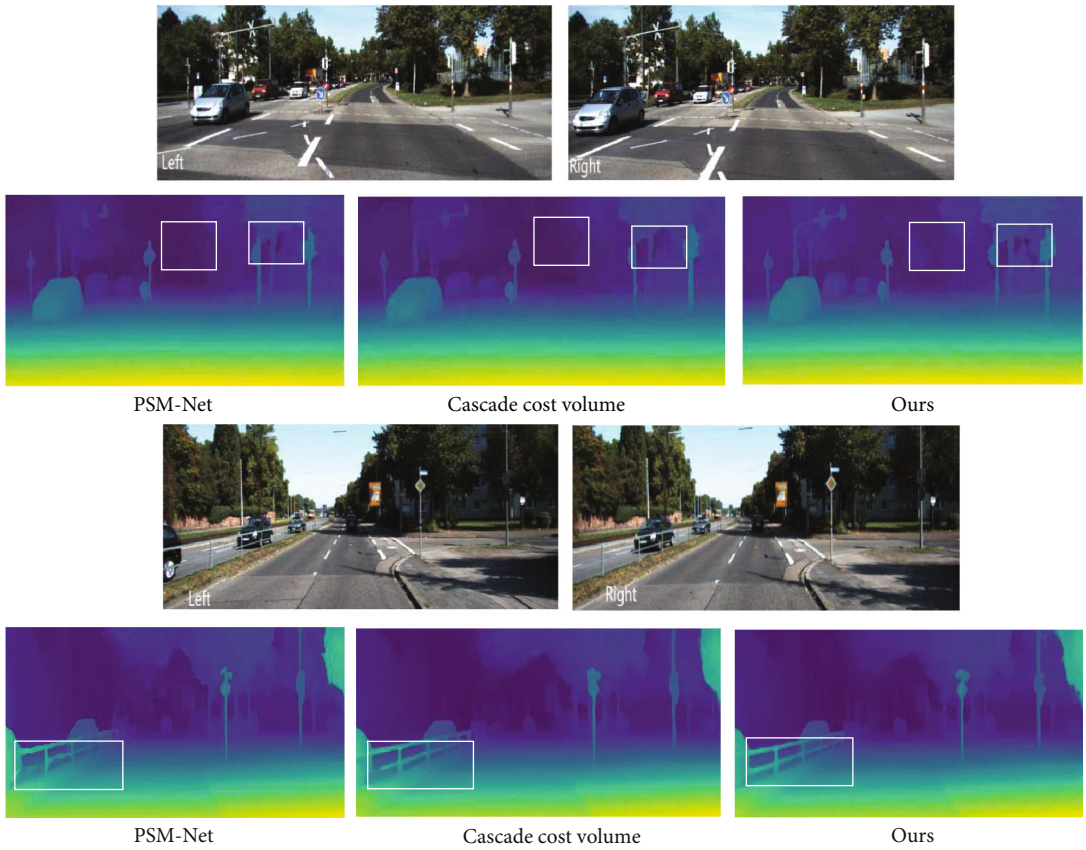


Figure 7: Results of disparity estimation for KITTI 2015 images. The top panel shows the input stereo image pair. For each input image, the disparity maps were obtained by PSM-Net, cascade cost volume, and our network.

Table 4: KITTI 2015 stereo matching network test results.

Model	EPE	>3px	Run time (s)
PSM-Net	0.88	**2.80**	1.53
GWCNet	0.74	3.20	1.43
Cascade cost volume	0.72	3.04	1.81
GC-Net	1.02	2.97	2.53
Ours	**0.68**	2.91	**1.06**

The bold data represents the optimal data result under this column.

Net and the Cascade Network. The matching results are reported by the KITTI evaluation server, and the best result can be obtained when the interval is 2. The GASPP network realizes robust results with less running time. It implies that in practical application such as mobile robots, our network will display fine real-time performance. Table 4 shows the comparison between our network and other stereo matching networks on the KITTI 2015 datasets.

6. Conclusions

This paper presents a group-based atrous convolution pyramid pooling module, which uses a densely atrous convolution to form multiscale receptive fields, reduce the loss of local information, and improve the matching accuracy. The dilation rate of GASPP is continuous to lower the hole part generated in an original atrous convolution operation. The dilation rate between groups is gradually increased to ensure the aggregation of multiscale contextual information to improve the accuracy of disparity estimation. Furthermore, we propose a tailored cascaded cost volume and add a 2D convolution to form a layered structure, to reduce memory usage and speed up the execution efficiency of our network. The experimental results demonstrate the effectiveness of the proposed method in terms of both disparity estimation accuracy and running efficiency. In our future work, as most of the end-to-end algorithms rely on datasets for learning, training, and verification, without using the images in real life, we would further investigate our test on data collected from real-world applications. To date, weak-supervised and multitask fusion networks have developed rapidly. Weak-supervised networks do not depend on large-scale datasets with ground truth values, which makes network training easier. The multitask fusion network is extraordinarily helpful for matching accuracy in small or ill-posed areas, which will be the focus on follow-up research.

Data Availability

The data used to support the findings of this study are available from the authors upon request.

Conflicts of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the Dalian University Research Platform Project Funding: Dalian Wise Information Technology of Med and Health Key Laboratory, and the National Natural Science Foundation of China (No. 11701061): Research on SA Algorithm for Nonconvex Stochastic Semidefinite Programming.

References

[1] J. D. S. Solak and E. D. Bolat, "A new hybrid stereovision-based distance-estimation approach for mobile robot platforms," *Computers & Electrical Engineering*, vol. 67, pp. 672–689, 2018.

[2] L. I. Xiu-Juan, W. Liu, and L. I. Shan-Hong, "Robust control algorithm of bionic robot based on binocular vision navigation," *Computer Science*, vol. 21, no. 44, pp. 318–322, 2017.

[3] T. Trzcinski, M. Christoudias, P. Fua, and V. Lepetit, "Boosting binary keypoint descriptors," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, CVPR2013 Portland, American, 2013.

[4] Z. Kun, M. Xiangxi, and B. Cheng, "Review of stereo matching algorithms based on deep learning," *Computational Intelligence and Neuroscience*, vol. 2020, 12 pages, 2020.

[5] H. Wang, M. Wu, Y. Zhang, and L. Zhang, "Effective stereo matching using reliable points based graph cut," in *2013 Visual Communications and Image Processing (VCIP)*, pp. 1–6, Kuching, Malaysia, 2013.

[6] A. Klaus, M. Sormann, and K. Karner, "Segment-based stereo matching using belief propagation and a self-adapting dissimilarity measure," in *18th International Conference on Pattern Recognition (ICPR'06)*, pp. 15–18, Hong Kong, China, 2006.

[7] Y. Ji, Y. Li, X. Sun, S. Yan, and N. Guo, "Stereo matching algorithm based on binocular vision," in *2020 7th International Forum on Electrical Engineering and Automation (IFEEA)*, pp. 843–847, Hefei, China, 2020.

[8] S. Mohammad and T. Morris, "Binary robust independent elementary feature features for texture segmentation," *Advanced Science Letters*, vol. 23, no. 6, pp. 5178–5182, 2017.

[9] H. Hirschmuller and D. Scharstein, "Evaluation of cost functions for stereo matching," in *2007 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–8, Minneapolis, MN, USA, 2007.

[10] K.-J. Yoon and I.-S. Kweon, "Locally adaptive support-weight approach for visual correspondence search," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol. 2, pp. 924–931, San Diego, CA, USA, 2005.

[11] J. Yang, D. Xing, Z. Hu, and T. Yao, "A two-branch network with pyramid-based local and spatial attention global feature learning for vehicle re-identification," *CAAI Transactions on Intelligence Technology*, vol. 6, no. 1, pp. 46–54, 2021.

[12] J. Žbontar and Y. LeCun, "Computing the stereo matching cost with a convolutional neural network," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1592–1599, CVPR 2015 Boston, MA, USA, 2015.

[13] L. 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. 40, no. 4, pp. 834–848, 2018.

[14] A. Kendall, H. Martirosyan, S. Dasgupta et al., "End-to-end learning of geometry and context for deep stereo regression," in *2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 66–75, ICCV2017 Venice, Italy, 2017.

[15] J. Chang and Y. Chen, "Pyramid stereo matching network," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5410–5418, CVPR2018 Salt Lake City, UT, USA, 2018.

[16] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 9, pp. 1904–1916, 2015.

[17] X. Song, X. Zhao, L. Fang, H. Hu, and Y. Yu, "EdgeStereo: an effective multi-task learning network for stereo matching and edge detection," *International Journal of Computer Vision*, vol. 128, no. 4, pp. 910–930, 2020.

[18] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," 2016, https://arxiv.org/abs/1511.07122.

[19] H. Dai, X. Zhang, Y. Zhao, H. Sun, and N. Zheng, "Adaptive disparity candidates prediction network for efficient real-time stereo matching," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, p. 1, 2021.

[20] J. Pang, W. Sun, J. S. Ren, C. Yang, and Q. Yan, "Cascade residual learning: a two-stage convolutional neural network for stereo matching," in *2017 IEEE International Conference on Computer Vision Workshops (ICCVW)*, pp. 878–886, ICCVW2017 Venice, Italy, 2017.

[21] F. Zhang, V. Prisacariu, R. Yang, and P. H. S. Torr, "GA-Net: guided aggregation net for end-to-end stereo matching," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 185–194, CVPR2019 Long Beach, CA, USA, 2019.

[22] X. Jia, W. Chen, and Z. Liang, "Bidirectional stereo matching network with double cost volumes," *IEEE Access*, vol. 9, pp. 19651–19658, 2021.

[23] Z. Liang, Y. Guo, Y. Feng et al., "Stereo matching using multi-level cost volume and multi-scale feature constancy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 1, pp. 300–315, 2021.

[24] G. Y. Nie, M. M. Cheng, Y. Liu et al., "Multi-level context ultra-aggregation for stereo matching," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3278–3286, CVPR2019 Long Beach, CA, USA, 2019.

[25] G. Yang, J. Manela, M. Happold, and D. Ramanan, "Hierarchical deep stereo matching on high-resolution images," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5510–5519, CVPR2019 Long Beach, CA, USA, 2019.

[26] L. C. Chen, G. Papandreou, F. Schroff, and H. Adam, *Rethinking atrous convolution for semantic image segmentation*, 2018.

[27] Q. Du, R. Liu, Y. Pan, S. Sun, S. Sun, and Z. Zheng, "Depth estimation with multi-resolution stereo matching," in *2019 IEEE Visual Communications and Image Processing (VCIP)*, pp. 1–4, Sydney, NSW, Australia, 2019.

[28] P. Wang, P. Chen, Y. Yuan et al., "Understanding convolution for semantic segmentation," in *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 1451–1460, Lake Tahoe, NV, USA, 2018.

[29] M. Yang, K. Yu, C. Zhang, Z. Li, and K. Yang, "DenseASPP for semantic segmentation in street scenes," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 3684–3692, CVPR2018 Salt Lake City, UT, USA, 2018.

[30] E. Shelhamer, J. Long, and T. Darrell, "Fully convolutional networks for semantic segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 4, pp. 640–651, 2017.

[31] X. Gu, Z. Fan, S. Zhu, Z. Dai, F. Tan, and P. Tan, "Cascade cost volume for high-resolution multi-view stereo and stereo matching," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2492–2501, CVPR2020 Seattle, WA, USA, 2020.

[32] X. Guo, K. Yang, W. Yang, X. Wang, and H. Li, "Group-wise correlation stereo network," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3268–3277, CVPR2019 Long Beach, CA, USA, 2019.

[33] C. Lu, H. Uchiyama, D. Thomas, A. Shimada, and R.-i. Taniguchi, "Sparse cost volume for efficient stereo matching," *Remote Sensing*, vol. 10, no. 11, p. 1844, 2018.

[34] G. Huang, Y. Gong, Q. Xu, K. Wattanachote, K. Zeng, and X. Luo, "A convolutional attention residual network for stereo matching," *IEEE Access*, vol. 8, pp. 50828–50842, 2020.

[35] N. Mayer, E. Ilg, P. Hausser et al., "A large dataset to train convolutional networks for disparity, optical flow, and scene flow estimation," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4040–4048, CVPR2016 Las Vegas, NV, USA, 2016.

[36] M. Menze, C. Heipke, and A. Geiger, "Joint 3d estimation of vehicles and scene flow," *ISPRS annals of the photogrammetry, remote sensing and spatial information sciences*, vol. II-3/W5, pp. 427–434, 2015.

Hindawi
Wireless Communications and Mobile Computing
Volume 2023, Article ID 9873492, 1 page
https://doi.org/10.1155/2023/9873492

Retraction

Retracted: Featureless Blood Pressure Estimation Based on Photoplethysmography Signal Using CNN and BiLSTM for IoT Devices

Wireless Communications and Mobile Computing

Received 10 October 2023; Accepted 10 October 2023; Published 11 October 2023

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. Li, L. N. Harfiya, and C. Chang, "Featureless Blood Pressure Estimation Based on Photoplethysmography Signal Using CNN and BiLSTM for IoT Devices," *Wireless Communications and Mobile Computing*, vol. 2021, Article ID 9085100, 10 pages, 2021.

Hindawi
Wireless Communications and Mobile Computing
Volume 2021, Article ID 9085100, 10 pages
https://doi.org/10.1155/2021/9085100

Research Article

Featureless Blood Pressure Estimation Based on Photoplethysmography Signal Using CNN and BiLSTM for IoT Devices

Yung-Hui Li,[1] **Latifa Nabila Harfiya**,[2] **and Ching-Chun Chang**[3]

[1]*AI Research Center, Hon Hai Research Institute, Taipei 114699, Taiwan*
[2]*Department of Computer Science and Information Engineering, National Central University, Taoyuan 32001, Taiwan*
[3]*Department of Computer Science, University of Warwick, Coventry CV47AL, UK*

Correspondence should be addressed to Latifa Nabila Harfiya; latifanaharfiya@gmail.com

Received 8 October 2021; Accepted 8 November 2021; Published 26 November 2021

Academic Editor: Chi-Hua Chen

Continuous blood pressure (BP) acquisition is critical to health monitoring of an individual. Photoplethysmography (PPG) is one of the most popular technologies in the last decade used for measuring blood pressure noninvasively. Several approaches have been carried out in various ways to utilize features extracted from PPG. In this study, we develop a continuous systolic and diastolic blood pressure (SBP and DBP) estimation mechanism without the need for any feature engineering. The raw PPG signal only got preprocessed before being fed to our model which mainly consists of one-dimensional convolutional neural network (CNN) and bidirectional long short-term memory (BiLSTM) network. We evaluate the resulting SBP and DBP value by the root-mean-squared error (RMSE) and mean absolute error (MAE). This study addresses the effectiveness of the model by outperforming the previous feature engineering-based methods. We achieve RMSE of 11.503 and 6.525 for SBP and DBP, respectively, and MAE of 7.849 and 4.418 for SBP and DBP, respectively. The proposed method is expected to substantially enhance the current efficiency of healthcare IoT (Internet of Things) devices in BP monitoring using PPG signals only.

1. Introduction

Blood pressure (BP) is a biomarker that interprets how much tension the blood exerted on a blood vessel wall for every unit area. The more tension the blood imposes on the blood vessel, the higher the BP value is. The measurement of BP occurs in the arterial blood vessels adjacent to the heart. This measurement is a direct function of ventricular contractions. BP can be measured as a function of the resistance through the blood vessels and blood flow [1]. Thus, BP dynamically fluctuates in response to the changes in diameter of the blood vessel, vessel length, and the viscosity of the blood. Hence, as the blood volume in the vessels becomes greater, so does the BP. All these changes are the consequences of a perplexing interchange between the environmental, physical, and emotional factors. Accordingly, BP might vary depending on the daily, hourly, or even minutely challenges of each individual [2].

The temporal dimensions and patterns characterizing the BP variations define the term BP variability (BPV). From a clinical perspective, BPV could be seen as a source of noise that creates difficulties in assessing the individual's "true" BP value. Evidence is now available to support its role also as an independent predictor of cardiovascular risk. While the BPV increases, the possibility of pharmacological treatment's target becomes higher as well [2]. On that account, monitoring continuous BP is critical in order to capture the absolute BP value of an individual.

Currently, there is one machine that can do accurate continuous measurement called Finapres [3] which uses a noninvasive method based on a photoplethysmographic system. The growth of Internet of Things (IoT) and wearable devices applied in healthcare industry makes it much easier to measure physiological signal in a noninvasive fashion, and it is undoubtedly good news for most patients. As they can do routine self-health monitoring, it helps them to get

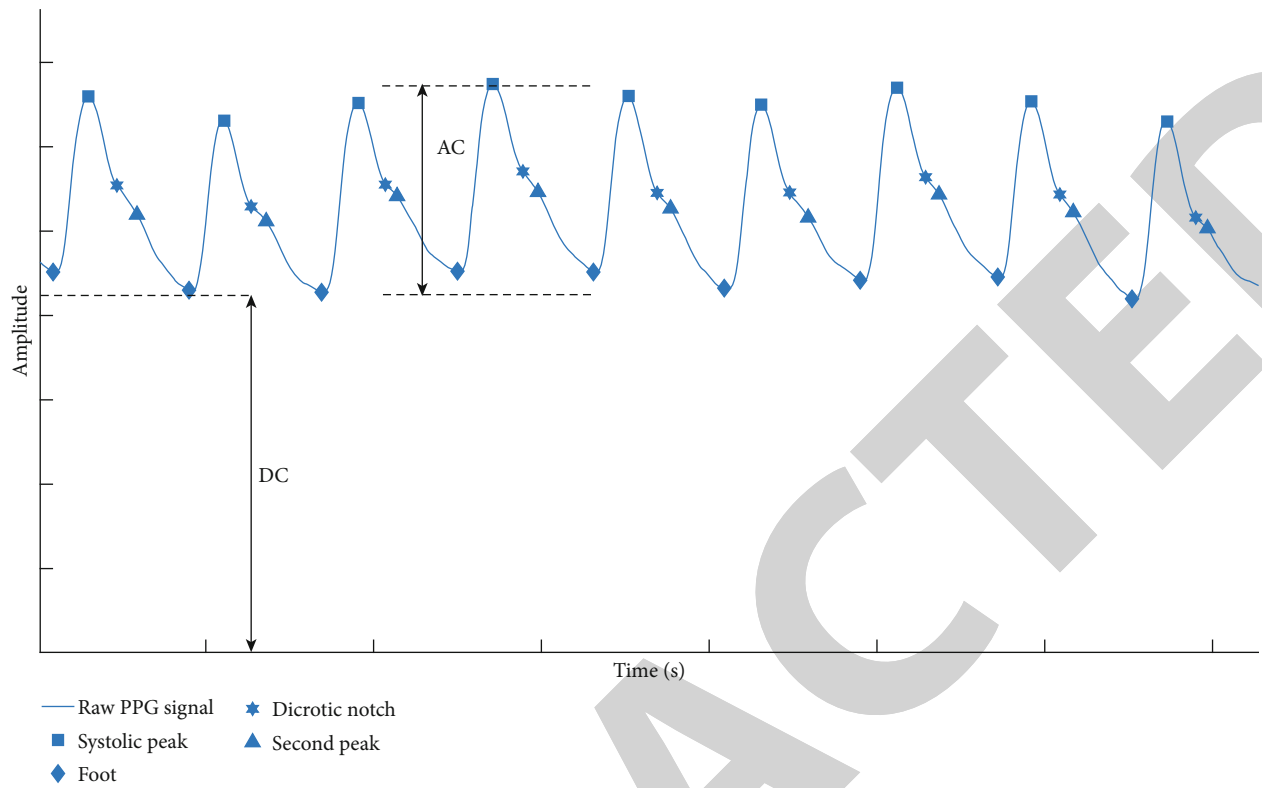


Figure 1: PPG signal's decomposition.

early warning about any abnormality for their health. In the last decade, numerous IoT-enabled wearable biosensor devices utilize Photoplethysmography (PPG) for monitoring the physiological conditions of a patient. In addition to the massive application in personal wearable IoT devices, PPG is also commonly applied in pulse oximetry due to its convenience and capacity to perform continuous readings [4]. Nevertheless, a PPG waveform discloses the cardiovascular and respiratory systems' activity of a patient in the corresponding time period [5].

A PPG waveform is designated with a pulsatile physiological waveform "AC." For every heartbeat, the cardiac synchronous alteration over the blood volume is reflected by this waveform. This waveform lays over a slowly varying baseline "DC" which contains lower frequency components. This part reflects the potential conditions related to respiration, thermoregulation, and skin tissue condition [6]. Blood is pumped by the heart to the periphery in each cardiac cycle. Amid the pressure that reaches the skin, arteries and arterioles are amplified in the subcutaneous tissue. The pressure pulse can be seen from the venous plexus upon a light reflex or disseminate detector device adhered to the skin as a secondary peak. On the other hand, the larger peak appears for each cardiac cycle which the blood volume alters due to the pressure pulse captured by illuminating the skin with a Light-Emitting Diode (LED) and photodetector, namely, photodiode, which measures how much is the transmitted or reflected light [7], as seen in Figure 1.

In Figure 1, parameters that are commonly utilized to generate features in PPG for BP estimation [7–13] are presented such as systolic peak, foot, dicrotic notch, and the second peak. Various approaches, namely, pulse transit time (PTT), pulse arrival time (PAT), and pulse wave velocity (PWV), are extracted using given parameters from two PPG sensors located on two distant sites. These parameters, however, may not always appear in the signal mostly due to the moving artifacts in the process of acquisition [14]. Automatic feature extraction from PPG signal is becoming a necessity since noises are hard to handle even with complex feature engineering [15]. Prior studies [13, 16] have successfully predicted BP using complex time series modelling such as long short-term memory (LSTM) network with a low error. These methods, however, tried to skip every defined range of signal with unhandled noise which is discontinuous in nature.

Herein, the purpose of this study is twofold. First, we develop a continuous BP estimation framework without plethoric concern about how to extract features. As the use of IoT devices for healthcare purposes provides benefits for people to monitor themselves, applying featureless framework for inferencing is expected to reliably lessen the response time and the computing cost. Second, we propose a robust deep learning model to do the automatic feature extraction as well as the BP estimation. Convolutional neural network (CNN) has been shown to be the state-of-the-art when it comes to automatic feature extraction while LSTM is an effective choice for analyzing time series data with an ability to handle long sequential data. PPG signal is obviously a one-dimensional signal which varies with time. This study, hence, will utilize the 1D CNN and bidirectional

LSTM (BiLSTM) network to train a deep learning model for the BP estimation. The output will be the estimated values of two types of BP, which are systolic blood pressure (SBP) and diastolic blood pressure (DBP). Related works using different methods are described in Section 2. We provide the detail of our proposed model in Section 3. We then present the result of the model's evaluation in Section 4 followed by the comparative analysis of the foregoing result and conclude it in Section 5.

2. Related Works

2.1. Feature-Based BP Estimation. BP is known to have a nonlinear relationship with PTT which is commonly obtained by measuring the time difference between the electrocardiogram (ECG) R peak and the maximum slope of the corresponding PPG signal [16]. Aside from PTT, various features extracted from PPG are found to be correlated with BP. The amplitude of systolic shown in Figure 1 indicates the pulsatile transformation in blood volume due to the arterial blood flow alongside the distal site. Moreover, the systolic amplitude is prompt to be a more appropriate parameter for BP estimation instead of PTT [17]. A number of features from PPG which have been proposed in [12, 14] are listed as follows (see Figure 2(a)):

(i) Augmentation index (AI) which equals to y/x

(ii) Large Artery Stiffness Index (LASI) or S3

(iii) Inflection Point Area ratio (IPA), denoted by S1, S2, S3, and S4

(iv) Cardiac period (CP)

(v) Systolic upstroke time (SUT)

(vi) Diastolic time (DT)

(vii) Systolic width (SW) at a % of maximum amplitude

(viii) Diastolic width (DW) at a % of maximum amplitude

The inference process is performed using regression-based supervised machine learning algorithms including support vector machine (SVM) and artificial neural network (ANN). Additional investigation based on new-time domain features has been done in [18]. In the study, it is also found that Womersley number which interprets the influence of fluid flow properties on BP affects the accuracy. The best prediction result is done using random forest (RF) with a genetic algorithm (GA) as the feature selection method for minimizing the computational cost. Deeper investigation to the PPG signal's derivatives not only alleviates the number of features but also reduces the estimation errors in our previous work [19]. From the original PPG signal, the first derivative (dPPG) and the second derivative (sdPPG) are computed to generate new parameters, i.e., the ascending and descending area of dPPG defined as dAA and dDA, respectively, and the ascending and descending area of sdPPG defined as sdAA and sdDA, respectively, as shown in Figure 2. In this work, a four-layered deep neural network

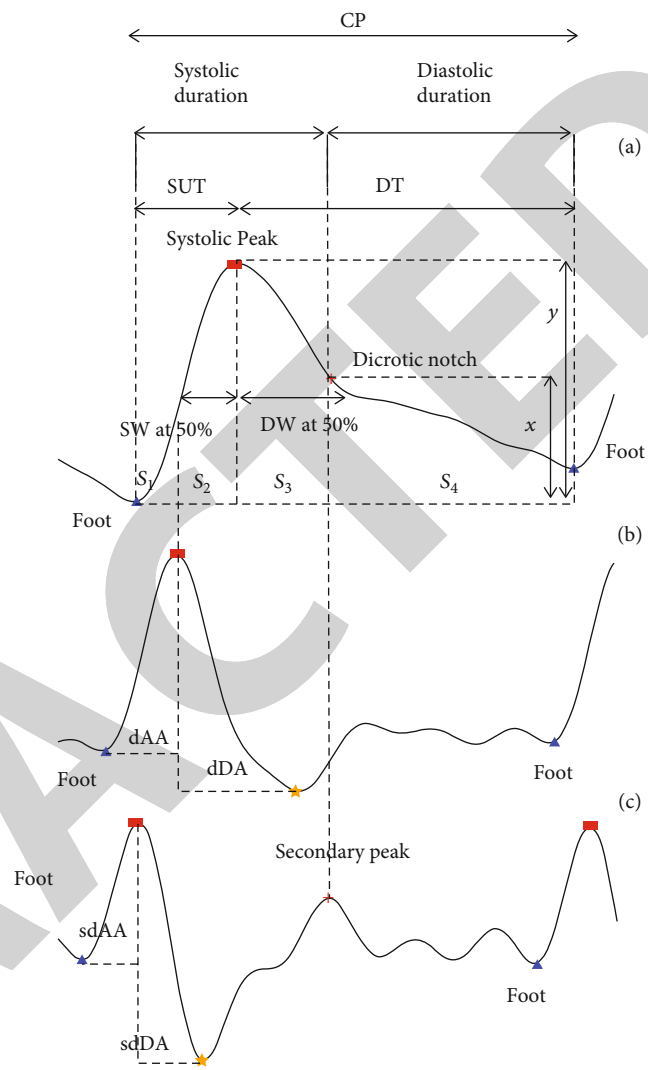


Figure 2: (a) PPG, (b) dPPG, and (c) sdPPG.

is suggested as the best algorithm to predict SBP and DBP from the input of 59 features. However, the size of the network is quite large which requires high computing resource.

2.2. Featureless-Based BP Estimation. The PPG waveform varies over subjects due to various influences such as age, drugs, and diseases. There are four typical waveform variations of PPG waveform [15], as shown in Figure 3. The ideal waveform is mostly found from cardiovascular disease-free people illustrated in Figure 3(a). Figures 3(b) and 3(c) show waveform with an indistinct and almost nonexistent dicrotic notch while Figure 3(d) shows an invisible dicrotic notch with diastolic duration that decays faster than the others. Thus, extracting handcraft features from nonideal waveform will be difficult to carry out in practice and automatic extraction of necessary features is proposed.

In [15], automatic extraction is performed using ANN with the input consisting of ECG and PPG signals. The output features are fed to a three-layered LSTM network to learn the generated features and predict the SBP and DBP. Instead of using multiple sensors for the data acquisition,

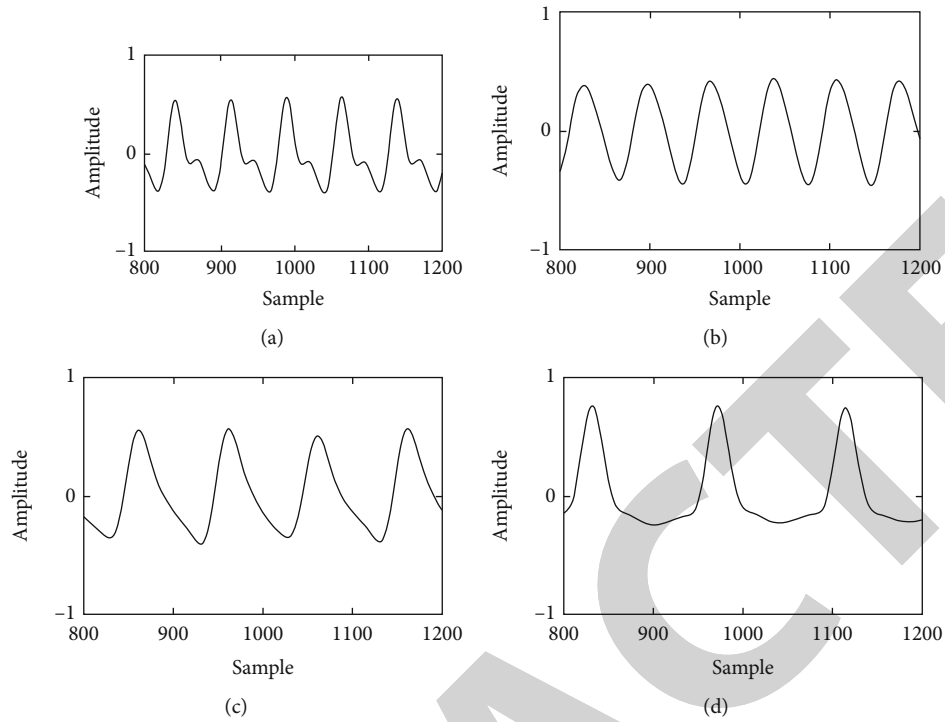


Figure 3: Waveform variations of PPG waveform including (a) ideal, (b), (c), and (d) nonideal [14].

automatic extraction from PPG signal only has been verified by [20] using a CNN which consists of an input layer, a convolutional layer, a pooling layer, and two fully connected layers. Using the PPG and its first and second derivatives, the estimation result is reported to be improved compared to the traditional method (i.e., applying multiple regression analysis of the pulse wave). However, large errors can still be found for the cases of extremely high or low SBP.

3. Materials and Methods

In this study, a new approach to estimate continuous blood pressure from PPG signal without feature engineering is proposed. Specifically, the data and structured steps to embody the proposed methodology are explained in this section.

3.1. Dataset. The PPG signal is obtained from Multiparameter Intelligent Monitoring in Intensive Care (MIMIC) II database [12, 21] which contains PPG records from more than 10,000 subjects with normal and abnormal cases. This dataset provides arterial blood pressure (ABP) signal from the related subjects as well. We extract the SBP and DBP values from the corresponding ABP signals and use them for the ground truth in the process of training and testing.

3.2. Preprocessing

3.2.1. Segmentation. Before signals are being trained with our model, signal segmentation is carried out. We define one segment of signal that begins at one PPG's foot to the following foot consecutively. Thus, we conduct a foot detection anterior to the segmentation.

3.2.2. Resampling. PPG signals are recorded from subjects with varying conditions. The waveforms exhibit varying frequency and the length of the PPG with different subjects. For signal length normalization, we avoid zero padding considering that the padded signal might contain zeros up to 50% of the final length which impacts the model negatively. Instead, we applied signal interpolation. Thus, each of the PPG segments is resampled to 700 data points in order to unify the length of all segments.

3.3. Partitioning. After performing preprocessing on the dataset, we obtain more than 100,000 resampled PPG segments. We then randomly select 50,165 segments and partitioned them into three sets. The first set is the training set which is 80% of the total selected data. The second set is the validation set which is 10% of the training set. Lastly, the third set is the testing set which is the rest 20% of the total selected data. Given that partitioning, we are certain that our model is trained and tested using completely disjointed data.

3.4. Evaluation Metrics. The root-mean-squared error (RMSE), mean absolute error (MAE), and also standard deviation (STD) of estimation error are used for the model evaluation on the disjointed test set. We also present an evaluation based on the Association for the Advancement of the Medical Instrumentation (AAMI) standard [22] and British Hypertension Society (BHS) standard [23].

3.5. Overall Estimator Network

3.5.1. CNN. The success of CNN in many tasks related to image segmentation, classification, retrieval, and also

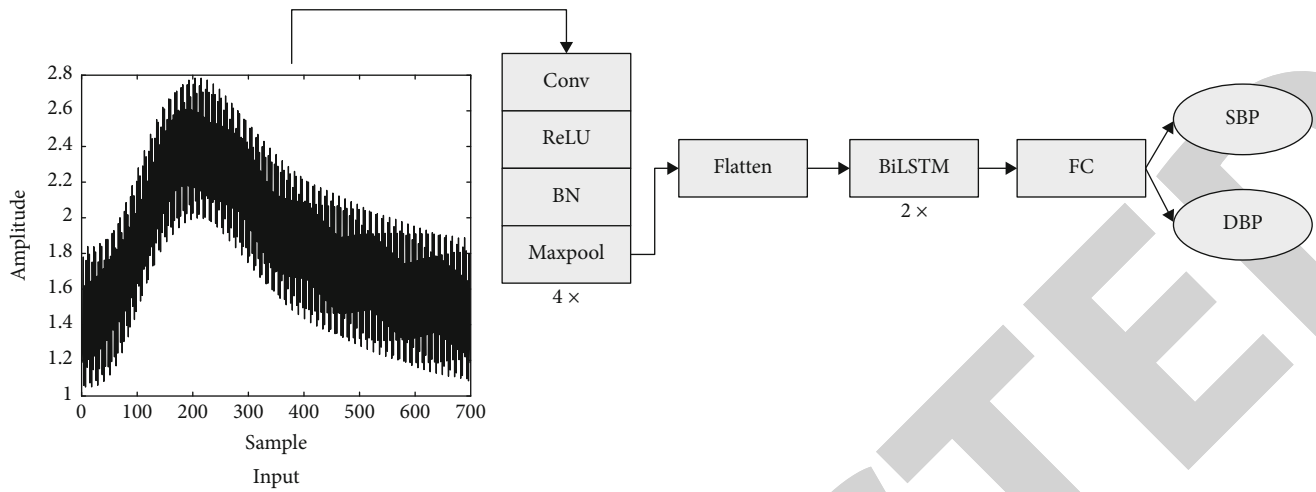


Figure 4: Illustration of the proposed network model.

captioning [24, 25] is because of its ability to exploit either the spatial or temporal correlation within the data [26–28]. In general, a typical CNN architecture includes convolution and pooling layer which is followed by a fully connected layer as the last layer. The convolutional layer works through slicing the input into small slices, commonly acknowledged as receptive fields. By slicing it into small pieces, the network is encouraged to understand feature motifs. The feature motifs may occur at various locations. However, the precise location becomes less relevant once important features are extracted, as long as its approximate position relative to others is retained. The pooling operation facilitates the network to extract a combination of features by summing up similar information in the neighbourhood. Thus, its components are making CNN a great option for automatic feature extractor.

3.5.2. BiLSTM. The memory cell that can maintain its state over time and its nonlinear gating units that thoroughly control the information flow is the key point behind the LSTM's success in resolving the long sequence problem [26]. While BiLSTM tries to connect two hidden layers of LSTM (i.e., the input sequence and the reverse copy of the corresponding sequence), it enhances the ability to learn longer dependency and subsequently improve the model performance. There are four modes how the BiLSTM connects the hidden layers, such as the following:

(i) Concat: concatenating the output of both layers

(ii) Mul: multiplying the output of both layers

(iii) Sum: adding the output of both layers

(iv) Ave: taking the average of both layers' output

Assume that having an input sequence with the size of $n \times t \times c$, which denotes number of batch size, number of timesteps, and number of states, respectively, the "mul," "sum," and "ave" mode will return an output size of $n \times t \times c$. Otherwise, the "concat" mode will return an output size of $n \times t \times 2c$ which is more informative since it does not lose any information from both input sequence (forward) and its reverse copy (backward). In this way, it allows the model to learn where to pick information and generates lower loss in the training process. Given that BiLSTM comprises of a forward LSTM and a backward LSTM layer, it performs better prediction significantly [29, 30]. The success of BiLSTM is also proved in BP estimation task, reported in [16]. Therefore, we adopt BiLSTM architecture in our BP estimation model.

Accordingly, the proposed network is a model consisting of two hierarchy levels. The lower hierarchy level uses CNN layers to extract necessary features. The upper level uses BiLSTM to do the estimation by learning the temporal relations among the features extracted in the lower hierarchy. Each of the resampled PPG segments is the input into the CNN layers. The output of CNN layers then will be fed up to the BiLSTM layers which then output a regression result of SBP and DBP. The general illustration is shown in Figure 4.

The proposed model comprises of four 1D CNN layers which are followed by rectified linear unit (ReLU) activation function, batch normalization (BN), and max pooling in each layer. The output from the last max-pooling layer is then being flatten to be the input for two BiLSTM layers with "concat" mode. The last layer of the proposed model is a fully connected (FC) layer which generates regression output of SBP and DBP value. This model is trained using MATLAB 2019B with one GPU (NVIDIA GeForce GTX 750 Ti) within 20 epochs. We set the batch size into 128 and the initial learning rate is 0.001 which is then decreased by a factor of 0.1 every 175 iterations. The detailed information about the proposed model along with the best hyperparameter setting is presented in Table 1.

4. Results and Discussion

The testing results from our proposed model are presented in Table 2. In the first four rows of Table 2, we compare our result with previous studies which are based on feature engineering method in terms of RMSE, MAE, and STD of

Table 1: Network architecture.

Layer	Filter size	#filter/hidden units	Input size
Input	—	—	$1 \times 700 \times 1$
Conv_1	1×350	2	$1 \times 700 \times 2$
ReLU_1	—	—	$1 \times 700 \times 2$
BN_1	—	—	$1 \times 700 \times 2$
Maxpool_1	1×175	—	$1 \times 700 \times 2$
Conv_2	1×175	10	$1 \times 700 \times 10$
ReLU_2	—	—	$1 \times 700 \times 10$
BN_2	—	—	$1 \times 700 \times 10$
Maxpool_2	1×25	—	$1 \times 700 \times 10$
Conv_3	1×25	20	$1 \times 700 \times 20$
ReLU_3	—	—	$1 \times 700 \times 20$
BN_3	—	—	$1 \times 700 \times 20$
Maxpool_3	1×10	—	$1 \times 700 \times 20$
Conv_4	1×10	40	$1 \times 700 \times 40$
ReLU_4	—	—	$1 \times 700 \times 40$
BN_4	—	—	$1 \times 700 \times 40$
Maxpool_4	1×5	—	$1 \times 700 \times 40$
Flatten	—	—	28000
BiLSTM_1	—	128	256
BiLSTM_2	—	350	700
FC	—	—	2

Table 2: Performance of different approaches.

Model	Input	SBP (mmHg)			DBP (mmHg)		
		RMSE	MAE	STD	RMSE	MAE	STD
SVM [12]	5 features from ECG & PPG	—	12.38	16.17	—	6.34	8.45
SVR [31]	35 features from PPG	10.9	8.54	—	5.8	4.34	—
NN [31]	35 features from PPG	11.6	13.4	—	5.9	6.9	—
RF [18]	>15 features from ECG & PPG	13.83	9.54	—	6.80	5.48	—
1D CNN + LSTM	Raw PPG signal	13.49	8.92	8.23	8.78	6.14	5.22
Proposed	Raw PPG signal	11.50	7.85	8.41	6.53	4.42	4.80

the estimation results. The first work [12] uses the MIMIC II dataset and uses 4,254 records for the experiments. Each record contains predefined features extracted from PPG and ECG signal such as pulse transit time (PTT) and heart rate (HR). The study uses regularized linear regression (RLR), artificial neural network (ANN), and support vector machine (SVM) approaches to do the prediction.

Here, we compare our result with the result from the SVM approach which is the best one. The second and third studies are from [31] which used merely 910 good PPG signals from the MIMIC II dataset. The study uses 35 features extracted from the obtained PPG to train a neural network (NN) and support vector regression (SVR) as the estimator. Given that the results are acceptable, we compare our result

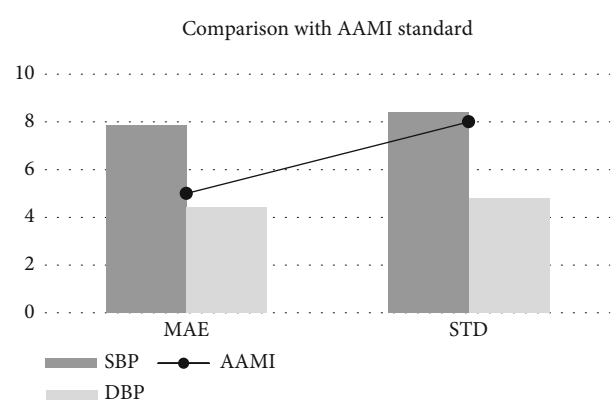


Figure 5: Comparison with AAMI standard.

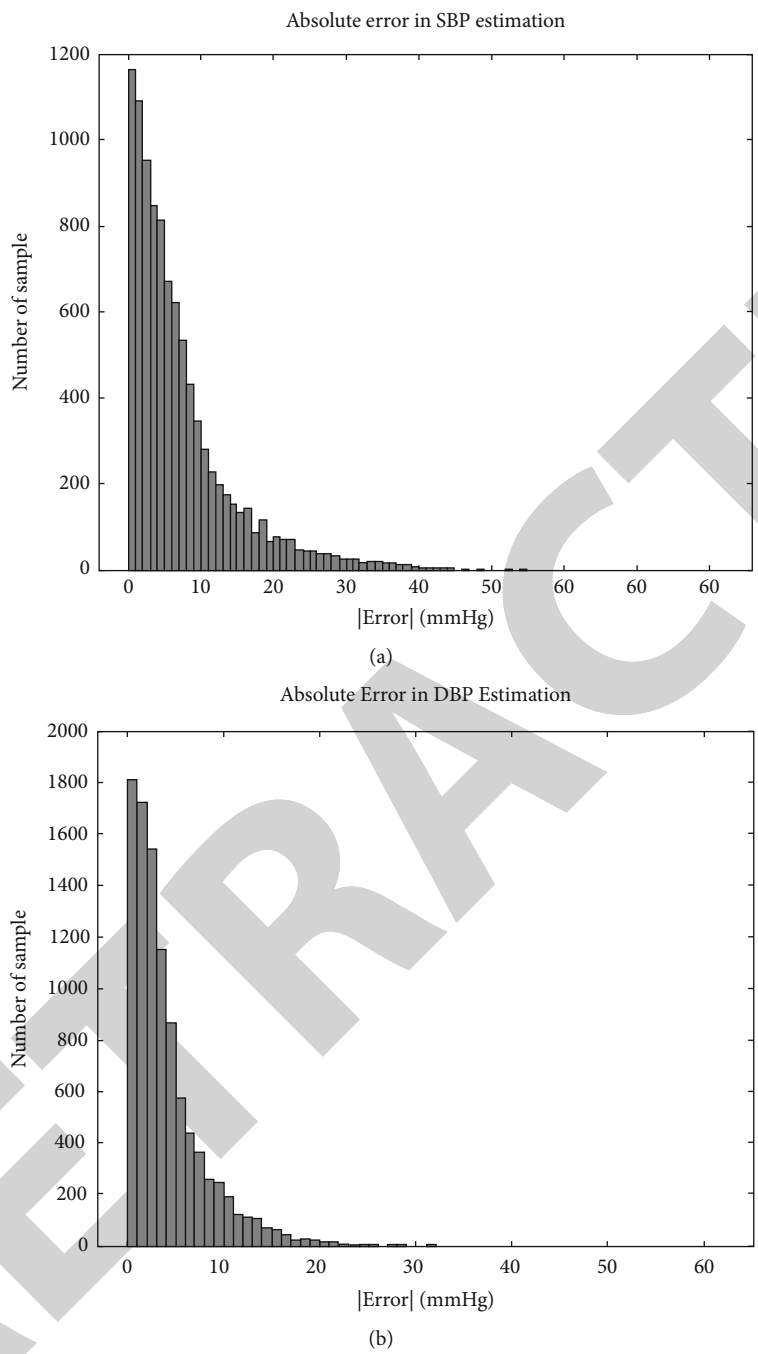


Figure 6: Histogram of absolute error in (a) SBP estimation and (b) DBP estimation.

to both of them. The results shown in the fourth row are depicted from [18] which utilized 43 features from ECG and PPG signal obtained from the MIMIC II dataset. To minimize estimation error, the genetic algorithm is used to discard negligible features. Several machine learning approaches are used as a predictor in this study, and random forest (RF) was reported as the most accurate predictor among them. To verify the benefit of BiLSTM in the network, we build a twin network that uses unidirectional LSTM instead of BiLSTM. With the same number of layers, hidden nodes, and settings, we can see how BiLSTM can affect the model performance.

The AAMI standard requires both SBP and DBP estimators to have mean error and standard deviation error below 5 and 8 mmHg, respectively, measured on a dataset consisting of more than 85 subjects. In our case, only the DBP estimator satisfies the AAMI standard while the SBP estimator slightly missed with the STD restriction as shown in Figure 5. Figure 6 presents the distribution of the absolute error of SBP and DBP estimation, respectively. The comparison result with the BHS standard concludes that our SBP estimator reaches grade C while our DBP estimator exceeds the standard with grade A which can be seen in Figure 7 with the criterion specifically presented in Table 3.

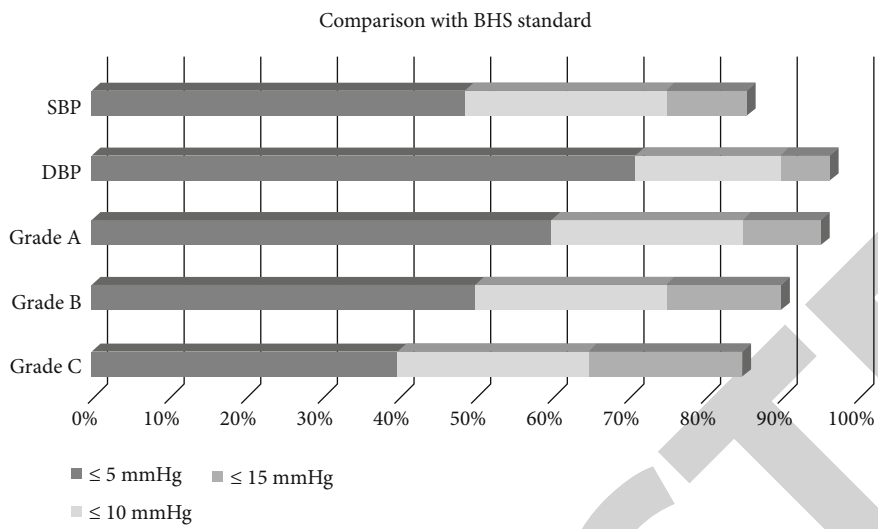


Figure 7: Comparison with BHS standard.

Table 3: Comparison with BHS standard.

Cumulative error		≤5 mmHg	≤10 mmHg	≤15 mmHg
Our result	SBP	48.69%	74.92%	85.49%
	DBP	70.84%	89.98%	96.38%
BHS	Grade A	60%	85%	95%
	Grade B	50%	75%	90%
	Grade C	40%	65%	85%

4.1. Comparison with Prior Studies. It is doubtless that conducting a fair comparison with prior studies is difficult due to the following reasons. Although all the studies being compared use the MIMIC II dataset, the number of subjects being used in each study are varying. Moreover, the evaluation metrics presented in every study are also different which cannot portray a comprehensive comparison. Nevertheless, we tried our best to summarize the existing work and compared the proposed method with them. Table 2 compares the performance of various existing methods with the proposed method, in terms of RMSE, MAE, and STD. We can see that the proposed method outperforms other methods except for SVR [31]. However, the method in [31] needs to select 35 features by the domain expert, while with the proposed method, we can directly feed the raw PPG signal into the system and get the result. The number of records that are used in this study is more than 50,000 segments, which are also chosen randomly from 100,000 segments by doing segmentation on signals of 5000 subjects, which is much larger than in [31] (910 records of good signal). Thus, we speculate that there may be high variance in the data. In this case, the higher error can be addressed due to this problem. It is also evident that using bidirectional LSTM can reduce the estimation error. Both SBP and DBP predictions are significantly improved with BiLSTM compared to 1D CNN + LSTM. We confirm that learning the information extracted from the convolution layer not only in a forward manner but also in a backward manner can help the framework to understand its pattern better.

4.2. Perspective. The method in [31] which achieves the least error in the comparison experiment can be treated as an approach of classic "feature-based" signal processing while the proposed method is an "end-to-end" machine learning technique which can be treated as "featureless" signal processing. It does not require prior knowledge about the specific domain and therefore saves a lot of extra costs and is more preferred in the deep learning community. Using a deep learning method with a featureless processing can also save time [32] which will be very practical to be applied at wearable devices. Although we focus on PPG signals only in this study, this "featureless" signal processing can be a starting point for the other application using one-dimensional signal such as ECG, BCG, etc. We believe that in the future, "end-to-end" training, which needs no prior domain knowledge in the loop, will become more popular as the amount of data and computational resources increases. The transition from "feature-based" to "featureless" signal processing will be a paradigm shift in the domain of biomedical signal processing.

5. Conclusions

Despite the fact that the PPG sensor is becoming very popular for measuring SBP and DBP noninvasively, PPG signal can be easily affected by noise, especially during the signal acquisition stage. The challenge of extracting feature from poor quality of PPG signals for doing the measurement has been concluded by our model which discards feature engineering process by applying 1D CNN and BiLSTM network. Our model achieved acceptable SBP and DBP estimation results in terms of RMSE, MAE, and STD of the estimation error. It also satisfies the AAMI standard on DBP estimation and achieves grade C and grade A for SBP and DBP estimation, respectively. Through its simplicity and sufficiency, the proposed model can be applied into healthcare IoT devices. Moreover, further investigation on the model optimization such as applying an attention mechanism is required to improve the model's performance and reduce the resulting error.

Data Availability

The MIMIC II dataset used to support the findings of this study have been deposited in the UCI Machine Learning repository (https://archive.ics.uci.edu/ml/datasets/Cuff-Less+Blood+Pressure+Estimation)

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

References

[1] H. Peterson, "Factors affecting blood pressure," *Access*, 2020, http://www.indianhills.edu/_myhills/courses/BIO178/documents/lu04_bp.pdf.

[2] G. Parati, G. S. Stergiou, E. Dolan, and G. Bilo, "Blood pressure variability: clinical relevance and application," *The Journal of Clinical Hypertension*, vol. 20, no. 7, pp. 1133–1137, 2018.

[3] P. Andriessen, R. L. M. Schoffelen, R. C. M. Berendsen et al., "Noninvasive assessment of blood pressure variability in preterm infants," *Pediatric Research*, vol. 55, no. 2, pp. 220–223, 2004.

[4] C. El-Hajj and P. A. Kyriacou, "A review of machine learning techniques in photoplethysmography for the non- invasive cuff-less measurement of blood pressure," *Biomedical Signal Processing and Control*, vol. 58, article 101870, 2020.

[5] J. L. Moraes, M. X. Rocha, G. G. Vasconcelos, J. E. Vasconcelos Filho, V. H. C. de Albuquerque, and A. R. Alexandria, "Advances in photopletysmography signal analysis for biomedical applications," *Sensors*, vol. 18, no. 6, p. 1894, 2018.

[6] J. Allen, "Photoplethysmography and its application in clinical physiological measurement," *Physiological Measurement*, vol. 28, no. 3, pp. R1–R39, 2007.

[7] F. Rundo, A. Ortis, S. Battiato, and S. Conoci, "Advanced bio-inspired system for noninvasive cuff-less blood pressure estimation from physiological signal analysis," *Computation*, vol. 6, no. 3, p. 46, 2018.

[8] S. Chen, Z. Ji, H. Wu, and Y. Xu, "A non-invasive continuous blood pressure estimation approach based on machine learning," *Sensors*, vol. 19, no. 11, p. 2585, 2019.

[9] X. Ding, B. P. Yan, Y.-T. Zhang, J. Liu, N. Zhao, and H. K. Tsang, "Pulse transit time based continuous cuffless blood pressure estimation: a new extension and a comprehensive evaluation," *Scientific Reports*, vol. 7, no. 1, article 11554, 2017.

[10] W. H. Lin, X. Li, Y. Li, G. Li, and F. Chen, "Investigating the physiological mechanisms of the photoplethysmogram features for blood pressure estimation," *Physiological Measurement*, vol. 41, no. 4, article 044003, 2020.

[11] P. Su, X. Ding, Y. Zhang, J. Liu, F. Miao, and N. Zhao, "Long-term blood pressure prediction with deep recurrent neural networks," in *2018 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)*, pp. 323–328, Las Vegas, NV, USA, 2018.

[12] M. Kachuee, M. M. Kiani, H. Mohammadzade, and M. Shabany, "Cuff-less high-accuracy calibration-free blood pressure estimation using pulse transit time," in *2015 IEEE International Symposium on Circuits and Systems (ISCAS)*, pp. 1006–1009, Lisbon, Portugal, 2015.

[13] F. P. Lo, C. X. Li, J. Wang, J. Cheng, and M. Q. Meng, "Continuous systolic and diastolic blood pressure estimation utilizing long short-term memory network," in *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 1853–1856, Jeju, Korea, 2017.

[14] Y. Kurylyak, F. Lamonaca, and D. Grimaldi, "A neural network-based method for continuous blood pressure estimation from a PPG signal," in *2013 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*, pp. 280–283, Minneapolis, MN, USA, 2013.

[15] M. S. Tanveer and M. K. Hasan, "Cuffless blood pressure estimation from electrocardiogram and photoplethysmogram using waveform based ANN-LSTM network," *Biomedical Signal Processing and Control*, vol. 51, pp. 382–392, 2019.

[16] Y.-H. Li, L. N. Harfiya, K. Purwandari, and Y.-D. Lin, "Real-time cuffless continuous blood pressure estimation using deep learning model," *Sensors*, vol. 20, no. 19, p. 5606, 2020.

[17] M. Elgendi, "On the analysis of fingertip photoplethysmogram signals," *Current Cardiology Reviews*, vol. 8, no. 1, pp. 14–25, 2012.

[18] G. Thambiraj, U. Gandhi, U. Mangalanathan, V. J. M. Jose, and M. Anand, "Investigation on the effect of Womersley number, ECG and PPG features for cuff less blood pressure estimation using machine learning," *Biomedical Signal Processing and Control*, vol. 60, article 101942, 2020.

[19] Y.-C. Hsu, Y.-H. Li, C.-C. Chang, and L. N. Harfiya, "Generalized deep neural network model for cuffless blood pressure estimation with photoplethysmogram signal only," *Sensors*, vol. 20, no. 19, p. 5668, 2020.

[20] S. Shimazaki, H. Kawanaka, H. Ishikawa, K. Inoue, and K. Oguri, "Cuffless blood pressure estimation from only the waveform of photoplethysmography using CNN," in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 5042–5045, Berlin, Germany, 2019.

[21] M. Saeed, M. Villarroel, A. T. Reisner et al., "Multiparameter intelligent monitoring in intensive care II: a public-access intensive care unit database," *Critical Care Medicine*, vol. 39, no. 5, pp. 952–960, 2011.

[22] American National Standard Manual, *Electronic or Automated Sphygmonanometers*, AAMI, 2003.

[23] E. O'Brien, J. Petrie, W. Littler et al., "The British Hypertension Society protocol for the evaluation of automated and semi-automated blood pressure measuring devices with special reference to ambulatory systems," *Journal of Hypertension*, vol. 8, no. 7, pp. 607–619, 1990.

[24] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," in *Proceedings of the IEEE international conference on computer vision*, pp. 2961–2969, Venice, Italy, 2017.

[25] F. Schroff, D. Kalenichenko, and J. Philbin, "Facenet: a unified embedding for face recognition and clustering," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 815–823, Boston, USA, 2015.

[26] A. Khan, A. Sohail, U. Zahoora, and A. S. Qureshi, "A survey of the recent architectures of deep convolutional neural networks," *Artificial Intelligence Review*, vol. 53, no. 8, pp. 5455–5516, 2020.

[27] L. Lo, H.-X. Xie, H.-H. Shuai, and W.-H. Cheng, "Facial chirality: using self-face reflection to learn discriminative features for facial expression recognition," in *The 2021 IEEE*

Hindawi
Wireless Communications and Mobile Computing
Volume 2023, Article ID 9825958, 1 page
https://doi.org/10.1155/2023/9825958

Retraction

Retracted: Age Label Distribution Learning Based on Unsupervised Comparisons of Faces

Wireless Communications and Mobile Computing

Received 11 July 2023; Accepted 11 July 2023; Published 12 July 2023

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. Li, Z. Deng, W. Xu, Z. Li, and H. Liu, "Age Label Distribution Learning Based on Unsupervised Comparisons of Faces," *Wireless Communications and Mobile Computing*, vol. 2021, Article ID 1996803, 7 pages, 2021.

Hindawi
Wireless Communications and Mobile Computing
Volume 2021, Article ID 1996803, 7 pages
https://doi.org/10.1155/2021/1996803

Research Article

Age Label Distribution Learning Based on Unsupervised Comparisons of Faces

Qiyuan Li,[1,2] **Zongyong Deng**,[1,3] **Weichang Xu**,[1,4] **Zhendong Li**,[1,4] **and Hao Liu**[1,4]

[1]*School of Information Engineering, Ningxia University, Yinchuan 750021, China*
[2]*School of Cyberspace Science and Technology, Beijing Institute of Technology, Beijing 100081, China*
[3]*College of Computer Science, Sichuan University, Chengdu 610065, China*
[4]*Collaborative Innovation Center for Ningxia Big Data and Artificial Intelligence Co-founded by Ningxia Municipality and Ministry of Education, Yinchuan 750021, China*

Correspondence should be addressed to Weichang Xu; xuwch@nxu.edu.cn and Hao Liu; liuhao@nxu.edu.cn

Received 22 August 2021; Accepted 16 October 2021; Published 13 November 2021

Academic Editor: Chi-Hua Chen

Although label distribution learning has made significant progress in the field of face age estimation, unsupervised learning has not been widely adopted and is still an important and challenging task. In this work, we propose an unsupervised contrastive label distribution learning method (UCLD) for facial age estimation. This method is helpful to extract semantic and meaningful information of raw faces with preserving high-order correlation between adjacent ages. Similar to the processing method of wireless sensor network, we designed the ConAge network with the contrast learning method. As a result, our model maximizes the similarity of positive samples by data enhancement and simultaneously pushes the clusters of negative samples apart. Compared to state-of-the-art methods, we achieve compelling results on the widely used benchmark, i.e., MORPH.

1. Introduction

Human face is a basic biological feature of human beings, and its image contains a lot of useful information, such as age, gender, identity, race, and emotion [1]. Face age estimation is aimed at using computer technology to predict the accurate age values for the given facial images. However, variations of the shape of the skull, the position of the facial features, wrinkles, lighting, expressions, and movements of videos likely give rises to bias prediction in the wild conditions [2]. Particularly when a small amount of training data is used, the accuracy of age prediction is generally not high.

Recently, although people have been working on age estimation research, the performance is still very limited. This is mainly affected by two factors. On the one hand, because the existing dataset is not complete enough, most methods are trained in a supervised way, which requires manual annotations. On the other hand, the relationship of face data and age labels is usually complexly heterogeneous and nonlinear [3, 4]. Hence, this urgently prompts us to propose robust and accurate facial age estimation particularly under unconstrained environments.

Conventional age estimation methods could be roughly categorized into two major ingredients: feature representation and age predictor. Feature representation-based methods [5–7] are aimed at seeking discriminative feature descriptors for ages based on the face images. Respectively, age predictor-based methods [8, 9] basically learn to classify the age ranker based on the input feature representation. Apart from that, label distribution has emerged as the widely employed and state-of-the-art methods such as [10–12]. The algorithm typically encodes a range of age labels to a symmetrical distribution, e.g., Gaussian or triangle distribution, reflecting the smoothness for high-performance age estimation. Nevertheless, they are constrained to take only fixed-structural form to model the ambiguous properties of age labels, which are usually nonrobust to complex cross-

population face data domains. In order to solve this problem, most scholars usually adopt feature fusion methods, such as [13, 14], but these methods seldom pay attention to the high correlation between adjacent samples and often require a lot of annotation data to achieve. Therefore, we propose a flexible unsupervised comparison of label distribution learning age estimation method, which can solve the above problems.

Similar to the wireless sensor network in the space to monitor and record the physical conditions of the environment and organize the collected data in a central location. In this article, we propose a label distribution learning method based on unsupervised comparison, dubbed UCLD, which typically models heterogeneous face aging data for robust face age estimation. Compared with the traditional fixed and inflexible label distribution methods, our method not only takes into account the high correlation between adjacent samples but also reduces the dependence of the model on the data. In this article, we believe that the learned distribution is determined by the relationship between the samples, as shown in Figure 1. Technically, we first construct the embedding space of each anchored sample based on the facial appearance information. Then, the age feature is extracted through the constraints of the two projection layers and the contrast loss. Our network structure uses the improved VGG-16 [15] for effective feature learning. Figure 2 illustrates the flow chart. In order to further evaluate the effectiveness of our proposed method, we conduct extensive experiments on two field datasets. Compared with the existing facial age estimation methods, it achieves significantly superior performance.

2. Methodology

In this section, we present a detailed description of our problem formulation, the proposed UCLD model, and finally its alternatively associated optimization procedure.

Considering the size and efficiency of the model, the convolutional neural network used in this article is an improved network from four aspects based on the VGG-16 [15] architecture. First, the three fully connected layers of the VGG-16 [15] architecture contain approximately 90% of the parameters of the entire model. In this paper, only two fully connected layers are used and the dimensionality is reduced sequentially, and the mixed layer constructed by the maximum pooling layer and the global average pooling layer is retained. Second, in order to further reduce the model size, the number of filters in each convolutional layer is reduced by half to make it thinner than the original VGG-16 [15] architecture. Third, in order to speed up the training, a batch normalization layer is added after each convolutional layer [17]. Finally, the pretraining model is obtained through the comparison learning module, and then, the label distribution learning module and the expectation regression module are added to jointly learn the age distribution. The algorithm will be described in detail in the following.

2.1. Problem Setting. Assume the input space $X = R^{h*w*c}$, where h, w, and c represent the height, width, and number of channels of the input image, respectively. The label $Y = R$ represents the actual age value. On the training set $\mathscr{D} = \{(x^i, y^i)\}_{i=1}^{N}$ with the number of samples N, define $x^i \in X$ as the ith input image, and $y^i \in Y$ as the corresponding age. The age estimation problem is to learn the mapping function $\mathscr{F} : X \longrightarrow Y$ in order to make the error between the predicted value $\hat{y}$ and the true value y as small as possible on a given input image x.

Gao et al. [18] defined $l = [0 : \triangle l : 100]$ as an ordered label vector, where $\triangle l$ is a fixed real number. Using an equal step size $\triangle l$ to quantify y, the probability density function of the normal distribution that generates the true value p through y and σ is

$$p^k = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(l_k - y)^2}{2\sigma^2}\right), \tag{1}$$

where σ is a hyperparameter and p^k is the probability that the true age is l_k years old. This article is aimed at maximizing the similarity between the true value p and the predicted value $\hat{p}$ generated by the convolutional neural networks.

2.2. Contrastive Loss. For a set of N randomly sampled sample pairs $\{x_k, y_k\}_{k=1\cdots N}$, the corresponding batch used for training consists of $2N$ sample pairs $\{x_l, y_l\}_{l=1\ldots 2N}$, where x_{2k} and y_{2k-1} are two random enhanced views of $x_{k(k=1\cdots N)}$ and $y_{2k-1} = y_{2k} = y_k$.

In the data processing of $2N$ extended samples, let $i \in I \equiv \{1\ldots 2N\}$ be the index of an arbitrary augmented sample, and let $j(i)$ be the index of the other augmented sample originating from the same source sample. In unsupervised contrastive learning [19–21], the loss takes the following form:

$$\mathscr{L}^{\text{self}} = \sum_{i\in I} \mathscr{L}_i^{\text{self}} = -\sum_{i\in I} \log \frac{\exp\left(Z_i \cdot Z_{j(i)}/\tau\right)}{\sum_{a\in A(i)} \exp(Z_i \cdot Z_a/\tau)}. \tag{2}$$

Here, $Z_l = \text{Proj}(\text{Enc}(x_l)) \in R^{\mathscr{D}_p}$, the $\cdot$ symbol denotes the inner product, $\tau \in R^+$ is a scalar temperature parameter, and $A(i) \equiv I \setminus \{i\}$. The index i is called the anchor, index $j(i)$ is called the positive, and the other $2(N-1)$ indices $(k \in A(i) \setminus \{j(i)\})$ are called the negatives. Note that for each anchor i, there is 1 positive pair and $2N-2$ negative pairs. The denominator has a total of $2N-1$ terms (the positive and negatives).

2.3. Label Distribution Learning. If the true ages of the two input images are similar, the two images are considered similar. In other words, input images with similar outputs are theoretically highly correlated. In order to use the features extracted from these correlations, the label distribution learning module quantifies the range of possible y values into labels in l.

Specifically, given the input image x and the corresponding label distribution p, it is assumed that $f = \mathscr{F}(x;\theta)$ is the activation of the last layer of the convolutional neural network, where θ represents the parameters of the convolutional neural network. A fully connected layer passes f to $x \in R^K$ through

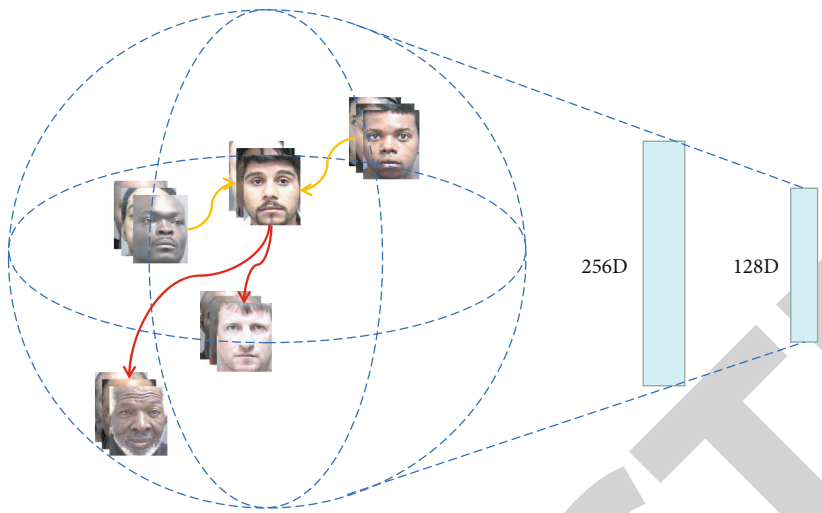


Figure 1: Demonstration of our insight. Our model is aimed at constructing a balanced embedding space, so that the anchor is closer to similar samples and farther away from different samples. Then, the age characteristics of the samples are extracted through two projection layers to make a robust age estimation.

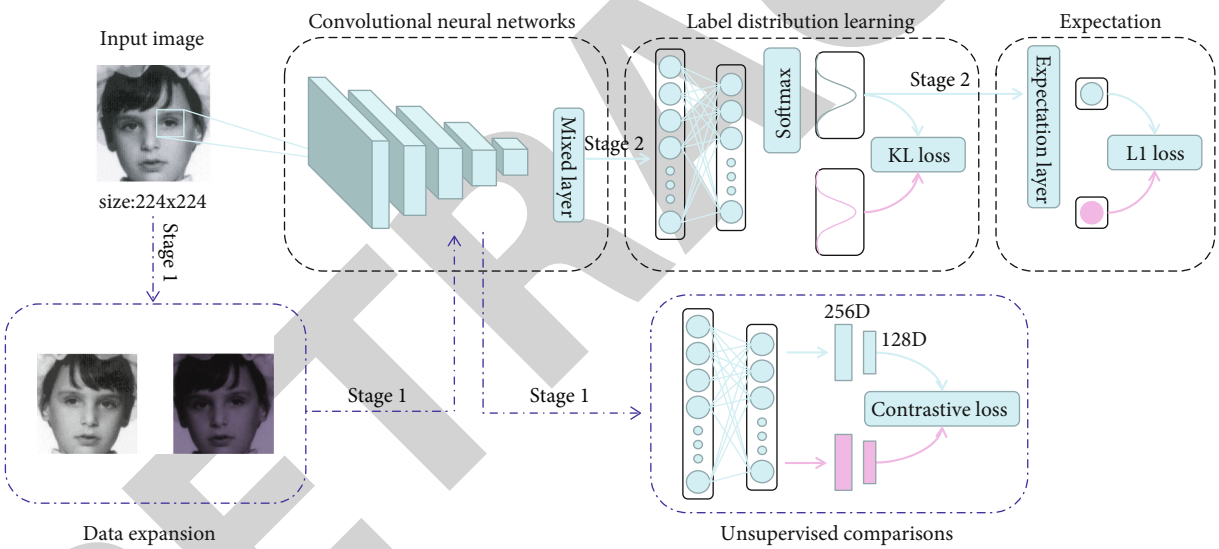


Figure 2: Flowchart of our UCLD. Our structure is divided into two stages. In the first stage, after data expansion of the image, the age samples are input into the preset CNN to get the normalized embedding of the image and then the vector embedded through the two projection layers is calculated and compared to the loss to obtain the ConAge model, which is the basis for the algorithm proposed in this paper. In the second stage, after obtaining these relevant depth features, they are projected into the average variance label distribution through a small linear layer, and the network parameters are optimized through backpropagation. At the same time, the mixed hyperparameters of the average variance label distribution are iterated through the widely used expectation-maximization optimization [16].

$$x = W^T f + b. \tag{3}$$

Then, we use the softmax function to convert x into a probability distribution as follows:

$$\widehat{p}_k = \frac{\exp(x_k)}{\sum_t \exp(x_t)}. \tag{4}$$

For a given input image, the goal of label distribution learning is to find θ, W, and b to generate $\widehat{p}$ similar to p.

Finally, the KL divergence is used as a measure of the difference between the real label and the predicted label. Therefore, the following loss function is defined on the training sample:

$$L_{ld} = \sum_k p_k \ln \frac{p_k}{\widehat{p}_k}. \tag{5}$$

2.4. Expectation Regression. Using only the label distribution learning module cannot accurately predict the age of the

Table 1: Face age estimation result table.

Method	Network	Dataset	MAE
DLDL-v2 (baseline)	TinyAge	FGNET	4.4676
	ThinAge		4.1322
UCLD	*ConAge*		*3.6046*
DLDL-v2 (baseline)	TinyAge	MORPH	2.5118
	ThinAge		2.3440
UCLD	*ConAge*		*2.2142*

Table 2: Comparison result table of different settings.

	ConAge*	ConAge	ConAge1	ConAge2	ConAge3
Label	Exist	None	None	None	None
Linear	2	2	1	2	2
Batch size	24	24	24	64	72
Epoch	120	120	120	120	120
Temp	0.07	0.07	0.07	0.07	0.07
MAE	2.3477	*2.2142*	2.2627	2.3185	2.3436

face. Therefore, this paper uses the expected regression module proposed in the DLDL-v2 [18] framework to improve the accuracy of face age prediction.

As shown in Figure 2, when the predicted value and label are obtained, the expected value is output:

$$\hat{y} = \sum_k \hat{p}_k l_k, \tag{6}$$

where $\hat{p}_k$ represents the predicted probability that the input image belongs to label l_k. Given the input image, the error between the expected value $\hat{y}$ and the true value y is minimized. The error metric uses the l_1 loss function, as shown in the following:

$$L_{er} = |\hat{y} - y|, \tag{7}$$

where $|\cdot|$ represents the absolute value.

2.5. Optimization. By jointly learning the label distribution and expected regression, the values of θ, W, and b can be obtained in a given data set $\mathscr{D}$. The final loss function is defined as a weighted combination of two loss functions L_{ld} and L_{er}.

$$L = L_{ld} + L_{er}, \tag{8}$$

where λ is the weight that weighs the importance of the two losses. Substituting (5), (6), and (7) into (8), we get

$$L = -\sum_k p_k \ln \hat{p}_k + \lambda \left| \sum_k \hat{p}_k l_k - y \right|. \tag{9}$$

In this framework, optimization variables include θ, W, and b. First, backpropagation through the network, and then use the stochastic gradient descent algorithm to optimize the parameters. The derivative of L with respect to $\hat{p}_k$ is

$$\frac{\partial L}{\partial \hat{p}_k} = -\frac{p_k}{\hat{p}_k} + \lambda l_k \operatorname{sign}(\hat{y} - y). \tag{10}$$

For any k and j, the derivative of the softmax function (4) is as follows:

$$\frac{\partial \hat{p}_k}{\partial x_j} = \hat{p}_k \left(\delta_{(k=j)} - \hat{p}_j \right). \tag{11}$$

Among them, if $k = j$, then $\delta_{(k=j)}$ is 1; otherwise, it is 0. Then,

$$\frac{\partial L}{\partial x} = \hat{p} - p + \lambda \operatorname{sign}(\hat{y} - y) \hat{p} \circ (l - \hat{y}). \tag{12}$$

Applying the chain rule to (3) again, the derivative of L with respect to θ, W, and b can be easily obtained

$$\frac{\partial L}{\partial W} = \frac{\partial L}{\partial x} f, \frac{\partial L}{\partial b} = \frac{\partial L}{\partial x}, \frac{\partial L}{\partial \theta} = \frac{\partial L}{\partial x} W^T \frac{\partial \mathscr{F}}{\partial \theta}. \tag{13}$$

Once θ, W, and b are known, in the forward network calculation, the age prediction value $\hat{y}$ of any face image x can be generated by (6), and finally, the age estimation of the face image is realized.

3. Experiments

In order to evaluate the effectiveness of this method, we conducted research results on two widely used datasets, including FGNET [22] and MORPH [23]. Due to wild conditions, face samples in these datasets often experience challenging situations. In order to illustrate the advantages of this model, we only use the MORPH dataset for model pretraining.

3.1. Datasets. The FG-NET dataset was constructed by Professor Lanitis of the University of Cyprus in Europe while studying the age estimation algorithm for faces. This dataset collected a total of 1002 facial images of 82 people through scanning. Each image provides 68 key points of face information, ranging from 0 to 69 years old. It is currently one of the most open real age datasets of the young people. For fair evaluation setting, we employed the leave-one-person-out (LOPO) protocol by following [9].

The MORPH dataset was constructed by Karl Ricanek Jr. of North Carolina State University and others when they studied face aging. The dataset consists of two parts: Album1 and Album2, which contain 1724 and 55608 face images, respectively. Album1 was collected from 1962 to 1998, and the age span was 15-68 years; Album2 was collected from 2003 to 2007, and the age span was 16-77. Since the number of collections of Album2 is significantly more than that of Album1, most scholars use Album2 for facial age estimation research. In order to make fair comparisons, we also use the

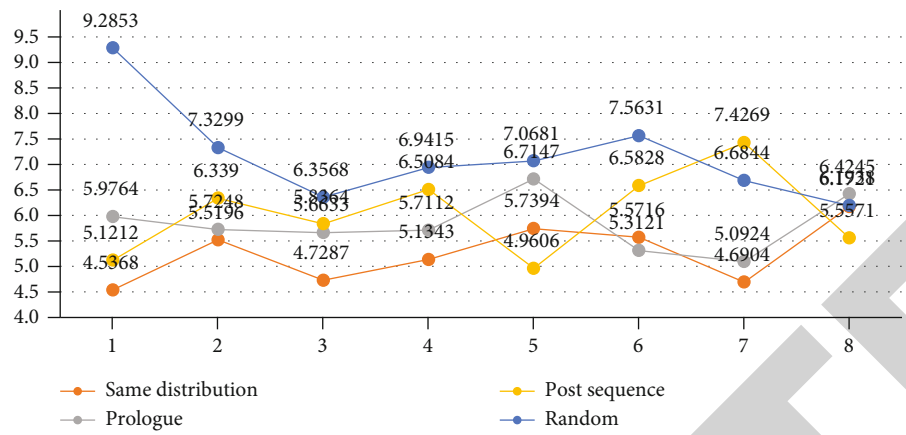


Figure 3: The comparison results of the four weakly supervised sampling methods on the TinyAge network architecture and the FG-NET dataset in 8 experiments.

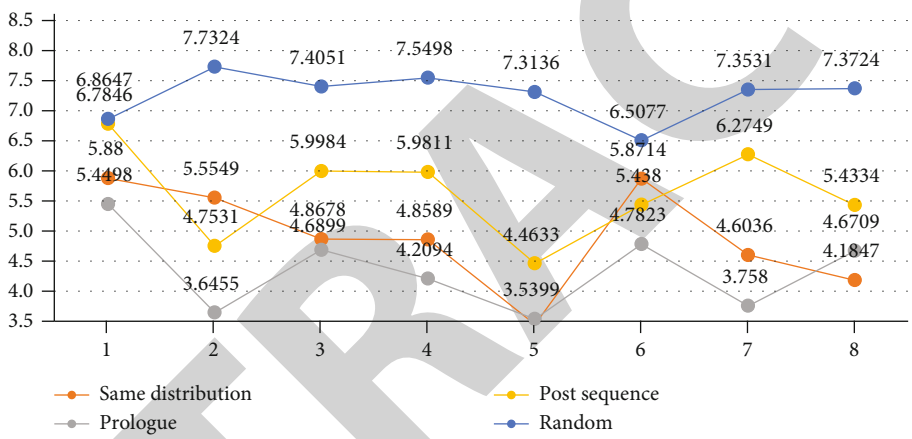


Figure 4: The comparison results of the four weakly supervised sampling methods on the ThinAge network architecture and the FG-NET dataset in 8 experiments.

Table 3: Weakly supervised face age estimation result table.

Method	Network	Dataset	MAE
DLDL-v2 (baseline)	ThinAge	FGNET (25%)	6.4146
UCLD	*ConAge*		*6.3342*
DLDL-v2 (baseline)	ThinAge	MORPH (25%)	2.8834
UCLD	*ConAge*		*2.6545*

Album2 dataset, where 80% of the data is used as the training set and 20% of the data is used as the test set.

3.2. Evaluation Metric. In the experiment, we use Mean Absolute Error (MAE) [24] to calculate the difference between the estimated age value and the true age value. Obviously, the smaller the value of MAE, the smaller the error between the predicted age and the true age, and the better the performance of the model, as shown in Table 1.

Please note that the DLDL-v2 [18] mentioned in this article is all source codes released by them. Compare our method with the experimental results of DLDL-v2 on the FGNET and MORPH datasets. Obviously, our method is more advantageous. In addition, we also changed the experimental settings several times as shown in Table 2.

Among them, linear represents the number of projection layers used. Despite using different settings, the experimental results of our method on the MORPH dataset still maintain the most advanced performance.

3.3. Implementation Details. For each face image, the size is adjusted to 224×224 before being input to the network. Then, select one of the five data enhancement methods: random horizontal flip, random zoom, random rotation, color distortion, and Gaussian blur to process the image. The comparative learning module of the network is used to generate a pretraining model on the MORPH dataset. The initial learning rate is set to 0.001, and it is reduced by 10 times every 30 iterations. After the pretraining is completed, delete the contrast learning module of the network and add the label distribution learning module and the expectation regression module to test the face age dataset. During the

test, the test image and its flipped copy are fed to the network, and its predicted value is averaged as the final age estimate.

In order to further evaluate the performance of the method proposed in this paper, the following weakly supervised experiments are completed. Regarding the sample order in fully supervised training as the original order, five sampling methods are proposed as follows:

(i) Sampling with the same distribution: that is, the probability of taking out 25% of the labeled data in the original sample interval is equal.

(ii) Preorder sampling: take the first 25% of the labeled data in the order of the original sample.

(iii) Postsampling: take the last 25% of the labeled data in the order of the original sample.

(iv) Random sampling: 25% of the labeled data is randomly selected from the original sample.

(v) Single sampling: that is, only different labeled data are retained in the original sample.

The TinyAge and ThinAge network architectures were applied to these five sampling methods, respectively, and eight tests were performed on the first face data file in the FG-NET dataset. The average MAE after 8 tests on the two networks with a single sampling method are 16.81 and 13.03, respectively. The test results of the other four sampling methods are shown in Figures 3 and 4.

Change the training dataset to a weakly supervised training dataset, and use only 25% of the labeled data to test the optimal ThinAge network architecture in DLDL-v2 and the ConAge network architecture proposed in this article. The experimental results are shown in Table 3.

It can be seen from the experimental results that our method has better performance than the DLDL-v2 framework regardless of whether it is fully supervised or weakly supervised. In addition, we have reached three conclusions: (1) traditional methods, such as DEX [25] and ODFL [25], process each age label independently without considering their previous correlation. Our unsupervised comparison method simulates the way humans observe things and can flexibly consider the relationship between age samples. (2) Some label distribution learning methods, such as LDL [11] and CPNN [11], only implement a fixed structural model on the age label distribution, which may lead to rigid adaptation to real-world facial aging data. Thanks to the comparative learning module, our method obtains more accurate semantic information, making subsequent test results more accurate. Particularly in a weakly supervised experimental setting, it can be seen that even if only a quarter of the data is used, the performance of our UCLD is better than most technical levels. This achievement is mainly because our model is less dependent on data.

4. Conclusion

In this article, in view of the high correlation between adjacent age samples and the strong dependence of existing methods on data, we combine contrast loss and label distribution learning to learn abstract representations in an unsupervised manner. An unsupervised contrast label distribution (UCLD) learning method is proposed, which is similar to the processing form of wireless sensor networks. Extensive experiments on two datasets have proved the effectiveness of the method, especially the MORPH dataset reflects the advanced nature of the method. In future work, we will focus on efficiently distinguishing similar images to solve the problem of age prediction accuracy.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Acknowledgments

This work was supported in part by the National Science Foundation of China under Grants 61806104 and 62076142, in part by the West Light Talent Program of the Chinese Academy of Sciences under Grant XAB2018AW05, and in part by the Youth Science and Technology Talents Enrolment Projects of Ningxia under Grant TJGC2018028.

References

[1] R. Angulu, J. R. Tapamo, and A. O. Adewumi, "Age estimation via face images: a survey," *EURASIP Journal on Image and Video Processing*, vol. 2018, no. 1, 2018.

[2] N. Ramanathan, R. Chellappa, and S. Biswas, "Age progression in human faces: a survey," *Journal of Visual Languages and Computing*, vol. 15, pp. 3349–3361, 2009.

[3] W. Li, J. Lu, J. Feng, C. Xu, J. Zhou, and Q. Tian, "Bridgenet: a continuity-aware probabilistic network for age estimation," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1145–1154, Long Beach, CA, USA, 2019.

[4] W. Shen, Y. Guo, Y. Wang, K. Zhao, B. Wang, and A. L. Yuille, "Deep regression forests for age estimation," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2304–2313, Salt Lake City, UT, USA, 2018.

[5] X. Geng, Z.-H. Zhou, and K. Smith-Miles, "Automatic age estimation based on facial aging patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 12, pp. 2234–2240, 2007.

[6] Yun Fu, Guodong Guo, and T. S. Huang, "Age synthesis and estimation via faces: a survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 11, pp. 1955–1976, 2010.

[7] J. Lu, V. E. Liong, and J. Zhou, "Costsensitive local binary feature learning for facial age estimation," *EEE Transactions on Image Processing*, vol. 24, no. 12, pp. 5356–5368, 2015.

[8] Z. Yu and D.-Y. Yeung, "Multi-task warped Gaussian process for personalized age estimation," in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 2622–2629, San Francisco, CA, USA, 2010.

Hindawi
Wireless Communications and Mobile Computing
Volume 2021, Article ID 1676879, 15 pages
https://doi.org/10.1155/2021/1676879

Research Article

A Compact Adaptive Particle Swarm Optimization Algorithm in the Application of the Mobile Sensor Localization

Wei-Min Zheng, Ning Liu, Qing-Wei Chai, and Shu-Chuan Chu

College of Computer Science and Engineering, Shandong University of Science and Technology, Qingdao 266590, China

Correspondence should be addressed to Shu-Chuan Chu; scchu0803@gmail.com

Received 12 September 2021; Accepted 20 October 2021; Published 9 November 2021

Academic Editor: Chi-Hua Chen

The mobile sensor network can sense and collect the data information of the monitored object in real time in the monitoring area. However, the collected information is meaningful only if the location of the node is known. This paper mainly optimizes the Monte Carlo Localization (MCL) in mobile sensor positioning technology. In recent years, the rapid development of heuristic algorithms has provided solutions to many complex problems. This paper combines the compact strategy into the adaptive particle swarm algorithm and proposes a compact adaptive particle swarm algorithm (cAPSO). The compact strategy replaces the specific position of each particle by the distribution probability of the particle swarm, which greatly reduces the memory usage. The performance of cAPSO is tested on 28 test functions of CEC2013, and compared with some existing heuristic algorithms, it proves that cAPSO has a better performance. At the same time, cAPSO is applied to MCL technology to improve the accuracy of node localization, and compared with other heuristic algorithms in the accuracy of MCL, the results show that cAPSO has a better performance.

1. Introduction

Metaheuristic algorithms are algorithms inspired by the life habits of various creatures in nature. Metaheuristic algorithms can effectively solve many problems in life [1] and are widely used in finance, transportation, physics, chemistry, military, and other fields [2–8]. The No Free Lunch Theorem [9, 10] proves that any optimization algorithm cannot suit all situations. Therefore, various metaheuristic algorithms and their improved algorithms are constantly being proposed to solve more complicated problems [11–13].

The particle swarm optimization (PSO) is one of the most important metaheuristic algorithms proposed by Kennedy and Eberhart [14, 15]. They observed and analyzed the foraging behavior of birds and then proposed this algorithm. There is a global optimal position and an individual optimal position in the particle swarm optimization algorithm. These two positions are updated according to the fitness value in each iteration, so that the algorithm is closer to the optimal solution of the problem. The characteristics of PSO, such as few parameters, simple structure, and fast search speed, make it applied to many fields. There are also many improved PSO algorithms, such as constricted particle swarm optimization (CPSO) [16], fully informed particle swarm optimization (FIPSO) [17], comprehensive learning particle swarm optimization (CLPSO) [18], intelligence single particle optimization (ISPO) [19], and adaptive particle swarm optimization (APSO) [20]. The PSO can solve many problems, such as optimizing neural networks [21], solving vehicle routing problems [22, 23], scheduling workflow scheduling [24], and locating wireless sensor nodes [25]. Based on the original PSO algorithm, the APSO algorithm introduces evolutionary state evaluation strategies, elite learning strategies, and system adaptive parameter strategies to improve the original PSO algorithm. The APSO solves the problem of slow convergence speed and easy to fall into the local optimum of the original PSO algorithm. This paper mainly tries to combine the compact strategy with APSO to improve the accuracy of mobile sensor localization.

The application of a mobile sensor network is to provide services to people when the location information of the node is known [26]. The data measured by the node without

location information is meaningless in many situations [27], such as forest fire detection [28]. Therefore, to make full use of the monitored data, it is necessary to know the location information of the node [29, 30]. Installing GPS for each node is the best way to solve this problem, but this way is expensive and energy-consuming. Therefore, a small number of nodes should be randomly selected to install GPS, and then, the positions of other nodes should be obtained by using positioning technology through the location of the GPS nodes [31]. The most significant difference between a mobile sensor and a fixed wireless sensor is its mobility [32, 33]. The mobility enables the sensor to collect effective information in the specified area better and solves the problem that the information in a certain area cannot be collected due to the damage of a specific location node. The deployment of mobile sensors is more convenient and does not require detailed design like fixed wireless sensor deployment nodes [34].

The compact idea is to use the behavior probability of the particle swarm to replace the position and velocity of each individual to express the particle swarm. The compact algorithm can effectively save memory space [20] and has applications in small robots [35], remote office [36], and space shuttle control [37]. Algorithms using compact ideas have also been continuously proposed, such as compact artificial bee colony (cABC) [38], compact sine cosine algorithm (cSCA) [39], compact bat algorithm (cBA) [40], and compact particle swarm optimization (cPSO) [41]. However, the combination of compact algorithm and APSO has not been mentioned. This paper hopes to combine the idea of compact with APSO and propose a cAPSO algorithm that uses small memory and fast convergence and applies it to mobile sensor localization.

The rest of the paper is organized as follows. The second section introduces the basis of related work and briefly introduces APSO algorithm and mobile sensor localization technology. The third section presents the improvement methods and steps of the algorithm. The fourth section tests the performance of the algorithm and compares it with similar algorithms. The fifth section applies the improved algorithm to mobile sensor localization. The sixth section gives the conclusion of this paper.

2. Related Work

This section will briefly introduce APSO and mobile sensor localization technology MCL.

2.1. Particle Swarm Optimization. The PSO imitates the foraging behavior of birds. The food location is unknown, and each bird is affected by the surrounding birds and keeps approaching the bird with the best position during the foraging process [14, 15]. Depending on the solution of the problem, the position of the optimal individual is constantly updated iteratively. Suppose the problem dimension is D, the current position of the i-th individual is $X_i = (X_{i1}, X_{i2}, \cdots, X_{iD})$, and the current speed of the i-th individual is $V_i = (V_{i1}, V_{i2}, \cdots, V_{iD})$. In each iteration, the particle swarm retains the global optimal position and the current optimal position, and each particle is affected by these two optimal particles and approaches these two particles. The iterative formulas for updating the position and velocity of the next generation of particles are shown in Equations (1) and (2).

$$V_i^{g+1} = wV_i^g + c_1 \times \text{rand}\,() \times \left(p\text{Best}_i - X_i^g\right) + c_2 \times \text{rand}\,() \times \left(g\text{Best} - X_i^g\right), \tag{1}$$

$$X_i^{g+1} = X_i^g + V_i^{g+1}, \tag{2}$$

where c_1 and c_2 are two learning factors, w is the inertia weight, and rand () is a random number between (0,1). The iterative update of $p\text{Best}_i$ and $g\text{Best}$ are performed through the fitness value comparison. The pseudo-code of the PSO is shown in Algorithm 1.

2.2. Adaptive Particle Swarm Optimization. The APSO is based on the original PSO by introducing state estimation strategy, elite learning strategy, and parameter adaptation strategy to improve it [20]. The improved algorithm can find the solution faster and more stable. The three strategies will be briefly introduced below.

2.2.1. State Estimation Strategy. The APSO divides the entire search process into four states, namely, exploration, exploitation, convergence, and jump-out, which are represented by S1, S2, S3, and S4. The division basis is the value of the evolution factor f. To calculate the evolution factor f, we must first calculate the Euclidean distance d_i of each particle. The calculation formula of d_i is shown in Equation (3).

$$d_i = \frac{1}{N-1} \times \sum_{j=1, j\neq i}^{N} \sqrt{\sum_{k=1}^{D} \left(X_i^k - X_j^k\right)^2}, \tag{3}$$

where d_i represents the Euclidean distance of the i-th particle, N represents the total number of particle swarms, and D represents the dimension of the problem.

After obtaining the Euclidean distance of each particle, we find the minimum value $d_{\min}$, maximum value $d_{\max}$, and optimal value $d_{g\text{best}}$. Then, calculate the evolution factor f by Equation (4).

$$f = \frac{d_{g\text{best}} - d_{\min}}{d_{\max} - d_{\min}}. \tag{4}$$

According to the value of the evolution factor f, the state is divided according to Figure 1. In addition, the state division is also related to the state of the previous iteration. For example, the value of the evolution factor f of this iteration is 0.55. It is assumed that the influence of the state of the previous iteration is not considered. In that case, the current state will be set as S1, but considering that the state of the previous iteration will have an impact on this state, so if the state of the previous iteration is S1 or S4, the current state will be set to S1. If the state of the previous iteration is S2 or S3, the current state will be set to S2.

```
while i < particles do
   Initialize the position X_i and velocity V_i of each particle
   Calculate the fitness value of each particle fitness(i)
   i = i +1
end
Initialize the pBest_i = X_i
Initialize the gBest = min(pBest_i)
for g =1 to iterMax do
      for i =1 to particles do.
         Update the V_i and X_i of the particles by Equations (1) and (2)
         Calculate the fitness value of the new particle fitness(i)
         if fitness(i) < fitness(pBest_i) then
            pBest_i = X_i
         end
         if fitness(i) < fitness(gBest) then
            gBest= X_i
         end
      end
end
```

Algorithm 1: The pseudo-code of the PSO.

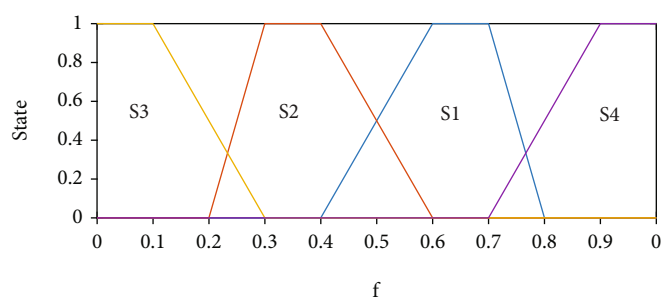


Figure 1: The division diagram of four states.

2.2.2. Parameter Adaptation Strategy. Three dynamic parameters are involved in APSO: inertial weight w, the individual learning factor c_1, and the global learning factor c_2. The weight of inertia changes with the evolutionary state. Its relationship with evolution factor f is shown in Equation (5).

$$w = \frac{1}{1 + 1.5e^{-2.6f}}. \tag{5}$$

APSO initializes w to 0.9. In the exploration state and the jump-out state, f is more extensive, which leads to a more oversized w, which is conducive to the global optimal search. In the convergence state and the exploitation state, f is more minor, which leads to a smaller w, which is conducive to the local convergence.

The c_1 and the c_2 are initialized to 2.0, and the two learning factors are adjusted according to the different evolutionary state. In the exploration state, increase c_1 and decrease c_2. This ensures that the individual learning factors play a leading role in helping the particles explore their own best individuals and avoid falling into the local optimum. In the exploitation state, slightly increase c_1 and slightly decrease c_2. Particles in the exploration phase gradually approach the local optimum. Increasing c_1 can more effectively enable the particles to explore around the individual optimum. Since the local optimum found may not be the global optimum, c_2 should be slightly reduced to prevent premature convergence because this will easily lead to the problem of the population falling into the local optimum. In the convergence stage, slightly increase c_1 and c_2. Increasing c_2 means that the particle swarm has found the global optimum at this stage, and the particle swarm can converge to this global optimum. A slight increase of c_1 is to prevent the learning factor from reaching the upper limit prematurely. If the upper bound is reached prematurely, the particles will treat the local optimum as the global optimum and quickly converge. In the jump-out state, reduce c_1 and increase c_2. It is helpful for the global optimum particle to jump out of the convergence zone and find a better position; other particles will follow the global optimal particle and converge to a better position. Table 1 shows the dynamic changes of the two learning factors in different states.

2.2.3. Elite Learning Strategy. The elite learning strategy makes the optimal global particles jump out of the convergence zone and finds a more superior position in the state

Table 1: Parameter adaptive strategy factor factors.

State	c_1	c_2
Exploration	Increase	Decrease
Exploitation	Increase slightly	Decrease slightly
Convergence	Increase slightly	Increase slightly
Jump-out	Decrease	Increase

of convergence. The elite learning strategy is determined by the elite learning rate σ. The calculation formula of σ is shown in Equation (6).

$$\sigma = \sigma_{\max} - (\sigma_{\max} - \sigma_{\min}) \times \frac{g}{\text{iterMax}}, \tag{6}$$

where g is the current number of iterations, $\sigma_{\max}$ and $\sigma_{\min}$ are the maximum and minimum values of the elite learning rate σ, taking 1.0 and 0.1, respectively, and iterMax is the maximum number of iterations. After obtaining the elite learning rate, we randomly select a dimension for Gaussian perturbations to get the global optimal particle out of the convergence region. The formula is shown in Equation (7).

$$T^d = T^d + \left(X^d_{\max} - X^d_{\min}\right) \times \text{Gaussian}\left(\mu, \sigma^2\right). \tag{7}$$

2.2.4. The Pseudo-Code of APSO. The APSO introduces the above three strategies on the basis of PSO to optimize, and the optimized APSO can find the optimal solution to the problem better and faster when solving the problem.

The pseudo-code of APSO that combines the three strategies is shown in Algorithm 2.

2.3. Mobile Sensor Localization. This section mainly introduces the MCL mobile node location method [42, 43]. The mobile nodes in the mobile sensor network move with random speed and random direction. The biggest advantage of mobile nodes over fixed nodes lies in their mobility. The mobile node solves the problem that the information in a certain area cannot be collected due to the damage of the node at a specific location. The MCL method is mainly divided into three stages: the initialization stage, the prediction stage, and the filtering stage [44]. In the initialization stage, the moving area and the maximum moving speed are specified for each node. In the prediction stage, a preliminary estimate of the location of the mobile node is made. The speed and direction of node movement are uncertain. The position after the movement is within a circle. The center of this circle is the position of the last node, and the radius is the product of the speed and the positioning interval time, as shown in Figure 2.

The filtration stage is the most critical stage of MCL. Firstly, according to the distance between the anchor node and the unknown node, MCL determines which anchor nodes are the one-hop beacon nodes of the unknown node and which anchor nodes are the two-hop beacon nodes of the unknown node. Then, MCL obtains the one-hop beacon node set B_1 and the two-hop beacon node set B_2. Secondly, randomly select points in possible areas, and filter the non-conforming points according to whether the selected points are within the range of one-hop and two-hop beacon nodes. The filter condition is shown in Equation (8).

$$\begin{aligned}\text{filer}(\text{node}) = {} & (\forall b_1 \in B_1, \text{distance}(\text{node}, b_1) \\ & \le R) \cup (\forall b_2 \in B_2, \text{distance}(\text{node}, b_2) \\ & \le 2R).\end{aligned} \tag{8}$$

The gray areas in Figure 3 are the sets of points that meet the filter conditions. The MCL locates the position of the unknown node in a small space according to the number of hops from the unknown node to the anchor node. Equation (8) filters out the points that fall in this small space. In order to prevent contingency, the coordinates of all points in this small space are averaged as the position of the unknown node initially predicted by MCL. Finally, the qualified points after filtering are estimated by Equation (9) to estimate the position of the unknown node.

$$\text{Position}(b) = \frac{\left(\sum_{i=1}^{N} \text{node}_i\right)}{N}, \tag{9}$$

where N represents the total number of points that meet the filter conditions, and node represents the position of node that meet the filter conditions.

3. Compact Adaptive Particle Swarm Optimization

This section mainly introduces the idea of compact strategy and how to apply compact strategy to adaptive particle swarm optimization algorithm.

3.1. Compact Strategy. The primary purpose of the compact strategy is to reduce memory usage without changing the performance of the original algorithm or even improving the performance of the original algorithm. The running speed will naturally be improved if the memory usage is reduced. The compact strategy uses PV perturbation vectors to represent the overall motion state of the population instead of simply using the position and velocity of each individual to represent the population state. The disturbance vector PV is defined as $\text{PV}^g = (\mu^g, \sigma^g)$, where μ represents the average value of the disturbance vector, σ represents the standard deviation of the disturbance vector, and g represents the current iteration update times.

The compact strategy ultimately returns a value between (0,1). The PV vector is composed of μ and σ, and the probability distribution function (PDF) can be calculated through μ and σ. Then, the cumulative distribution function (CDF) can be calculated by PDF. The calculation formulas of PDF and CDF are in Equations (10) and (11).

$$\text{PDF} = \frac{e^{-(x-\mu)^2/2\sigma^2 \times \sqrt{2/\pi}}}{\sigma \times \left(\text{erf}\left(\mu + 1/\sqrt{2\sigma}\right) - \text{erf}\left(\mu - 1/\sqrt{2\sigma}\right)\right)}, \tag{10}$$

```
while i < particles do
   Initialize the position X_i and velocity V_i of each particle
   Calculate the fitness value of each particle fitness(i)
   i = i +1
end
Initialize the pBest_i = X_i
Initialize the gBest = min(pBest_i)
for g =1 to iterMax do
      for i =1 to particles do
            Update the V_i and X_i of the particles by Equations (1) and (2)
            Calculate the fitness value of the new particle fitness(i)
            if fitness(i) < fitness(pBest_i) then
               pBest_i= X_i
            end.
            if fitness(i) < fitness(gBest) then
               gBest= X_i
            end
end
Calculate the d_i by Equation (3)
Calculate the f by Equation (4)
Determine the evolutionary state S by Figure 1
w is adjusted according to Equation (5)
c_1 and c_2 are adjusted according to Table 1
if evolutionary state S == convergence state S3 then
         σ = σ_max − (σ_max − σ_min) × g/iterMax
         T^d = T^d + (X^d_max − X^d_min) × Gaussian(μ, σ^2)
         if fitness(T) < fitness(gBest) then
               gBest= T
         end
      end
end
```

Algorithm 2: The pseudo-code of APSO.

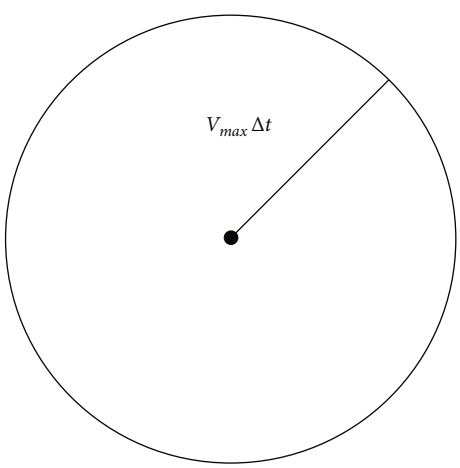


Figure 2: Rang of node movement.

$$\mathrm{CDF} = \int_{-\infty}^{x} \frac{e^{-(x-\mu)^2/2\sigma^2 \times \sqrt{2/\pi}}}{\sigma \times \left(\mathrm{erf}\left(\mu + 1/\sqrt{2\sigma}\right) - \mathrm{erf}\left(\mu - 1/\sqrt{2\sigma}\right)\right)} dx. \tag{11}$$

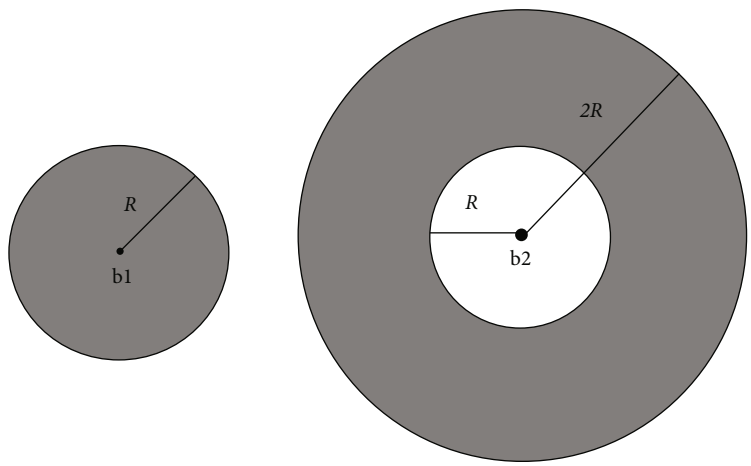
The CDF value range of the cumulative distribution function is (0,1), which is also the value range returned by the compact strategy. Taking the standard normal distribution as an example, PDF and CDF of the standard normal distribution are shown in Figure 4.

Another important content of the compact strategy is the iterative update of the PV disturbance vector. The compact strategy is based on comparison, and the winner and loser are obtained through a competitive game mechanism. Then, update it through the update iteration strategy. The update formulas are shown in Equations (12) and (13).

$$\mu^{g+1} = \mu^{g} + \frac{1}{N_p}(\mathrm{winner} - \mathrm{loser}), \tag{12}$$

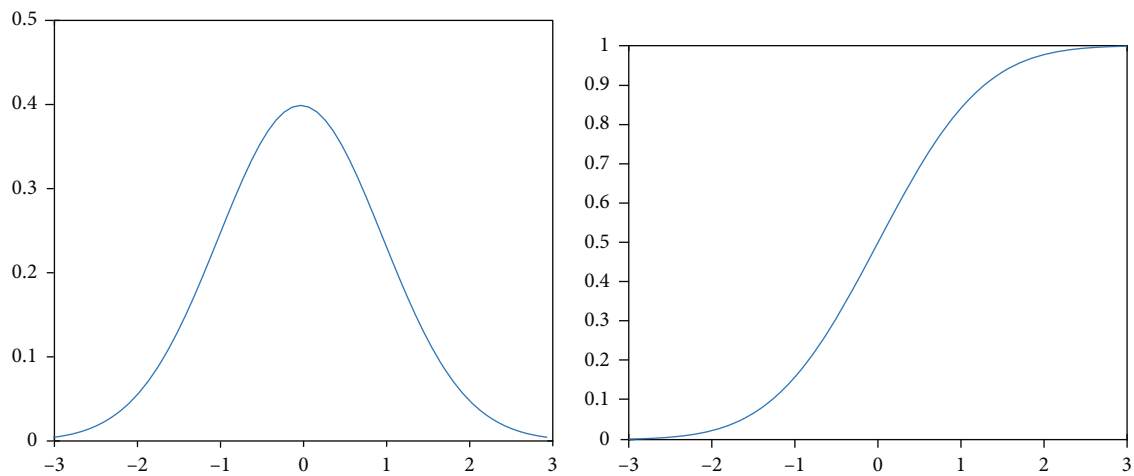
$$\sigma^{g+1} = \sqrt{(\sigma^{g})^2 + (\mu^{g})^2 + \frac{2}{N_p}\left(\mathrm{winner}^2 - \mathrm{loser}^2\right)}, \tag{13}$$

where g represents the current iteration times and N_p represents the number of virtual populations. The mean value μ in the PV disturbance vector is generally set to 0, and the standard deviation σ in the PV disturbance vector is generally set to 10 to avoid the contingency of the local optimum during initialization. After a large number of experiments, it has been proven that the effect achieved



(a) One-hop beacon node locations (b) Two-hop beacon node locations

Figure 3: One-hop and two-hop beacon node locations.



(a) Standard normal distribution PDF (b) Standard normal distribution CDF

Figure 4: PDF and CDF of standard normal distribution.

when the number of virtual populations N_p is set to 300 is the best. So N_p is generally set to 300.

3.2. Implementation of the Compact Strategy in APSO. The compact strategy is based on comparison. In other words, there must be a game mechanism of size two so that the winner and loser can be generated, and then, μ and σ in the PV disturbance vector can be updated. APSO can meet this condition through the comparison of fitness values, so APSO meets the primary conditions of the compact strategy, and the compact strategy can be combined with APSO to achieve the goal of reducing memory and improving operating speed. For example, if there are 30 particles in the particle swarm and the dimension to solve the problem is also 30 dimensions, then APSO needs 900 storage units to store the position information of each particle in each dimension. However, cAPSO only needs 60 storage units to store the particle swarm in each dimension using PV perturbation vectors, which significantly saves storage space. The saving of storage space reduces the number of reads and writes to the memory, and the speed of the algorithm will also be improved accordingly.

The cAPSO expresses the position of the particle swarm through the overall probability distribution, and the value range of the compact strategy is between (0,1), assuming that y is the return value of the compact strategy. Then, the return value of the compact strategy should be corresponded to the actual position range by Equation (14).

$$X = \frac{1}{2} \times (X_{\max} - X_{\min}) \times y + \frac{1}{2} \times (X_{\max} - X_{\min}). \quad (14)$$

The evolution factor f in the original APSO is calculated by Equations (3) and (4). However, the distribution probability is used to represent the position of the

```
Initialize the PV(μ, σ) disturbance vector parameter
Initialize a particle swarm position X = X_min + (X_max − X_min) × rand ()
Initialize a particle swarm velocity V = V_min + (V_max − V_min) × rand ()
Initialize pBest = X
Initialize gBest = X
for g =1 to iterMax do
        pBest = 1/2 × (X_max − X_min) × compact(μ, σ) + 1/2 × (X_max − X_min)
        Update the V and X of the particles by Equations (1) and (2)
        Calculate the fitness value of the new particle fitness(X)
        (winner, loser) = compare(fitness(pBest), fitness(X))
        Update the PV disturbance vector by Equations (12) and (13)
        Calculate the f by Equation (15)
        Determine the evolutionary state S by Figure 1
        w is adjusted according to Equation (5)
        c_1 and c_2 are adjusted according to Table 1
        if evolutionary state S == convergence state S3 then
           σ = σ_max − (σ_max − σ_min) × g/iterMax
           T^d = T^d + (X^d_max − X^d_min) × Gaussian(μ, σ^2)
           (winner, loser) = compare(fitness(gBest), fitness(X))
           gBest= winner
        end
end
```

Algorithm 3: The pseudo-code of cAPSO.

particle swarm in cAPSO, so the calculation of the evolution factor f is replaced by Equation (15).

$$f = \frac{\mu_{\text{mean}} - \mu_{\text{min}}}{\mu_{\text{max}} - \mu_{\text{min}}}. \tag{15}$$

The cAPSO compares the fitness value of the position of this iteration generated by the PV disturbance vector with the fitness value of the optimal position of the present individual and generates winner and loser. The generated winner and loser use the Equations (12) and (13) to update the disturbance vector PV and then generate the position probability distribution of the particle swarm in the next iteration.

The pseudo-code of cAPSO is shown in Algorithm 3.

4. The Performance Test of cAPSO

In this section, the cAPSO algorithm is mainly tested in 28 test functions of CEC2013 [45]. This paper compares the cAPSO with common heuristic algorithms and common compact algorithms. The 28 test functions in CEC2013 include 8 mixed functions, 15 multimodal functions, and 5 unimodal functions. These 28 functions are very representative. The 28 functions are represented by f1 to f28. Every experiment keeps the common parameters consistent during the comparison process to ensure the fairness of the comparison.

4.1. Performance Comparison of cAPSO and Common Heuristic Algorithms. In this section, the cAPSO is compared with genetic algorithm (GA) [46], differential evolution algorithm (DE) [47], whale optimization algorithm (WOA) [48], bat algorithm (BA) [49], and sine cosine algorithm (SCA) [50] on 28 test functions of CEC2013. At the same time, the overall performance of each algorithm compared with the cAPSO algorithm was measured at a significant level $\alpha = 0.05$ under the Wilcoxon signed rank test. Twenty tests are carried out on each test function, and then, the average value is taken to avoid the occurrence of chance. To ensure the fairness of the algorithm during testing, the dimension of the problem is set to 50, the number of populations is set to 60, and the number of iterations is set to 3000, and the search range requirement in CEC2013 is [-100,100]. Table 2 shows the performance comparison of cAPSO and common heuristic algorithms. In addition, Table 2 also shows that each algorithm is measured under the Wilcoxon signed rank test, with a significance level of $\alpha = 0.05$. The symbol "$>$" represents that the performance of the cAPSO is better than the other heuristic algorithm, the symbol "$=$" represents that the performance of the cAPSO is the same as the other heuristic algorithm, and the symbol "$<$" represents that the performance of the cAPSO is worse than the other heuristic algorithm. The last row of Table 2 summarizes the comparison results of all test functions.

Table 2 shows that compared with DE, the test performance of cAPSO is better than DE in 20 functions, the same as DE in 2 functions, and worse than DE in 6 functions. Compared with BA, the test performance of cAPSO is better than BA in 21 functions, the same as BA in 1 function, and worse than BA in 6 functions. Compared with GA, WOA, and SCA, cAPSO has the same performance as these three algorithms in function f8, but it is better than others in other functions. It can be seen that the performance of the cAPSO algorithm combined with the compact strategy is greatly improved compared with the common heuristic algorithm.

In order to further describe the effect of the algorithm, this paper uses the convergence curve for evaluation. Since the convergence of many algorithms is very similar, the performance is not obvious in the convergence curve, so this

Table 2: The performance comparison between cAPSO and common heuristic algorithms and the Wilcoxon signed rank test of each algorithm at the significance level $\alpha = 0.05$.

Function	GA		DE		BA		WOA		SCA		cAPSO
f1	1.61*E*-05	(>)	-1.40*E*-03	(=)	-1.39*E*-03	(>)	-1.36*E*-03	(>)	2.70*E*-04	(>)	-1.40*E*-03
f2	5.34*E*-09	(>)	2.39*E*-08	(>)	5.19*E*-06	(<)	8.10*E*-07	(>)	5.07*E*-08	(>)	7.33*E*-06
f3	7.38*E*-19	(>)	1.87*E*-10	(>)	5.10*E*-08	(<)	4.11*E*-10	(>)	1.09*E*-11	(>)	3.69*E*-09
f4	6.55*E*-05	(>)	1.10*E*-05	(>)	1.98*E*-04	(>)	6.18*E*-04	(>)	6.37*E*-04	(>)	7.64*E*-03
f5	8.99*E*-04	(>)	-9.86*E*-02	(>)	-9.96*E*-02	(>)	-8.07*E*-02	(>)	2.44*E*-03	(>)	-1.00*E*-03
f6	2.88*E*-04	(>)	-8.54*E*-02	(<)	-8.40*E*-02	(>)	-6.66*E*-02	(>)	1.21*E*-03	(>)	-8.20*E*-02
f7	6.09*E*-06	(>)	-6.67*E*-02	(>)	9.86*E*-02	(>)	-3.18*E*-02	(>)	-6.06*E*-02	(>)	-6.71*E*-02
f8	-6.79*E*-02	(=)	-6.79*E*-02	(=)	-6.79*E*-02	(=)	-6.79*E*-02	(=)	-6.79*E*-02	(=)	-6.79*E*-02
f9	-5.19*E*-02	(>)	-5.31*E*-02	(>)	-5.33*E*-02	(>)	-5.29*E*-02	(>)	-5.26*E*-02	(>)	-5.42*E*-02
f10	2.18*E*-04	(>)	-4.31*E*-02	(>)	-4.96*E*-02	(>)	-2.17*E*-02	(>)	3.42*E*-03	(>)	-4.98*E*-02
f11	2.11*E*-03	(>)	-2.79*E*-02	(>)	7.61*E*-02	(>)	4.02*E*-02	(>)	3.09*E*-02	(>)	-3.95*E*-02
f12	1.96*E*-03	(>)	1.22*E*-02	(>)	8.98*E*-02	(>)	6.50*E*-02	(>)	4.56*E*-02	(>)	7.37*E*-01
f13	2.03*E*-03	(>)	2.29*E*-02	(<)	1.16*E*-03	(>)	8.43*E*-02	(>)	5.53*E*-02	(>)	2.79*E*-02
f14	1.64*E*-04	(>)	2.13*E*-03	(>)	8.93*E*-03	(>)	8.92*E*-03	(>)	1.36*E*-04	(>)	1.98*E*-03
f15	1.61*E*-04	(>)	1.45*E*-04	(>)	8.81*E*-03	(<)	1.15*E*-04	(>)	1.44*E*-04	(>)	1.01*E*-04
f16	2.05*E*-02	(>)	2.04*E*-02	(>)	2.02*E*-02	(<)	2.03*E*-02	(>)	2.03*E*-02	(>)	2.02*E*-02
f17	5.26*E*-03	(>)	5.14*E*-02	(>)	2.65*E*-03	(>)	1.47*E*-03	(>)	1.31*E*-03	(>)	3.51*E*-02
f18	5.29*E*-03	(>)	8.74*E*-02	(>)	2.94*E*-03	(>)	1.51*E*-03	(>)	1.40*E*-03	(>)	8.54*E*-02
f19	2.59*E*-07	(>)	5.21*E*-02	(>)	5.68*E*-02	(>)	6.74*E*-02	(>)	4.29*E*-04	(>)	5.02*E*-02
f20	6.25*E*-02	(>)	6.23*E*-02	(<)	6.25*E*-02	(>)	6.25*E*-02	(>)	6.24*E*-02	(>)	6.25*E*-02
f21	1.26*E*-04	(>)	1.23*E*-03	(<)	1.53*E*-03	(>)	1.86*E*-03	(>)	4.67*E*-03	(>)	1.60*E*-03
f22	1.86*E*-04	(>)	3.63*E*-03	(<)	1.27*E*-04	(>)	1.28*E*-04	(>)	1.54*E*-04	(>)	4.59*E*-03
f23	1.81*E*-04	(>)	1.55*E*-04	(>)	1.22*E*+04	(<)	1.39*E*-04	(>)	1.61*E*-04	(>)	1.26*E*-04
f24	2.08*E*-03	(>)	1.37*E*-03	(>)	1.45*E*-03	(>)	1.41*E*-03	(>)	1.43*E*-03	(>)	1.36*E*-03
f25	1.76*E*-03	(>)	1.48*E*-03	(>)	1.47*E*-03	(<)	1.53*E*-03	(>)	1.55*E*-03	(>)	1.51*E*-03
f26	1.79*E*-03	(>)	1.52*E*-03	(<)	1.69*E*-03	(>)	1.65*E*-03	(>)	1.59*E*-03	(>)	1.64*E*-03
f27	4.68*E*-03	(>)	3.29*E*-03	(>)	3.50*E*-03	(>)	3.54*E*-03	(>)	3.66*E*-03	(>)	3.18*E*-03
f28	1.72*E*-04	(>)	2.77*E*-03	(>)	1.08*E*-04	(>)	8.79*E*-03	(>)	6.84*E*-03	(>)	2.73*E*-03
>/=/<	27/1/0		20/2/6		21/1/6		27/1/0		27/1/0		

paper selects several scattered representative curves for display. Figure 5 shows the convergence process of the algorithm on some test functions. The horizontal axis represents iteration times, and the vertical axis represents the fitness value of different algorithms. The smaller the fitness value, the better the performance on this test function. It can be seen from Figure 5 that the performance of the proposed cAPSO algorithm is better than other heuristic algorithms on f4, f9, f11, f17, and f27. But the performance of the test function on f13 and f22 is worse than DE, and the performance on f17 is not as good as BA.

4.2. Performance Comparison of cAPSO and Common Compact Algorithms. In this section, the cAPSO is compared with cPSO [41], cABC [38], cSCA [39], and cBA [40] on 28 test functions of CEC2013. At the same time, the overall performance of each algorithm compared with the cAPSO algorithm was measured at a significant level $\alpha = 0.05$ under the Wilcoxon signed rank test. Twenty tests are carried out on each test function, and then, the average value is taken to avoid the occurrence of chance. To ensure the fairness of the algorithm during testing, the dimension of the problem is set to 50, the number of populations is set to 60, the number of iterations is set to 3000, the virtual number of populations is set to 300, and the search range requirement in CEC2013 is [-100,100]. Table 3 shows the performance comparison of cAPSO and common compact algorithms. In addition, Table 3 also shows that each algorithm is measured under the Wilcoxon signed rank test, with a significance level of $\alpha = 0.05$. The symbols ">," "=," and "<" have the same meaning as in Section 4.1. The last row of Table 3 summarizes the comparison results of all test functions.

Table 3 shows that compared with cPSO, the test performance of cAPSO is better than cPSO in 24 functions, the same as cPSO in 2 functions, and worse than cPSO in 2 functions. Compared with cBA, the test performance of cAPSO is better than cBA in 20 functions, the same as cBA in 3 functions, and worse than cBA in 5 functions.

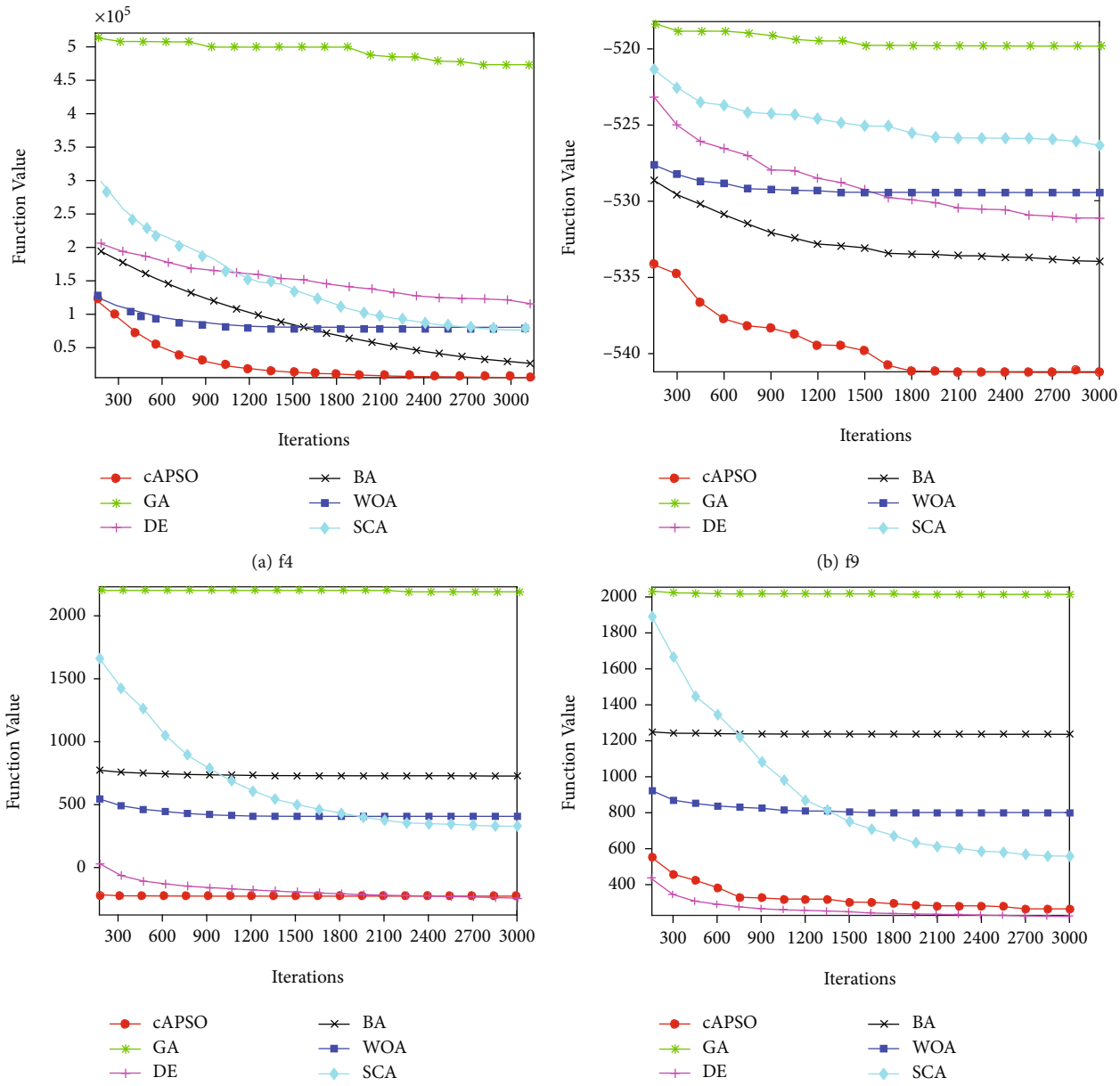


Figure 5: Continued.

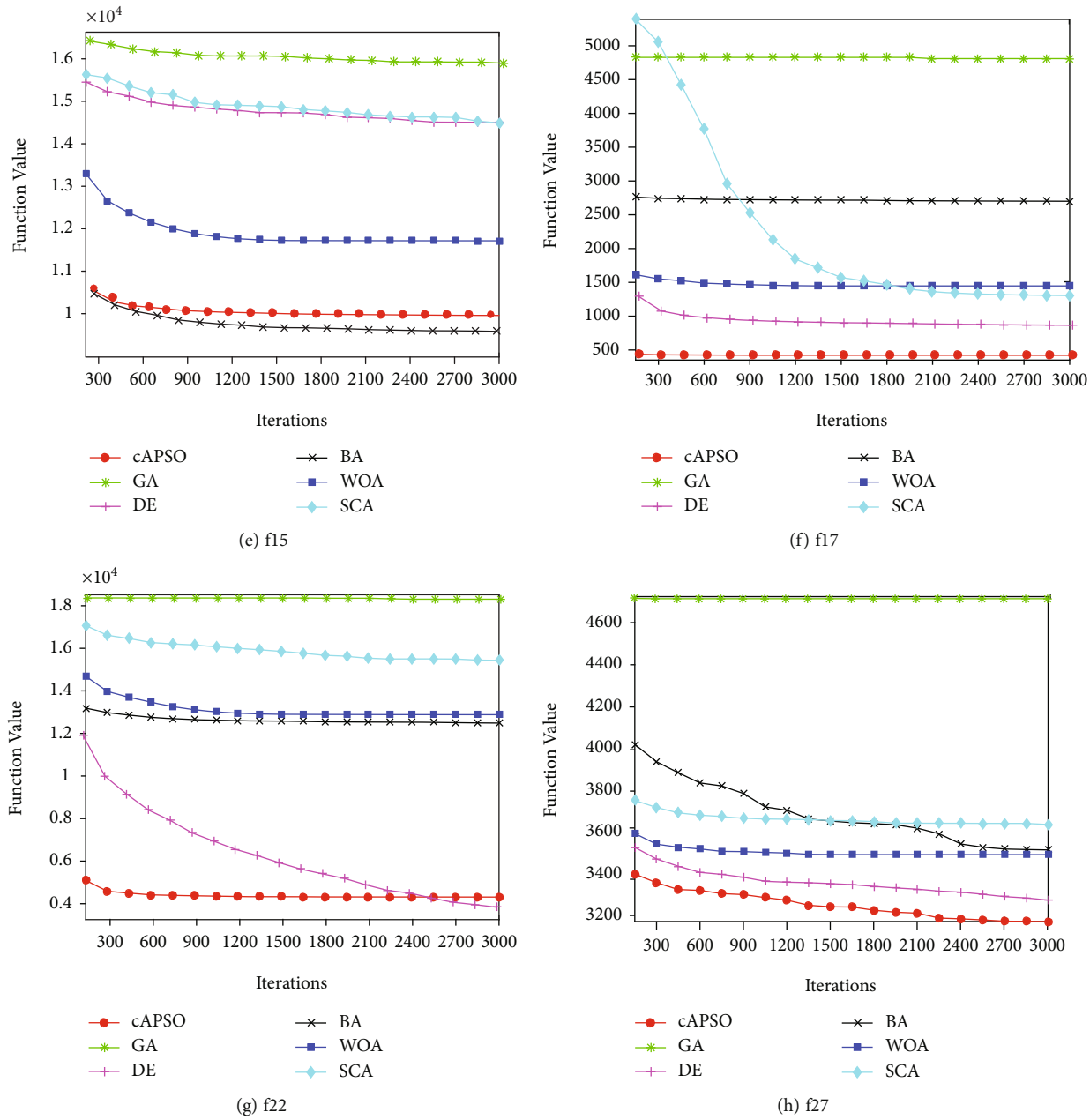


Figure 5: Performance comparison between cAPSO and common heuristic algorithms.

Compared with cABC and cSCA, cAPSO has the same performance as these two algorithms in function f8, but it is better than these two algorithms in other functions. It can be seen that cAPSO has strong competitiveness in compact algorithms and has advantages over other compact algorithms in performance.

As in Section 4.1, Figure 6 shows representative convergence curves of the proposed cAPSO and other compact algorithms. It can be seen from Figure 6 that the performance of the proposed cAPSO algorithm is better than other compact algorithms on f12, f15, f20, and f25. But the performance of the test function on f16 is not as good as cBA, and the performance on f23 is worse than cPSO.

5. Application of cAPSO in Mobile Sensor Localization

This section mainly applies cAPSO to the mobile sensor localization technology MCL and compares it with the original MCL, WOA-based MCL, and BA-based MCL under different anchor node numbers and different communication radius. It takes a lot of time and computing power to directly find a position with a small error through the MCL technology. A position with a large error is initially obtained through MCL technology, and then, cAPSO is applied around the obtained position for further optimization. The cAPSO broadcasts nodes around it, and the broadcast nodes move according to the idea of the cAPSO. The position after

Table 3: The performance comparison between cAPSO and common compact algorithms and the Wilcoxon signed rank test of each algorithm at the significance level $\alpha = 0.05$.

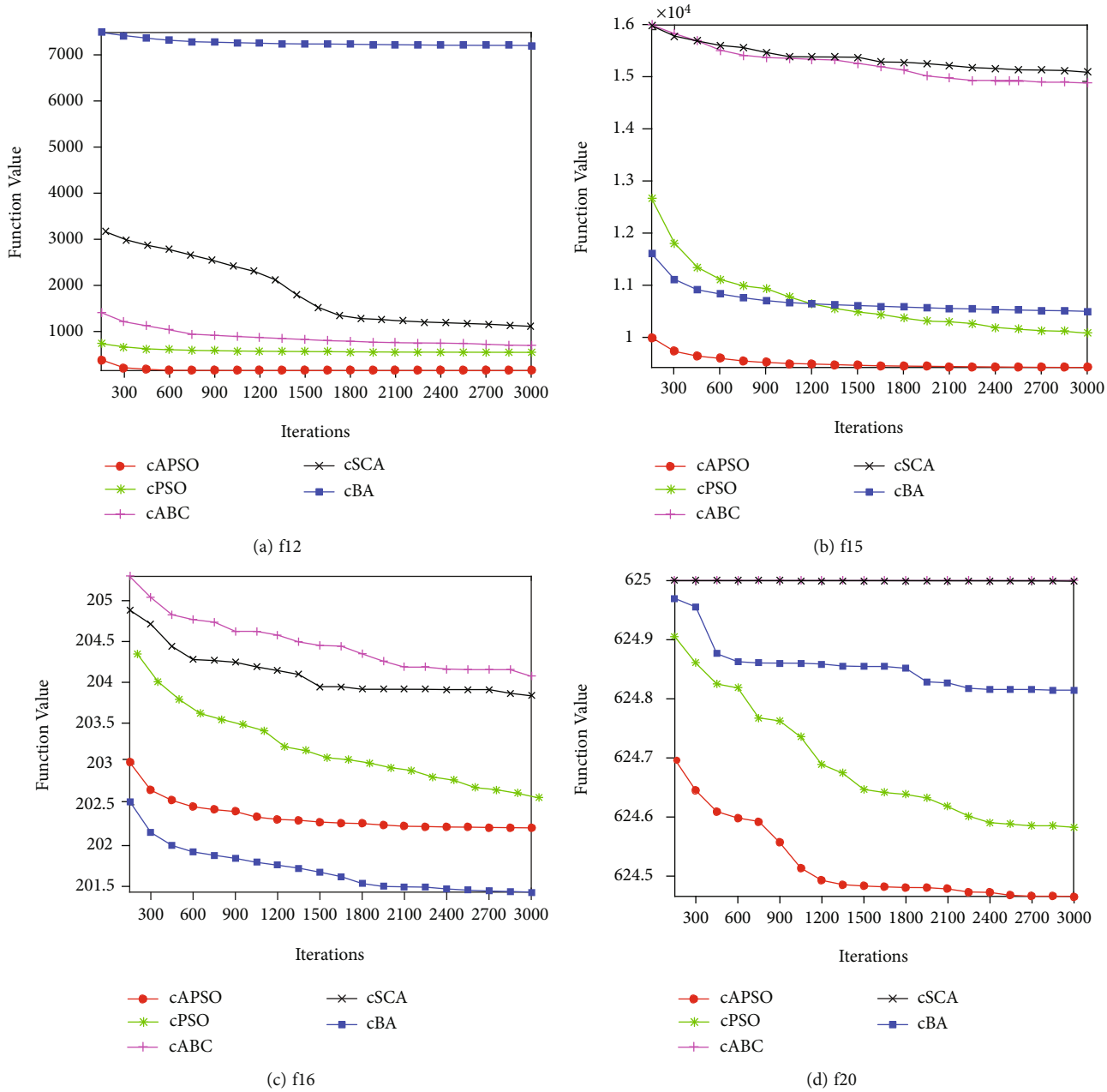
Function	cPSO		cABC		cSCA		cBA		cAPSO
f1	-1.40*E*+03	(=)	2.50*E*+04	(>)	5.20*E*+04	(>)	-1.40*E*+03	(=)	-1.40*E*+03
f2	2.54*E*+07	(>)	7.05*E*+08	(>)	1.02*E*+09	(>)	3.31*E*+06	(<)	8.07*E*+06
f3	7.45*E*+09	(>)	2.07*E*+11	(>)	2.52*E*+15	(>)	5.34*E*+09	(>)	3.78*E*+09
f4	3.76*E*+03	(<)	1.44*E*+05	(>)	1.84*E*+05	(>)	2.87*E*+05	(>)	8.68*E*+03
f5	-9.88*E*+02	(>)	5.94*E*+03	(>)	1.44*E*+04	(>)	-1.00*E*+03	(=)	-1.00*E*+03
f6	-8.05*E*+02	(>)	1.28*E*+03	(>)	3.13*E*+03	(>)	-8.49*E*+02	(<)	-8.30*E*+02
f7	-5.68*E*+02	(>)	-4.55*E*+02	(>)	8.72*E*+03	(>)	3.26*E*+12	(>)	-6.75*E*+02
f8	-6.79*E*+02	(=)	-6.79*E*+02	(=)	-6.79*E*+02	(=)	-6.79*E*+02	(=)	-6.79*E*+02
f9	-5.37*E*+02	(>)	-5.24*E*+02	(>)	-5.26*E*+02	(>)	-5.11*E*+02	(>)	-5.41*E*+02
f10	-4.76*E*+02	(>)	4.31*E*+03	(>)	7.25*E*+03	(>)	-4.99*E*+02	(<)	-4.98*E*+02
f11	3.33*E*+02	(>)	4.32*E*+02	(>)	4.35*E*+02	(>)	2.84*E*+03	(>)	-3.96*E*+02
f12	5.47*E*+02	(>)	6.96*E*+02	(>)	9.21*E*+02	(>)	7.20*E*+03	(>)	1.52*E*+02
f13	4.48*E*+02	(>)	8.20*E*+02	(>)	1.14*E*+03	(>)	8.65*E*+03	(>)	2.65*E*+02
f14	9.05*E*+03	(>)	1.45*E*+04	(>)	1.60*E*+04	(>)	8.59*E*+03	(>)	2.11*E*+03
f15	1.01*E*+04	(>)	1.50*E*+04	(>)	1.52*E*+04	(>)	1.06*E*+04	(>)	9.48*E*+03
f16	2.03*E*+02	(>)	2.04*E*+02	(>)	2.04*E*+02	(>)	2.01*E*+02	(<)	2.02*E*+02
f17	1.06*E*+03	(>)	1.91*E*+03	(>)	1.45*E*+03	(>)	1.50*E*+04	(>)	3.51*E*+02
f18	1.07*E*+03	(>)	2.28*E*+03	(>)	1.59*E*+03	(>)	1.65*E*+04	(>)	9.16*E*+02
f19	5.47*E*+02	(>)	6.82*E*+04	(>)	1.86*E*+04	(>)	8.46*E*+02	(>)	5.02*E*+02
f20	6.25*E*+02	(>)	6.25*E*+02	(>)	6.25*E*+02	(>)	6.25*E*+02	(>)	6.24*E*+02
f21	1.61*E*+03	(>)	5.58*E*+03	(>)	5.10*E*+03	(>)	1.62*E*+03	(>)	1.58*E*+03
f22	1.26*E*+04	(>)	1.66*E*+04	(>)	1.81*E*+04	(>)	1.10*E*+04	(>)	4.31*E*+03
f23	1.29*E*+04	(<)	1.72*E*+04	(>)	1.68*E*+04	(>)	1.40*E*+04	(>)	1.32*E*+04
f24	1.42*E*+03	(>)	1.42*E*+03	(>)	1.52*E*+03	(>)	1.69*E*+03	(>)	1.37*E*+03
f25	1.57*E*+03	(>)	1.60*E*+03	(>)	1.57*E*+03	(>)	1.65*E*+03	(>)	1.50*E*+03
f26	1.61*E*+03	(>)	1.64*E*+03	(>)	1.71*E*+03	(>)	1.40*E*+03	(<)	1.59*E*+03
f27	3.54*E*+03	(>)	3.65*E*+03	(>)	3.81*E*+03	(>)	5.28*E*+03	(>)	3.18*E*+03
f28	4.16*E*+03	(>)	7.40*E*+03	(>)	8.71*E*+03	(>)	9.08*E*+04	(>)	2.18*E*+03
>/=/<	24/2/2		27/1/0		27/1/0		20/3/5		

move is compared with the historical optimal position to update the optimal position. After a certain number of iterations, an optimal position is obtained as the final position of the MCL positioning technology. Mobile sensor positioning technology is a technology that uses the information of anchor nodes to estimate the location of unknown nodes, so the position error becomes the key to the technology. The smaller the position error, the more conducive to the accuracy of the information. Applying heuristic algorithms to MCL technology can better locate unknown nodes and reduce localization errors. The more accurate position coordinates are iteratively updated by the error function. The error function is defined as Equation (16).

$$\text{error} = \frac{\sum_{j=1}^{M}\left(\sum_{i=1,i\neq j}^{N}\sqrt{\left(x_i - x_j\right)^2 + \left(y_i - y_j\right)^2} - D_{ij}\right)}{M}, \tag{16}$$

where (x_i, y_i) represents the estimated location of the unknown node i, (x_j, y_j) represents the location of the anchor node j, M represents the total number of unknown nodes, and N represents the total number of anchor nodes. D_{ij} represents the distance between each unknown node i and each anchor node j. This paper assumes that the anchor node j can obtain the distance between the anchor node j and the unknown node i by the strength of the signal received from the unknown node i. The smaller the error value, the higher the accuracy of localization.

5.1. Influence Analysis of Different Anchor Nodes. In the simulation experiment, 200 nodes are randomly scattered within a range of 200 m × 200 m. The number of anchor nodes is 5, 10, 15, 20, 25, 30, 35, and 40, so the number of unknown nodes is 195, 190, 185, 180, 175, 170, 165, and 160. The communication radius is set to 50 m, and the simulation experiments with the same parameters were tested 20 times, and the average value was calculated as the error



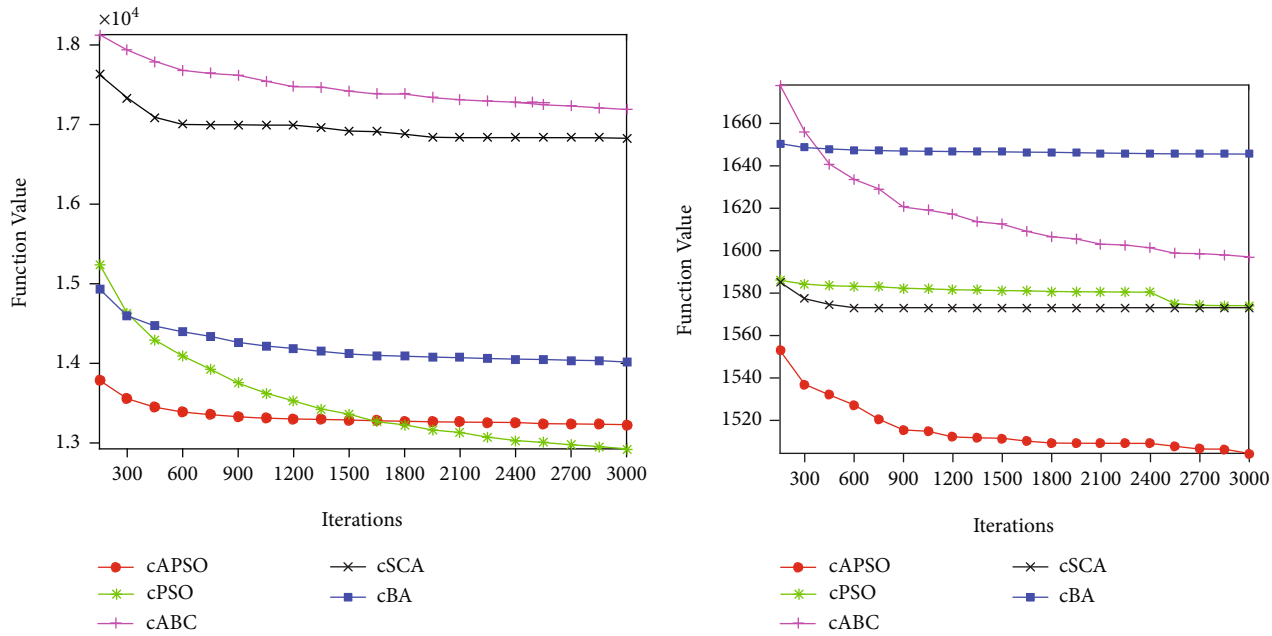
(a) f12

(b) f15

(c) f16

(d) f20

Figure 6: Continued.



(e) f23 (f) f25

Figure 6: Performance comparison between cAPSO and common compact algorithms.

Table 4: Experimental results of the localization error of different anchor nodes.

Anchor	MCL		cAPSO MCL		WOA MCL		BA MCL	
	Mean	Std	Mean	Std	Mean	Std	Mean	Std
5	22.0428	2.4357	12.5510	2.4495	12.6486	2.4621	12.5604	2.4466
10	15.5169	1.2365	6.8567	1.0768	6.9593	1.0758	6.8692	1.0787
15	11.4610	0.8525	3.4532	0.7155	3.5635	0.7134	3.4695	0.7231
20	3.4568	0.7152	2.0050	0.4179	2.1148	0.4147	2.0199	0.4179
25	7.6420	0.5372	1.1901	0.3881	1.3114	0.3877	1.2074	0.3878
30	1.2074	0.3878	0.6402	0.1856	0.7603	0.1832	0.6591	0.1852

Table 5: Experimental results of the localization error of different communication radius.

Radius	MCL		cAPSO MCL		WOA MCL		BA MCL	
	Mean	Std	Mean	Std	Mean	Std	Mean	Std
10	17.7767	3.7450	8.4470	3.2895	8.5827	3.2828	8.4567	3.2881
20	13.2522	0.7495	4.9535	0.5931	5.0626	0.5932	4.9637	0.5932
30	9.5178	0.5585	2.7765	0.4599	2.8918	0.4639	2.7952	0.4622
40	7.8710	0.4849	1.8598	0.3892	1.9788	0.3893	1.8773	0.3889
50	1.8773	0.3889	1.1117	0.3746	1.2293	0.3744	1.1303	0.3743

result. Apply cAPSO, WOA, and BA algorithms to the simulation experiment, respectively. The experimental results are shown in Table 4.

Table 4 shows that when the total number of nodes is 200 and the communication radius is 50 m, the larger the number of anchor nodes, the more accurate the location of unknown nodes and the smaller the error value. Table 4 clearly shows that the positioning error after optimization by the heuristic algorithm is much smaller than the original MCL. In the combined different heuristic algorithms, it can be clearly seen that although the cAPSO algorithm is not better than other algorithms in standard deviation every time, and it has better results than other heuristic algorithms in error mean.

5.2. Influence Analysis of Different Communication Radius. The simulation experiment randomly scatter 200 nodes in the range of 200 m × 200 m, the number of anchor nodes is set to 30, and the communication radius is set to 10 m, 20 m, 30 m, 40 m, and 50 m. The simulation experiments with the same parameters were tested 20 times, and the average value was calculated as the error result. Apply cAPSO,

WOA, and BA algorithms to the simulation experiment, respectively. The experimental results are shown in Table 5.

Table 5 shows that when the total number of nodes is 200 and the number of anchor nodes is 30, the greater the communication radius, the more accurate the location of unknown nodes, and the smaller the error value. Table 5 clearly shows that the positioning error after optimization by the heuristic algorithm is much smaller than that of the original MCL. In the combined different heuristic algorithms, it can also be clearly seen that the cAPSO algorithm is better than other heuristic algorithms in the comparison of the error mean.

6. Conclusion

In this paper, an improved APSO algorithm combined with compact strategy is proposed and applied to mobile sensor localization. The compact strategy no longer stores the position of each particle in each dimension but describes the distribution characteristics of the particles in each dimension through the operation of probability model. The compact strategy can reduce the use of memory very well. This paper tests the performance of cAPSO on 28 test functions of CEC2013 and compares it with the common heuristic algorithms GA, DE, BA, WOA, and SCA and the common compact strategy heuristic algorithms cPSO, cABC, cSCA, and cBA. The comparison results show that cAPSO has better test performance. Finally, cAPSO is applied to mobile sensor localization technology MCL, and it is also compared with WOA-based MCL and BA-based MCL. The results show that MCL based on cAPSO is more effective in solving this problem.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare no conflict of interest.

Acknowledgments

This project was funded by the National Key Research and Development Program of China under Grant No. 11974373.

Supplementary Materials

Reference URL of compact strategy thought: https://blog.csdn.net/Liu_Ning_666/article/details/118308477. Reference URL of many improved PSO algorithms: https://blog.csdn.net/Liu_Ning_666/article/details/120174723 *(Supplementary Materials)*

References

[1] I. Boussaïd, J. Lepagnot, and P. Siarry, "A survey on optimization metaheuristics," *Information Sciences*, vol. 237, pp. 82–117, 2013.

[2] L. Tonutti, B. Dalla Costa, H. Decolatti, G. Mendow, and C. Querini, "Determination of kinetic constants for glycerol acetylation by particle swarm optimization algorithm," *Chemical Engineering Journal*, vol. 424, article 130408, 2021.

[3] A. Kokhanovskiy, E. Kuprikov, A. Bednyakova, I. Popkov, S. Smirnov, and S. Turitsyn, "Inverse design of mode-locked fiber laser by particle swarm optimization algorithm," *Scientific Reports*, vol. 11, pp. 1–9, 2021.

[4] F. Zhou, S. Bian, and D. Wang, "Optimal dispatching of microgrid based on improved particle swarm optimization," *Journal of Physics: Conference Series. IOP Publishing*, vol. 1871, article 012141, 2021.

[5] E. Bas, E. Egrioglu, and E. Kolemen, "Training simple recurrent deep artificial neural network for forecasting using particle swarm optimization," *Granular Computing*, pp. 1–10, 2021.

[6] J. Jafari-Asl, B. S. Kashkooli, and M. Bahrami, "Using particle swarm optimization algorithm to optimally locating and controlling of pressure reducing valves for leakage minimization in water distribution systems," *Sustainable Water Resources Management*, vol. 6, no. 4, pp. 1–11, 2020.

[7] H. C. Huang, S. C. Chu, J. S. Pan, C. Y. Huang, and B. Y. Liao, "Tabu search based multi-watermarks embedding algorithm with multiple description coding," *Information Sciences*, vol. 181, no. 16, pp. 3379–3396, 2011.

[8] Z. Meng and J. S. Pan, "QUasi-Affine TRansformation Evolution with External ARchive (QUATRE-EAR): an enhanced structure for differential evolution," *Knowledge-Based Systems*, vol. 155, pp. 35–53, 2018.

[9] D. H. Wolpert and W. G. Macready, "No free lunch theorems for optimization," *IEEE Transactions on Evolutionary Computation*, vol. 1, no. 1, pp. 67–82, 1997.

[10] T. Joyce and J. M. Herrmann, "A review of no free lunch theorems, and their implications for metaheuristic optimisation," *Nature-inspired algorithms and applied optimization*, vol. 44, pp. 27–51, 2018.

[11] P. C. Song, S. C. Chu, J. S. Pan, and H. Yang, "Simplified Phasmatodea population evolution algorithm for optimization," *Complex & Intelligent Systems*, pp. 1–19, 2021.

[12] J. S. Pan, N. Liu, S. C. Chu, and T. Lai, "An efficient surrogate-assisted hybrid optimization algorithm for expensive optimization problems," *Information Sciences*, vol. 561, pp. 304–325, 2021.

[13] J. Wang, B. Pan, Q. R. Wang, and Q. Ding, "A chaotic key expansion algorithm based on genetic algorithm," *Journal of Information Hiding and Multimedia Signal Processing*, vol. 10, pp. 289–299, 2019.

[14] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proceedings of ICNN'95-international conference on neural networks*, vol. 4, pp. 1942–1948, Perth, WA, Australia, 1995.

[15] R. Eberhart and J. Kennedy, "A new optimizer using particle swarm theory," in *MHS'95. Proceedings of the Sixth International Symposium on Micro Machine and Human Science*, pp. 39–43, Nagoya, Japan, 1995.

[16] M. Clerc and J. Kennedy, "The particle swarm - explosion, stability, and convergence in a multidimensional complex space," *IEEE Transactions on Evolutionary Computation*, vol. 6, no. 1, pp. 58–73, 2002.

[17] R. Mendes, J. Kennedy, and J. Neves, "The fully informed particle swarm: simpler, maybe better," *IEEE Transactions on Evolutionary Computation*, vol. 8, no. 3, pp. 204–210, 2004.

[18] J. J. Liang, A. K. Qin, P. N. Suganthan, and S. Baskar, "Comprehensive learning particle swarm optimizer for global optimization of multimodal functions," *IEEE Transactions on Evolutionary Computation*, vol. 10, no. 3, pp. 281–295, 2006.

[19] J. Zhou, Z. Ji, and L. Shen, "Simplified intelligence single particle optimization based neural network for digit recognition," in *2008 Chinese Conference on Pattern Recognition*, pp. 1–5, Beijing, China, 2008.

[20] Z. H. Zhan, J. Zhang, Y. Li, and H. S. H. Chung, "Adaptive particle swarm optimization," *IEEE transactions on systems, man, and cybernetics. Part B (Cybernetics)*, vol. 39, no. 6, pp. 1362–1381, 2009.

[21] P. Zhang, Z. Cui, Y. Wang, and S. Ding, "Application of BPNN optimized by chaotic adaptive gravity search and particle swarm optimization algorithms for fault diagnosis of electrical machine drive system," *Electrical Engineering*, pp. 1–13, 2021.

[22] Y. Marinakis, M. Marinaki, and A. Migdalas, "A multi-adaptive particle swarm optimization for the vehicle routing problem with time windows," *Information Sciences*, vol. 481, pp. 311–329, 2019.

[23] J. S. Pan, X. Wang, S. C. Chu, and T. Nguyen, "A multi-group grasshopper optimisation algorithm for application in capacitated vehicle routing problem," *Data Science and Pattern Recognition*, vol. 4, pp. 41–56, 2020.

[24] Z. J. Wang, Z. H. Zhan, W. J. Yu et al., "Dynamic group learning distributed particle swarm optimization for large-scale optimization and its application in cloud workflow scheduling," *IEEE transactions on cybernetics*, vol. 50, pp. 2715–2729, 2019.

[25] H. Wu, J. Liu, Z. Dong, and Y. Liu, "A hybrid mobile node localization algorithm based on adaptive MCB-PSO approach in wireless sensor networks," *Wireless Communications and Mobile Computing*, vol. 2020, 17 pages, 2020.

[26] R. Abdul Razak, S. Sukumar, and H. Chung, "Scalar field estimation with mobile sensor networks," *International Journal of Robust and Nonlinear Control*, vol. 31, no. 9, pp. 4287–4305, 2021.

[27] Y. Lv, L. Chen, D. Zhang, and J. Liu, "Research on indoor mobile node localization based on wireless sensor networks," *Engineering Letters*, vol. 29, 2021.

[28] C. I. Wu, H. Y. Kung, C. H. Chen, and L. C. Kuo, "An intelligent slope disaster prediction and monitoring system based on WSN and ANP," *Expert Systems with Applications*, vol. 41, no. 10, pp. 4554–4562, 2014.

[29] J. S. Pan, L. Kong, T. W. Sung, P. W. Tsai, and V. Snášel, "α-Fraction first strategy for hierarchical model in wireless sensor networks," *Journal of Internet Technology*, vol. 19, pp. 1717–1726, 2018.

[30] J. S. Pan, L. Kong, T. W. Sung, P. W. Tsai, and V. Snášel, "A clustering scheme for wireless sensor networks based on genetic algorithm and dominating set," *Journal of Internet Technology*, vol. 19, pp. 1111–1118, 2018.

[31] Q. W. Chai, S. C. Chu, J. S. Pan, and W. M. Zheng, "Applying adaptive and self assessment fish migration optimization on localization of wireless sensor network on 3-D Te rrain," *Journal of Information Hiding and Multimedia Signal Processing*, vol. 11, pp. 90–102, 2020.

[32] Z. G. Du, J. S. Pan, S. C. Chu, H. J. Luo, and P. Hu, "Quasi-affine transformation evolutionary algorithm with communication schemes for application of RSSI in wireless sensor networks," *IEEE Access*, vol. 8, pp. 8583–8594, 2020.

[33] J. Wu, M. Xu, F. F. Liu, M. Huang, L. Ma, and Z. M. Lu, "Solar wireless sensor network routing algorithm based on multi-objective particle swarm optimization," *Journal of Information Hiding and Multimedia Signal Processing*, vol. 12, pp. 1–11, 2021.

[34] X. Xue and J. S. Pan, "A compact co-evolutionary algorithm for sensor ontology meta-matching," *Knowledge and Information Systems*, vol. 56, no. 2, pp. 335–353, 2018.

[35] S. X. Yang and C. Luo, "A neural network approach to complete coverage path planning," *IEEE transactions on systems, man, and cybernetics. Part B (Cybernetics)*, vol. 34, no. 1, pp. 718–724, 2004.

[36] G. Cheung, W. T. Tan, and T. Yoshimura, "Real-time video transport optimization using streaming agent over 3G wireless networks," *IEEE Transactions on Multimedia*, vol. 7, no. 4, pp. 777–785, 2005.

[37] P. G. Norman, "The new AP101S general-purpose computer (GPC) for the space shuttle," *Proceedings of the IEEE*, vol. 75, no. 3, pp. 308–319, 1987.

[38] T. K. Dao, T. S. Pan, T. Nguyen, and S. C. Chu, *A Compact Artificial Bee Colony Optimization for Topology Control Scheme in Wireless Sensor Networks*, 2015.

[39] A. M. Wazwaz, "The sine-cosine method for obtaining solutions with compact and noncompact structures," *Applied Mathematics and Computation*, vol. 159, no. 2, pp. 559–576, 2004.

[40] T. Nguyen, J. S. Pan, and T. K. Dao, "A compact bat algorithm for unequal clustering in wireless sensor networks," *Applied Sciences*, vol. 9, no. 10, p. 1973, 2019.

[41] F. Neri, E. Mininno, and G. Iacca, "Compact particle swarm optimization," *Information Sciences*, vol. 239, pp. 96–121, 2013.

[42] J. Eriksson, L. Girod, B. Hull, R. Newton, S. Madden, and H. Balakrishnan, "The pothole patrol: using a mobile sensor network for road surface monitoring," in *Proceedings of the 6th international conference on Mobile systems, applications, and services*, pp. 29–39, New York, USA, 2008.

[43] N. Heo and P. K. Varshney, "Energy-efficient deployment of intelligent mobile sensor networks," *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, vol. 35, pp. 78–92, 2005.

[44] L. Hu and D. Evans, "Localization for mobile sensor networks," in *Proceedings of the 10th annual international conference on Mobile computing and networking*, pp. 45–57, New York, USA, 2004.

[45] J. Liang, B. Qu, P. Suganthan, and A. G. Hernández-Díaz, "Problem definitions and evaluation criteria for the CEC 2013 special session on real-parameter optimization," *Computational Intelligence Laboratory, Zhengzhou University, Zhengzhou, China and Nanyang Technological University, Singapore, Technical Report*, vol. 201212, pp. 281–295, 2013.

[46] S. Mirjalili, "Genetic algorithm," in *Evolutionary algorithms and neural networks*, pp. 43–55, Springer, 2019.

[47] S. Das and P. N. Suganthan, "Differential evolution: a survey of the state-of-the-art," *IEEE Transactions on Evolutionary Computation*, vol. 15, pp. 4–31, 2011.

[48] S. Mirjalili and A. Lewis, "The whale optimization algorithm," *Advances in Engineering Software*, vol. 95, pp. 51–67, 2016.

[49] X. S. Yang and X. He, "Bat algorithm: literature review and applications," *International Journal of Bio-inspired computation*, vol. 5, no. 3, pp. 141–149, 2013.

[50] S. M. Mirjalili, S. Z. Mirjalili, S. Saremi, and S. Mirjalili, "Sine cosine algorithm: theory, literature review, and application in designing bend photonic crystal waveguides," *Nature-inspired optimizers*, vol. 811, pp. 201–217, 2020.

Hindawi
Wireless Communications and Mobile Computing
Volume 2021, Article ID 8213946, 10 pages
https://doi.org/10.1155/2021/8213946

Research Article

Real-Time Precise Human-Computer Interaction System Based on Gaze Estimation and Tracking

Junhao Huang,[1,2] **Zhicheng Zhang**,[3] **Guoping Xie**,[4] **and Hui He**[1]

[1]*Advanced Institute of Natural Sciences, Beijing Normal University, Zhuhai 519087, China*
[2]*Engineering Research Center of Intelligent Technology and Educational Application, Ministry of Education, Beijing 100875, China*
[3]*Department of Radiation Oncology, Stanford University, Stanford, CA 94305, USA*
[4]*State Information Center, Beijing, China*

Correspondence should be addressed to Hui He; 986293685@qq.com

Received 22 August 2021; Revised 5 October 2021; Accepted 21 October 2021; Published 8 November 2021

Academic Editor: Chi-Hua Chen

Noncontact human-computer interaction has an important value in wireless sensor networks. This work is aimed at achieving accurate interaction on a computer based on auto eye control, using a cheap webcam as the video source. A real-time accurate human-computer interaction system based on eye state recognition, rough gaze estimation, and tracking is proposed. Firstly, binary classification of the eye states (opening or closed) is carried on using the SVM classification algorithm with HOG features of the input eye image. Second, rough appearance-based gaze estimation is implemented based on a simple CNN model. And the head pose is estimated to judge whether the user is facing the screen or not. Based on these recognition results, noncontact mouse control and character input methods are designed and developed to replace the standard mouse and keyboard hardware. Accuracy and speed of the proposed interaction system are evaluated by four subjects. The experimental results show that users can use only a common monocular camera to achieve gaze estimation and tracking and to achieve most functions of real-time precise human-computer interaction on the basis of auto eye control.

1. Introduction

The wireless sensor network (WSN) consists of many sensors like visual sensors, thermal sensors, and various others. Sensor nodes are widely used for nonstop sensing, event detection, position sensing, and many other things, including helping the disabled with interfaces [1]. Capturing the eye tracking signals can help the disabled improve their quality of life by noncontact human-computer communicating with visual sensors; for example, this can be applied to the eye control wheelchair for the disabled [2]. Eye tracking refers to eye scanning, gaze, blinking, etc. And eye movement is closely related to brain activity. Therefore, the process of human learning and cognition can be studied through eye movement [3]. This is how the eye tracking technique was born. Eye tracking can obtain the focus of sight in real time and be applied to analyze the user's eye movement in reading [3] or in a critical state [4], so as to infer the user's content of interest in reality [5, 6]. Furthermore, it can also be used for the prediction and treatment of patients with brain neurological diseases [7, 8].

Eye-based human-computer interaction has a variety of other potential applications, especially on the WSN [9], such as smart home monitoring based on the Internet of Things [10]. Furthermore, sound users will accept eye control interaction as an additional means in the future [11], for instance, the combination of eye control interaction and virtual reality technology [12] and the small game platform with outdoor noncontact control by eye tracking technology [13].

Classified by devices, there are three mainstream interactive methods of eye tracking in recent years, including electrooculography [14], infrared fundus imaging [15], and appearance-based methods. Among them, electrooculography and infrared fundus imaging technology have been sufficiently mature. Appearance-based methods estimate the gaze by the shape and texture properties of the eye or the position of the pupil relative to canthus. These methods do not rely on good hardware

configuration, which makes them suitable for implementation on platforms without high-resolution cameras or additional light sources [16]. Methods for gaze estimation include methods based on feature points around the eyes [17], methods based on pupil position [18], and methods based on deep learning, such as CNN proposed in recent years [19].

In appearance-based researches, literature [20] trained the machine learning algorithm by combining the azimuth information with the geometric features of the eyes of which the angle error of the gaze tracking was about 4°. In literature [18], a pupil detection algorithm based on color intensity change was used to detect the center of the pupil in order to estimate the user's sight. Divided into 25 areas of the screen, the gaze estimation accuracy is 78%. In literature [21], 4 cameras with 5×5 pixel resolution were used to train the neural network for gaze tracking, and the minimum angle error was about 1.79°. Although the four cameras used are more complex to construct than a single webcam, the study reveals the great potential of the appearance-based approaches to gaze estimation using a DCNN. Literature [22] established a large data set of gaze tracking and achieved an average error of 10.5° in prediction across data sets by using a DCNN model. Literature [23] trained a CNN with images of both eyes and face posture, and the prediction errors of 1.71 cm and 2.53 cm can be achieved without calibration on the mobile phone and tablet terminals. Literature [24] used the CNN to predict the blinking behavior in nine directions and used it to control the nine-grid keyboard whose input speed could reach 20 letters per minute.

In summary, eye tracking has been widely used for decades in vision research, language, and usability. However, most prior research has focused on large desktop displays using specialized eye trackers that are expensive [25]. However, simple devices limit the resolution of images so that the accuracy of gaze estimation is often unsatisfactory, which makes it difficult to perform high-precise interactive operations. Thus, most research adopts interaction models that divide functions according to the roughly effective area of the screen, where the eye is gazing and uses specially designed interfaces. On account of the limited interactive function, these models face problems such as poor user experience, low operation efficiency, and low degree of freedom of interaction.

Therefore, this work designed a simple eye movement control human-computer interaction system, which can completely replace the mouse and keyboard hardware to operate the computer. This system is characterized by a high degree of freedom of interaction, and it can be carried out directly on the graphical user interface, such as the computer screen with a normal webcam.

2. Materials and Methods

2.1. Basic Methods

2.1.1. Eye Image Acquisition. The method of obtaining the user's eye image in this work is shown in Figure 1. First, each frame is extracted from the video read from the webcam, and the frame size is 640×480. For a single frame image, graying and flipping are performed. When the face is not recognized, face detection is performed with all frames. Then, the template matching method is used to track the face in the subsequent frames, marked with a tracking box.

Based on the cropped face image corresponding to the tracking box, 68 facial landmarks including eye corners, eyebrow, mouth center, and face contour are obtained by detecting the facial feature points [26]. Using the location of the eye, the two eyes' images can be cropped separately. Compared with the method of using the pupil to locate the eye [27], the positions of the four corners of the eye are determined based on the whole facial contour. So, eye tracking will not affect the stability of the eyes' positioning. This process is also not easily affected by the environment such as the illumination change. As a result, the recognition accuracy is improved.

Then, to enhance the eye image, an edge-preserving filter is used for denoising, and power transform is used to enhance the overall contrast of the region and weaken the influence of shadow and illumination change. Finally, two eye images with the size of 36×36 pixels are obtained. Both images will be used as the data source for the next processing.

2.1.2. Eye Movement Recognition and Gaze Tracking

(1) Eye State (Opening or Closed) Recognition. Eye state (opening or closing) recognition is used to reduce the gaze tracking error and as a means of interaction. The SVM (Support Vector Machine) classification algorithm with the HOG (Histogram of Oriented Gradient) features of the input eye image is used to classify the eye states: opening and closing, without distinguishing between the left eye and the right eye.

The parameters of the HOG are as follows: detection window (36, 36), cell size (18, 18), block size (18, 18), block step size (18, 18), and bin number of the histogram 9, and as a result, the feature descriptor of 36 dimensions is obtained. The SVM uses the Gaussian kernel function, and other parameters are default.

The recognition of the closed state and the duration need to know the approximate frame rate. For example, if 30 frames per second and 15 consecutive frames are classified as closed state, it means the duration is 0.5 seconds.

(2) Gaze Estimation and Tracking Based on CNN. First, a brief architecture based on CNN is established by reference to the iTracker model [23] for the estimation task, shown in Figure 2. It consists of a convolutional layer with 20 convolution cores with a size of 5×5 pixels, followed by a max-pooling layer with a size of 2×2 pixels and a concatenate layer followed by a final full-connection layer. The features of the two input images are fused through the concatenate layer, and the two outputs are obtained by the full connected layer, that is, the normalized coordinate value (x, y) of the predicted fixation on the screen.

2.1.3. Head Pose Estimation. Head pose estimation can be used to judge whether the user is facing the screen or not. It can also be used to control the page scrolling or operate the mouse direction. Using 68 facial feature points, the pitch angle, yaw angle, and roll angle can be estimated accurately [28]. However,

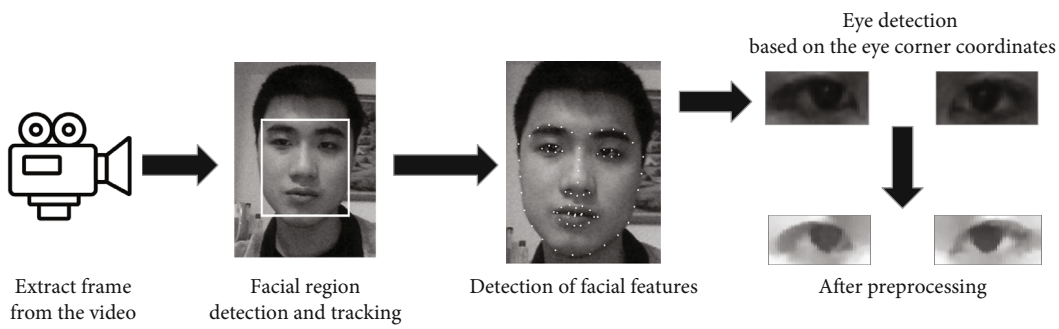


Figure 1: Eye image acquisition processing.

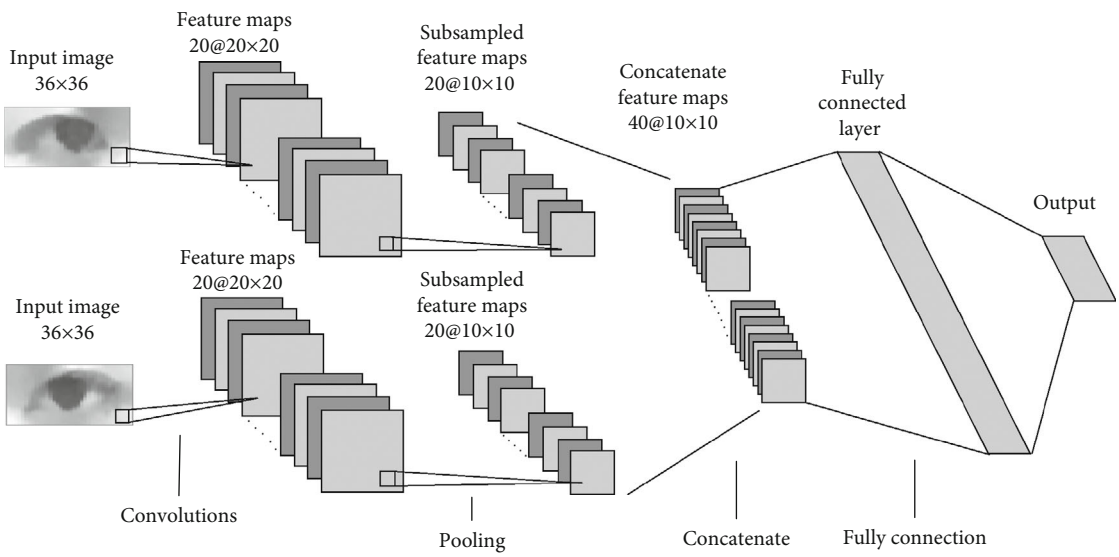


Figure 2: Architecture of the proposed CNN-based model for gaze estimation and tracking.

it only needs to roughly judge whether the user's head turns beyond a certain angle in this work. As shown in Figure 3(a), when the feature point P in the center of the nose is located in the area of the left quarter face, the head is considered to the left. Similarly, when the feature point P is located in the right quarter of the facial region, the head is considered to the right. As shown in Figure 3(b), the center of gravity of P1, P3, and P4 and the relative positions with P2 are calculated, and two thresholds are set to determine whether the head has an up or down deflection.

2.2. Design of the Interactive System

2.2.1. The System Operation Process. A problem of human-computer interaction through appearance-based gaze estimation is Midas contact [29]. The eyes not only act as an important sensory channel but also can provide motion response to control the computer [30]. Therefore, designing an appropriate identification process can greatly reduce the errors of false touch. The proposed system interaction process is shown in Figure 4.

In particular, in Step I, the interaction function should be suspended in the following situations: the user's face has left the screen, the face is too far or too close to the screen, and the face moves too fast. In Step II, when the head obviously deviates from the screen, the interaction function is suspended. In Step III, the SVM+HOG method is used to identify the eye state in each frame. If any one eye is classified as closed at least, the gaze estimation in Step IV will be not performed in this round.

2.2.2. Techniques for Noncontact Mouse Control. Assume that the user's actual point of gaze on the screen is called APOG and the user's predicted point of gaze on the screen is called PPOG. The position of the mouse on the screen is called POM.

Generally speaking, the user's gaze can be accurately calculated when using a high-precision camera; the distance between APOG and PPOG is very small. In this case, we only need to set the POM directly on the PPOG and a certain dwell time to trigger click, so as to complete the function of the hardware mouse. In recent years, there are also methods to detect EEG signals and visual dwell time to accurately distinguish the user's operation intention [31, 32]. In contrast, the result of appearance-based gaze estimation is rough. Although the user's APOG is stable, there may be large differences between the PPOG of adjacent frames, resulting in the user's inability to stably select

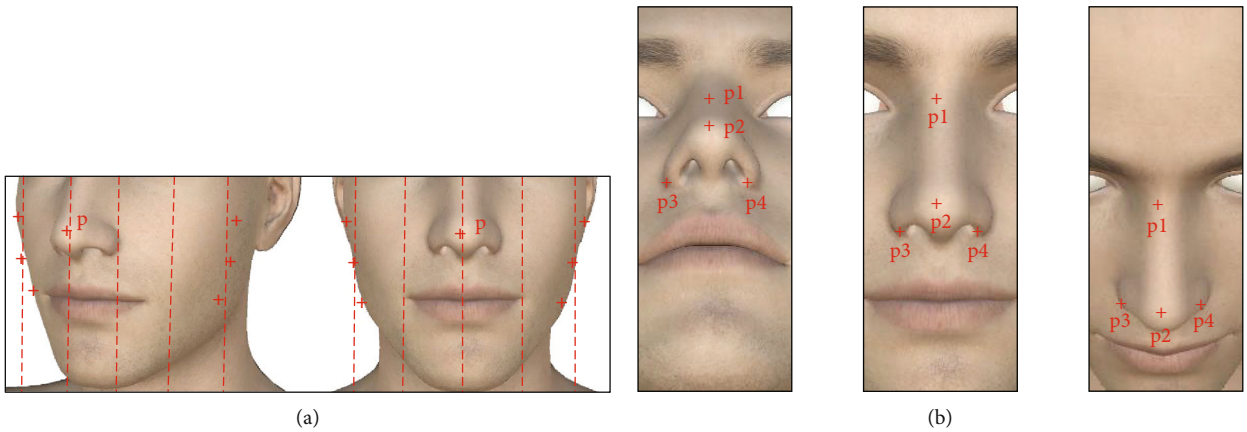


Figure 3: Diagram of head posture estimation: (a) lateral rotation direction estimation; (b) longitudinal rotation direction estimation.

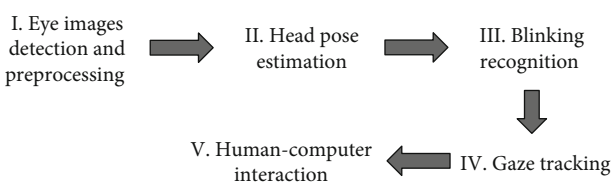


Figure 4: The system operation process.

the target on the screen. This is the biggest challenge of using appearance-based gaze estimation for human-computer interaction.

This work is aimed at designing a rule of accurate mouse movements, which allows the user to control PPOG by adjusting APOG, and then indirectly controlling the POM through this rule to achieve accurate and stable control. The general idea is, first of all, to use the mouse with high sensitivity to let the user use the gaze to control the mouse to move near the target location. Then, the user can move the mouse slowly to the top of the target with the gaze. Finally, one eye blinking is used to trigger the click to complete a round of targets from selection to click operation. There are three ways to control the mouse movement based on gaze estimation; the scheme is described in detail as follows.

The distance of mouse movement is determined by $D_i\,(i = 1, 2, 3)$, and the calculation method is shown in

$$D_i = L_i \times W_i, \tag{1}$$

where L_i refers to the reference distance of mouse movement, and different mouse movement modes have different meanings. W_i is the sensitivity of mouse movement.

(1) Moving Based on Gaze Estimation Named Gaze Movement. A moment before moving the mouse in a frame of the video stream is shown in Figure 5. The circle in the figure represents the PPOG, and the distance between the mouse and the PPOG is expressed by L_1. Finally, the mouse will move with distance D_1 in the direction of PPOG.

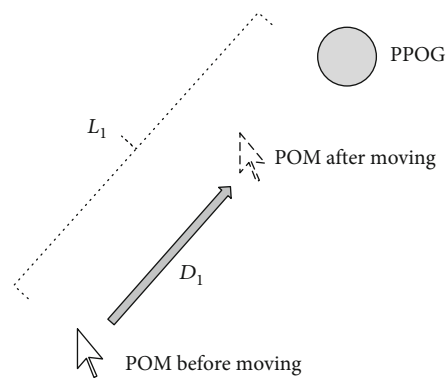


Figure 5: Moving sketching based on gaze estimation.

Gaze movement is a kind of movement mode in which the mouse sensitivity W_1 will gradually decrease. It is used to manipulate the mouse to reach or near the target to click. Ideally, at the beginning of the mouse position reset, the mouse will follow the PPOG. Then, the sensitivity W_1 decreases gradually, the mouse will still move to the PPOG, but the moving speed is slower and slower. When W_1 drops to 0, POM is fixed and cannot move anymore. At this time, if POM is already upon the target, you can click the target by blinking. If there is still some distance away from the target, and it will be difficult to move the POM through gaze movement, the following two techniques can be used to fine-tune the POM. The user can choose one of the following two methods according to their habits to continue to adjust the POM.

(2) Moving Relative to the Screen Center. This is a method to adjust the POM slightly, and sensitivity W_2 is fixed to a relatively small value. As shown in Figure 6, it calculates the distance L_2 between the screen center coordinate and the PPOM, and the mouse moves with the distance D_2 in the direction parallel to the PPOM.

(3) Moving with Remote Control. This is also a way to adjust the POM slightly. Sensitivity W_3 is also fixed to a relatively small value. The mouse can only move in four directions

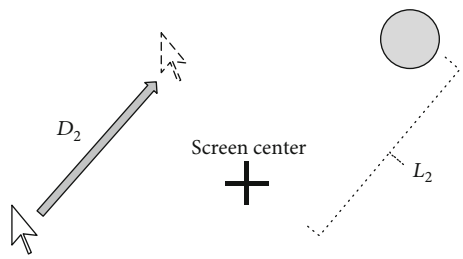


Figure 6: The diagram of moving relative to the screen center.

in this way, as shown in Figure 7; the screen is divided into four areas: up, down, left, and right. The mouse moves to the corresponding gaze area. In the figure, the PPOG appears in the area above the screen, and the distance between the PPOG and the Y axis of the screen center is expressed by L_3, so the mouse moves up along D_3.

2.2.3. Character Input

(1) Input by a Virtual Keyboard. A translucent virtual keyboard will be displayed on the screen while typing in a text box, as shown in Figure 8. This function is identical to that of the hardware keyboard. The user can click the keys on the virtual keyboard by the proposed noncontact mouse controls (see Section 2.2.2). Each key is placed in a square of about 1.5 cm $*$ 1.5 cm, and the entering speed is related to the user's proficiency in clicking a 1.5 cm $*$ 1.5 cm target. In addition, the "FN" key in the lower left corner of the keyboard is designed as a combination key input mode. Users need to click once at the beginning and end of the input combination key. For example, when inputting the capital letter "a" with "shift+a," the order of clicking is "FN," "shift," "a," "FN."

(2) Input by the Gaze Tracking Coding. This method implements character input through trajectories of different eye movements and the dwell time [33, 34]. The specific proposed strategy is described as follows.

The screen is divided into four regions as shown in Figure 9; each region corresponds to a number, namely, "1," "2," "3," and "4." When the user's gaze stays in one of the regions for a period of time, the corresponding number is entered. Every combination of some or all of these four numbers represents one character on the keyboard. For example, the combination "123" expresses the letter "a." The user's gaze only needs to stay in Regions (1), (2), and (3) for one second, respectively, that is, at least 3 seconds to finish the letter "a" input. Of course, the user can shorten the required gazing dwell time for each number (symbol) by multiple training so as to speed up the typing.

To cover all keys on the keyboard in this way, it is necessary to design a coding algorithm with the numbers "1," "2," "3," and "4." We use an idea similar to the Quad Huffman tree to encode for each key on the normal keyboard. The code of any key will not be the prefix of other keys' codes. Adjust the length of each code according to the utilization frequency of the key; the complete coding result is shown in Figure 10.

In practical application, the input process will be displayed in real time on the screen. Take the letter "H" input as an example, the process and the contents of the screen prompt are shown in Figure 11.

It should be noted that the logic of the character input designed in this paper is the same as the hardware keyboard; users can use "shift" and "caps lock" to enter different characters in the same key. Then, two ways are designed for the combination key input: (1) For general two-key combination, add several special codes of commonly used keys, such as shift, alt, and ctrl. Specifically, input a common key code after a special key code input. That is, the double-key combination command is executed. (2) For key combination of not fewer than two keys, first, input a special code to represent multikey combination, and then, put the ordinary key in a queue. You should enter the special code again to end the input; meanwhile, all keys in the queue are executed as key combinations.

2.2.4. Triggering Methods of Other Interactive Commands

(1) Application of Eye State Recognition.

(1) Left single closed for 0.5 s: click with the left mouse button

(2) Left single closed for 1 s: double-click with the left mouse button

(3) Right single closed for 0.5 s: right click

(4) Closed for 1 s: reset the mouse position when the mouse is moving

(5) Blinking eyes three times in 5 seconds: turn on the keyboard or turn off the keyboard

(2) Application of Head Pose Recognition.

(1) Head up: roll up

(2) Head down: roll down

(3) Head left: delete characters on the character input mode

(4) Head right: turn the eye control on or off

3. Results and Discussion

3.1. Calibration. As shown in Figure 12, a laptop with a front camera or a desktop with an external USB camera is used as the experimental equipment; the camera resolution should be 640 × 480 and above. The user is facing the screen, about half a meter away from the camera.

Before the interaction, we need to record the face feature information when the user is facing the computer screen, the gaze tracking training model, and the eye state (opening or closed) training model. The process of collecting this information and training is called calibration. In order to improve the

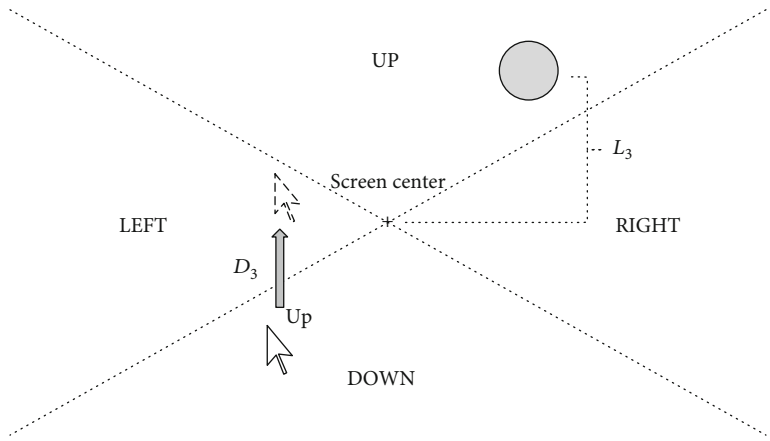


Figure 7: The diagram of moving with remote control.

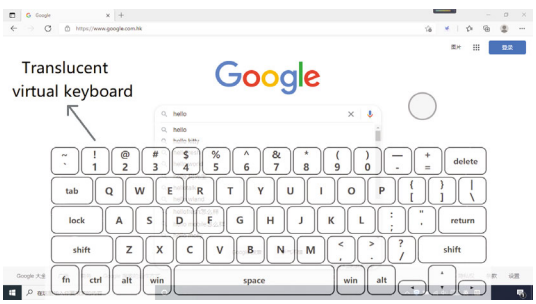


Figure 8: The proposed virtual keyboard.

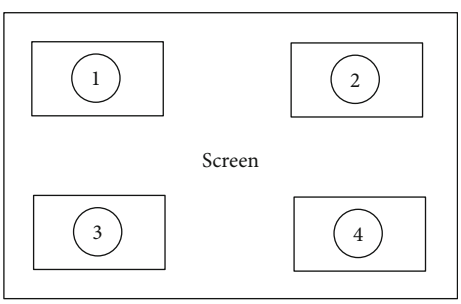


Figure 9: Valid input regions of the screen.

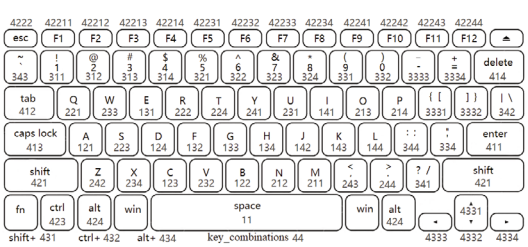


Figure 10: Coding results of each key on the normal keyboard.

recognition accuracy, the user should spend a few minutes completing the calibration step before interaction. The calibration function of the program is divided into three steps.

Step 1. The facial feature points are collected after the user faces the screen for three seconds to estimate the user's head pose.

Step 2. Collect the training data of gaze tracking. The mouse cursor moves according to similar tracks in Figure 13. The number represents the operation order, and the arrow direction is the direction of the mouse movement. The cursor can be controlled by the program or manually according to the actual situation. During the acquisition, the user should avoid head movements, blinking, and other behaviors. When the mouse moves, the program collects the user's eye image and the mouse coordinates as a training sample. After collection, all images are preprocessed to deal with the impact of the environment, and the coordinates of the mouse are normalized to achieve screen size adaptation. All samples were then sent to training; in general, 700-1000 samples are required to train a gaze tracking CNN-based model that meets the requirements. The training model does not need to be retrained when the appearance of the user's eye, the experimental environment, and equipment are not changed much.

More than 500 samples were collected according to the above moving trajectory, and the tracking effect was tested after calibration. The subjects were asked to look at 12 circular targets on the screen in turn. When the subjects looked at each circular target, 10 PPOGs of continuous gaze estimation were recorded. Results are shown in Figure 14; the average error between the PPOG and the APOG is about 2 cm~3 cm.

Step 3. The user closes his eyes for three seconds to collect his/her closed eye images and to train the open and closed eyes binary classification model of SVM with the HOG features of the user's opening eye image collected in the previous step.

The experimental results show that the classification accuracy of a single image is about 90% when the sample size reaches 2000. And this accuracy meets the requirements, because 15 consecutive images are classified as a closed eye state to trigger the interactive command, and the probability of false touch is very low.

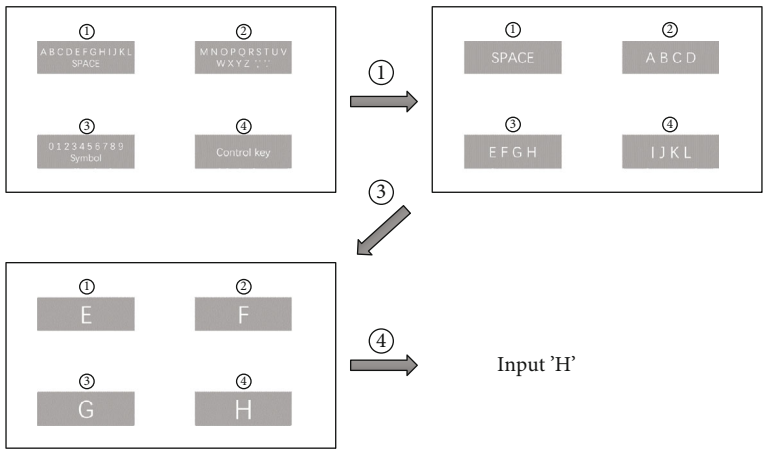


Figure 11: The process and the screen prompt for the input of letter "H."

Figure 12: The optional experimental equipment.

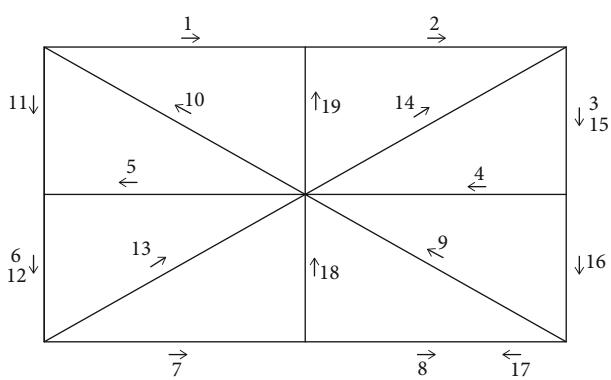


Figure 13: An example of gaze data calibration.

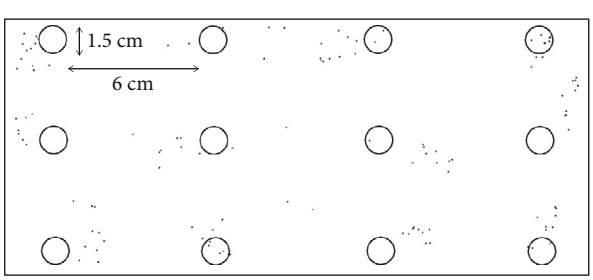


Figure 14: Calibrated gaze data.

3.2. Training for the Interaction. Four 20-year-old college students were selected to participate in the test. None of the four users had been exposed to eye tracking-based interaction, and they need to be trained before the formal test. The training includes click, moving mouse, and typing through auto eye control. After explaining how to operate and what should be paid attention to, the training program began.

Take the training of controlling the mouse based on gaze estimation as an example: the training is divided into three stages; the first stage is to be familiar with the methods (see Section 3.1) of the cursor's operation. In the second stage, the program randomly generates blocks with different positions and sizes on the screen. The user needs to control the cursor to move to the specified block. When the cursor touches the block, it disappears and the next block is generated. In the third stage, blocks are also generated. The difference is that the user needs to blink and click correctly before the next block is generated. Each subject was trained for 20 minutes a day on average, and the next test could be carried out after training for more than three days.

3.3. Interoperability Test of the Proposed Interaction System

3.3.1. Test for the Noncontact Mouse Control. As shown in Figure 15, place 25 blocks of the same length and width on the screen with the same intervals as the targets. The subject needs to use the methods described above to control the mouse with eye movements to click on these targets. Squares with two different sizes are prepared, which are 2 cm × 1 cm and 1 cm × 0.5 cm. Four subjects were asked to select each square line by line to click; the time for every subject was recorded. If one square is clicked correctly, the subject can continue to click the next one in sequence. If all the squares are clicked correctly, you will stop counting time. Each size of the squares is tested three times; that is, 75 correct clicks are required. The experimental result recording after three times is shown in Tables 1 and 2. The highest and lowest accuracy on the smaller click square is 100% and 80%, respectively. And the slowest tester spent 10.85 seconds for one click. It can be seen for the two tables, on

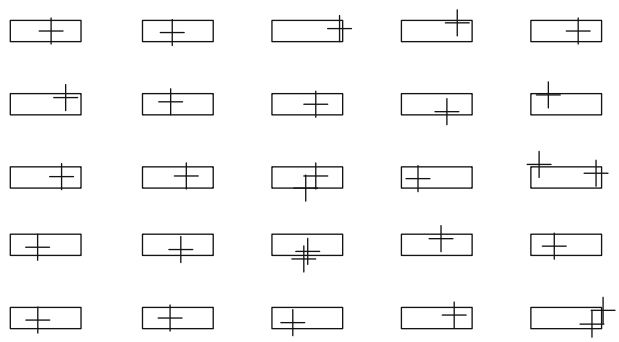


Figure 15: A target click test design.

Table 1: Test results of the proposed mouse click in the square (2 cm × 1 cm).

Subject	Total time (s)	Accuracy (%)	Speed (s per click)
1	417	100.0	5.56
2	556	89.3	7.41
3	489	97.5	6.52
4	507	93.3	6.76

Table 2: Test results of the proposed mouse click in the square (1 cm × 0.5 cm).

Subject	Total time (s)	Accuracy (%)	Speed (s per click)
1	585	100.0	7.80
2	760	80.0	10.13
3	639	96.0	8.52
4	814	90.6	10.85

the square (2 cm × 1 cm), that the speed of the noncontact mouse click basically meets the real-time requirement.

3.3.2. Character Input Test. Let the user use the two proposed typing methods (see Section 3.2) to type a sentence with 56 characters including spaces. If the input character is wrong, you need to delete the character and reenter it. The testing results are shown in Tables 3 and 4.

3.4. Result Analysis and Discussion. According to the experimental results, the average distance between CNN's PPOG and the user's APOG is 2.5 cm after calibration. It can be seen that the average error of the model is 3.18° when the distance is 45 cm between the user and the screen. In fact, compared with other appearance-based gaze estimation models [35], the performance of this result is not the best. However, this just shows the universality of the proposed interaction mode in the upper application of other appearance-based gaze estimation models. For example, combined with the proposed interaction rules, the model [23], which does not need additional calibration with large error, can still achieve accurate human-computer interaction applications.

The results of the interaction test show that the overall accuracy of the eye control mouse is 93.33%, and the click speed is 7.9 seconds. In terms of typing, the overall accuracy

Table 3: Results of character input test based on the virtual keyboard.

Subject	Total time (s)	Accuracy (%)	Speed (s per character)
1	454	98.2	8.10
2	529	91.0	9.44
3	468	94.6	8.35
4	480	96.4	8.57

Table 4: Results of character input test based on the character coding.

Subject	Total time (s)	Accuracy (%)	Speed (s per character)
1	384	98.2	6.85
2	443	96.4	7.91
3	428	98.2	7.64
4	422	94.6	7.53

of the virtual keyboard is 95.05%, and the speed is 8.6 seconds per character. The overall accuracy of characters input based on gaze tracking is 96.85%, and the speed is 7.48 seconds per character.

Compared to several appearance-based eye tracking methods proposed in recent years, this work has the advantages of flexible interaction and complete interaction functions, etc. The specific comparison results are shown in Table 5.

The experimental results also show that the accuracy and speed largely depend on subjects' proficiency. For example, Test No. 1 practiced for 3 hours before the test, which was 2 hours more than other subjects, so he got the best results. Basically, the subject can use the three mouse movement methods to make the mouse click anywhere on the screen. But it will take more time to click near the edge of the screen than at the center.

In typing, compared with the virtual keyboard, the character input method based on eye tracking coding was faster. But the character input speed still cannot fully meet the requirements of real-time interaction. A feasible way to speed up typing is to introduce animation effect into the input process to guide the user to quickly and accurately move the line of sight to enter characters.

Furthermore, the results show that the size of the screen has an impact on interaction efficiency. Generally speaking, a large screen has a wide range of gaze estimation, and it has advantages in the accuracy and speed of character input. However, it may take more time to challenge the final landing point of the mouse if you want to select a target. Therefore, the screen size should not be too large or too small. Our experiments indicated that the appropriate size is 14 to 19 inches.

In addition, it should be noted that subjects who use eye tracking interaction intensively in a short period of time will be tired. Most of the subjects experienced fatigue about 10 minutes after using the system for the first time. And they developed dizziness, nausea, and other symptoms 15 minutes later. They had to put off the test for a day. After a variety of

Table 5: Comparison results.

Literatures	How to interact	Key methods	Accuracy	Speed	Comparison conclusions
[18]	The screen is divided into 3×3 or 5×5, which can monitor the user's gaze on the screen in real time and can be used as 9 or 25 interactive channels	Improved pupil detection algorithm based on color intensity	3×3: 94% 5×5: 78%	In real time	It had poor accuracy and insufficient flexibility for practical applications.
[24]	Use the nine key inputs widely applied in straight board mobile phones	Using CNN to estimate the gaze in nine directions, representing nine keys	90%	20 letters per minute for a skilled user	It was twice as fast compared to our work, but it had insufficient flexibility and limited functions.
[26]	Move the mouse through the head posture, and control the mouse switch, click and scroll through blinking and other facial expressions	Head posture estimation, face detection, and expression recognition	Not mentioned	Not mentioned	It requires various operations to adjust the position of the head when moving the mouse, which reduces the interaction efficiency and usability.

test training, the intensive training can last for 20 minutes without fatigue.

This work also arranged the daily testing task with the proposed human-computer interaction system, including browsing the web and watching videos. All the subjects were able to complete the tasks successfully.

4. Conclusions

In order to achieve simple, affordable, and accurate auto eye control human-computer interaction application, this work applied a webcam to obtain eye image and a lightweight CNN to complete the appearance-based gaze estimation. Using the results of gaze estimation, eye control mouse and character input methods are designed to replace the traditional human-computer interaction devices such as the mouse and keyboard hardware to complete the accurate interactive operation with the computer. The benefits of our method are as follows: (1) the proposed human-computer interaction system enables ordinary people to operate the computer when both hands are occupied. (2) It can also effectively help people with behavioral disorders to solve the input dilemma in human-computer interaction. (3) It realizes the goal of accurate interaction with computers by using imprecise gaze tracking. However, if a completely unfamiliar user uses this system, he needs to spend time to understand the principles of the proposed interaction methods and master the skills before he can skillfully achieve various interactive operations.

Although the interactive system in this work has a certain threshold, it is difficult to meet the requirements of real-time interaction for a long time. However, simple equipment requirements make the proposed interactive system easy to popularize, especially for some people with behavioral disorders with limited living conditions. An ordinary laptop with a camera allows them to complete simple interactive tasks such as browsing videos and information like ordinary people in their spare time.

In the future, we will study the optimization accurate eye positioning method for the user wearing glasses to extend the system to a more general situation. For people with behavioral disorders, the setting and interactive operation of various commands need to be in line with their habits to make the system more accessible. Therefore, developing an interactive system for people with behavioral disorders is also the follow-up task of this study. Finally, the research results can be extended to regulate driving behavior and more noncontact interactive control applications.

Data Availability

All data included in this study are available upon request by contact with the corresponding author.

Conflicts of Interest

There are no conflicts of interest.

Acknowledgments

This work is supported by grants from the Zhuhai Basic and Applied Basic Research Foundation (ZH22017003200027PWC).

References

[1] O. Palagin, V. Romanov, I. Galelyuka, V. Hrusha, and O. Voronenko, "Wireless smart biosensor for sensor networks in ecological monitoring," in *2017 9th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications*, pp. 679–683, Bucharest, Romania, 2017.

[2] M. Dahmani, M. E. H. Chowdhury, A. Khandakar et al., "An intelligent and low-cost eye-tracking system for motorized wheelchair control," *Sensors*, vol. 20, no. 14, article 3936, 2020.

[3] N. Dirix, H. Vander Beken, E. de Bruyne, M. Brysbaert, and W. Duyck, "Reading text when studying in a second language: an eye-tracking study," *Reading Research Quarterly*, vol. 55, no. 3, pp. 371–397, 2020.

[4] F. Guo, M. Li, Y. Chen, J. Xiong, and J. Lee, "Effects of highway landscapes on drivers' eye movement behavior and emergency reaction time: a driving simulator study," *Journal of Advanced Transportation*, vol. 2019, 9 pages, 2019.

[5] Z. Li, S. Zhao, L. Su, and C. Liao, "Study on the operators' attention of different areas in university laboratories based on eye movement tracking technology," *IEEE Access*, vol. 8, pp. 8262–8267, 2020.

[6] Y. Tamura and K. Takemura, "Estimating focused object using smooth pursuit eye movements and interest points in the real world," *The Adjunct Publication of the 32nd Annual ACM Symposium on User Interface Software and Technology*, pp. 21–23, 2019.

[7] H. Y. Lai, G. Saavedra-Pena, C. G. Sodini, V. Sze, and T. Heldt, "Measuring saccade latency using smartphone cameras," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 3, pp. 885–897, 2020.

[8] J. Rapela, T. Y. Lin, M. Westerfield, T. P. Jung, and J. Townsend, "Assisting autistic children with wireless EOG technology," in *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 3504–3506, San Diego, CA, USA, 2012.

[9] M. Khamis, F. Alt, and A. Bulling, "The past, present, and future of gaze-enabled handheld mobile devices: survey and lessons learned," in *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services*, pp. 1–17, Barcelona, Spain, 2018.

[10] A. Bissoli, D. Lavino-Junior, M. Sime, L. Encarnação, and T. Bastos-Filho, "A human–machine interface based on eye tracking for controlling and monitoring a smart home using the internet of things," *Sensors*, vol. 19, no. 4, p. 859, 2019.

[11] R. Jacob and S. Stellmach, "What you look at is what you get," *Interactions*, vol. 23, no. 5, pp. 62–65, 2016.

[12] V. Clay, P. König, and S. Koenig, "Eye tracking in virtual reality," *Journal of Eye Movement Research*, vol. 12, no. 1, pp. 1–18, 2019.

[13] K. Pfeuffer, M. Vidal, J. Turner, A. Bulling, and H. Gellersen, "Pursuit calibration: making gaze calibration less tedious and more flexible," in *Proceedings of the 26th annual ACM symposium on User interface software and technology*, pp. 261–270, St. Andrews Scotland, United Kingdom, 2013.

[14] W. D. Chang, "Electrooculograms for human–computer interaction: a review," *Sensors*, vol. 19, no. 12, article 2690, 2019.

[15] S. Bi, Y. Gu, J. Zou, L. Wang, C. Zhai, and M. Gong, "High precision optical tracking system based on near infrared trinocular stereo vision," *Sensors*, vol. 21, no. 7, p. 2528, 2021.

[16] A. Kar and P. Corcoran, "A review and analysis of eye-gaze estimation systems, algorithms and performance evaluation methods in consumer platforms," *IEEE Access*, vol. 5, pp. 16495–16519, 2017.

[17] Y. L. Wu, C. T. Yeh, W. C. Hung, and C. Y. Tang, "Gaze direction estimation using support vector machine with active appearance model," *Multimedia Tools and Applications*, vol. 70, no. 3, pp. 2037–2062, 2014.

[18] C. Zheng and T. Usagawa, "A rapid webcam-based eye tracking method for human computer interaction," in *2018 International Conference on Control, Automation and Information Sciences*, pp. 133–136, Hangzhou, China, 2018.

[19] A. A. Akinyelu and P. Blignaut, "Convolutional neural network-based methods for eye gaze estimation: a survey," *IEEE Access*, vol. 8, pp. 142581–142605, 2020.

[20] D. Cazzato, F. Dominio, R. Manduchi, and S. M. Castro, "Real-time gaze estimation via pupil center tracking," *Paladyn, Journal of Behavioral Robotics*, vol. 9, no. 1, pp. 6–18, 2018.

[21] M. Tonsen, J. Steil, Y. Sugano, and A. Bulling, "InvisibleEye," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 1, no. 3, pp. 1–21, 2017.

[22] X. Zhang, Y. Sugano, M. Fritz, and A. Bulling, "Mpiigaze: real-world dataset and deep appearance-based gaze estimation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 1, pp. 162–175, 2019.

[23] K. Krafka, A. Khosla, P. Kellnhofer et al., "Eye tracking for everyone," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2176–2184, Las Vegas, NV, USA, 2016.

[24] C. Zhang, R. Yao, and J. Cai, "Efficient eye typing with 9-direction gaze estimation," *Multimedia Tools and Applications*, vol. 77, no. 15, pp. 19679–19696, 2018.

[25] N. Valliappan, N. Dai, E. Steinberg et al., "Accelerating eye movement research via accurate and affordable smartphone eye tracking," *Nature Communications*, vol. 11, no. 1, pp. 1–12, 2020.

[26] K. Meena, M. Kumar, and M. Jangra, "Controlling mouse motions using eye tracking using computer vision," in *2020 4th International Conference on Intelligent Computing and Control Systems*, pp. 1001–1005, Madurai, India, 2020.

[27] J. Xiahou, H. He, K. Wei, and Y. She, "Integrated approach of dynamic human eye movement recognition and tracking in real time," in *2016 International Conference on Virtual Reality and Visualization*, pp. 94–101, Hangzhou, China, 2016.

[28] P. B. M. Thomas, T. Baltrušaitis, P. Robinson, and A. J. Vivian, "The Cambridge face tracker: accurate, low cost measurement of head posture using computer vision and face recognition software," *Translational Vision Science & Technology*, vol. 5, no. 5, pp. 1–7, 2016.

[29] R. Menges, C. Kumar, K. Sengupta, and S. Staab, "eyegui: a novel framework for eye-controlled user interfaces," in *Proceedings of the 9th Nordic Conference on Human-Computer Interaction*, pp. 1–6, Gothenburg Sweden, 2016.

[30] C. Kumar, R. Menges, and S. Staab, "Eye-controlled interfaces for multimedia interaction," *IEEE Multimedia*, vol. 23, no. 4, pp. 6–13, 2016.

[31] S. Han, R. Liu, C. Zhu et al., "Development of a human computer interaction system based on multi-modal gaze tracking methods," in *2016 IEEE International Conference on Robotics and Biomimetics*, pp. 1894–1899, Qingdao, China, 2016.

[32] M. Zhao, H. Gao, W. Wang, and J. Qu, "Research on human-computer interaction intention recognition based on EEG and eye movement," *IEEE Access*, vol. 8, pp. 145824–145832, 2020.

[33] P. P. Banik, M. K. Azam, C. Mondal, and M. A. Rahman, "Single channel electrooculography based human-computer interface for physically disabled persons," in *2015 International Conference on Electrical Engineering and Information Communication Technology*, pp. 1–6, Savar, Bangladesh, 2015.

[34] M. Yildiz and H. Ö. Ülkütaş, "A new PC-based text entry system based on EOG coding," *Advances in Human-Computer Interaction*, vol. 2018, 8 pages, 2018.

[35] Y. Cheng, H. Wang, Y. Bao, and F. Lu, "Appearance-based gaze estimation with deep learning: a review and benchmark," pp. 1–21, 2021, https://arxiv.org/abs/2104.12668.

Hindawi
Wireless Communications and Mobile Computing
Volume 2021, Article ID 9374908, 14 pages
https://doi.org/10.1155/2021/9374908

Research Article

Low-Rank and Spectral-Spatial Sparse Unmixing for Hyperspectral Remote Sensing Imagery

Fan Li

Jiangxi Province Key Laboratory of Water Information Cooperative Sensing and Intelligent Processing, School of Information Engineering, Nanchang Institute of Technology, Nanchang 330099, China

Correspondence should be addressed to Fan Li; lifan@nit.edu.cn

Received 19 September 2021; Accepted 8 October 2021; Published 31 October 2021

Academic Editor: Chi-Hua Chen

Sparse unmixing is an important technique for hyperspectral data analysis. Most sparse unmixing algorithms underutilize the spatial and spectral information of the hyperspectral image, which is unfavourable for the accuracy of endmember identification and abundance estimation. We propose a new spectral unmixing method based on the low-rank representation model and spatial-weighted collaborative sparsity, aiming to exploit structural information in both the spatial and spectral domains for unmixing. The spatial weights are incorporated into the collaborative sparse regularization term to enhance the spatial continuity of the image. Meanwhile, the global low-rank constraint is employed to maintain the spatial low-dimensional structure of the image. The model is solved by the well-known alternating direction method of multiplier, in which the abundance coefficients and the spatial weights are updated iteratively in the inner and outer loops, respectively. Experimental results obtained from simulation and real data reveal the superior performance of the proposed algorithm on spectral unmixing.

1. Introduction

Hyperspectral remote sensing is considered to be a major breakthrough in remote sensing technology due to its high spectral resolutions and simultaneous acquisition of both images and spectra of objects [1, 2]. Hyperspectral remote sensing images provide rich earth observation information, which shows great application potential in mineral resource exploration, precision agriculture, and environmental monitoring [3]. However, the insufficient spatial resolution of hyperspectral images leads to the emergence of a large number of mixed pixels, which influences the fine interpretation of hyperspectral images [4]. Spectral unmixing is an effective way to deal with mixed pixels. It obtains pure spectral signals of ground objects (endmembers) in the imaging area and their corresponding proportions (abundance) [5].

The observed spectrum of the mixed pixel is a mixture of endmember spectra, and unmixing is essentially an optimization problem for a given spectral mixture model. The linear mixture model (LMM) assumes that each pixel is formed by linearly combining spectral signatures of endmembers [3]. Many hyperspectral unmixing algorithms based on LMM have been proposed, among which sparse unmixing methodology has received extensive attention [6, 7]. Sparse unmixing is a semisupervised spectral unmixing method, which takes the spectral library as a priori and assumes that the spectra of the endmembers participating in the hyperspectral image can be found in the library [5]. Generally, the number of endmembers in each mixed pixel is very small relative to the number of spectral signatures in the library; that is, the corresponding abundance is sparse. Therefore, the sparse regression algorithm is adopted to select the optimal endmember subset from the spectral library and simultaneously estimate the fractional abundances [8].

Sparse unmixing can be regarded as solving the combinatorial optimization problem by a constrained sparse linear regression algorithm, in which sparsity-inducing term promotes the sparsity of the abundance solutions. For example, the sparse unmixing via variable splitting and augmented Lagrangian (SUnSAL) uses the ℓ_1 norm to characterize the

sparsity of each abundance vector [5]. Furthermore, the collaborative SUnSAL (CLSUnSAL) uses the $\ell_{2,1}$ norm to enforce the joint sparsity among all pixels [9].

The iterative reweighted ℓ_1 norm minimization is confirmed to enhance the sparsity of solutions [10, 11]. In the work of [12], a weighting factor is introduced in the sparse regularization term to reduce the solution space. The double reweighted sparse unmixing algorithm (DRSU) employs the double weighted ℓ_1 norm to simultaneously reduce the nonzero rows in the abundance matrix corresponding to the actual endmembers and the nonzero values in each abundance vector [13].

Hyperspectral data presents a three-dimensional cube structure with 2-dimensional spatial information and 1-dimensional spectral information of the target area [14, 15]. Taking spatial information as a priori knowledge greatly improves the unmixing performance, so it has become one of the research hotspots in the field of hyperspectral unmixing [16–18]. For instance, the total variation (TV) regularizer is proposed to integrate the spatial-context information [19, 20], and the TV regularization term is added to the SUnSAL algorithm to construct the SUnSAL-TV algorithm [19]. Moreover, the local collaborative sparse unmixing algorithm (LCSU) focuses on the local collaborative characteristics of the image, which imposes different spatial weights on each local region of the abundance map [21].

Low rankness is another inherent characteristic of the abundance matrix, which provides a new perspective for exploring the spatial structure of the data [22, 23]. The high spatial correlation of pixels in the image reflects in the high spectral similarity of these pixels. The similar pixels usually share the same endmembers and similar abundance maps, which means that the corresponding abundance vectors are linearly dependent; that is, the abundance matrix is low rank or approximately low rank. The low-rank constraint helps to obtain a low-rank approximation of the abundance matrix, which captures the global spatial data structure. The alternating direction sparse and low-rank unmixing (ADSpLRU) algorithm first attempts to combine sparse and low-rank constraints, which considers the spatial correlation of the pixels in a 3×3 sliding window [24]. The joint-sparse-blocks and low-rank unmixing (JSpBLRU) algorithm divides the abundance matrix into flexible regular blocks with one row and multiple columns and applies joint sparsity to the blocks, and the low-rank regularizer is merged into the block sparse model synchronously [25].

In this paper, a low-rank and spectral-spatial sparse unmixing algorithm is proposed, which provides the low-rank constraint and the collaborative sparse constraint incorporating neighborhood spatial weights. The multiconstraint strategy effectively exploits the spatial and spectral information of the image and improves the accuracy of the abundance estimation.

The rest of this paper is structured as follows. Section 2 introduces the sparse unmixing model and its variants. Section 3 presents the low-rank and spectral-spatial sparse unmixing model and its optimization algorithm in detail. In Section 4, the effectiveness of the proposed algorithm is demonstrated by experiments on simulated and real hyperspectral data. Section 5 draws conclusions and suggests possible future research.

2. Related Works

Sparse unmixing replaces the endmember set used in LMM with a large spectral library and searches the library for the combination of spectral signatures that best represents the hyperspectral images [5]. Let $\mathbf{Y} \in \mathbb{R}^{l \times n}$ be a hyperspectral image including n pixels with l bands. The sparse unmixing model can be expressed as

$$\mathbf{Y} = \mathbf{AX} + \mathbf{N}, \tag{1}$$

where $\mathbf{A} \in \mathbb{R}^{l \times m}$ is a spectral library including m spectral signatures, $\mathbf{X} \in \mathbb{R}^{m \times n}$ is the abundance matrix corresponding to the spectral library $\mathbf{A}$, and $\mathbf{N} \in \mathbb{R}^{l \times n}$ is the noise or model error. The abundance matrix $\mathbf{X}$ reflects the proportion of spectral signatures from the library $\mathbf{A}$ in each pixel. According to the physical meaning of abundance, two constraints are considered: abundance nonnegativity constraint (ANC) and abundance sum-to-one constraint (ASC), which can be written as $\mathbf{X} \geq 0$and $\mathbf{1}^T\mathbf{X} = \mathbf{1}$.

In fact, there are only a few entries from the spectral library $\mathbf{A}$ act on the mixed pixels, which results in many zeros in the relevant fractional abundances. That is, the abundance matrix $\mathbf{X}$ is sparse. Based on sparse and redundant representation theory, the unmixing problem can be modeled as follows:

$$\begin{aligned} \min_{\mathbf{X}} \frac{1}{2} \|\mathbf{Y} - \mathbf{AX}\|_F^2 + \lambda \|\mathbf{X}\|_0 \\ \text{s.t.} \quad \mathbf{X} \geq 0, \end{aligned} \tag{2}$$

where $\|\cdot\|_F$ is the Frobenius norm, λ is a nonnegative regularization parameter, and $\|\mathbf{X}\|_0$ indicates the number of nonzero values in $\mathbf{X}$. Note that the ASC is not strictly conducted in the model (2), which is ascribed to the drawbacks mentioned in [5].

The optimization problem related to model (2) is NP-hard. Under the condition of restricted isometric property (RIP), ℓ_0 norm can be relaxed to ℓ_1 norm [26]. The model (2) is modified as

$$\begin{aligned} \min_{\mathbf{X}} \frac{1}{2} \|\mathbf{Y} - \mathbf{AX}\|_F^2 + \lambda \|\mathbf{X}\|_{1,1} \\ \text{s.t.} \quad \mathbf{X} \geq 0, \end{aligned} \tag{3}$$

where $\|\mathbf{X}\|_{1,1} = \sum_{j=1}^{n} \|\mathbf{x}_j\|_1$, $\mathbf{x}_j$ is the jth column vector of $\mathbf{X}$. The SUnSAL algorithm efficiently solves model (3) [19]. However, the ℓ_1 norm measures the sparsity of each abundance vector independently instead of considering the joint sparsity of all pixels in the scene. The mixed norm $\ell_{2,1}$ norm is proposed to improve sparsity among the rows of $\mathbf{X}$ collaboratively. The corresponding optimization problem is described as

$$\begin{aligned} \min_{\mathbf{X}} \frac{1}{2} \|\mathbf{Y} - \mathbf{AX}\|_F^2 + \lambda \|\mathbf{X}\|_{2,1} \\ \text{s.t.} \quad \mathbf{X} \geq 0, \end{aligned} \tag{4}$$

where $\|\mathbf{X}\|_{2,1} = \sum_{i=1}^{m} \|\mathbf{x}_i\|_2$, $\mathbf{x}_i$ is the ith row vector of $\mathbf{X}$. The optimization problem related to (4) is solved by the CLSUnSAL algorithm. As an improvement of the SUnSAL algorithm, it is aimed at finding a structured solution with only a few nonzero lines. In other words, it promotes the row sparsity of the abundance matrix $\mathbf{X}$.

The abovementioned unmixing models underutilize the spatial structure features of hyperspectral images. The rich spatial information contained in the image, which has an important impact on the unmixing performance, is not incorporated into these models. To this end, we propose a new sparse unmixing algorithm to address the aforementioned limitation.

3. The Proposed Low-Rank and Spectral-Spatial Sparse Unmixing Algorithm

3.1. The Proposed LRSSU Model. Inspired by the idea of iterative weighting [27], the spatial weight based on neighborhood is introduced into $\ell_{2,1}$ norm sparse regularization term, which integrates the spatial and spectral information. Due to the advantage of the low-rank representation model in maintaining the low-dimensional structure of the image, a low-rank regularization term is incorporated to explore the global spatial correlation of the hyperspectral data. Combining the sparse and low-rank constraints, a new sparse unmixing model with low-rank constraint is proposed as follows:

$$\min_{\mathbf{X}} \frac{1}{2} \|\mathbf{Y} - \mathbf{A}\mathbf{X}\|_F^2 + \lambda \left\|\mathbf{H}_{\text{spa}} \odot \mathbf{X}\right\|_{2,1} + \tau \text{ rank } (\mathbf{X}) \quad \text{s.t.} \quad \mathbf{X} \geq 0, \tag{5}$$

where $\mathbf{H}_{\text{spa}} \in \mathbb{R}^{m \times n}$ is the spatial weight matrix, $\odot$ is the Hadamard product, rank $(\mathbf{X})$ is the rank of the abundance matrix $\mathbf{X}$, and λ and τ are nonnegative regularization parameters. $\mathbf{H}_{\text{spa}}$ is updated iteratively as follows:

$$\mathbf{H}_{\text{spa},ij}^{(t+1)} = \frac{\mathbf{1}}{f_{\mathcal{N}(p)}\left(x_p^{(t)}\right) + \varepsilon}, \tag{6}$$

where $\mathbf{H}_{\text{spa},ij}^{(t)}$ represents the element in the ith row and jth column of $\mathbf{H}_{\text{spa}}$ at the tth iteration. p represents the pixel corresponding to fractional abundance x_{ij}, $x_p = (x_{1j}, x_{2j}, \cdots, x_{mj})^T$ is the abundance vector for pixel p, $\mathcal{N}(p)$ represents the neighbor pixel set centered on p in the image, and $f_{\mathcal{N}(\cdot)}(\cdot)$ denotes a function that explores the spatial correlation of pixels in a neighborhood system $\mathcal{N}(\cdot)$. In this paper, the 8 neighbor pixels of p constitute the neighborhood system $\mathcal{N}_8(p)$; then, the function $f(\cdot)$ is defined as

$$f_{\mathcal{N}_8(p)}\left(x_p\right) = \frac{\sum_{q \in \mathcal{N}_8(p)} \rho_q x_q}{\sum_{q \in \mathcal{N}_8(p)} \rho_q}, \tag{7}$$

where q is any pixel in the 8 neighbor pixel set of p, x_q is the abundance vector corresponding to q, $\rho_q = 1/(d_E(p, q))$ denotes the neighborhood weight with regard to q, and $d_E(p, q)$ denotes the Euclidean distance between pixel q and its central pixel p.

The rank of $\mathbf{X}$ is closely relevant to the similarity of abundance vectors, and the reduced rank abundance matrix consists of fewer linearly independent vectors accordingly. When $\ell_{2,1}$ norm is used to promote the row sparsity of $\mathbf{X}$, the low-rank regularizer helps to determine the endmembers from the spectral library more accurately. Nevertheless, it is difficult to calculate the rank of the matrix directly. The nuclear norm $\|\mathbf{X}\|_*$ is the tightest convex relaxation of the rank of $\mathbf{X}$. Therefore, we replace rank $(\mathbf{X})$ with $\|\mathbf{X}\|_*$ and rewrite model (5) as

$$\min_{\mathbf{X}} \frac{1}{2} \|\mathbf{Y} - \mathbf{A}\mathbf{X}\|_F^2 + \lambda \left\|\mathbf{H}_{\text{spa}} \odot \mathbf{X}\right\|_{2,1} + \tau \|\mathbf{X}\|_* \quad \text{s.t.} \quad \mathbf{X} \geq 0, \tag{8}$$

where $\|\mathbf{X}\|_* = \sum_i \sigma_i(\mathbf{X})$ and $\sigma_i(\mathbf{X})$ is the ith singular value of $\mathbf{X}$.

The weighted nuclear norm minimization (WNNM) can enhance the sparsity of the singular vector by imposing different weights on each singular value [28]. Specifically, smaller singular values that carry less valid information are assigned larger weights and then are greatly reduced or even reduced to zero. On the contrary, larger values are assigned smaller weights, so that the main information is well preserved. The weighted nuclear norm $\|\mathbf{X}\|_{\mathbf{b},*}$ is adopted to improve the robustness and consistency of model (8), which is expressed as

$$\|\mathbf{X}\|_{\mathbf{b},*} = \sum_{i=1}^{r} b_i \sigma_i(\mathbf{X}), \tag{9}$$

where $\mathbf{b} = (b_1, b_2, \cdots)$ is the nonnegative weight vector and b_i is the weight related to the ith singular value of the matrix $\mathbf{X}$. At the kth iteration, b_i is updated as follows:

$$b_i^{(k)} = \frac{1}{\sigma_i\left(\mathbf{X}^{(k)}\right) + \varepsilon}, \tag{10}$$

where $\varepsilon > 0$ is a small constant added to avoid singularity and $\mathbf{X}^{(k)}$ represents the abundance matrix $\mathbf{X}$ at the kth iteration.

In summary, the low-rank and spectral-spatial sparse unmixing model (LRSSU) is expressed as follows:

$$\min_{\mathbf{X}} \frac{1}{2} \|\mathbf{Y} - \mathbf{A}\mathbf{X}\|_F^2 + \lambda \left\|\mathbf{H}_{\text{spa}} \odot \mathbf{X}\right\|_{2,1} + \tau \|\mathbf{X}\|_{\mathbf{b},*} \quad \text{s.t.} \quad \mathbf{X} \geq 0. \tag{11}$$

3.2. Solution of the Optimization Problem. In this section, an iterative optimization scheme of internal and external dual circulation based on the alternating direction method of multipliers (ADMM) [29] is proposed. The external

circulation updates the spatial weight matrix following equation (6), and the internal circulation solves the abundance matrix in the framework of ADMM. The flow of solving model (11) is described in detail as follows.

The auxiliary matrix variables $\mathbf{U}$, $\mathbf{V}_1$, $\mathbf{V}_2$, $\mathbf{V}_3$, and $\mathbf{V}_4$ are introdued, and model (11) is equivalently converted to the following form:

$$\begin{aligned} \min_{\mathbf{U},\mathbf{V}_1,\mathbf{V}_2,\mathbf{V}_3,\mathbf{V}_4} & \frac{1}{2}\|\mathbf{Y}-\mathbf{V}_1\|_F^2+\lambda\left\|\mathbf{H}_{\text{spa}}\odot\mathbf{V}_2\right\|_{2,1}+\tau\|\mathbf{V}_3\|_{\mathbf{b},*}+\iota_{R+}(\mathbf{V}_4) \\ \text{s.t.} \quad & \mathbf{U}=\mathbf{X}, \\ & \mathbf{V}_1=\mathbf{A}\mathbf{U}, \\ & \mathbf{V}_2=\mathbf{U}, \\ & \mathbf{V}_3=\mathbf{U}, \\ & \mathbf{V}_4=\mathbf{U}, \end{aligned} \tag{12}$$

where $\iota_{R+}(\mathbf{X})=\sum_{i=1}^{n}\iota_{R+}(\mathbf{x}_i)$ is the indicator function; i.e., $\iota_{R+}(\mathbf{x}_i)$ is zero if $\mathbf{x}_i$ is nonnegative and $+\infty$ otherwise. Equation (12) can be expressed compactly as

$$\begin{aligned} & \min_{\mathbf{U},\mathbf{V}} g(\mathbf{U},\mathbf{V}) \\ \text{s.t.} \quad & \mathbf{G}\mathbf{U}+\mathbf{B}\mathbf{V}=0, \end{aligned} \tag{13}$$

where $g(\mathbf{U},\mathbf{V})=(1/2)\|\mathbf{Y}-\mathbf{V}_1\|_F^2+\lambda\|\mathbf{H}_{\text{spa}}\odot\mathbf{V}_2\|_{2,1}+\tau\|\mathbf{V}_3\|_{\mathbf{b},*}+\iota_{R+}(\mathbf{V}_4)$, $\mathbf{G}=(\mathbf{A},\mathbf{I},\mathbf{I},\mathbf{I})^{\mathbf{T}}$, $\mathbf{B}=\text{diag}\,(-I)$, and $\mathbf{V}=(\mathbf{V}_1,\mathbf{V}_2,\mathbf{V}_3,\mathbf{V}_4)^{\mathbf{T}}$.

The augmented Lagrangian of equation (13) is written as

$$\mathscr{L}(\mathbf{U},\mathbf{V},\mathbf{D})=g(\mathbf{U},\mathbf{V})+\frac{\mu}{2}\|\mathbf{G}\mathbf{U}+\mathbf{B}\mathbf{V}-\mathbf{D}\|_F^2, \tag{14}$$

where μ is a nonnegative penalty parameter, $\mathbf{D}=(\mathbf{D}_1,\mathbf{D}_2,\mathbf{D}_3,\mathbf{D}_4)^{\mathbf{T}}$ is the Lagrange multipliers with regards to $\mathbf{G}\mathbf{U}+\mathbf{B}\mathbf{V}=0$.

The optimization problem (13) can be solved by minimizing $\mathscr{L}(\mathbf{U},\mathbf{V},\mathbf{D})$ in the framework of ADMM. We update the variables $\mathbf{U}$ and $\mathbf{V}$ sequentially by solving the following subproblems:

$$\begin{cases} \mathbf{U}^{(k+1)}=\arg\ \min_{\mathbf{U}}\mathscr{L}\left(\mathbf{U},\mathbf{V}^{(k)},\mathbf{D}^{(k)}\right), \\ \mathbf{V}^{(k+1)}=\arg\ \min_{\mathbf{V}}\mathscr{L}\left(\mathbf{U}^{(k+1)},\mathbf{V},\mathbf{D}^{(k)}\right). \end{cases} \tag{15}$$

The $\mathbf{U}$-subproblem has the closed-form solution, which is obtained by taking the partial derivative of $\mathbf{U}$. The variable $\mathbf{U}$ is calculated as follows:

$$\begin{aligned} \mathbf{U}^{(k+1)}=\arg\ \min_{\mathbf{U}} & \frac{\mu}{2}\left\|\mathbf{A}\mathbf{U}-\mathbf{V}_1^{(k)}-\mathbf{D}_1^{(k)}\right\|_F^2 \\ & +\frac{\mu}{2}\left\|\mathbf{U}-\mathbf{V}_2^{(k)}-\mathbf{D}_2^{(k)}\right\|_F^2+\frac{\mu}{2}\left\|\mathbf{U}-\mathbf{V}_3^{(k)}-\mathbf{D}_3^{(k)}\right\|_F^2 \\ & +\frac{\mu}{2}\left\|\mathbf{U}-\mathbf{V}_4^{(k)}-\mathbf{D}_4^{(k)}\right\|_F^2 \\ = & \left(\mathbf{A}^{\mathbf{T}}\mathbf{A}+3\mathbf{I}\right)^{-1}\left(\mathbf{A}^{\mathbf{T}}\left(\mathbf{V}_1^{(k)}+\mathbf{D}_1^{(k)}\right)+\mathbf{V}_2^{(k)}+\mathbf{D}_2^{(k)}\right. \\ & \left.+\mathbf{V}_3^{(k)}+\mathbf{D}_3^{(k)}+\mathbf{V}_4^{(k)}+\mathbf{D}_4^{(k)}\right). \end{aligned} \tag{16}$$

The $\mathbf{V}$-subproblem is split into four independent optimization problems with regard to $\mathbf{V}_1$, $\mathbf{V}_2$, $\mathbf{V}_3$, and $\mathbf{V}_4$. The solution of $\mathbf{V}_1$ is calculated as follows:

$$\begin{aligned} \mathbf{V}_1^{(k+1)} &= \arg\ \min_{\mathbf{V}_1}\frac{1}{2}\|\mathbf{Y}-\mathbf{V}_1\|_F^2+\frac{\mu}{2}\left\|\mathbf{A}\mathbf{U}^{(k+1)}-\mathbf{V}_1-\mathbf{D}_1^{(k)}\right\|_F^2 \\ &= \frac{1}{1+\mu}\left[\mathbf{Y}+\mu\left(\mathbf{A}\mathbf{U}^{(k+1)}-\mathbf{D}_1^{(k)}\right)\right]. \end{aligned} \tag{17}$$

The optimization problem related to $\mathbf{V}_2$ is

$$\mathbf{V}_2^{(k+1)}=\arg\ \min_{\mathbf{V}_2}\lambda\left\|\mathbf{H}_{\text{spa}}\odot\mathbf{V}_2\right\|_{2,1}+\frac{\mu}{2}\left\|\mathbf{U}^{(k+1)}-\mathbf{V}_2-\mathbf{D}_2^{(k)}\right\|_F^2. \tag{18}$$

The rth row of $\mathbf{V}_2$ is written as $\mathbf{V}_{2,r}$; then, the closed-form solution of problem (18) is

$$\mathbf{V}_{2,r}^{(k+1)}=\textbf{vect-soft}\left(\mathbf{U}_r^{(k+1)}-\mathbf{D}_{2,r}^{(k)},\frac{\lambda}{\mu}\mathbf{H}_{\text{spa},r}\right), \tag{19}$$

where $\textbf{vect-soft}(y,\tau)=y(\max\{\|y\|_2-\tau,0\}/\max\{\|y\|_2-\tau,0\}+\tau)$ denotes the soft threshold function for row vector. $\mathbf{V}_3$ is updated according to the following optimization problem:

$$\begin{aligned} \mathbf{V}_3^{(k+1)} &= \arg\ \min_{\mathbf{V}_3}\tau\|\mathbf{V}_3\|_{\mathbf{b},*}+\frac{\mu}{2}\left\|\mathbf{U}^{(k+1)}-\mathbf{V}_3-\mathbf{D}_3^{(k)}\right\|_F^2 \\ &= \mathbf{SVT}_{\mathbf{b},\tau/\mu}\left(\mathbf{U}^{(k+1)}-\mathbf{D}_3^{(k)}\right), \end{aligned} \tag{20}$$

where $\mathbf{SVT}_{\mathbf{b},\omega}(\cdot)$ denotes the singular value threshold function and the weight vector $\mathbf{b}$ is calculated according to equation (10). The singular value decomposition (SVD) of $\mathbf{X}$ is written as $\mathbf{X}=\widehat{\mathbf{U}}\mathbf{Diag}(\sigma_1,\cdots,\sigma_r)\mathbf{V}\wedge^{\mathbf{T}}$, then $\mathbf{SVT}_{\mathbf{b},\omega}(\mathbf{X})=\widehat{\mathbf{U}}\mathbf{Diag}(\max(\sigma_1-\omega b_1,0)),\cdots,(\max(\sigma_r-\omega b_r,0))\mathbf{V}\wedge^{\mathbf{T}}$. In the same vein, we get the solution of $\mathbf{V}_4$ as follows:

$$\begin{aligned} \mathbf{V}_4^{(k+1)} &= \arg\ \min_{\mathbf{V}_4}\iota_{R+}(\mathbf{V}_4)+\frac{\mu}{2}\left\|\mathbf{U}^{(k+1)}-\mathbf{V}_4-\mathbf{D}_4^{(k)}\right\|_F^2 \\ &= \max\left(\mathbf{U}^{(k+1)}-\mathbf{D}_4^{(k)},0\right). \end{aligned} \tag{21}$$

1: **Initialization:**
2: $k, t = 0$, choose $\mu, \lambda, \tau, \varepsilon > 0$, $\mathbf{U}^{(0)}$, $\mathbf{V}_1^{(0)}$, $\mathbf{V}_2^{(0)}$, $\mathbf{V}_3^{(0)}$, $\mathbf{V}_4^{(0)}$, $\mathbf{D}_1^{(0)}$, $\mathbf{D}_2^{(0)}$, $\mathbf{D}_3^{(0)}$, $\mathbf{D}_4^{(0)}$
3: **Repeat:**
4: $\mathbf{H}_{spa,ij}^{(t+1)} = \mathbf{1}/(f_{\mathcal{N}(p)}(\mathbf{x}_p^{(t)}) + \varepsilon)$, $\mathbf{x}_p = (x_{1j}, x_{2j}, \cdots, x_{mj})^T$ is the jth column of $(\mathbf{U}^{(t)} - \mathbf{V}_2^{(t)})$
5: **Repeat:**
6: $\mathbf{U}^{(k+1)} \longleftarrow (\mathbf{A}^{\mathbf{T}}\mathbf{A} + 3\mathbf{I})^{-1}(\mathbf{A}^{\mathbf{T}}(\mathbf{V}_1^{(k)} + \mathbf{D}_1^{(k)}) + \mathbf{V}_2^{(k)} + \mathbf{D}_2^{(k)} + \mathbf{V}_3^{(k)} + \mathbf{D}_3^{(k)} + \mathbf{V}_4^{(k)} + \mathbf{D}_4^{(k)})$
7: $\mathbf{V}_1^{(k+1)} \longleftarrow 1/(1+\mu)[\mathbf{Y} + \mu(\mathbf{A}\mathbf{U}^{(k+1)} - \mathbf{D}_1^{(k)})]$
8: $\mathbf{V}_{2,r}^{(k+1)} \longleftarrow \mathbf{vect\text{-}soft}(\mathbf{U}_r^{(k+1)} - \mathbf{D}_{2,r}^{(k)}, (\lambda/\mu)\mathbf{H}_{spa,r})$
9: $\mathbf{V}_3^{(k+1)} \longleftarrow \mathbf{SVT}_{\mathbf{b},\tau/\mu}(\mathbf{U}^{(k+1)} - \mathbf{D}_3^{(k)})$
10: $\mathbf{V}_4^{(k+1)} \longleftarrow \max(\mathbf{U}^{(k+1)} - \mathbf{D}_4^{(k)}, 0)$
11: **Update Lagrange multipliers:**
12: $\mathbf{D}_1^{(k+1)} \longleftarrow \mathbf{D}_1^{(k)} - \mathbf{A}\mathbf{U}^{(k+1)} + \mathbf{V}_1^{(k+1)}$
13: $\mathbf{D}_2^{(k+1)} \longleftarrow \mathbf{D}_2^{(k)} - \mathbf{U}^{(k+1)} + \mathbf{V}_2^{(k+1)}$
14: $\mathbf{D}_3^{(k+1)} \longleftarrow \mathbf{D}_3^{(k)} - \mathbf{U}^{(k+1)} + \mathbf{V}_3^{(k+1)}$
15: $\mathbf{D}_4^{(k+1)} \longleftarrow \mathbf{D}_4^{(k)} - \mathbf{U}^{(k+1)} + \mathbf{V}_4^{(k+1)}$
16: **Update iteration:**$k \longleftarrow k+1$
17: $\mathbf{U}^{(t+1)} \longleftarrow \mathbf{U}^{(k)}$
18: $\mathbf{D}_2^{(t+1)} \longleftarrow \mathbf{D}_2^{(k)}$
19: **Update iteration:**$t \longleftarrow t+1$
20: **Until** some stopping criterion is satisfied.

Algorithm 1: Pseudocode of the LRSSU algorithm.

Finally, the Lagrange multipliers are updated as follows:

$$\mathbf{D}^{(k+1)} = \mathbf{D}^{(k)} - \left(\mathbf{G}\mathbf{U}^{(k+1)} + \mathbf{B}\mathbf{V}^{(k+1)}\right). \tag{22}$$

To sum up, the complete scheme for the proposed LRSSU model is presented in Algorithm 1, which consists of internal and external dual circulations. The external circulation updates the spatial weight matrix in step 4, which is set up to 100 iterations. The internal circulation updates the variables and Lagrange multipliers in ADMM framework from step 6 to step 15 and sets 5 iterations empirically [10, 30, 31].

The convergence of Algorithm 1 is difficult to prove. Nevertheless, the residual of the optimization problem decreases with the increase of iterations. Figure 1 shows the residual $\|\mathbf{G}\mathbf{U}^{(t)} + \mathbf{B}\mathbf{V}^{(t)}\|_F$ as a function of the number of iterations for the complete algorithm. It can be observed that the residual arrived at a certain level close to zero after about 20 iterations of the external circulation. Therefore, the stopping criterion of the algorithm is set as the residual error reaches the error tolerance or the number of iterations reaches the upper limit, which guarantees the convergence of the algorithm.

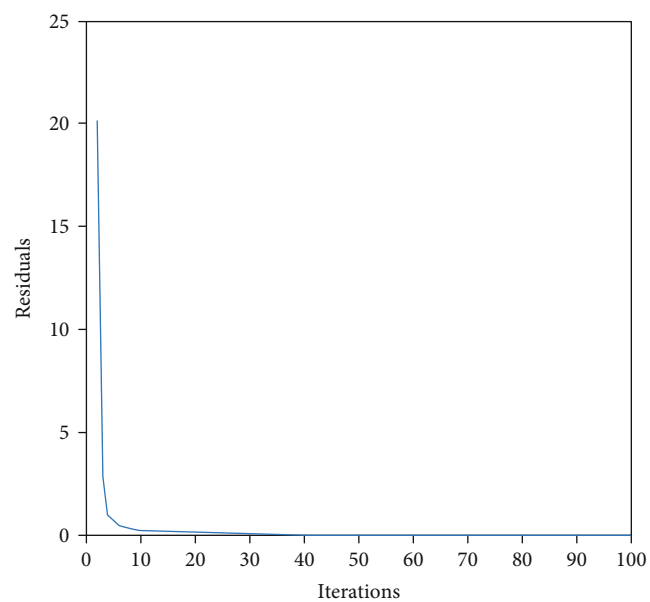


Figure 1: The residual $\|\mathbf{G}\mathbf{U}^{(t)} + \mathbf{B}\mathbf{V}^{(t)}\|_F$ as a function of the number of iterations for the complete algorithm.

4. Experimental Analysis

In this section, the performance of the proposed LRSSU algorithm will be discussed in simulated and real hyperspectral data experiments. In the simulated data experiment, the signal-to-reconstruction error (SRE) is used to quantitatively evaluate the unmixing accuracy. The SRE (dB) is defined as follows:

$$\text{SRE (dB)} = 10 \cdot \log_{10}\left(E\left(\|\mathbf{x}\|_2^2\right)/E\left(\|\mathbf{x} - \hat{\mathbf{x}}\|_2^2\right)\right), \tag{23}$$

where $E(\cdot)$ is the expectation function, $\hat{\mathbf{x}}$ is the estimated abundance vector, $\mathbf{x}$ is the true abundance vector. In addition, the probability of success p_s is another quantitative

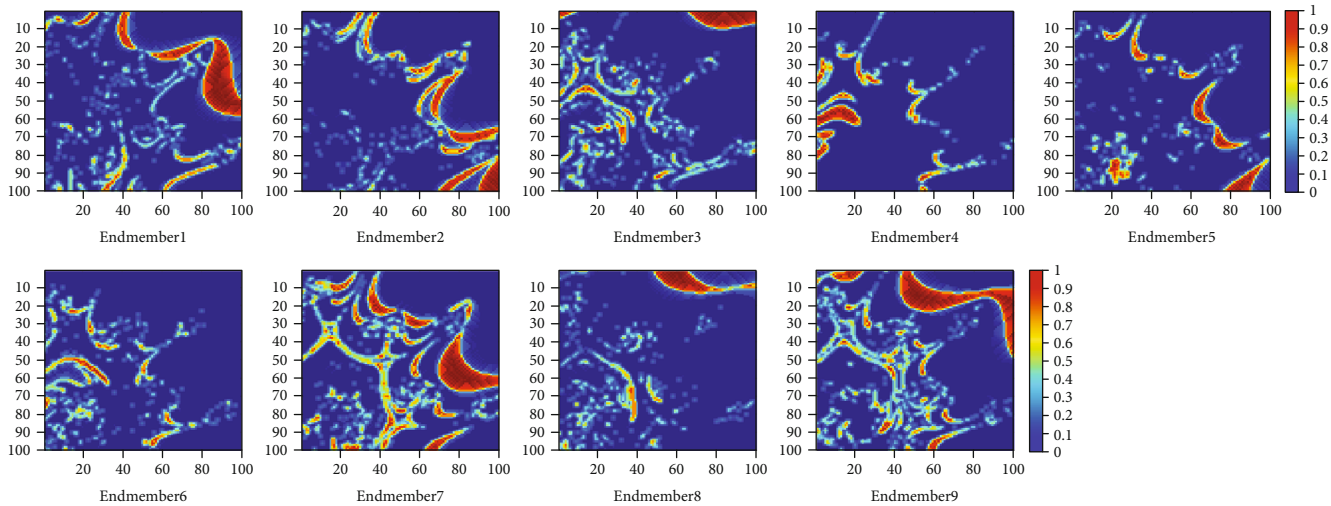


Figure 2: True fractional abundances of the endmembers in the simulated data.

Table 1: SRE (B) and p_s results after applying different unmixing algorithms to the simulated data (the optimal parameter settings for which the reported values were achieved are shown in the parentheses).

Algorithm	SNR = 30 dB		SNR = 40 dB		SNR = 50 dB	
	SRE (dB)	p_s	SRE (dB)	p_s	SRE (dB)	p_s
SUnSAL	6.3967	0.6303	11.5562	0.8871	18.9222	0.9991
	($\lambda = 8e$-3)		($\lambda = 2e$-3)		($\lambda = 3e$-4)	
CLSUnSAL	6.6157	0.7269	14.7819	0.9999	26.6827	1
	($\lambda = 3e$-1)		($\lambda = 2e$-2)		($\lambda = 7e$-3)	
SUnSAL-TV	9.0378	0.7822	15.4433	0.9861	25.3656	1
	($\lambda = 4e$-3; $\lambda_{TV} = 2e$-3)		($\lambda = 6e$-5; $\lambda_{TV} = 9e$-4)		($\lambda = 5e$-5; $\lambda_{TV} = 9e$-5)	
DRSU	14.3296	0.9465	26.0656	1	34.4967	1
	($\lambda = 2e$-3)		($\lambda = 6e$-4)		($\lambda = 1e$-4)	
JSpBLRU	15.0546	0.9928	24.8702	1	34.0073	1
	($\lambda = 2.3$; $\tau = 1.4$)		($\lambda = 4e$-1; $\tau = 2e$-1)		($\lambda = 2e$-1; $\tau = 4e$-2)	
LRSSU	19.8380	0.9989	27.9738	1	36.4791	1
	($\lambda = 4e$-2; $\tau = 8e$-2)		($\lambda = 2e$-2; $\tau = 1e$-2)		($\lambda = 4e$-3; $\tau = 2e$-3)	

evaluation indicator, which appraises the stability of the results [5] . p_s is defined as follows: $p_s \equiv P(\|\mathbf{x}\wedge - \mathbf{x}\|^2/\|\mathbf{x}\|^2 \leq \text{threshold})$. The results will be considered successful, if the relative error power of results is less than an appropriate threshold, usually 3.16 (5 dB). The larger the SRE (dB) or p_s, the better the performance of the algorithm. For fair comparison, the parameters of all algorithms in the experiment are tuned to the optimal in advance, and the maximum number of iterations is set to 500. The proposed LRSSU algorithm is compared with the classic SUnSAL, CLSUnSAL, and SUnSAL-TV algorithms, as well as the advanced DRSU algorithm and the JSPBLRU algorithm considering the low-rank constraint.

4.1. Experiment on Simulated Data. The spectral library $\mathbf{A} \in \mathbb{R}^{221\times 222}$ used in the simulated experiment is composed of 222 spectral curves randomly selected from the splib06 spectral library provided by the United States Geological Survey (USGS) (http://speclab.cr.usgs.gov/spectral.lib06.), with 221 spectral bands. The reflectance values are evenly distributed in the range of 0.4-2.5 μm. The abundance map containing 100×100 pixels is generated form the HyperMix standard dataset, where ANC and ASC are implemented. The data imitates the distribution features of surface objects in nature, such as surface water, coastline, vegetation, and mountains, so it is widely used to test the unmixing performance of different algorithms. The simulated data cube is synthesized by nine endmembers selected from the library $\mathbf{A}$, and then, white Gaussian noise with different signal-to-noise ratios (SNR) is added. The true fractional abundances of the nine endmembers are displayed in Figure 2.

Table 1 lists the value of SRE (dB) and p_s obtained by different algorithms on the simulated data at SNR = 30 dB, 40 dB, and 50 dB. From Table 1, we observe that the proposed LRSSU algorithm achieves higher SRE (dB) and p_s than other comparison algorithms at all noise levels. Compared with SUnSAL, CLSUnSAL, and SUnSAL-TV, the LRSSU algorithm shows great advantages in stability and

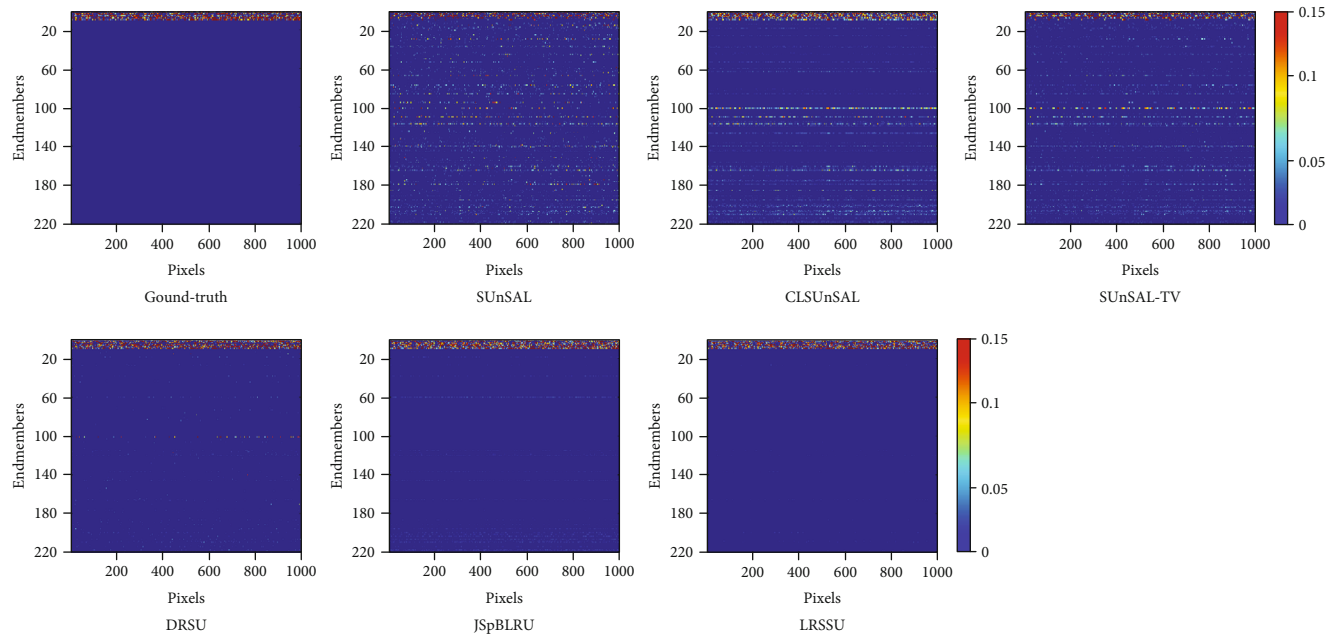


Figure 3: The ground-truth abundance and the estimated abundance maps obtained from the simulated data under noise ratio of SNR = 30 dB.

accuracy, which proves its superior unmixing performance. In the case of high noise, the LRSSU and JSPBLRU algorithms maintain high p_s, which indicates that minimizing the rank of the abundance matrix can effectively improve the robustness of the unmixing algorithm. The SRE (dB) obtained by the LRSSU algorithm is higher than that of the DRSU and JSpBLRU algorithms, which demonstrates that the local spatial weighting strategy plays an important role in improving the accuracy of unmixing.

To illustrate the effect of the proposed LRSSU algorithm on endmember identification, Figure 3 exhibits estimated abundances obtained by each algorithm on the simulated data with SNR = 30 dB. For visual clarity, Figure 3 displays only 1000 pixels selected from the image. Each colored line in the figure represents a nonzero row vector in the abundance matrix, corresponding to an actual endmember. It can be seen from the figure that the abundance map obtained by the LRSSU algorithm is consistent with the ground-truth abundance. More specifically, the endmembers identified by the LRSSU algorithm from the spectral library are the same as the true ones, with the least outliers. The abundance maps estimated by the SUnSAL, CLSUnSAL, and SUnSAL-TV algorithms emerge a mass of false endmember lines and interference noises, which do not match the actual endmembers well. The DRSU algorithm confuses some actual endmembers with other highly similar entries in the spectral library, resulting in endmember mismatches, and it is susceptible to noise. The results obtained by the JSpBLRU algorithm are disturbed by abundance vectors with small values corresponding to false endmembers. It can be concluded that the proposed LRSSU algorithm further enhances the row sparsity of the abundance matrix and improves the accuracy of identifying endmembers from the spectral library.

Figure 4 exhibits the abundance maps of the ninth endmember estimated by different algorithms from the simulated data at SNR = 30 dB. The other endmembers exhibit similar behavior, so they are not shown. For a more intuitive comparison, Figure 4 also shows the difference maps between the estimated abundance obtained by each algorithm and the true abundance. The abundance map obtained by the proposed LRSSU algorithm is close to the true distribution, which reveals a significant advantage than other algorithms. The SUnSAL, CLSUnSAL, and SUnSAL-TV algorithms estimate the abundances inaccurately, and the unmixing results are obviously different from the true abundance map. Compared with the DRSU and JSpBLRU algorithms, the LRSSU algorithm retains more spatial detail information of the abundance map, showing better spatial smoothness and less noise. It is proved that the combination of spectral-spatial collaborative sparse and low-rank constraints enhances the sparsity of the solution, preserves the finer structure and texture information of the image, promotes the spatial consistency between pixels, and improves the accuracy of unmixing.

The setting of regularization parameters has an important impact on the unmixing performance of the algorithm. To analyze the sensitivity of parameters in the proposed LRSSU algorithm, Figure 5 shows the SRE (dB) as functions of parameters λ and τ. We observe that when the noise level decreases, the value of the optimal parameter decreases accordingly. Moreover, the parameter λ affects the accuracy of solutions more than the parameter τ. Overall, when the regularization parameters λ and τ vary in a large range, SRE (dB) is still optimal or suboptimal. This is helpful to achieve the best performance of the algorithm and improve its practical application.

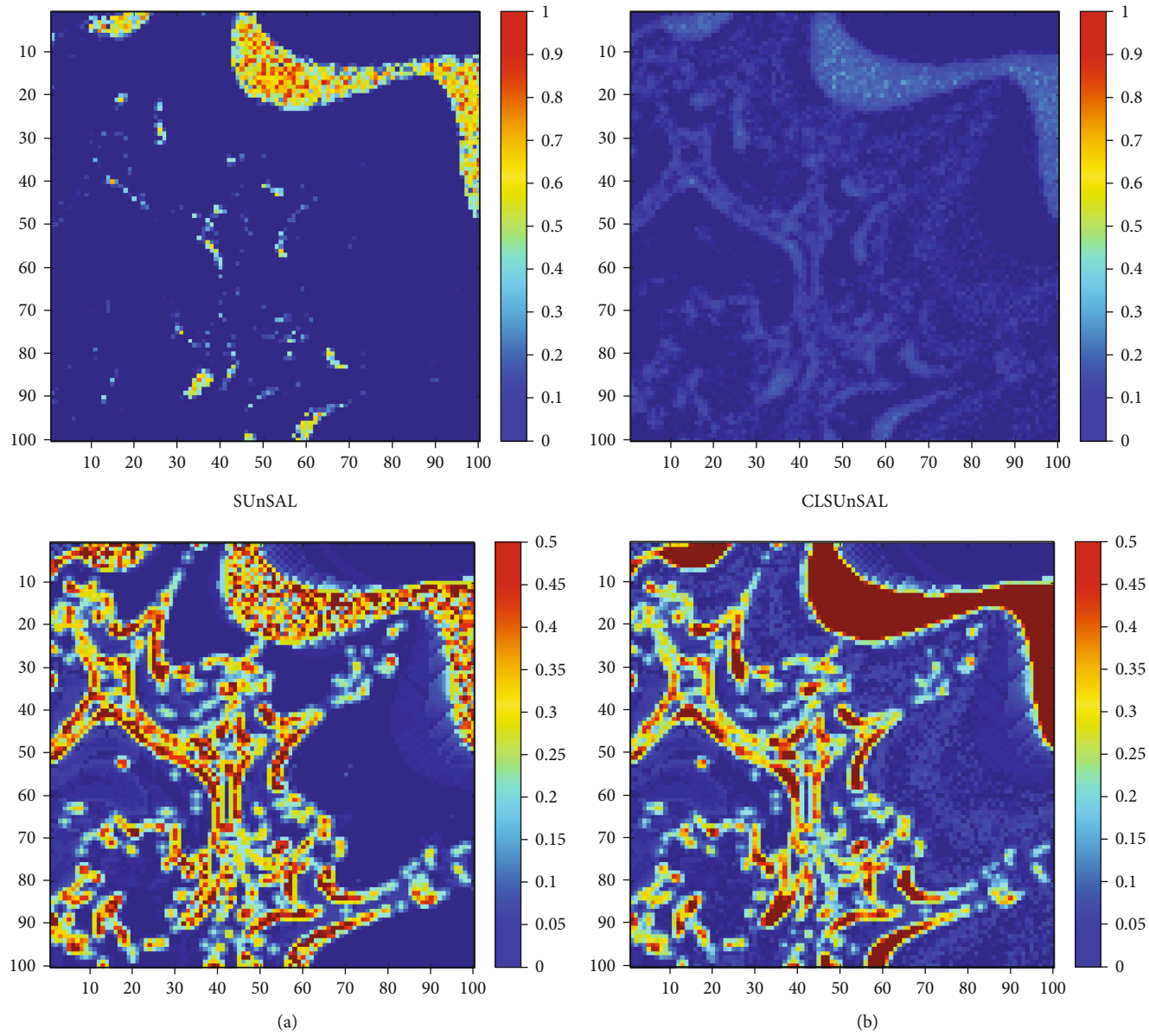


Figure 4: Continued.

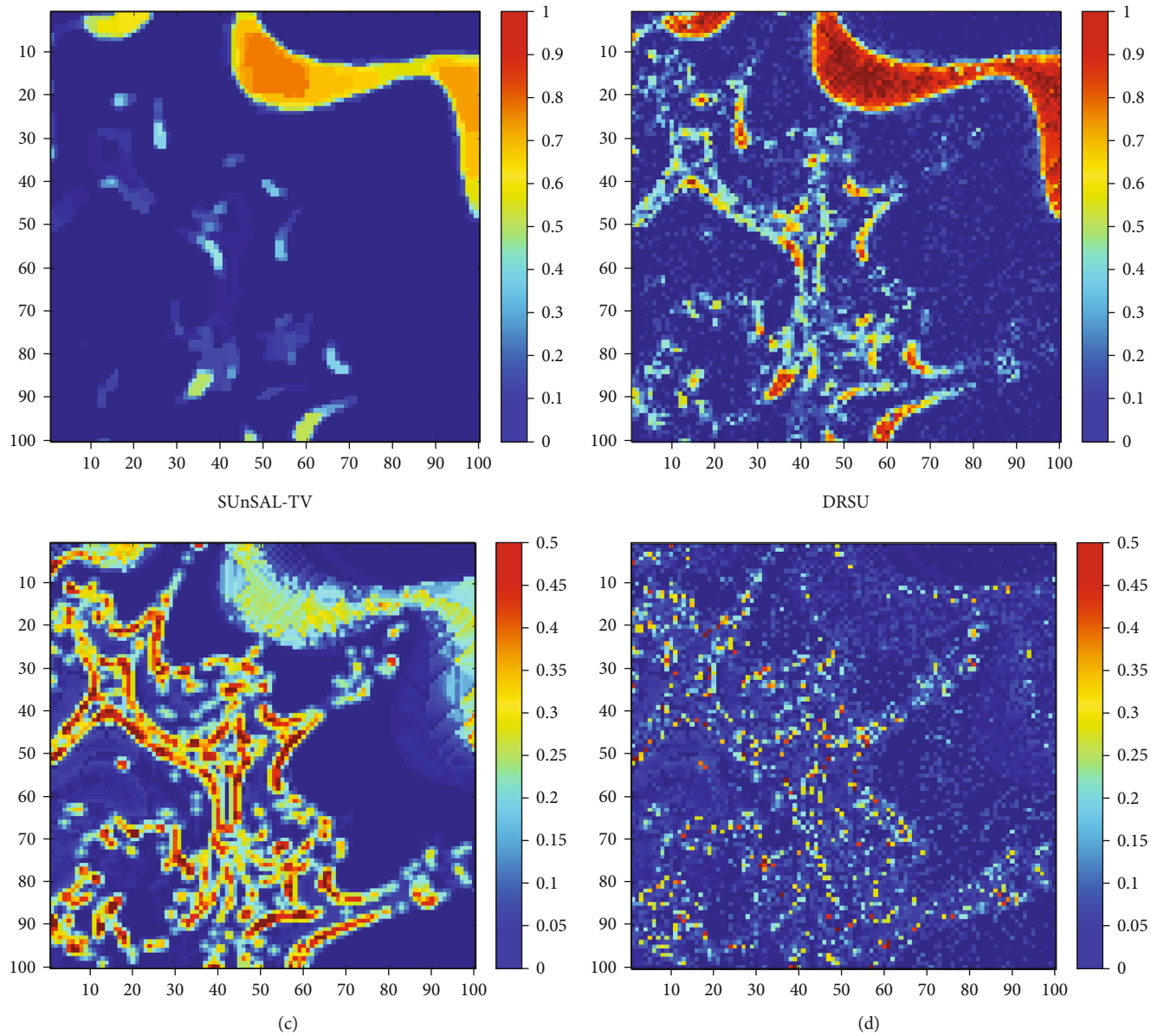


Figure 4: Continued.

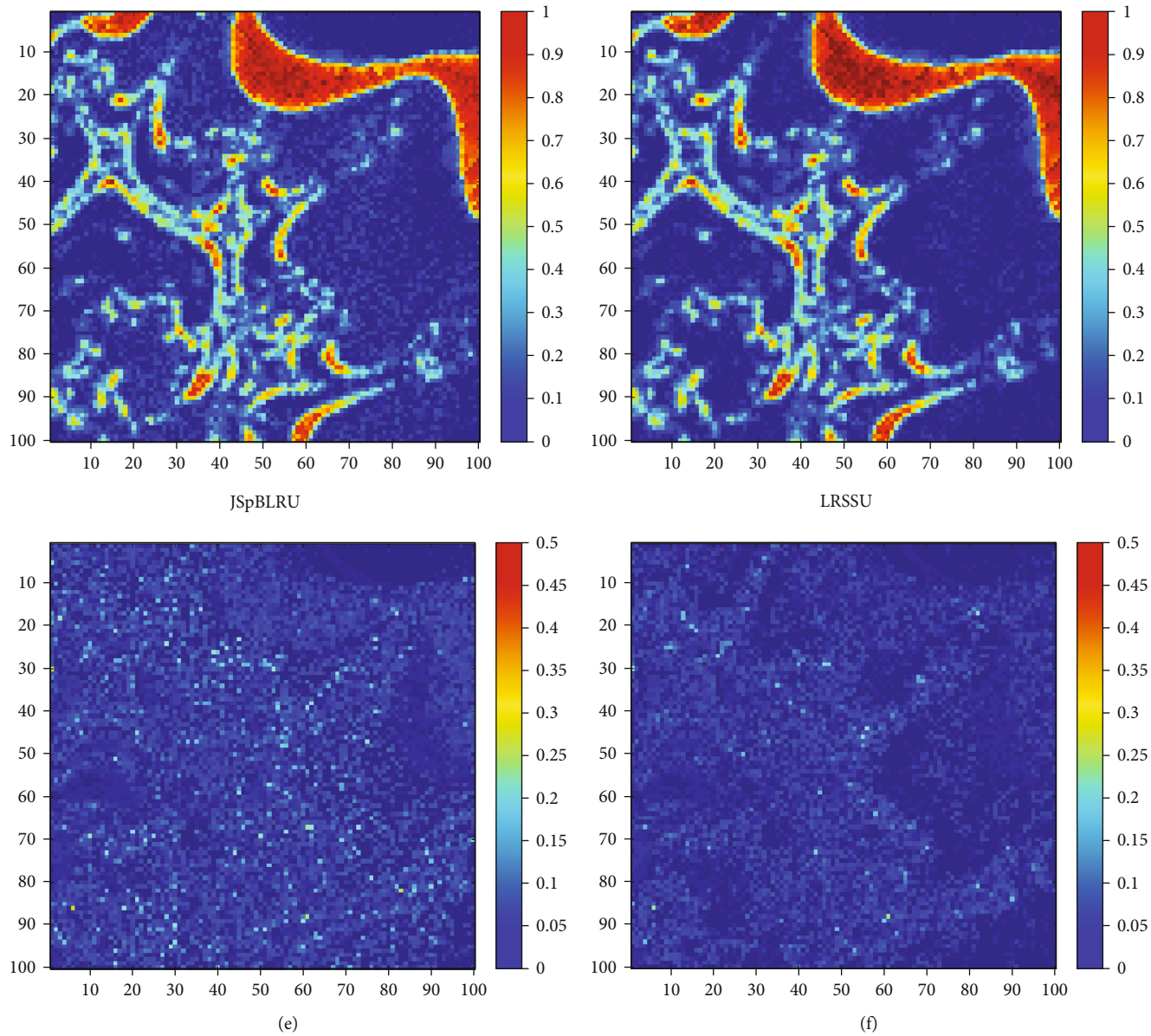


Figure 4: Abundance maps obtained for the ninth endmember from the simulated data under noise ratio of SNR = 30 dB. Difference maps between the ground-truth abundance and the results obtained by (a) SUnSAL, (b) CLSUnSAL, (c) SUnSAL-TV, (d) DRSU, (e) JSpBLRU, and (f) LRSSU.

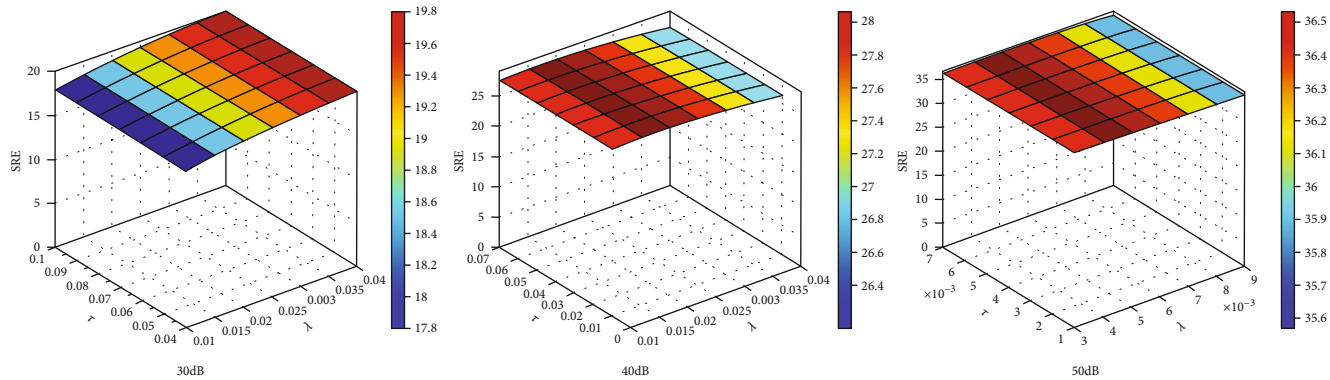


Figure 5: SRE (B) as functions of parameters λ and τ for the simulated data at SNR = 30 dB.

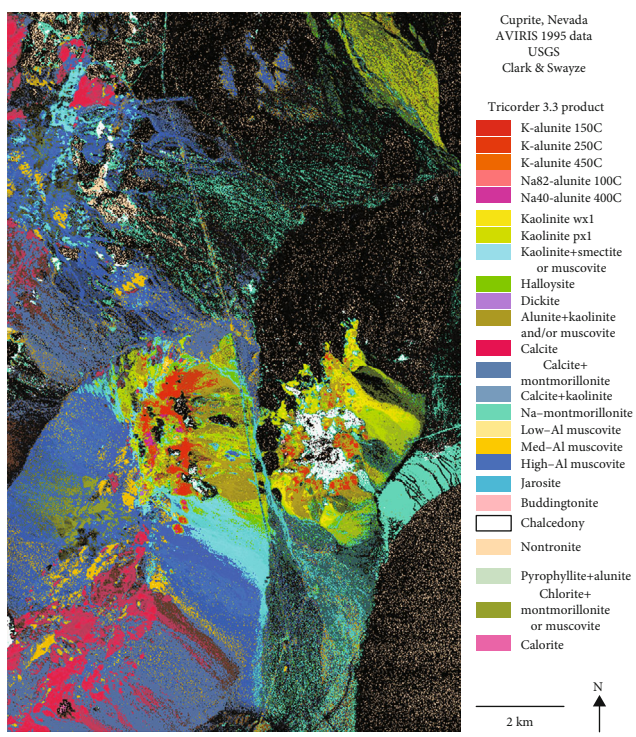


Figure 6: USGS mineral map of Cuprite mining district in Nevada [5].

4.2. Experiment on Real Data. The Cuprite mining dataset collected by the Airborne Visible Infrared Imaging Spectrometer (AVIRIS) (http://aviris.jpl.nasa.gov/html/aviris.freedata.html.) is used in the real data experiment to evaluate the unmixing performance of the proposed LRSSU algorithm. The part utilized in the experiment is a subset of the scene, with 100×100 pixels, including 224 bands in the range of 0.4-2.5 μm, with a spectral resolution of 10 nm. The experiment excludes the low SNR and low water absorption bands; that is, the 1-2, 105-115, 150-170, and 223-224 bands are removed, leaving 188 spectral bands. The spectral library utilized in the experiment is selected from the USGS library, which contains 240 spectral signatures representing different minerals. Low SNR and low water absorption bands are removed from the library in the same way, and 188 bands are left. Since the true abundance maps of the Cuprite dataset are not available, we will refer to the mineral classification map (as shown in Figure 6) generated by Tricorder software (http://speclab.cr.usgs.gov/PAPER/tetracorder.) to qualitatively analyze the performance of different unmixing algorithms [32]. The experiment meets the maximum number of iterations to ensure the convergence of each unmixing algorithm.

The regularization parameters involved in SUnSAL, CLSUnSAL, and DRSU algorithms are empirically set to $\lambda = 0.001$, $\lambda = 0.01$, and $\lambda = 0.0001$, and the parameters related to SUnSAL-TV, JSpBLRU, and LRSSU algorithms are empirically set to $(\lambda = 0.001, \lambda_{\mathrm{TV}} = 0.001)$ and $(\lambda = 0.05, \tau = 0.02)$, and $(\lambda = 0.08, \tau = 0.02)$, respectively. Figure 7 takes three typical minerals: alunite, buddingtonite, and

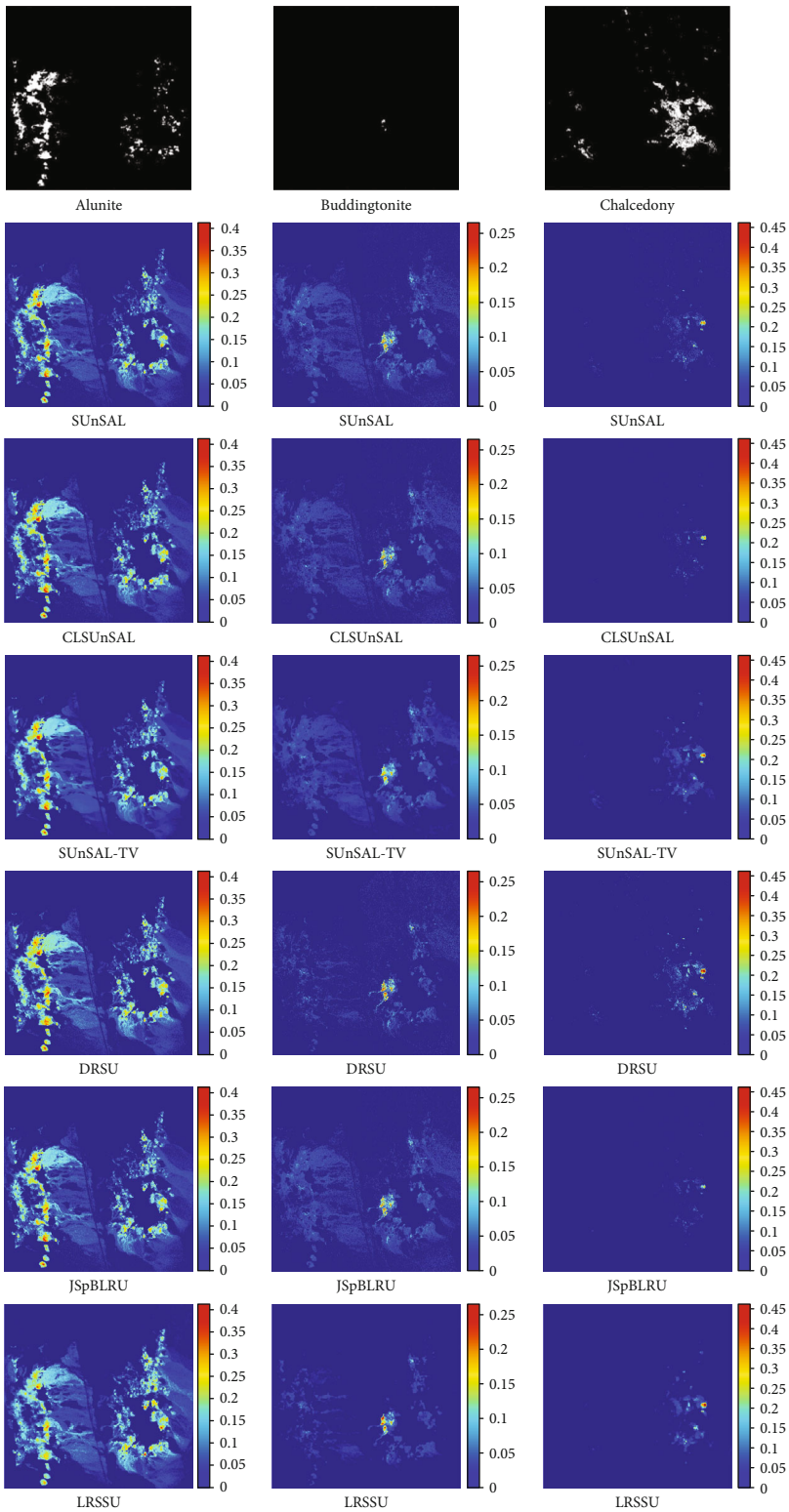


Figure 7: Fractional abundance maps estimated by different unmixing algorithms for the Cuprite subscene.

chalcedony, as examples to show the abundance maps obtained by unmixing the Cuprite dataset with the above algorithms. From Figure 7, we can observe that the abundance maps obtained by the SUnSAL and CLSUnSAL algorithms are noisy, especially in buddingtonite. The abundance maps obtained by the SUnSAL-TV algorithm are obviously oversmooth, such as alunite and buddingtonite. The abundance maps obtained by the DRSU algorithm have poor spatial consistency, such as chalcedony. In addition, the abundance maps estimated by the LRSSU algorithm are less disturbed by noise than the JSPBLRU algorithm that also adopts the low-rank constraint, such as buddingtonite, and exhibit better spatial consistency, such as chalcedony. The abundance maps of the proposed LRSSU algorithm are closest to the Tricorder reference maps among all comparison algorithms. The qualitative analysis for the real data experiment confirms that the LRSSU algorithm is effective in unmixing real hyperspectral data and gets better unmixing results.

5. Conclusions and Future Work

In this paper, we propose a new sparse reduced-rank regression model for hyperspectral unmixing, which merges spatial and spectral information into the regularizer to enhance the interpretation of hyperspectral data. A local spatial weighting factor is introduced in the collaborative sparse unmixing model, which promotes the spatial smoothness of the image. Moreover, a low-rank constraint is imposed on the abundance estimation to enhance the row sparsity of the abundance matrix, which to some extent overcomes the influence of the high correlation of spectral signatures in the library. The proposed LRSSU model is solved by an ADMM-based the inner and outer loop scheme, the inner loop updates fractional abundances, and the outer loop calculates the spatial weights. The scheme accelerates the convergence of the algorithm. Simulated and real hyperspectral data experiments reveal that the proposed LRSSU algorithm enhances the identification of actual endmembers in the spectral library. Compared with SUnSAL, CLSUnSAL, SUnSAL-TV, DRSU, and JSPBLRU algorithms, the proposed LRSSU algorithm improves the accuracy of abundance estimation and shows the advantages of the spectral-spatial multiple constraint strategy. In the future, we will extend the constraint strategy to blind unmixing and further explore the spectral-spatial structure information of the hyperspectral data cube represented by a third-order tensor.

Data Availability

The United States Geological Survey (USGS) library used in simulated and real data experiments is available online at http://speclab.cr.usgs.gov/spectral.lib06. The Airborne Visible Infrared Imaging Spectrometer (AVIRIS) Cuprite dataset used in the real data experiment is available online at http://aviris.jpl.nasa.gov/html/aviris.freedata.html.

Conflicts of Interest

The author declares that there are no conflicts of interest.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grant 61865012, in part by Basic Science and Technology Research Project of National Key Laboratory of Science and Technology on Automatic Target Recognition under Grant WDZC2020 5500204, in part by Key Research and Development Program of Jiangxi Province under Grant 20202BBGL73081 and Grant 20181ACG70022, and in part by the Science and Technology Project of Jiangxi Provincial Department of Education under Grants GJJ170992 and GJJ190956.

References

[1] J. M. Bioucas-Dias, A. Plaza, G. Camps-Valls, P. Scheunders, N. M. Nasrabadi, and J. Chanussot, "Hyperspectral remote sensing data analysis and future challenges," *IEEE Geoscience and Remote Sensing Magazine*, vol. 1, no. 2, pp. 6–36, 2013.

[2] L. Drumetz, M. A. Veganzones, R. Marrero et al., "Binary partition tree-based local spectral unmixing," *Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, 2014.

[3] J. M. Bioucas-Dias, A. Plaza, N. Dobigeon et al., "Hyperspectral unmixing overview: geometrical, statistical, and sparse regression-based approaches," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 5, no. 2, pp. 354–379, 2012.

[4] D. Landgrebe, "Hyperspectral image data analysis," *IEEE Signal Processing Magazine*, vol. 19, no. 1, pp. 17–28, 2002.

[5] M.-D. Iordache, J. M. Bioucas-Dias, and A. Plaza, "Sparse unmixing of hyperspectral data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 6, pp. 2014–2039, 2011.

[6] C. Deng, S. Zhang, W. T. ShengqianWang, and Z. Wu, "Sparse hyperspectral unmixing based on smoothed l0 regularization," *Infrared Physics and Technology*, vol. 67, pp. 306–314, 2014.

[7] F. Li, S. Zhang, B. Liang, C. Deng, C. Xu, and S. Wang, "Hyperspectral sparse unmixing with spectral-spatial low-rank constraint," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 6119–6130, 2021.

[8] L. Qi, J. Li, Y. Wang, Y. Huang, and X. Gao, "Spectral-spatial-weighted multiview collaborative sparse unmixing for hyperspectral images," *IEEE Transactions on Geoscience and Remote Sensing*, pp. 1–14, 2020.

[9] M.-D. Iordache, J. M. Bioucas-Dias, and A. Plaza, "Collaborative sparse regression for hyperspectral unmixing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 1, pp. 341–354, 2014.

[10] S. Zhang, J. Li, H. Li, C. Deng, and A. Plaza, "Spectral-spatial weighted sparse regression for hyperspectral image unmixing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 6, pp. 3265–3276, 2018.

[11] H. Li, R. Feng, L. Wang, Y. Zhong, and L. Zhang, "Superpixel-based reweighted low-rank and total variation sparse unmixing for hyperspectral remote sensing imagery," *IEEE Transactions on Geoscience and Remote Sensing*, pp. 1–19, 2020.

[12] J. Liu, J. Zhang, Y. Gao, C. Zhang, and Z. Li, "Enhancing spectral unmixing by local neighborhood weights," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 5, no. 5, pp. 1545–1552, 2012.

[13] R. Wang, H.-C. Li, W. Liao, and A. Pizurica, "Double reweighted sparse regression for hyperspectral unmixing," in *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pp. 6986–6989, Beijing, China, July 2016.

[14] G. Martin and A. Plaza, "Spatial-spectral preprocessing prior to endmember identification and unmixing of remotely sensed hyperspectral data," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 5, no. 2, pp. 380–395, 2012.

[15] J. Li, C. Wu, R. Song et al., "Deep hybrid 2-d-3-d CNN based on dual second-order attention with camera spectral sensitivity prior for spectral super-resolution," *IEEE Transactions on Neural Networks and Learning Systems*, 2021.

[16] O. Eches, N. Dobigeon, and J.-Y. Tourneret, "Enhancing hyperspectral image unmixing with spatial correlations," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 11, pp. 4239–4247, 2011.

[17] X. Wang, J.-S. Pan, and S.-C. Chu, "A parallel multi-verse optimizer for application in multilevel image segmentation," *IEEE Access*, vol. 8, pp. 32018–32030, 2020.

[18] S. Zhang, G. Zhang, F. Li et al., "Spectral-spatial hyperspectral unmixing using nonnegative matrix factorization," *IEEE Transactions on Geoscience and Remote Sensing*, pp. 1–13, 2021.

[19] M.-D. Iordache, J. M. Bioucas-Dias, and A. Plaza, "Total variation spatial regularization for sparse hyperspectral unmixing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 11, pp. 4484–4502, 2012.

[20] W. He, H. Zhang, and L. Zhang, "Total variation regularized reweighted sparse nonnegative matrix factorization for hyperspectral unmixing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 7, pp. 3909–3921, 2017.

[21] S. Zhang, J. Li, K. Liu, C. Deng, L. Liu, and A. Plaza, "Hyperspectral unmixing based on local collaborative sparse regression," *IEEE Geoscience and Remote Sensing Letters*, vol. 13, no. 5, pp. 631–635, 2016.

[22] Q. Qu, N. M. Nasrabadi, and T. D. Tran, "Abundance estimation for bilinear mixture models via joint sparse and low-rank representation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 7, pp. 4404–4423, 2014.

[23] L. Sun, F. Wu, T. Zhan, W. Liu, J. Wang, and B. Jeon, "Weighted nonlocal low-rank tensor decomposition method for sparse unmixing of hyperspectral images," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, pp. 1174–1188, 2020.

[24] P. V. Giampouras, K. E. Themelis, A. A. Rontogiannis, and K. D. Koutroumbas, "Simultaneously sparse and low-rank abundance matrix estimation for hyperspectral image unmixing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 8, pp. 4775–4789, 2016.

[25] J. Huang, T. Huang, L. Deng, and X. Zhao, "Joint-sparse-blocks and low-rank representation for hyperspectral unmixing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 4, pp. 2419–2438, 2019.

[26] R. G. Baraniuk, V. Cevher, M. F. Duarte, and C. Hegde, "Model-based compressive sensing," *IEEE Transactions on Information Theory*, vol. 56, no. 4, pp. 1982–2001, 2010.

[27] E. J. Candes and T. Tao, "Near-optimal signal recovery from random projections: universal encoding strategies?," *IEEE Transactions on Information Theory*, vol. 52, no. 12, pp. 5406–5425, 2006.

[28] S. Gu, L. Zhang, W. Zuo, and X. Feng, "Weighted nuclear norm minimization with application to image denoising," in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2862–2869, Columbus, OH, USA, June 2014.

[29] J. M. Bioucas-Dias and M. Figueiredo, "Alternating direction algorithms for constrained sparse regression: application to hyperspectral unmixing," in *2010 2nd Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing*, pp. 1–4, Reykjavik, Iceland, June 2010.

[30] S.-C. Chu, D. Zhi-Gang, Y.-J. Peng, and J.-S. Pan, "Fuzzy hierarchical surrogate assists probabilistic particle swarm optimization for expensive high dimensional problem," *Knowledge-Based Systems*, vol. 220, p. 106939, 2021.

[31] P.-C. Song, J.-S. Pan, and S.-C. Chu, "A parallel compact cuckoo search algorithm for three-dimensional path planning," *Applied Soft Computing*, vol. 94, p. 106443, 2020.

[32] R. Clark, G. Swayze, K. Livo et al., "Imaging spectroscopy: Earth and planetary remote sensing with the USGS Tetracorder and expert systems," *Journal of Geophysical Research*, vol. 108, no. E12, pp. 5131–5135, 2003.

Hindawi
Wireless Communications and Mobile Computing
Volume 2021, Article ID 9632277, 10 pages
https://doi.org/10.1155/2021/9632277

Research Article

An Optimal Allocation Method of Power Multimodal Network Resources Based on NSGA-II

Ao Xiong,[1] **Yuanzheng Tong**,[1] **Shaoyong Guo**,[1] **Yanru Wang**,[2] **Sujie Shao**,[1] **and Lin Mei**[3]

[1]*State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing 100876, China*
[2]*Beijing FibrLINK Communications Co., Ltd., Beijing 100070, China*
[3]*State Grid Henan Information & Telecommunication Company, Zhengzhou 450052, China*

Correspondence should be addressed to Yuanzheng Tong; tongyz@bupt.edu.cn

Received 23 July 2021; Accepted 4 October 2021; Published 26 October 2021

Academic Editor: Chi-Hua Chen

Basic services for power business were provided by the power multimodel network providers. However, because the power multimodal network is usually complex and changeable, the service of power business is often unstable. This problem can be solved by a suitable network resource optimization method. Therefore, how to design a network resource optimization method that seeks a compromise between multiple performance indicators that achieve the normal operation of power multimode networks is still extremely challenging. An optimal allocation method of power multimodal network resources based on NSGA-II was proposed by this paper. Firstly, the power multimodal network-resource model is established, and the problems existing in the resource optimization process are analyzed. Secondly, preprocessing technology and indirect coding technology are applied to NSGA-II, which solves the coding problem and convergence problem of the application of genetic algorithm to the optimization of network resource allocation. Finally, the simulation results show that, compared with the control algorithm, this method has further optimized the various indicators of the resource allocation of the power multimodal network, and the performance has been improved by more than 6%.

1. Background

New business needs with large connections and wide coverage have gradually arisen by the research and development of power multimodal networks. However, the current power multimodal network is difficult to meet the business processing needs, which will result in a significant increase in network transmission pressure and computing load. In addition, in the power multimode network, the network structure is relatively complicated, many different types of communication equipment are needed, and the connection methods between the devices and the information conversion methods are different, resulting in the entire power multimode network [1]. The network structure is very complicated which causes the power multimodal network structure to become very complicated. At the same time, the power system is the infrastructure of daily life, which requires the power multimode network to have higher stability. Therefore, the power multimode network is also required to be in the process of working without various interruptions and sudden changes, which means higher flexibility and reliability.

Optimizing the software and hardware in the power multimodal network to improve the quality of service and meet the business needs is necessary to solve the above problems. At the software level, reasonable resource optimization can be performed on multiple indicators such as link delay, reliability, and resource occupancy distribution of the power multimodal network [2]. On the one hand, establishing a reasonable power multimodal network resource model is needed for common resource optimization methods, and on the other hand, reasonable optimization methods for different resource optimization scenarios are also needed.

Therefore, a suitable and reasonable power multimodal network resource model should be established firstly. Some

existing model schemes only consider single indicators, such as delay and reliability [3], while others consider various network indicators comprehensively, but only part of the indicators can be optimized due to the limitation of optimization methods [4]. Secondly, choosing a reasonable method for the scene of the power multimodal network is necessary. Because network resource optimization is an NP-hard problem, heuristic algorithms are proposed to obtain approximate optimal solutions, but it is easy to fall into local optimal results. In addition, among the heuristic algorithms, some heuristic algorithms such as the ACS algorithm were originally designed for discrete problems and are not suitable for network resource optimization. The GA algorithm designed for continuous problems is considered an effective way to solve the problem of network resource optimization. Among them, the NSGA-II algorithm uses fast nondominated sorting with elite strategy, which solves the main shortcomings of NSGA and achieves fast and accurate search performance, so it is suitable for complex and multiobjective optimization problems [5].

In summary, in view of the above-mentioned resource optimization problem in the power multimodal network, this paper proposes a multimodal network resource optimization model and designs a power multimodal network resource optimization allocation method based on NSGA-II to optimize the model. The main contributions of this paper are listed as follows.

(1) From the perspective of resource association and business relevance, the indicators of resource expenditure, link reliability, resource occupancy, and distribution in the power multimode network are analyzed, and the power multimode network resource optimization model is designed to ensure service quality of the power multimode network

(2) Based on the NSGA-II algorithm, a power multimodal network resource optimization allocation method is designed and implemented to solve the resource optimization problem. By introducing indirect coding technology and preprocessing technology, the algorithm is guaranteed convergence speed while the network resources are optimized

The rest of this paper is organized as follows. Section 2 discusses the related work. Section 3 builds a power multimodal network optimization model. Section 4 proposes a multiobjective optimization method based on NSGA-II. Section 5 verifies the performance of the algorithm through simulation, and the last chapter summarizes the full text and draws conclusions.

2. Related Work

In this section, we focus on two important stages in the power multimode network, which are optimized allocation of network resources and multiobjective optimization.

2.1. Optimized Allocation of Network Resources. The optimization of network resource allocation is mainly at the software level by recontrolling and adjusting service paths in multimodal network logic resources to improve network resource utilization and reduce service response time, so as to meet the service quality of various power services. Reference [6] proposes an entropy-based business uniform distribution algorithm, which analyzes business traffic according to the business characteristics of the power communication network. The distribution of business traffic reflects the operating status of the network business, and the business information entropy is used as a measure of the uniformity of network business distribution. Then, use the information entropy as the objective function to get the algorithm for optimizing global business routing. Reference [7] proposes a hierarchical QoS routing method based on network resource reservation by calculating the required routing and determining the allocation sequence of the link resources to be reserved, thereby minimizing the response time. Reference [8] discusses the online convex optimization problem including the adversarial loss function and adversarial constraints and proposes an improved online saddle point (MOSP) method and applies it to the dynamic network resource allocation problem. Reference [9] proposed an integrated analytic hierarchy process (IAHP), which established an optimization model under the premise of considering factors such as resource cost, connectivity, and reliability, to obtain a relatively balanced solution under the condition of satisfying constraints.

In short, the current solutions for network resource optimization are mostly focused on optimizing network service paths, which means changing routing options through different methods to ensure network service quality. However, the optimization results of the above methods are relatively simple, which lack consideration of power multimodal network scenarios. Moreover, the optimization measures of some methods are often fixed when the network is initialized, and there is no suitable solution to the problem of dynamic network resource allocation.

2.2. Multiobjective Optimization. Network resource optimization problems can be abstracted as multiobjective optimization problems to solve. The existing multiobjective optimization problems usually have three solving methods. The first one is to take a specific objective as the optimization objective and other objectives as constraints, thereby simplifying the multiobjective optimization problem to a single objective optimization problem [3]. The second method is to assign different weights to each goal according to the importance of each goal and use the weight method to get the aggregate objective function through weighted summation of each goal, thereby simplifying the multiobjective optimization problem to a single-objective optimization problem [4]. The third method is multiobjective collaborative optimization. The optimal solution set composed of many Pareto optimal solutions is first jointly optimized according to multiple indicators and then compared and selected according to the optimization direction [10]. Because the optimization requirements for various indicators in multimodal networks are different, this paper adopts a multiobjective collaborative optimization method to optimize network resource allocation.

The evolutionary algorithm of multiobjective collaborative optimization mainly includes multiobjective genetic algorithm, artificial immune algorithm, multiobjective PSO constraint algorithm, and ant colony algorithm [10]. The operation of artificial immune algorithm lacks stability [11], and the research of multiobjective PSO constraint algorithm is still in its infancy [12]. Multiobjective ant colony algorithm will appear premature stagnation, and the control parameters are difficult to determine [13], so the multiobjective genetic algorithm with good robustness and superiority is adopted by this paper [14].

Multiobjective genetic algorithms are widely used in various fields. Reference [15] uses genetic algorithm to solve the MOTSP problem and generates an approximate optimal solution. In Reference [16], from the viewpoint of the minimum actual power loss of the system, genetic algorithms are used to evaluate the impact of the optimal configuration of different types of dg DLMs in the distribution network, and the similarities and differences between different distributed power sources are studied. Reference [17] proposed a multiobjective optimization method that combines artificial neural network and genetic algorithm. Firstly, artificial neural network is used to predict the properties of nanofluids, and then, genetic algorithm is used for multiobjective optimization. As a classic multiobjective collaborative optimization genetic algorithm, NSGA-II adds an elite strategy on the basis of NSGA, adopts the concept of congestion, and reduces the complexity of the algorithm. Reference [4] uses a special fitness function for NSGA-II and uses a method to improve solution diversity. Reference [18] proposes an optimized classification model that constructs linear equations based on classification problems, which can better spread the solution and has higher classification accuracy and robustness.

It can be seen from the above work that there are few researches on power multimodal network resource optimization. In addition, the existing multiobjective optimization model is also difficult to meet the needs of power services that have different QoS requirements for various indicators.

3. Power Multimodal Network Model

As shown in Figure 1, the power multimodal network resource optimization architecture is mainly divided into three layers: a trusted monitoring platform for network resources, multimodal network links, and multimodal network resources. Under the precondition of accurate perception and recognition of multiple services through network monitoring technology, the trusted monitoring platform for network resources uses QoS as a comprehensive evaluation index for network services to analyze multiple network indicators such as delay and reliability, so as to implement unified resource scheduling for multimodal network resources through resource joint optimization strategies.

In general, the physical resources of a multimodal network are relatively fixed. Therefore, the resource optimization of power multimodal network mainly adjusts logical resource allocation through multiobjective collaborative optimization to ensure network service quality. There are many addressing methods for multimodal networks, such as spatial addressing based on content identification, addressing based on geographic location, and addressing based on identity identification [1], which can all be abstracted as the path change of each node in the network. Therefore, this problem can be abstracted as a multiobjective optimization problem of specific business under multimodal network extension [19]. The network topology can be represented as graph $G(V, E)$, where $V = \{v_1, v_2, v_3 \cdots v_n\}$ is a collection of nodes and $E = \{e_1, e_2, e_3 \cdots e_n\}$ is a collection of node links.

In a multimode network with multiple services, different optimization indicators and requirements are required by different service providers and users [20]. Therefore, according to the network index requirements of different services, the following three types of optimization objectives are considered by this paper: resource overhead of integrated delay and cost, link reliability, and resource occupation distribution. Mapped to the above network model, that is, for any service from the source node to the destination node, a link $L(s, d) = \{l_{s,i}, l_{i,j} \cdots l_{k,d}\}$ should be found to satisfy QoS, which reduce resource overhead, reach better reliability, improve link resource utilization, balance network load, and improve network throughput.

3.1. Resource Overhead. In the network topology, the composition of the end-to-end total service delay of the service resource chain is very complicated, which is mainly composed of three parts: link delay, node processing delay, and queuing delay. Considering that the processing delay and queuing delay of nodes in the network link are usually below the microsecond level, while the link delay can normally reach the millisecond level, the impact of link delay on service delay is mainly considered by this paper.

$$\mathrm{Delay}(s, d) = \sum_{i=1}^{n} \sum_{j=1}^{n} \mathrm{delay}(i, j) * \sigma(i, j), \tag{1}$$

where $\mathrm{Delay}(s, d)$ is the total delay of the service link, $\mathrm{delay}(i, j)$ is the delay of the link from node i to node j, and $\sigma(i, j)$ indicates whether the link is selected, which is

$$\sigma(i, j) = \begin{cases} 1, & l_{i,j} \in L(s, d), \\ 0, & l_{i,j} \notin L(s, d). \end{cases} \tag{2}$$

At the same time, for a service path, the cost of the service path is measured by the product of the proportion of service bandwidth in each link and the cost of each link.

$$\mathrm{Cost}(s, d) = \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{B(s, d)}{\mathrm{bandwidth}(i, j)} * \mathrm{cost}(i, j) * \sigma(i, j), \tag{3}$$

where $\mathrm{Cost}(s, d)$ is the overall cost of the service link, $B(s, d)$ is the bandwidth required for the service, $\mathrm{bandwidth}(i, j)$ is the remaining bandwidth of each link, and $\mathrm{cost}(i, j)$ is the cost overhead from node i to node j.

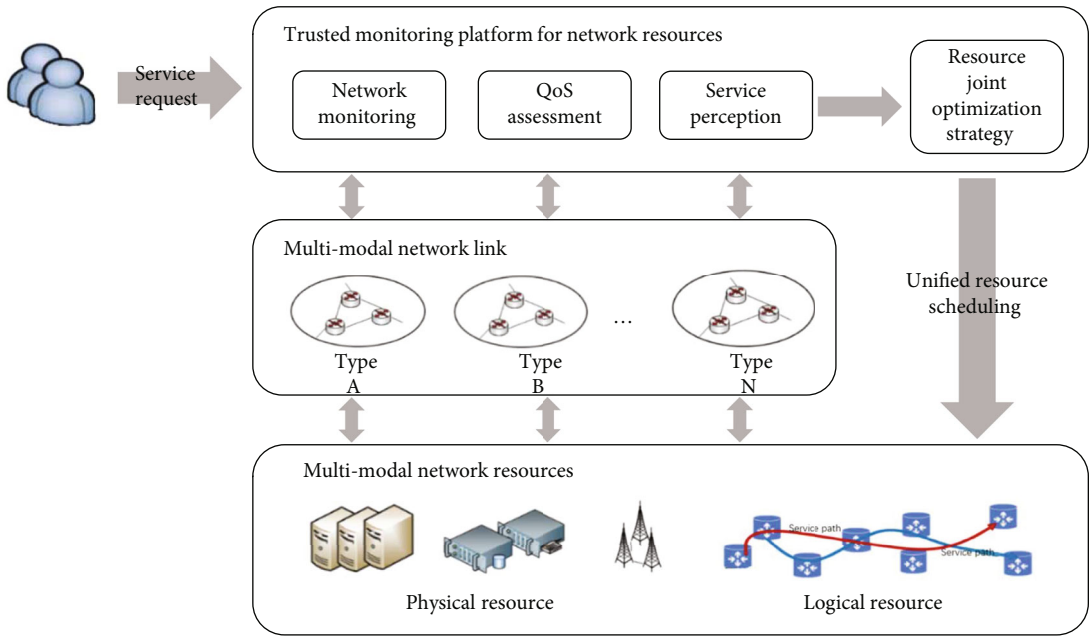


Figure 1: Power multimodal network resource optimization architecture.

The impact of delay time and the cost of reserved resources on the link are considered by this paper comprehensively to measure the network resource cost of the selected path through the resource consumption function as follows.

$$\text{Resource}(s, d) = \text{Delay}(s, d) * \text{Cost}(s, d), \tag{4}$$

where $\text{Resource}(s, d)$ is the resource consumption function.

3.2. Link Reliability. According to whether the path matrix is connected and the reliability of each corresponding link, the reliability index of each section is multiplied to calculate the overall reliability index of the link.

$$\text{Reliability}(s, d) = \prod_{l_{i,j} \in L} \text{reliability}(i, j). \tag{5}$$

Among them, $\text{Reliability}(s, d)$ is the overall reliability of the service link, and $\text{reliability}(i, j)$ is the reliability of the link from node i to node j.

3.3. Resource Occupation Distribution. The proportion of service bandwidth in each link is used to measure resource occupancy in this paper. At the same time, in order to distribute resources evenly on each link, the variance of the proportion of service bandwidth in each link, which is presented by $\delta^2(s, d)$, is used to measure the distribution of occupied resources.

Suppose $\text{Utilization}(i, j) = B(s, d)/\text{bandwidth}(i, j)$ is the resource occupation distribution from node i to node j, then

$$\begin{aligned} \text{Utilization}(s, d) &= \frac{\sum_{l_{i,j} \in L} \text{Utilization}(i, j)}{\sum_{l_{i,j} \in L} \sigma(i, j)}, \\ \delta^2(s, d) &= \sum_{l_{i,j} \in L} \left(\text{Utilization}(i, j) - \text{Utilization}(s, d)\right)^2, \end{aligned} \tag{6}$$

where $B(s, d)$ is the bandwidth required by the current service from the start node s to the destination node d, $\text{bandwidth}(i, j)$ is the bandwidth from node i to node j, so $\text{Utilization}(i, j)$ represents the bandwidth utilization from node i to node j, and $\delta^2(s, d)$ represents the resource utilization variance from node s to node d.

3.4. Optimization Goals. The goal of multiobjective network resource optimal allocation is to find the optimal solution set as much as possible on the premise of ensuring business QoS indicators according to corresponding business requirements. Therefore, meeting the following conditions should be needed by the selected path:

$$\begin{cases} \text{Min}\ (\text{Resource}(s, d)), \\ \text{Max}\ (\text{Reliability}(s, d)), \\ \text{Min}\ \left(\delta^2(s, d)\right). \end{cases} \tag{7}$$

It can be seen from the above formula that the multimodal network resource optimization results usually hope to obtain the minimum resource consumption, maximum reliability, minimum resource occupation distribution, which is a multiobjective optimization problem with a set of optimal solutions composed of many Pareto optimal

solutions, so users can choose the corresponding link according to their business needs.

4. An Optimal Allocation Method of Power Multimodal Network Resources Based on NSGA-II

The general operation process of genetic algorithm is mainly composed of several steps such as coding, initial population generation, fitness function determination, selection method, crossover, and mutation processing.

Aiming at the power multimodal network scenario, a power multimodal network resource optimization allocation method, which designs specific schemes for each step based on the NSGA-II genetic algorithm, is realized by this paper to meet the power business needs.

4.1. Population Initialization

4.1.1. Pretreatment. Before the population is initialized, the input data can be preprocessed to quickly filter out the data links that do not meet the bandwidth requirements, so as to reduce the operation scale of the algorithm. This preprocessing step traverses the network topology and takes the bandwidth required by the business as the standard. All data links smaller than this bandwidth are regarded as nonconnected links. The path is removed from the network topology diagram, so the network links all meet the bandwidth requirements in the final network and the bandwidth constraints will no longer be considered in subsequent studies.

4.1.2. Population Coding. There are usually two ways to initialize population: binary encoding and floating-point encoding. The binary code uses a two-dimensional matrix to represent the data link and uses $1, 0$ to indicate whether the link is connected. Usually, a compressed matrix storage method is used to reduce the space occupation, but this kind of representation method is more difficult to verify data connectivity and decoding, so this article uses floating-point encoding. Considering that the length of the network transmission path is not fixed, this article uses an indirect encoding method based on priority encoding [21]. The chromosome based on priority coding does not directly represent the selection path of the current individual, but only the priority of the current node being selected, and the true path corresponding to the chromosome needs to be obtained through a decoding operation.

4.2. Congestion Function. In the genetic evolution algorithm, the fitness function is used to determine the degree of individual adaptation to the "environment," so as to screen outstanding individuals to produce a new generation of populations. The fitness value in the evolutionary algorithm is the function value corresponding to the optimization goal, and the inferior individuals are eliminated by calculating the fitness value.

This algorithm, which draws on the idea of elite selection in the NSGA-II algorithm, firstly calculates the nondominated stratum where each individual in the population is located to achieve population reduction operations and then compares the degree of crowding. Individuals with small nondominated tiers and high crowdedness enter the next iteration to generate new individuals. The calculation formula for congestion $F_i(s, d)$ is

$$F_i(s, d) = \sum_{j=1}^{m} \left(\alpha_j \left| f_j^{i+1} - f_j^{i-1} \right| \right). \tag{8}$$

In the above formula, f_j^{i+1} represents the jth objective function value of the $i+1$th individual and α_j indicates the weight of the jth objective function value.

Because different service providers and users have different optimization indicators and requirements in network resource optimization, so different weights are assigned to each indicator to meet business needs according to different businesses, and they need to meet $\sum_{j=1}^{m} \alpha_j = 1$.

4.3. Population Selection. In this paper, the best retention selection method is used to select the population. First, the entire population is selected by the roulette method to select the genetic algorithm. The selection probability of the ith individual is $P_i = F_i / \sum_{k=1}^{n} F_k$ (where n is the population size), and then, the structure of the most adaptable individual is completely copied to the next-generation population to complete the selection of the entire population in the current population.

4.4. Population Crossover and Mutation

4.4.1. Population Crossing. Since the general crossover mutation may cause the data link to be blocked and have a negative optimization impact, this paper uses the sequential crossover method to carry out the population crossover operation as follows [14].

(1) Randomly select the start and end positions of several genes in a pair of parental chromosomes (the positions of the two chromosomes are the same)

(2) Generate a progeny based on the selected genes, and ensure the position of the selected gene in the progeny same as the parent

(3) Finally, find out the position of the gene selected in the first step in another parent, and then, put the rest of the genes in order in the offspring generated in the previous step so that a new offspring is produced

It should be noted that this crossover operation will also generate two offspring. The generation process of the other offspring is exactly the same. The genotype position selected in the first step is the same, and only the two parent chromosomes need to be exchanged.

4.4.2. Population Variation. The population generated after fitness calculation, selection, and crossover may converge to the local optimal solution instead of the global optimal solution. Therefore, mutation operation is needed to promote the population to jump out of the local optimal situation. In this paper, the chromosome segment reverse

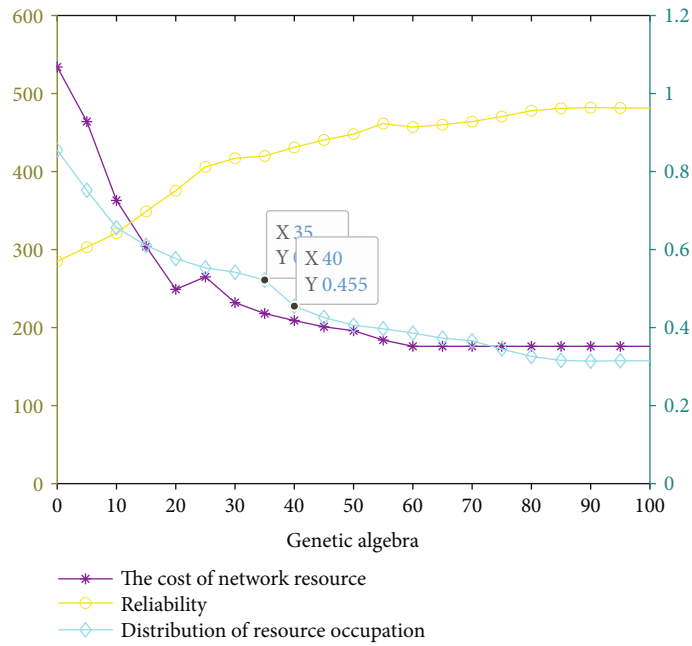


Figure 2: The change trend of various indicators in the genetic process.

transformation is performed on the new individuals generated by the crossover with a certain probability, and the result of the mutation will check whether the link is connected. If the link is connected, join the population; if not, perform a population selection operation to replace the path.

4.5. Restrictions. Assuming that each node in the network has enough buffer space to store packet data, the algorithm should meet the following conditions during each genetic evolution in order to ensure that the network is not congested and the selected path starts from the source node s and ends at the destination node d:

(1) The selected network link should avoid network congestion as much as possible, namely, Utilization $(s, d) < 1$

(2) At least one optional link $L(s, d)$ exists bandwidth$_l$ $< B, l \in L(s, d)$

4.6. Algorithm Flow. According to the above model construction and multiobjective optimization algorithm selection, the algorithm process of the NSGA-II-based multiobjective network resource optimization allocation method is as follows:

(1) Initialize the random network and scale-free network populations, and set the population size and the maximum number of evolutions

(2) Preprocess the initial network topology map to filter out network links that do not meet the bandwidth requirements

(3) Starting from the second generation, merge the parent and offspring to form a large population, perform quick nondominated sorting on it, and then, calculate the crowding degree for each nondominated level individual, and finally according to the nondominated relationship and crowding degree to select suitable individuals to form a new parent population

(4) Decode the population individuals to get the true network path

(5) Perform selection, crossover, and mutation operations to produce new offspring populations

(6) If the genetic algebra exceeds the set value, go to Step 7; otherwise, go to Step 3

(7) Output the optimal solution, and choose according to the specific network business requirements

5. Simulation Analysis

In this section, we evaluate the optimal allocation method of power multimodal network resources based on NSGA-II and compare it with other algorithms.

5.1. Results of Resource Allocation. In the power multimodal network, the network nodes and the links are very complex and difficult to abstract. Therefore, this paper uses 20 network nodes as examples to test, and the link bandwidth, delay, reliability, and other data are randomly generated within a reasonable range. The specific parameters in the experiment are set as follows: the number of network nodes is 20, the initial population size is 200, the genetic algebra is 100, the crossover probability XOVR = 0.8, the mutation probability $p_m = 0.1$, the crossover probability $p_c = 0.9$, the source node is 1, and the purpose node is 20. Using the above resource optimization method to carry out simulation

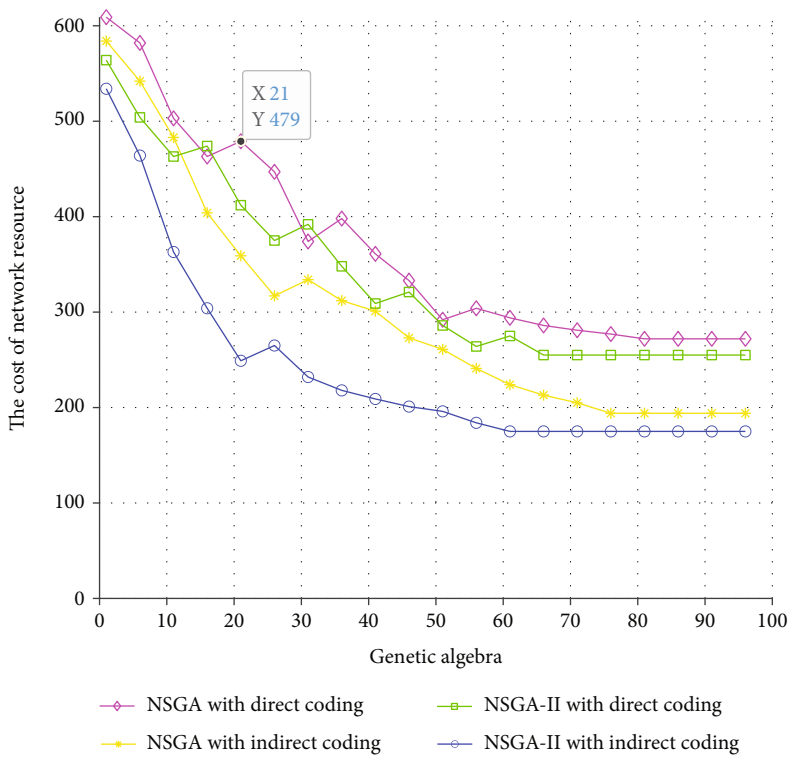


Figure 3: Changes in the resource expenditure of each experimental group in the genetic process.

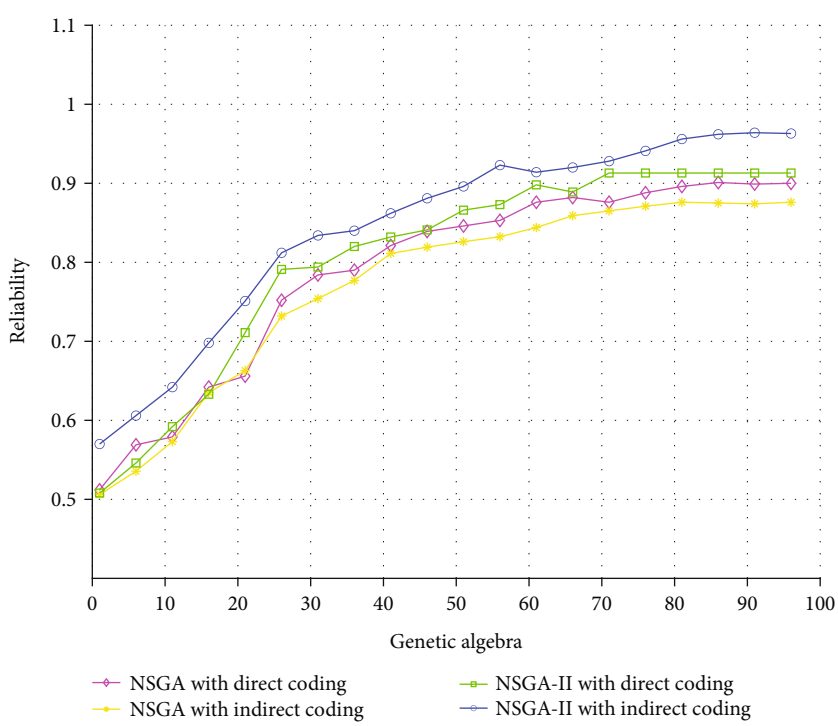


Figure 4: Changes in the link reliability of each experimental group in the genetic process.

experiments, the multiobjective network resource optimization result is shown in Figure 2.

Because the algorithm preprocesses remove the network links that do not meet the bandwidth requirements and simplify the calculation scale of the genetic algorithm before running, the entire process can be completed faster than usual. From Figure 2, it can be seen that the resource cost is continuously reduced to a stable state during the genetic process, indicating that the genetic algorithm can stably optimize the population, obtain the optimal value, and

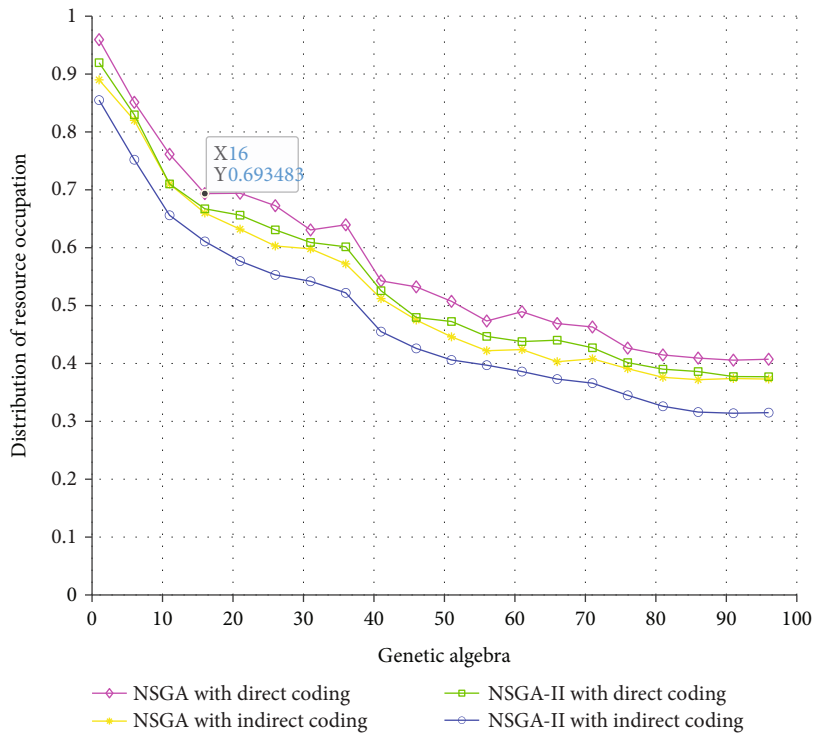


Figure 5: Variation trend of the distribution of resources occupied by each experimental group in the genetic process.

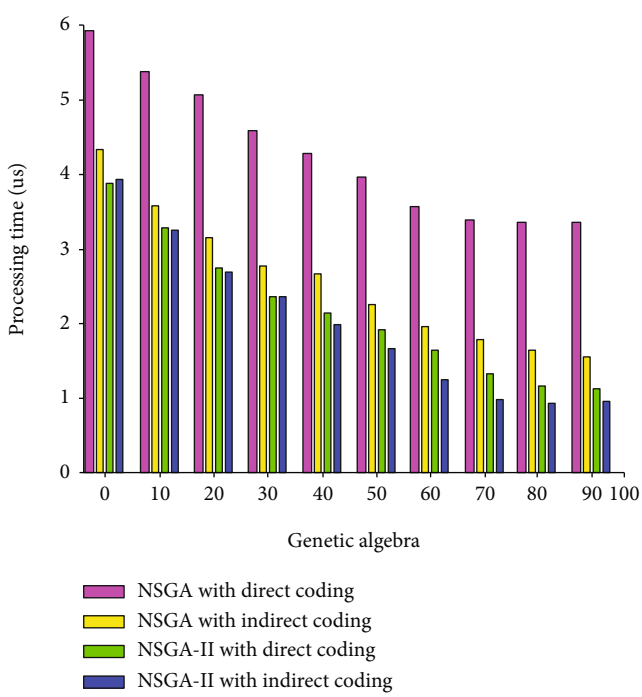


Figure 6: Convergence comparison between this genetic algorithm and other genetic algorithms.

quickly jump out of the local optimal solution to reach the state of the global optimal solution during the entire process.

Because the mutation probability in this experiment is $p_m \in (0, 1)$, $p_c \in (0, 1)$, and the genetic algorithm that retains the current optimal solution before selection can converge to the global optimal solution, it can be seen that the algorithm can converge to the global optimal solution [22].

5.2. Algorithm Performance. The performance of the algorithm is analyzed in two aspects. The first part is the change trend of various optimization indicators of the algorithm and other algorithms in the genetic process, and the second part is the convergence comparison between the algorithm and other algorithms.

In order to verify the effectiveness and convergence of the algorithm, we selected the classic multiobjective genetic algorithm NSGA and NSGA-II genetic algorithm to compare with the method in this paper. In order to verify the effectiveness of the method, we selected four experimental groups, arranged as follows. In the case of the same model, experimental group 1 adopts NSGA with direct encoding, experimental group 2 adopts NSGA with indirect encoding, experimental group 3 adopts NSGA-II with direct encoding, and experimental group 4 adopts the method of this article.

In Figure 3, the optimization results of the above four experimental groups for resource overhead are compared. It can be seen that disconnected paths are often generated in the cross-mutation process of the population in experimental group 1 and experimental group 3 with direct coding, which leads to negative optimization and unstable optimization results. However, experimental group 2 and experimental group 4 occasionally have negative optimization, but they will quickly stabilize and produce final results.

Figure 4 shows the change trend of the link reliability of each experimental group with the genetic algebra. It can be seen that the reliability of all experimental groups is gradually increasing. However, experimental group 1 and

experimental group 3 showed a fluctuating upward trend, and the optimization results were not stable enough. Both experimental group 2 and experimental group 4 grow steadily, but experimental group 2 tends to fall into the state of local optimal solution due to the limitation of the algorithm, so on the whole, experimental group 4 can reach the optimal solution quickly and steadily.

In Figure 5, the change trend of the resource occupation distribution of each experimental group is compared. It can be seen that all the experimental groups showed a downward trend as a whole. But with the increase of genetic algebra, the optimization value slows down. This is because with the increase of the genetic algebra, the population individuals gradually tend to the optimal solution, and each time the genetic change is not large, the overall optimization slowdown is slowed down. It can be seen from the figure that the NSGA-II-based power multimodal network resource optimization allocation method proposed in this paper reduces the distribution of resource occupation to the greatest extent and has a better optimization effect than other experimental groups.

Figure 6 is a comparison of the convergence between different algorithms. Although the algorithm has a longer running time due to preprocessing at the beginning, the network chain that does not meet the bandwidth requirements is removed in the process, which simplifies the calculation scale of the subsequent genetic algorithm. So the entire process can be completed faster than others, which further shows that the algorithm has high convergence and good reliability.

6. Conclusion

Unstable power business services are caused by the complex and changeable power multimodal network, which is difficult to meet the QoS requirements of different businesses. In order to ensure the normal operation of power business and improve resource utilization, an optimal allocation method of power multimodal network resources based on NSGA-II was proposed by this paper. Firstly, the power multimodal network-resource model is established, and the problems existing in the resource optimization process are analyzed. Secondly, preprocessing technology and indirect coding technology are applied to NSGA-II, which solves the coding problem and convergence problem of the application of genetic algorithm to the optimization of network resource allocation. Finally, the simulation results show that, compared with the control algorithm, this method has further optimized the various indicators of the resource allocation of the power multimodal network, and the performance has been improved by more than 6%.

However, the addition of preprocessing makes the algorithm often take a lot of time in the early stage of operation. In addition, because the model pays more attention to the fast response in time, it takes up a lot of memory space during the operation. This is the direction that needs attention and optimization in future work. We will consider improving the preprocessing algorithm to reduce the initial running time of the algorithm, and by optimizing the genetic method of NSGA-II, we will improve the power multimodal network resource optimization allocation method to reduce the memory space occupation during the algorithm operation.

Data Availability

The result data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by State Grid Corporation of China science and technology project "Key technology and application of new multimode intelligent network for State Grid" (5700-202024176A-0-0-00).

References

[1] H. U. Yuxiang, Y. I. Peng, S. U. N. Penghao, and W. U. Jiangxing, "Research on the full-dimensional defined polymorphic smart network," *Journal on Communications*, vol. 40, no. 8, pp. 1–12, 2019.

[2] M. Cremene, M. Suciu, D. Pallez, and D. Dumitrescu, "Comparative analysis of multi-objective evolutionary algorithms for QoS-aware web service composition," *Applied Soft Computing*, vol. 39, pp. 124–139, 2016.

[3] K. Deb, *Multi-Objective Optimization Using Evolutionary Algorithms*, John Wiley & Sons, 2001.

[4] A. Konak, D. W. Coit, and A. E. Smith, "Multi-objective optimization using genetic algorithms: a tutorial," *Reliability Engineering & System Safety*, vol. 91, no. 9, pp. 992–1007, 2006.

[5] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," *IEEE Transactions on Evolutionary Computation*, vol. 6, no. 2, pp. 182–197, 2002.

[6] C. U. I. Limin, S. U. N. Jingyue, L. I. Shanjun, and S. O. N. G. Guanglei, "An algorithm for business resource uniform distribution in power communication network based on entropy," *Power System Technology*, vol. 9, pp. 3066–3073, 2017.

[7] O. S. Yeremenko, O. V. Lemeshko, O. S. Nevzorova, and A. M. Hailan, "Method of hierarchical QoS routing based on network resource reservation," in *2017 IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON)*, pp. 971–976, IEEE, Kyiv, UKraine, 2017.

[8] T. Chen, Q. Ling, and G. B. Giannakis, "An online convex optimization approach to proactive network resource allocation," *IEEE Transactions on Signal Processing*, vol. 65, no. 24, pp. 6350–6364, 2017.

[9] Y. T. Leong, J. Y. Lee, R. R. Tan, J. J. Foo, and I. M. L. Chew, "Multi-objective optimization for resource network synthesis in eco-industrial parks using an integrated analytic hierarchy process," *Journal of Cleaner Production*, vol. 143, pp. 1268–1283, 2017.

[10] N. Gunantara, "A review of multi-objective optimization: methods and its applications," *Cogent Engineering*, vol. 5, no. 1, p. 1502242, 2018.

[11] D. Dasgupta, *Artificial immune systems and their applications*, Springer Science & Business Media, 2012.

[12] H. Garg, "A hybrid PSO-GA algorithm for constrained optimization problems," *Applied Mathematics and Computation*, vol. 274, pp. 292–305, 2016.

[13] Y. Gao, H. Guan, Z. Qi, Y. Hou, and L. Liu, "A multi-objective ant colony system algorithm for virtual machine placement in cloud computing," *Journal of Computer and System Sciences*, vol. 79, no. 8, pp. 1230–1242, 2013.

[14] T. Murata and H. Ishibuchi, "MOGA: multi-objective genetic algorithms," *IEEE international conference on evolutionary computation*, vol. 1, pp. 289–294, 1995.

[15] T. George and T. Amudha, "Genetic algorithm based multi-objective optimization framework to solve traveling salesman problem," in *Advances in Computing and Intelligent Systems*, pp. 141–151, Springer, Singapore, 2020.

[16] B. Singh, V. Mukherjee, and P. Tiwari, "GA-based multi-objective optimization for distributed generations planning with DLMs in distribution power systems," *Journal of Electrical Systems and Information Technology*, vol. 4, no. 1, pp. 62–94, 2017.

[17] S. A. Bagherzadeh, M. T. Sulgani, V. Nikkhah, M. Bahrami, A. Karimipour, and Y. Jiang, "Minimize pressure drop and maximize heat transfer coefficient by the new proposed multi-objective optimization/statistical model composed of "ANN + genetic algorithm" based on empirical data of CuO/-paraffin nanofluid in a pipe," *Physica A: Statistical Mechanics and its Applications*, vol. 527, p. 121056, 2019.

[18] B. Zhao, Y. Xue, B. Xu, T. Ma, and J. Liu, "Multi-objective classification based on NSGA-II," *International Journal of Computing Science and Mathematics*, vol. 9, no. 6, pp. 539–546, 2018.

[19] B. Tan, H. Ma, and Y. Mei, "A NSGA-II-based approach for service resource allocation in cloud," in *2017 IEEE Congress on Evolutionary Computation (CEC)*, pp. 2574–2581, IEEE, Donostia, Spain, 2017.

[20] A. S. MSofia and K. P. Ganesh, "Multi-objective task scheduling to minimize energy consumption and makespan of cloud computing using NSGA-II," *Journal of Network and Systems Management*, vol. 26, no. 2, pp. 463–485, 2018.

[21] L. Lin and M. Gen, "Priority-based genetic algorithm for shortest path routing problem in OSPF," in *Intelligent and Evolutionary Systems*, pp. 91–103, Springer, Berlin, Heidelberg, 2009.

[22] R. H. Hwang, W. Y. Do, and S. C. Yang, "Multicast routing based on genetic algorithms," *Journal of Information Science and Engineering*, vol. 16, no. 6, pp. 885–901, 2000.

Hindawi
Wireless Communications and Mobile Computing
Volume 2021, Article ID 5991767, 12 pages
https://doi.org/10.1155/2021/5991767

Research Article

Radio Station Background Noise Detection Based on Time-Frequency Domain Electromagnetic Spectrum

Fuzhai Wang,[1] Zhenjia Chen,[2] Xuanfeng Chen,[2] and Ting Chen[2]

[1]*Transport Planning and Research Institute of MOT, Beijing, China*
[2]*Hainan University, Haikou 570228, China*

Correspondence should be addressed to Zhenjia Chen; zjchen@hainanu.edu.cn

Received 5 September 2021; Revised 24 September 2021; Accepted 28 September 2021; Published 15 October 2021

Academic Editor: Chi-Hua Chen

The electromagnetic spectrum resource is one of the important national resources. It is a physical channel for wireless communication between ships and between ships and radio stations. Good communication quality must be guaranteed, so it is urgent to monitor and analyze the environmental background noise of the electromagnetic spectrum. The estimation of the radio frequency signal coverage in the target area during the monitoring process is of great significance to the study of electromagnetic spectrum resource management and control. This paper estimates the upper envelope and lower envelope of the background noise of the target frequency band based on the electromagnetic spectrum data in the time-frequency domain and combines the forward difference algorithm to estimate the background noise envelope curve. We set up fixed detection nodes and mobile detection nodes for specific construction areas and collect time-frequency spectrum data of electromagnetic spectrum in multiple locations. The instantaneous frequency spectrum and the collected data of a specific frequency point are compared, and it is difficult to judge whether there is a valid signal. This paper is based on the time-frequency domain electromagnetic spectrum data in the construction area of the project and estimates the background noise of the coast station frequency band in the current environment. It is based on the energy gradient estimation of the time-frequency domain spectrum, and the effective signal of the target frequency band is obtained and combines the noise envelope and the effective signal location to improve the estimation result of the background noise envelope. The experimental results show that the background noise estimation algorithm can reflect the changes in the noise floor of different target frequency bands.

1. Introduction

The electromagnetic spectrum resource is a nonrenewable national resource. Reasonable management and control of spectrum resources is the basic guarantee for wireless communication services and security. The safety management of maritime and ship traffic is particularly important, and maritime communication services are inseparable from the support of electromagnetic spectrum resources. In order to ensure the safety of ship traffic and meet the ever-increasing demand for maritime communications, maritime radio plays a very critical role. The wireless walkie-talkies that carry out wireless communication between the ship on the sea and the shore are called marine wireless walkie-talkies. It is also called a coast station. The background noise detection of the radio station is to ensure the communication quality between the receiving station and the sending station. We need to detect the background noise in the area near the receiving station to find the interference source of the radio channel. We collect spectrum data on the coast and estimate the environmental background noise. The estimated results play a vital role in studying how to improve the communication quality in the sea and on the shore and can effectively improve the safety of maritime navigation.

2. Related

Coast station refers to a land station dedicated to ship transportation and shipping management services. Its main function is to timely detect the distress alarm from the ship and quickly transfer the alarm to the relevant search and rescue coordination center. It launches a shore-to-ship distress

alert, and the search and rescue coordination center or search and rescue department can quickly obtain the alert information. It uses the radio interface between the land public communication network or the dedicated communication network and the ship station to play a switching role in the search and rescue coordination communication. Commonly used electromagnetic spectrum sensing methods include matched filter detection, cyclostationary feature detection, and energy detection (ED). Compared with the other two methods, the ED method has higher real-time performance and reliability. Due to its simplicity and applicability and its low calculation and implementation costs, ED constitutes the preferred method of spectrum sensing in cognitive radio [1]. In order to improve the detection probability, it uses a priori information related to the spectrum usage of the main user is used to enhance the spectrum sensing performance. [2] proposes a threshold setting method based on approximate analysis to achieve target detection probability or false alarm probability. [3] studied energy detectors based on dual thresholds for cooperative spectrum sensing mechanisms in cognitive radio networks. The setting method of the noise energy threshold determines the detection probability of the effective signal. [4] studied energy detectors based on dual thresholds for cooperative spectrum sensing mechanisms in cognitive radio networks. The author improves the detection probability of spectrum sensing by increasing the spatial dimension of spectrum detection. We propose a combined collaborative spectrum detection method to reduce the background noise of the overall detection system [5]. In large-scale cognitive radio networks, secondary users cannot share the public spectrum due to the limited coverage of primary users. [6] proposed a diffusion adaptive learning algorithm based on the correlation entropy cooperation strategy to realize public spectrum estimation.

In addition to energy detection methods, many scholars are also studying other spectrum sensing methods. [7] proposed a spectrum sensing scheme based on adaptive dual thresholds. The author compares it with the spectrum sensing technology based on the cyclostationary feature detector. At 8 dB signal-to-noise ratio (SNR), the detection probability of MED is 36.1% higher than the cyclostationary feature detection method. [8] have proposed that duty cycle (DC) and channel occupancy (COR) measurements are used to characterize the availability of white space in cognitive radio systems (CRS). [9] studied the joint impact of RF impairments on the spectrum sensing of CR systems based on energy detection in a multichannel environment. In particular, assuming Rayleigh fading, the author provides novel closed-form expressions to evaluate the probability of detection and false alarms. [10] proposed an improved channel occupancy estimation (iCOR) method that uses a high false alarm probability to increase sensitivity without the overestimation usually associated with high false alarm probability. The electromagnetic spectrum analysis method based on the time-frequency domain can effectively improve the estimation of background noise. [11] proposed a random multipath model for receiving signals, in which both the transmitter and receiver have directional antennas and are located in the same rectangular room. In indoor positioning applications, it is necessary to eliminate the electromagnetic interference of household appliances, and the background noise estimation is more important. Some scholars considered that the existing spectrum sensing algorithms are difficult to apply to practical applications and lead to waste of channels and energy resources, and [12] proposed an energy-saving spectrum-sensing algorithm. This algorithm can maximize the energy utilization efficiency under the premise of detecting sufficient channel available time, which is of great significance for reducing the cost of network deployment.

In order to provide meaningful data on spectrum usage, the occupancy measurement describing the utilization of a specific frequency band should be performed in a specific area rather than a single location. [13] described a comprehensive method for measuring and analyzing spectrum occupancy. [14] proposed a spectrum scanning method based on Bayesian inference to estimate the channel occupancy. This method takes into account the false detection probability and detection probability of the spectrum sensor to make the estimation of channel occupancy more accurate. [8] analyzed the problem of estimating the main channel COR based on the spectrum sensing decision and derived a tight closed-form expression for the required sensing sample size. However, different background noise distributions will be generated in different environments, which will directly affect the accuracy of channel estimation. One of the important judgment conditions for spectrum occupancy is noise threshold estimation, also called background noise estimation. Electromagnetic spectrum database as an important means to describe the electromagnetic environment. [15] optimized the spectrum sensing based on the IEEE 1900.6 standard to support the spectrum database. Especially in a complex environment, background noise estimation will directly affect the reliability of the electromagnetic spectrum database. The accurate estimation of background noise is one of the important indicators to improve the accuracy of spectrum sensing. [16] combined time-frequency domain electromagnetic spectrum data to solve the problem of accurately modeling the spectrum occupancy patterns of real radio communication systems, which is an important aspect of cognitive radio (CR) network research. [17] studied the spatiotemporal opportunity detection problem of spectrum heterogeneous cognitive radio network. The author proposes that at a given time, secondary users (SU) in different locations may encounter different spectrum access opportunities. The estimation of the background noise envelope of the time-frequency domain spectrum will directly affect the occupancy of the target frequency band determined by the secondary user.

3. Worked

We focus on electromagnetic spectrum detection methods based on time-frequency domain energy detection data. The method is simple and applicable, and it has higher real-time performance and reliability. We propose a target frequency band background noise estimation based on the time-frequency domain electromagnetic spectrum detection method. It estimates the fluctuation range of background

noise according to the maximum hold method and the minimum hold method and combines the forward difference algorithm to estimate the background noise envelope curve. Finally, it combines the collected electromagnetic spectrum data to analyze and estimate the background noise of the target frequency band.

4. Electromagnetic Spectrum Detection Based on Time-Frequency Domain

Coastal stations provide ships with daily public communication services, acting as a relay between users and ships, broadcasting navigation warnings, weather forecasts, and other maritime emergency and safety information for ships. Its operating frequency is assigned by the International Telecommunication Union (ITU), and the identification of coast stations consists of 9 decimal digits starting with 00. The spectrum distribution consists of the energy of each frequency point, it is affected by free space propagation loss and special geographic environment, and the detected instantaneous electromagnetic spectrum data will fluctuate to a certain extent. It is necessary to preprocess the electromagnetic spectrum data in the time-frequency domain to accurately obtain the background noise of the target frequency band.

4.1. Energy Detection and Preprocessing. The collected radio frequency data undergoes fast Fourier transform (FFT) to obtain energy values at different frequency points. The energy detection method can be expressed as

$$x(t) = \begin{cases} n(t), & H_0, \\ h(t)s(t) + n(t), & H_1, \end{cases} \tag{1}$$

where H_0 is defined as noise, and H_1 is defined as effective signal. $n(t)$ is noise data, $s(t)$ is the signal data, and $h(t)$ is the channel parameter. t represents a certain frequency point. $x(t)$ represents the energy detection value at frequency t after the target frequency band passes through FFT.

(1) represents the result of an energy detection in the target frequency band. The effective signal and background noise can be distinguished by the threshold setting method. When processing the energy distribution data of the target frequency band, many researchers will use the horizontal line threshold method to distinguish between noise and effective signal. The selection of the threshold directly affects the probability of false detection (P_{fa}) and the probability of missed detection (P_m) of spectrum sensing. It is assumed that the set threshold is T. When T is slightly greater than $n(t)$, most of the noise can be accurately determined. However, $n(t)$ will fluctuate due to environmental factors, and there are differences in $n(t)$ in different frequency bands. As T increases, the probability of a weak signal being judged as noise increases. As T decreases, the probability of noise being judged as a valid signal increases (see Figure 1). Therefore, it sets an appropriate T to ensure lower P_{fa} and P_m, as shown in the shaded part in Figure 1. In this way, the detection probability of spectrum sensing can be improved.

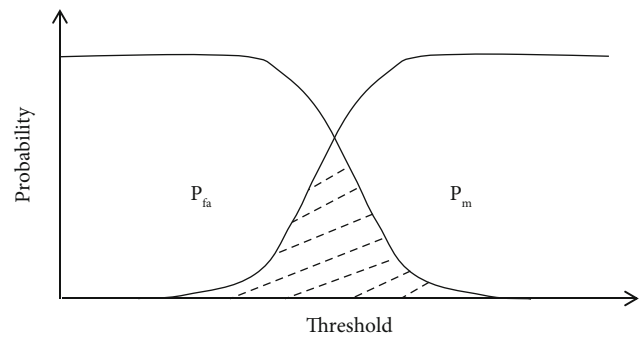


Figure 1: In the process of spectrum sensing, the relationship between P_{fa}, P_m, and the threshold setting.

In a complex environment, the fluctuation of the spectrum energy in the noise frequency band will be relatively large. This will result in a higher P_{fa} in spectrum detection. In wireless communication systems, noise can be described by a normal distribution. Therefore, the fluctuation of noise can be smoothed by averaging the data of multiple frames of spectrum detection. It assumes that n frames of time-frequency domain spectrum data are collected. It can be expressed as

$$x_{aver}(t) = \frac{1}{n}\sum_{i=1}^{n} x(t), \tag{2}$$

where i defines the spectrum data collected in the time-frequency domain spectrum of the i-th frame.

In order to calculate the background noise envelope of the target frequency band in the current environment, we propose to estimate the lower envelope of the background noise with the minimum hold method of the electromagnetic spectrum in the time-frequency domain and estimate the upper envelope of the background noise with the maximum hold method of the electromagnetic spectrum in the time-frequency domain. The upper envelope estimate of background noise can be expressed as

$$x_{\max}(t) = \max\,(x(t)),\ t = 1, \cdots, n. \tag{3}$$

The lower envelope estimate of the background noise can be expressed as

$$x_{\min}(t) = \min\,(x(t)),\ t = 1, \cdots, n. \tag{4}$$

If the signal existing in the target frequency band is a discontinuous signal, the minimum hold method is used to estimate the noise envelope. The maximum preservation method estimates the maximum threshold of noise fluctuations, which can avoid setting an excessively high noise threshold resulting in a higher P_m.

4.2. Background Noise Estimation. After it obtains the upper envelope and lower envelope of the background noise, the effective signal needs to be filtered out. The forward difference algorithm to estimate the adaptive threshold curve

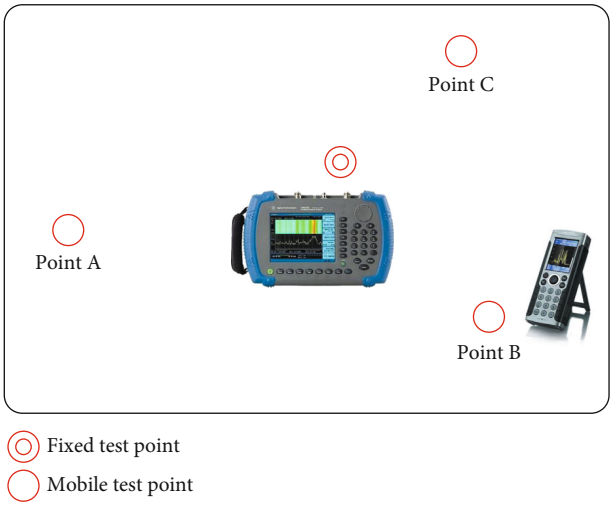


Figure 2: Schematic diagram of spectrum detection equipment and signal source location.

(ATC) is used to estimate the noise envelope [5]. The signal of the envelope is effectively filtered through the adjustment variable D. According to (2), calculating the average value of N frames of data is smoother than the instantaneous frequency spectrum. In the instantaneous frequency spectrum, the effective signal can be seen as a pulse waveform. There are two slopes with larger absolute values before and after the center frequency. It assumes that the difference between adjacent spectral data is defined as D_v, the forward calculation of the spectral difference distribution between adjacent frequency points. It can be expressed as

$$|\text{data}_\text{i} - \text{data}_{i-1}|_N = D_v. \tag{5}$$

It sets the difference threshold of the frequency spectrum, and after comparing and calculating the difference between adjacent frequency spectrums, it judges the front and back edge pulses of the signal. The spectral energy of the signal is filtered out, and the output is the noise curve. It can be expressed as

$$n_i = \begin{cases} \text{data}_i, & D_v \le D, \\ \text{data}_{i-1}, & D_v > D, \end{cases} \tag{6}$$

where the frequency point is defined as i, the difference threshold is defined as D, and the output corresponding to each frequency point in the noise curve is defined as n_i.

Instantaneous spectrum data is difficult to improve the detection efficiency of effective signals. In order to effectively estimate the background noise, we propose a recognition and positioning method combined with effective signals to improve the detection probability of the spectrum. According to the time-frequency domain spectrum data, the gradient of the adjacent spectrum is calculated, and the rising edge and the falling edge boundary of the effective signal are obtained. It assumes that r is the RSSI at the frequency f and time t, and the time-frequency domain energy matrix can be obtained. It can be expressed as

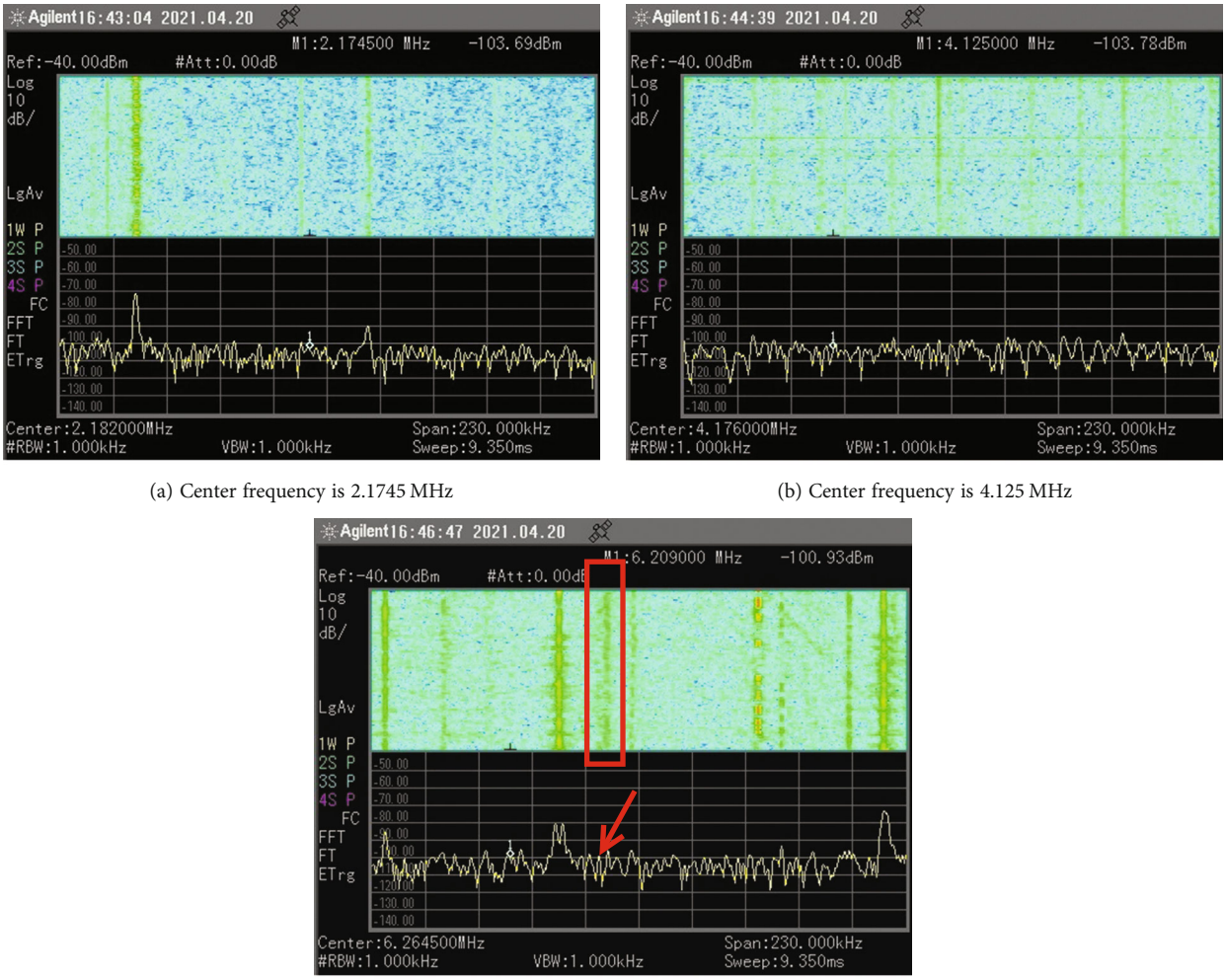
Table 1: Measurement data of mobile monitoring points.

Frequency (kHz)	Point A (dBm)	Point B (dBm)	Point C (dBm)
2174.5	-83.0	-106.6	-103.6
2182	-102.3	-101.2	-104.3
2187.5	-96.4	-102.5	-103.2
4125	-94.3	-103.4	-106.1
4174	-103.5	-106.4	-106.4
4177.5	-104.6	-104.9	-105.5
4207.5	-105.2	-103.1	-102.8
6209	-86.1	-106.3	-106.0
6215	-98.0	-104.2	-105.7
6264.5	-99.3	-106.7	-106.3
6268	-104.8	-105.4	-107.0
6312	-103.2	-107.0	-105.7
8258	-100.6	-106.9	-103.5
8291	-106.1	-108.9	-107.0
8376.5	-102.8	-104.0	-107.0
8395	-106.2	-106.9	-107.0
8414.5	-105.1	-106.2	-107.0
12260	-101.7	-105.2	-107.0
12290	-100.2	-108.4	-107.0
12302	-98.3	-103.1	-107.0
12335	-101.4	-105.6	-107.0
12510.5	-104.2	-105.6	-107.0
12520	-104.9	-106.6	-107.0
12546.5	-107.0	-105.9	-107.0
12577	-107.0	-106.3	-107.0
16420	-107.0	-105.2	-107.0
16516	-107.0	-105.6	-107.0
16695	-103.5	-106.8	-107.0
16762	-102.1	-105.0	-107.0
16804.5	-104.4	-107.0	-107.0
18795	-107.0	-107.0	-107.0
18819	-107.0	-107.0	-107.0
22039	-107.0	-107.0	-107.0

$$N = \begin{bmatrix} N_{0,0} & \cdots & N_{0,f-1} \\ \vdots & \ddots & \vdots \\ N_{t-1,0} & \cdots & N_{t-1,f-1} \end{bmatrix}. \tag{7}$$

There are generally two methods for gradient calculation, including horizontal gradient and vertical gradient. It can be expressed as

$$G_{f,t} = \begin{cases} N_{i+1,j} - N_{i,j} \\ N_{i,j+1} - N_{i,j} \end{cases}, i = 0, \cdots, t\, j = 0, \cdots, f. \tag{8}$$



(a) Center frequency is 2.1745 MHz

(b) Center frequency is 4.125 MHz

(c) Center frequency is 6.209 MHz

Figure 3: 2 MHz-6 MHz frequency band time-frequency domain spectrum (bandwidth is 230 kHz).

As time changes, the frequency spectrum of the signal is mainly distributed on a certain frequency point. Therefore, we use the transverse gradient to get the energy gradient matrix. It can be expressed as

$$G_{f,t} = \begin{bmatrix} N_{0,1} - N_{0,0} & \cdots & N_{0,t-1} - N_{0,t-2} & 0 \\ \vdots & \ddots & \vdots & \vdots \\ N_{t-1,1} - N_{t-1,0} & \cdots & N_{t-1,f-1} - N_{t-1,f-2} & 0 \end{bmatrix}. \tag{9}$$

Finally, it judges whether it is the upper edge or the lower edge of the valid signal according to the magnitude of the gradient value. The effective signal of the target frequency band can be marked, and the background noise area can be obtained at the same time.

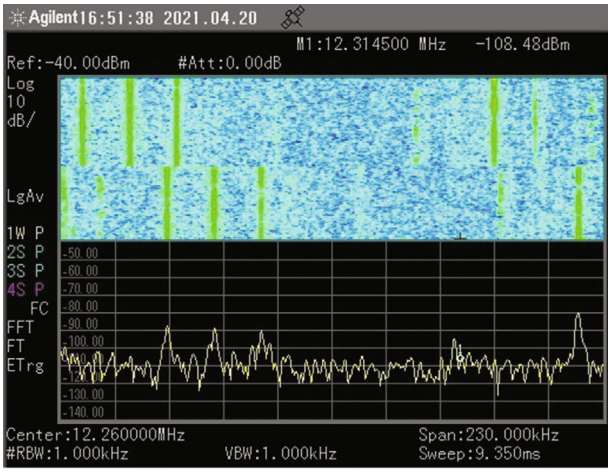
5. Detection and Analysis

The coast station has 34 operating frequency points, which are distributed between the 2 MHz and 23 MHz frequency band. The test equipment are Agilent N9342C RF analyzer, portable handheld field strength meter, Beidou positioning terminal, notebook computer, and other related equipment. The conditions of the test site are as follows:

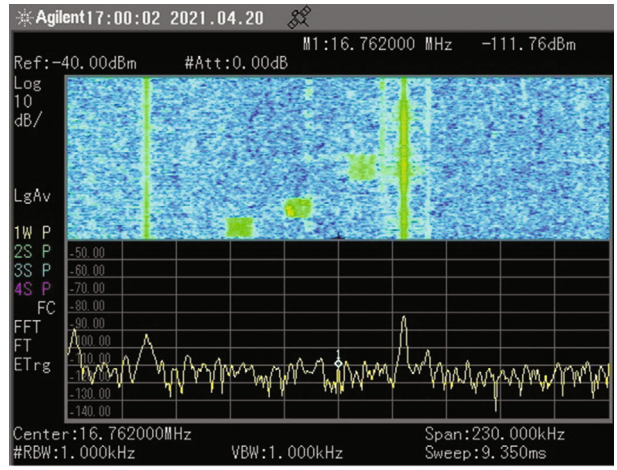
(1) In the field of 1 antennas, set up a fixed test point and continue testing

(2) In the construction project area, set up a total of 3 mobile test points A/B/C, mainly to test the background noise in the temporary antenna area to be built (point A/B) and the short-distance construction area (point C). Point A/B/C is the planned test location, and a suitable location nearby can be selected for testing according to the actual situation of the site

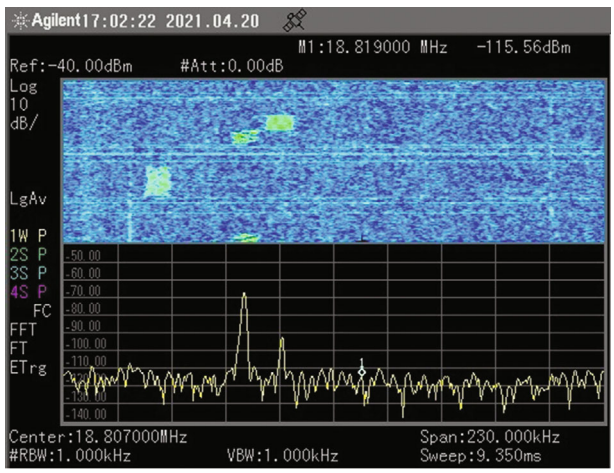
(3) It is finally determined according to the actual situation, and the fixed test point is set in the radio room (because it needs to be connected to the radio's short-wave antenna). There are 4 mobile test points (including inside the station). The test points are the top of the radio



(a) Center frequency is 12.26 MHz



(b) Center frequency is 16.762 MHz



(c) Center frequency is 18.819 MHz

Figure 4: 12 MHz-18 MHz frequency band time-frequency domain spectrum (bandwidth is 230 kHz).

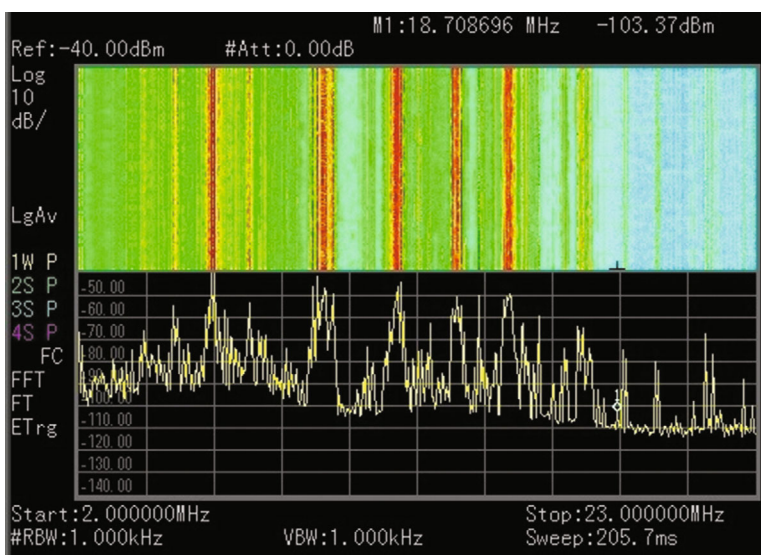
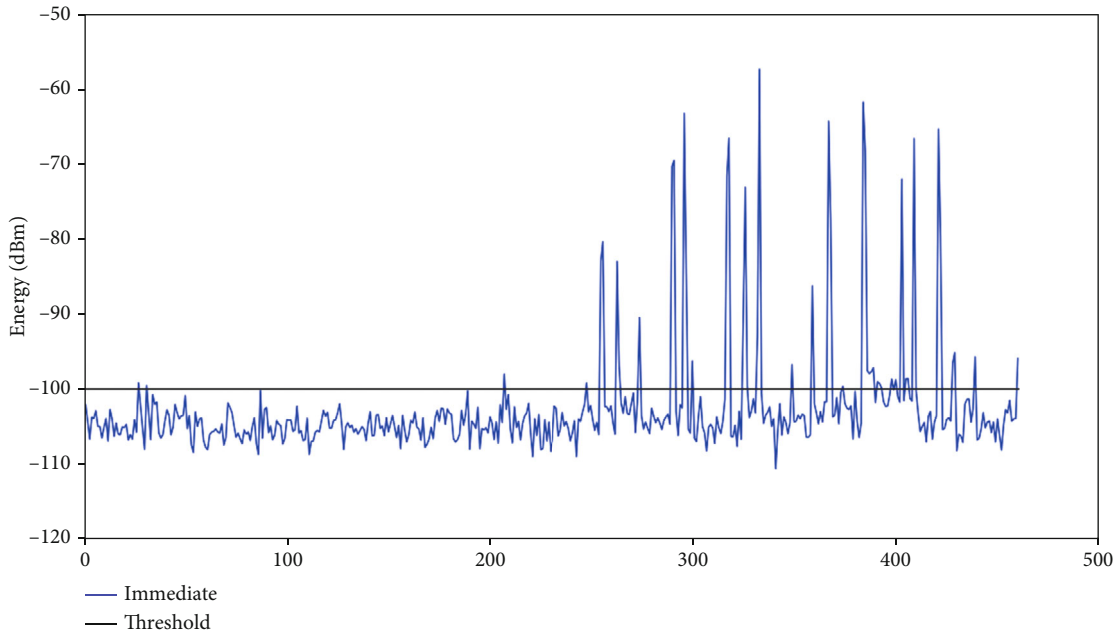
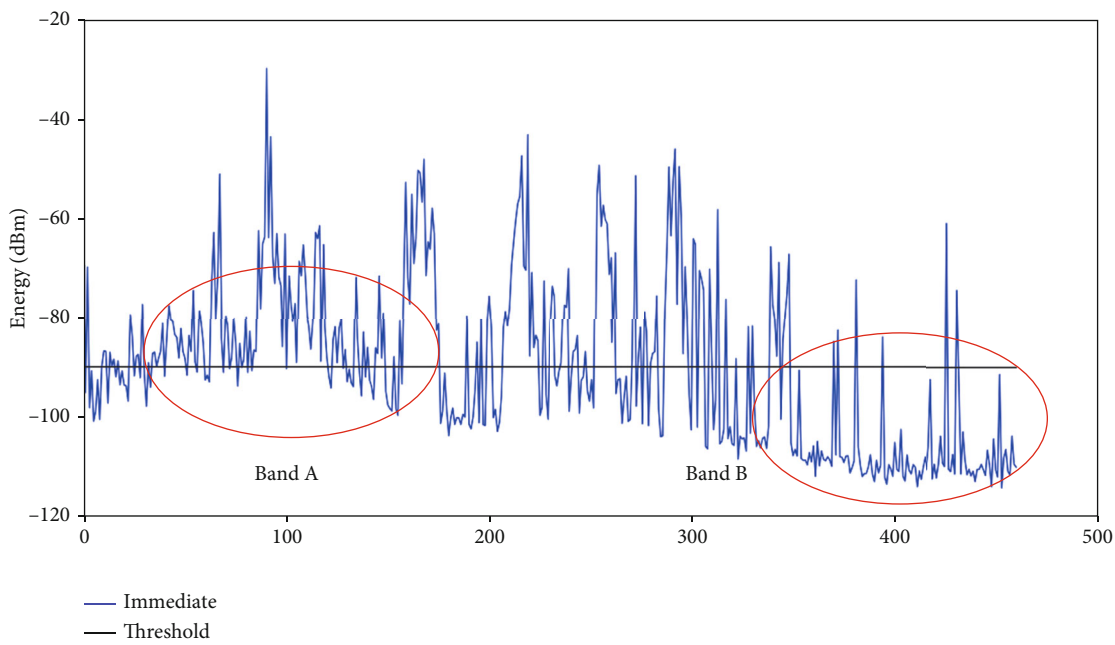


Figure 5: The time-frequency domain spectrum data of the 2 MHz-23 MHz frequency band acquired by the Agilent N9342C RF analyzer.



(a) 60 MHz-110 MHz



(b) 2 MHz-23 MHz

Figure 6: Spectrum data of different frequency bands.

(4) Computer room, point A on the southwest side of the construction project, point B on the southeast side of the construction project, and point C on the northeast side of the construction project

At the mobile monitoring points, a handheld RF field strength meter is used for testing. Every four hours, a portable field strength meter is used to test various mobile monitoring points. The mobile measurement records the signal strength of each frequency point. The influence of background noise is determined according to different energy fluctuations of frequency points (see Figure 2). The test results are shown in Table 1. The signal strength in the 2 MHz-8 MHz frequency band is -90 dBm to -100 dBm, and the signal strength in the 12 MHz-22 MHz frequency band is -100 dBm to -110 dBm. It is based on the test data analysis of the other three moving points, and the signal strength of the entire test frequency band is -100 dBm to -110 dBm. The 4207.5 kHz frequency of the mobile monitoring point A is selected. After comparing the test data, it can

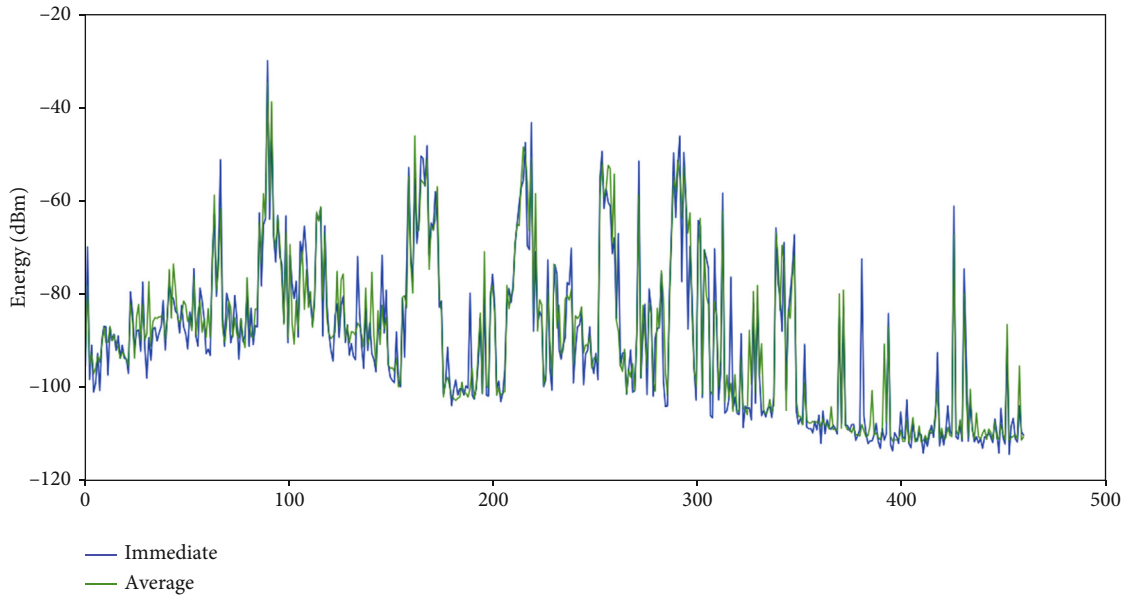


Figure 7: Instantaneous frequency spectrum and mean frequency spectrum.

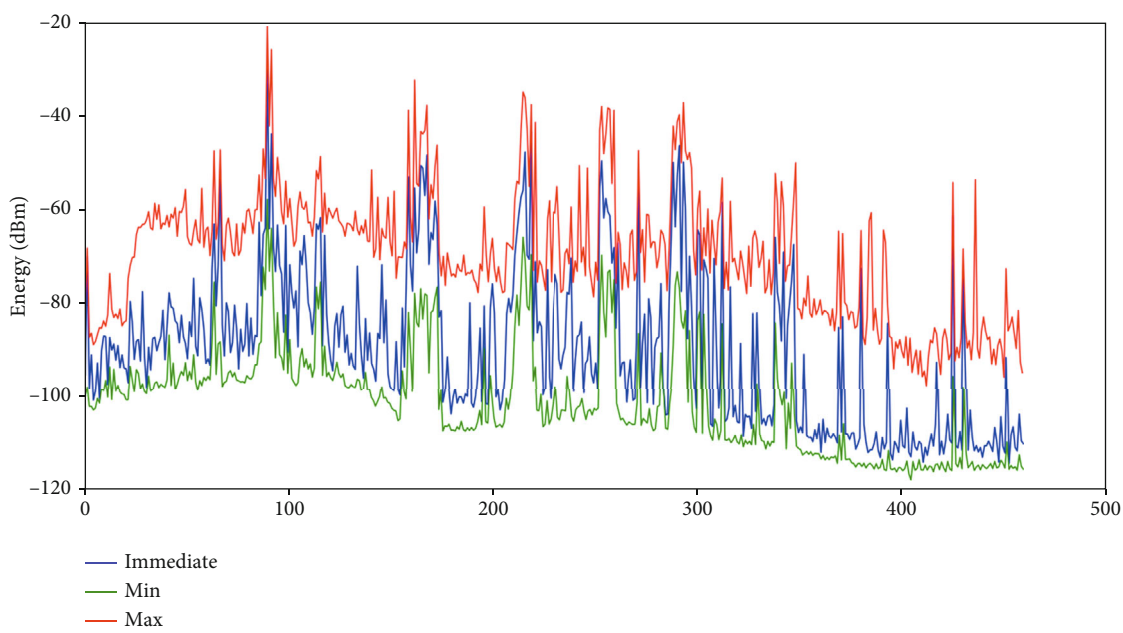


Figure 8: The upper envelope and lower envelope of the background noise.

be preliminarily determined that the background noise is about -95 dBm (2 MHz -8 MHz frequency band) and -105 dBm (12 MHz-22 MHz frequency band). We collected the time-frequency domain spectrum with center frequency of 2.1745 MHz, 4.125 MHz, 6.209 MHz, 12.26 MHz, 16.762 MHz, and 18.819 MHz, the sampling bandwidth is 230 kHz, and the sampling interval is 0.5 kHz, to ensure that it can cover all the signal frequency points in Table 1.

The background noise of the 2.1745 MHz, 4.125 MHz, and 6.209 MHz frequency bands is about -100 dBm (see Figure 3). The background noise envelope in the frequency band does not fluctuate much. The frequency point position of the effective signal can also be clearly determined from the time-frequency domain spectrum. The linear threshold can be used to quickly determine the noise and the effective signal. The background noise of the 12.26 MHz, 16.762 MHz, and 18.819 MHz frequency bands is about -105 dBm (see Figure 4). Compared to the 2 MHz-6 MHz frequency band, the overall background noise of this frequency band is reduced by 5 dB. If the same decision threshold T is used, the probability of missed detection of weak signals in this frequency band is higher, which will ultimately affect the

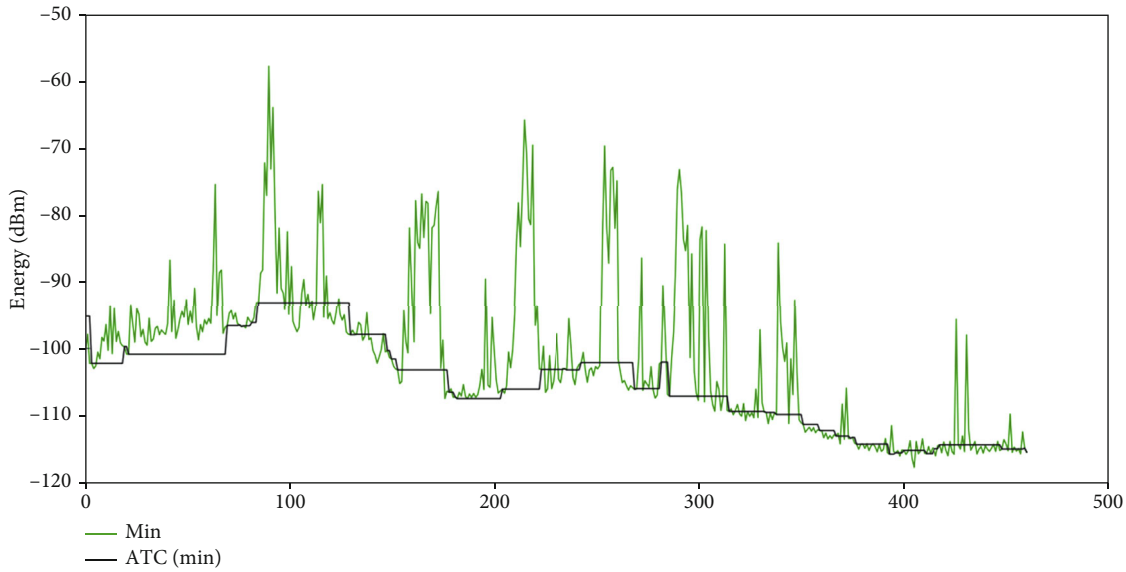


Figure 9: The upper envelope and lower envelope of the background noise.

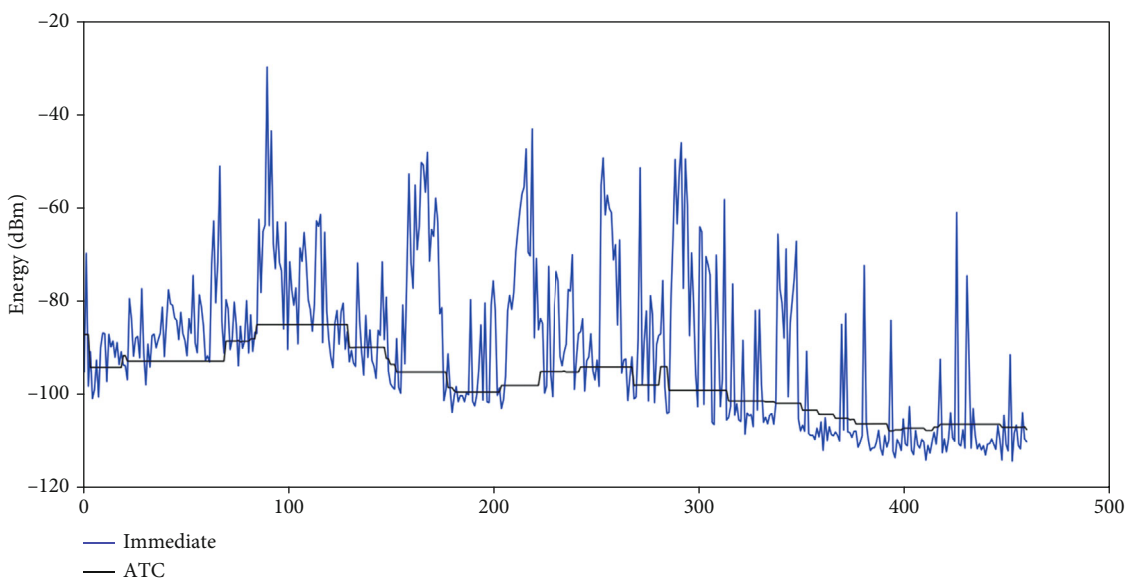
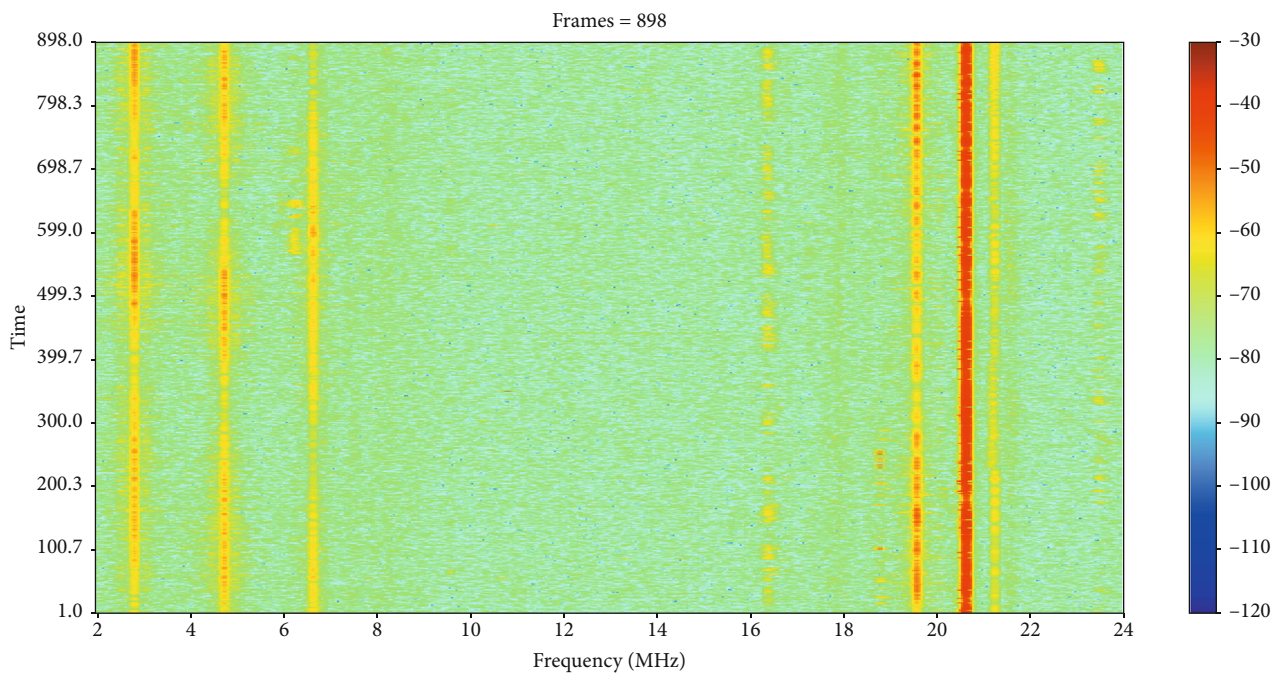


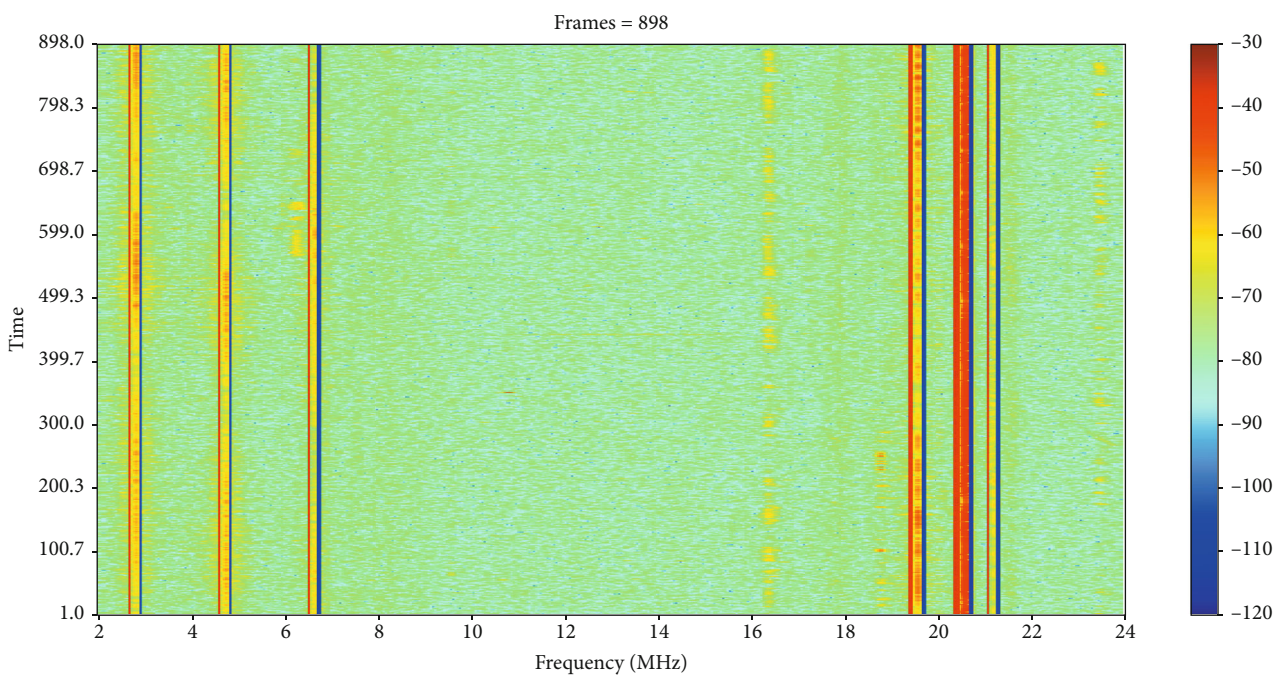
Figure 10: The upper envelope and lower envelope of the background noise.

overall detection probability. The time-frequency domain spectrum is also called waterfall chart. The horizontal axis represents frequency, the vertical axis represents time, and the color in the figure represents the received signal strength (received signal strength indication). The darker the color (red), the stronger the RSSI, and the lighter the color (blue), the weaker the RSSI. Comparing the RSSI in the 2 MHz-6 MHz frequency band and the 12 MHz-18 MHz frequency band, the overall background noise energy of the lower frequency band is higher. For low-frequency bands, the energy of background noise fluctuates over time, which will bury weak signals in the noise and increase the probability of missed detection. It is difficult to identify the effective signal through the instantaneous frequency spectrum, as shown in the position of the red arrow in Figure 3(c). However, through the distribution of the time-frequency domain spectrogram, it can be clearly found that there is a signal at the channel, as shown in the position of the red box in Figure 3(c). In the higher frequency band, the background noise energy is lower, so it is easier to judge the weak signal. However, a fixed energy threshold cannot be set, because it is not suitable for signal detection in the full frequency spectrum.

At fixed monitoring points, Agilent spectrum analyzers are used to collect electromagnetic spectrum data in the time-frequency domain. The parameters of the spectrum analyzer are set as follows: in order to cover all frequency points, the span SPAN is set to 230 kHz, the resolution bandwidth (RBW) is 1 kHz, the view bandwidth (VBW) is 1 kHz, the reference level is -40 dBm, the attenuation is 0 dB, and the sampling interval is 100 ms. The collected



(a) Time-frequency domain spectrum data



(b) Effective signal for energy gradient estimation

Figure 11: Time-frequency domain spectral energy gradient and effective signal estimation.

spectrum data is saved in a CSV file with a size of about 8 MB as 1500 frames of data. The time-frequency domain wide-band scanning range of the spectrum analyzer is 2 MHz-23 MHz, and the result is shown in Figure 5. In the lower frequency band and the middle frequency band, the signal reception strength is strong, and multiple signal pulses can be clearly observed. The signal surroundings are accompanied by higher noise energy, and the noise floor is about -80 dBm to -90 dBm. The noise floor in the higher frequency band is about -100 dBm to -110 dBm.

The instantaneous frequency spectrum of a certain frequency band of HF and VHF is shown in Figure 6. The background noise of the VHF spectrum is relatively flat compared to the HF spectrum. For the VHF spectrum, assuming $T = -100$ dBm, most of the noise can be judged, as shown in Figure 6(a). For the high-frequency spectrum,

suppose $T = -90$ dBm, as shown in Figure 6(b). Then, the noise of band A is judged to be a valid signal, and P_{fa} is higher. Weak signals in band B are all judged to be noise, and P_m is higher. The traditional linear threshold setting method is not suitable for wide-band spectrum scanning tasks due to its own characteristics. Moreover, the value of the threshold requires a large amount of a priori spectrum detection data to determine, and the actual application process is relatively tedious and time-consuming and laborious.

Because the mobile monitoring node has poor timeliness when measuring the instantaneous spectrum, it cannot accurately analyze the electromagnetic spectrum environment. Therefore, it is necessary to analyze the background noise of the target frequency band based on the electromagnetic spectrum data in the time-frequency domain. In order to reduce the influence of noise fluctuations, it performs average smoothing processing on multiframe detection data according to (2). The data after smoothing is shown in Figure 7.

It is necessary to reduce the energy fluctuation changes in the noise frequency band. Then calculate the upper envelope and lower envelope of the background noise according to (3) and (4) (see Figure 8). The envelope under background noise can basically fit the background noise curve of the instantaneous spectrum. The envelope on the background noise can be used as the maximum threshold for the background noise threshold setting. For different frequency bands, the background noise envelope estimation based on the time-frequency domain electromagnetic spectrum can dynamically estimate the noise floor.

According to (5) and (6), the effective signal of the lower envelope under the background noise is filtered out (see Figure 9). Among them, the parameter of D is set to 0.282 dBm.

At this time, it adjusts the background noise according to the maximum threshold of the upper envelope and finally gets the background noise envelope curve in the target frequency band. Among them, the background noise envelope is increased by 8 dB. The obtained background noise envelope curve can describe the noise floor of different frequency bands (see Figure 10).

As for the signal reception, it is better at night than during the day. From the perspective of noise floor, there is a difference of about 20 dB between the noise floor tested during the day and the noise floor tested at night. The reception situation of fixed monitoring points is better than that of mobile monitoring points, and the reception situation is worse near the construction test points, which can be known from the data received by different monitoring nodes. For spectrum data in the time-frequency domain, the background noise estimation of the instantaneous spectrum has certain limitations. It combines the time-frequency domain energy gradient matrix, and the effective signal of the target frequency band can be estimated quickly and effectively. According to (9) and the signal edge estimation method, the result of the effective signal estimation of the time-frequency domain spectrum is obtained (see Figure 11). Figure 11(a) shows the time-frequency domain spectrum data of a frequency band. The upper edge of the valid signal is marked with a red line, and the lower edge is marked with a blue line, as shown in Figure 11(b).

Although the energy gradient estimation method can estimate the effective signal in combination with the time-frequency domain spectrum data, there is still a certain probability of missed detection for the detection of discontinuous signals. This is also a problem that will be emphatically solved in the follow-up research.

6. Conclusions

Detecting and estimating the background noise of the electromagnetic spectrum environment of coast stations play an important role in measuring the quality of wireless communication in this area. We propose to estimate the upper envelope and lower envelope of the background noise by the maximum hold method and the minimum hold method, respectively. Combined with the forward difference algorithm, the background noise envelope curve is estimated. After the time-frequency domain electromagnetic spectrum data is processed, the noise floor of different target frequency bands is estimated, and the communication quality of the coast station is judged. This method combined with the time-frequency domain energy gradient estimation, the effective signal of the target frequency band can be quickly estimated, and according to the estimation of the background noise envelope and the effective signal, the background noise based on the time-frequency domain spectrum can be accurately determined. In the later period, we will study radio interference source detection and identification methods based on time-frequency domain data.

Data Availability

The data involved in this article are all obtained through the equipment Agilent N9342C RF analyzer.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The authors would like to thank the research was sponsored by the National Natural Science Foundation of China (62101166) and the Innovative Research Projects for Graduate Students in Ordinary Colleges and Universities in Hainan Province (Hys2020-216).

References

[1] M. López-Benítez and F. Casadevall, "Improved energy detection spectrum sensing for cognitive radio," *IET Communications*, vol. 6, no. 8, pp. 785–796, 2012.

[2] K. Umebayashi, K. Hayashi, and J. J. Lehtomäki, "Threshold-setting for spectrum sensing based on statistical information," *IEEE Communications Letters*, vol. 21, no. 7, pp. 1585–1588, 2017.

[3] J. J. Lehtomäki, R. Vuohtoniemi, and K. Umebayashi, "On the measurement of duty cycle and channel occupancy rate," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 11, pp. 2555–2565, 2013.

[4] Q.-T. Vien, H. X. Nguyen, R. Trestian, P. Shah, and O. Gemikonakli, "A hybrid double-threshold based cooperative spectrum sensing over fading channels," *IEEE Transactions on Wireless Communications*, vol. 15, no. 3, pp. 1821–1834, 2016.

[5] Z. Chen and Y. Zhang, "Cooperative energy detection algorithm based on background noise and direction finding error," *AEU - International Journal of Electronics and Communications*, vol. 95, pp. 326–341, 2018.

[6] M. Hajiabadi, H. Khoshbin, and G. Abed Hodtani, "Cooperative spectrum estimation over large-scale cognitive radio networks," *IET Signal Processing*, vol. 11, no. 8, pp. 1006–1014, 2017.

[7] A. Bagwari and G. S. Tomar, "Multiple energy detection vs cyclostationary feature detection spectrum sensing technique," in *2014 Fourth International Conference on Communication Systems and Network Technologies*, pp. 178–181, Bhopal, India, 2014.

[8] M. Lopez-Benitez and J. Lehtomaki, "On the sensing sample size for the estimation of primary channel occupancy rate in cognitive radio," in *IEEE Wireless Conference and Networking Conference (WCNC 2016)*, Doha, Qatar, 2016.

[9] A.-A. A. Boulogeorgos, N. D. Chatzidiamantis, and G. K. Karagiannidis, "Energy detection spectrum sensing under RF imperfections," *IEEE Transactions on Communications*, vol. 64, no. 7, pp. 2754–2766, 2016.

[10] J. J. Lehtomaki, M. Lopez-Benitez, K. Umebayashi, and M. Juntti, "Improved channel occupancy rate estimation," *IEEE Transactions on Communications*, vol. 63, no. 3, pp. 643–654, 2015.

[11] T. Pedersen, "Stochastic multipath model for the in-room radio channel based on room electromagnetics," *IEEE Transactions on Antennas and Propagation*, vol. 67, no. 4, pp. 2591–2603, 2019.

[12] E. Pei, L. Li, W. Cheng, and J. Jiang, "An energy efficient spectrum sensing algorithm in EH-HCRSN," *IEEE Access*, vol. 6, pp. 63021–63032, 2018.

[13] M. Höyhtyä, A. Mämmelä, M. Eskola et al., "Spectrum occupancy measurements: a survey and use of interference maps," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 4, pp. 2386–2414, 2016.

[14] M. Riahi Manesh, S. Subramaniam, H. Reyes, and N. Kaabouch, "Real-time spectrum occupancy monitoring using a probabilistic model," *Computer Networks*, vol. 124, pp. 87–96, 2017.

[15] O. Holland, B. Bochow, and K. Katzis, "IEEE 1900.6b sensing support for spectrum databases," in *2015 IEEE Conference on Standards for Communication and Networking(CSCN)*, pp. 199–205, Tokyo, Japan, 2015.

[16] M. López-Benítez and F. Casadevall, "Time-dimension models of spectrum usage for the analysis, design, and simulation of cognitive radio networks," *IEEE Transactions on Vehicular Technology*, vol. 62, no. 5, pp. 2091–2104, 2013.

[17] Q. Wu, G. Ding, J. Wang, and Y.-D. Yao, "Spatial-temporal opportunity detection for spectrum-heterogeneous cognitive radio networks: two-dimensional sensing," *IEEE Transactions on Wireless Communications*, vol. 12, no. 2, pp. 516–526, 2013.

Hindawi
Wireless Communications and Mobile Computing
Volume 2021, Article ID 7656877, 11 pages
https://doi.org/10.1155/2021/7656877

Research Article

Antitamper Image Watermarking Based on Cellular Network Topology for IoT-Themed Mobile Forensics

Xiao-zhu Xie,[1] Ching-Chun Chang,[2,3] Zhong-Liang Yang,[2] and Li Li[3]

[1]*School of Computer and Information Engineering, Xiamen University of Technology, Xiamen 361024, China*
[2]*Department of Electronic Engineering, Tsinghua University, Beijing, China*
[3]*College of Computer Science and Technology, Hangzhou Dianzi University, Hangzhou, China*

Correspondence should be addressed to Ching-Chun Chang; ccc@fcu.edu.tw

Received 6 August 2021; Accepted 2 September 2021; Published 12 October 2021

Academic Editor: Chi-Hua Chen

The Internet of Things (IoT) connects physical and digital worlds with mobile devices, accompanied by a surge in cybersecurity issues. With the rapid adoption of mobile devices, mobile forensics emerges as a new interdisciplinary field that concerns many forms of sabotage and cybercrime in the context of mobile computing. One of the most common cyberattacks is tampering. Digital watermarking is a tamper-evident technique used to protect data integrity. In this paper, we present an antitamper image watermarking scheme designed for mobile communications with low computational cost. A reference matrix based on cellular network topology is introduced to guide the watermark embedding and extraction processes. This reference matrix serves as a lookup table to reduce computational complexity, thereby enabling efficient implementation on mobile devices. Our scheme is aimed at offering high accuracy in detecting and localizing tampered regions. We also achieve a high watermarking capacity while leaving the visual quality of the carrier images nearly unharmed. Experimental results validate the effectiveness of our scheme against various types of simulated forgery including cropping and copy/paste attacks.

1. Introduction

The Internet of Things (IoT) bridges the gap between the physical and digital worlds and, at the same time, brings new cybersecurity challenges [1]. One of the most common cyberattacks is data tampering, which may take place during transmission across the Internet. Data integrity is crucial to many applications in IoT, and hence, it is important to develop an authentication scheme compatible with the IoT environment. Watermarking is an established technique used to protect copyright and authenticate the integrity of images [2, 3]. It embeds the secret information (i.e., Watermark) into the protected image invisibly. On the receiver side, the watermark is extracted and used for the proof of ownership or the authentication. Watermarking schemes can be broadly categorized as either robust watermarking schemes, typically designed to protect copyright, or fragile watermarking schemes, where the aim is to protect integrity. Furthermore, attack resisting and tampering localization are additional metrics used to evaluate the performance of watermarking schemes.

Since the embedding of watermark inevitably decreases the image quality, a trade-off must be made between the marked image quality and the ability to perform attack resisting and tampering localization. We propose an image tampering detection scheme based on blind and fragile watermarking scheme that maintains good image quality. The main contributions of our scheme can be summarized as below:

(i) This paper designs a cellular network reference matrix, based on which a blind and fragile watermarking scheme is proposed

(ii) The proposed scheme achieves a high embedding capacity of 1.403 bpp and maintains a good image quality of 47.16 dB

(iii) The proposed scheme can effectively resist attacks including cropping attack and copy and paste

attack, and it provides high accuracy tampering localization with tampering detection rate (TDR) of 100% after one dilation operation for cropping attacks

The structure of this paper is organized as follows. Section 2 provides a literature review. Section 3 introduces a preliminary work by which our cellular network matrix is inspired. Section 4 presents the proposed scheme, followed by the experimental results and analysis in Section 5. Finally, conclusions are drawn in Section 6.

2. Related Work

In robust watermarking schemes, watermark information, e.g., logos or copyright information, is embedded into the host image in an imperceptible way. The receiver can extract the watermark with accuracy even if the marked image has been tampered with. In other words, the verification of copyright can be guaranteed even if the marked image is subjected to malicious attacks. Since the frequency domain is more robust than the spatial domain against most attacks, robust watermarking schemes are mainly conducted in the frequency domain. In [4], the authors provide a robust watermarking scheme based on integer wavelet transform (IWT). It embeds the hash value of ROI (region of interest) into ROIN (region of noninterest) using IWT. Though it performs well on verifying the integrity of ROI, it is only applicable for images with small region of interest. Liu et al. [5] proposed a robust watermarking scheme by embedding the watermark into the LL subband after discrete wavelet transform (DWT) decomposition. Since the LL subband would hardly be changed when suffering the malicious attacks, it has strong robustness. However, it does not provide the tampering localization. Abdulrahman and Ozturk [6] applied discrete cosine transform (DCT) and DWT to each color component of an RGB image, then embed the watermark into four DWT bands. This approach is resistant to linear and nonlinear attacks. Singh and Bhatnagar [7] first transform the host image into the integer DCT domain before dividing it into nonoverlapping blocks in which the watermark is embedded using singular value decomposition. The scheme in [7] is robust against not only common attacks but also geometric attacks. However, the computational complexity of [7] is relatively high and not suitable for mobile computing environment. Yi et al. [8] generated a binary watermark by conducting double random-phase encoding on the host image itself and then embedded it into the DCT coefficients of the protected image. This scheme provides a robust authentication technique that can resist noise attacks, filtering attacks, partial occlusion attacks, etc. Yet, it fails to accomplish authentication if the error rate of the second phase key exceeds 20%. Robust watermarking schemes can protect against a wide range of attacks; however, image quality is relatively low, and the tampered region cannot be located.

Fragile watermarking schemes are typically designed for integrity authentication. Any slight modification to the marked image can be detected sensitively, and the tampered region can be reliably located. In this light, most existing schemes are based in the spatial domain. In early schemes, a watermark is generated with a secret key and a designated function. At the verifier side, the embedded watermark is extracted and compared with the original watermark. If there exists any difference, it indicates that the marked image has been tampered with. For instance, Yeung and Mintzer [2] use a secret key to generate a binary valued function which ensures that the extracted watermark will be the same as the embedded one if there exists no tampering. However, the scheme in [2] cannot specify the tampered region. Later, some block-based fragile watermarking schemes with tampering localization are proposed. Qin et al. [9] embed the authentication bits into the least significant bit (LSB) of the central pixel in one block and propose a block-wise fragile watermarking scheme with tampering localization. It provides tampering localization, at the cost of the image quality. Zhang et al. [10] generate a binary watermark by performing a local binary pattern (LBP) operation on nonoverlapping blocks of the original image, then embed the watermark using LSB substitution. Similar to [10], Gul and Ozturk [11] generate a watermark by performing the SHA-256 hash function on some divided blocks, then embedding the watermark in the LSBs of other blocks. Bhalerao et al. [12] propose a watermark using the secure hashing algorithm SHA-1 hash function. Though the methods in [8–12] can provide tampering localization and resist various attacks, they are based on blocks, which means they detect and locate the tampering region in the unit size of a block, resulting in a low resolution.

To improve the accuracy of tampering detection, some pixel-wise watermarking schemes are proposed. Prasad and Pal [13] take two pixels as a unit, then compute the watermark from the two most significant bits (MSBs) of each pixel using Hamming code, and finally embed them in the LSBs of the same pixels. Gong et al. [14] firstly generate two watermarks, i.e., the diffusion watermark and the authentication watermark, then arbitrarily embed them into the two LSBs of the cover image. Both watermarks are sensitive to alteration of the cover image, and the authentication watermark provides tampering localization at pixel level. Similarly, Memon and Alzahrani [15] generate a fragile watermark and a robust watermark, then, respectively, embed them into the region of interest (ROI) and the region of noninterest (RONI) of CT scan images. These schemes give good performance on tampering detection and localization; however, the image quality of the marked image is low.

3. Preliminary Work

Chang et al. [16] proposed the turtle shell matrix-based data hiding scheme in 2014. They first put forward the concept of turtle shell matrix, which is sized 256×256 and composed of turtle shells. The turtle shell is defined as a hexagon containing eight distinct digits, including two digits inside the hexagon, and six digits on the edges. The construction rules are as follows: (1) select a number from 0 to 7 to initialize the element with the coordinate of (0, 0); (2) the values of elements in the same row increase in steps of "1" modulo 7, and the values of elements in the same column increase in alternate steps of "2" and "3" modulo 7. An example of turtle shell matrix can be seen in Figure 1.

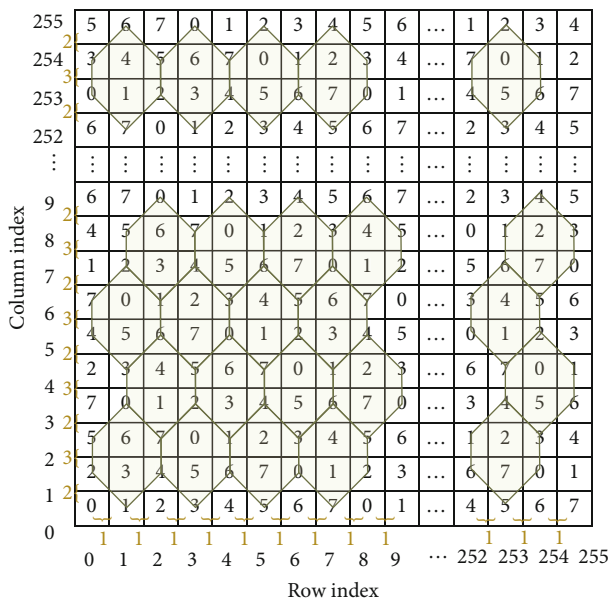


Figure 1: An example of turtle shell matrix.

Looking into the matrix, we can see that for any element, eight distinct digits ranging from 0 to 7 can be found nearby. In this light, Chang et al. regarded one pixel pair as the coordinate of the matrix and embedded a base-8 digit into one pixel pair. Interested readers can refer to [16] for more details about finding the closest element whose value equals to the to-be-embedded digit.

4. Proposed Scheme

We design a cellular network matrix, with which we propose a blind and fragile watermarking scheme for image tampering detection. Figure 2 shows a flowchart of the proposed scheme. The watermark is secretly embedded into the original image using the cellular network matrix to generate the watermarked image. During transmission in a public channel, the watermarked image may be tampered with by a malicious attacker. On the receiver side, the tampered region can be detected using the cellular network matrix, and the watermark can be extracted. In the following sections, cellular network matrix construction, data embedding, data extraction, and tampering detection are described successively.

4.1. Cellular Network Matrix Construction. Inspired by the turtle shell matrix [17], we design the cellular network reference matrix $M = [m(i,j)]_{N\times N}$ $(i, j = 0, 1, 2, \cdots, N)$ according to the following rules, where N is determined by the bit depth of the cover image.

(1) Set the value of an element to "Null" when the sum of its row and column is even. For instance, $m(0,0)$, $m(1,1)$, $m(1,3)$, and $m(2,2)$ are "Null"

(2) Otherwise, the values of elements in the same row increase in steps of 1 modulo 7, and the values of elements in the same column increase in steps of 4 modulo 7

It should be noted that the values of elements with the coordinate of (1, 0) and (0, 1) should be initialized first. And the difference between two values is 5 modulo 7. For instance, if $m(1,0)$ is valued 1, then $m(0,1)$ gets the value of 6; if $m(1,0)$ is valued 4, then $m(0,1)$ gets the value of 2. Once the initial values are given, the matrix is determined. And the initial values are referred as auxiliary information which is shared with the data hider and the receiver for data embedding and extraction. Since one initial value is represented using three bits, the auxiliary information occupies six bits totally. The auxiliary information can be self-embedded into the host image using LSB (least significant bit) substitution in our algorithm, as the information is so little that it barely affects the performance of the proposed scheme. After initialization, the rules can be formulated as

$$m(i,j) = \begin{cases} \text{Null}, & \text{if } (i+j) \text{ is even}, \\ m(i-2,j)+1 \mod 7, & \text{if } i \geq 3 \text{ and } (i+j) \text{ is odd and } i \text{ is odd}, \\ m(i,j-2)+4 \mod 7, & \text{if } j \geq 3 \text{ and } (i+j) \text{ is odd and } j \text{ is odd}. \end{cases} \quad (1)$$

For easy description, we refer to elements containing null values as "null elements" and elements that contain valid values as "valid elements." Figure 3 shows two examples of the matrix. Looking into the generated cellular network matrix, two important features can be observed:

(1) The four direct neighbors of any valid element are null elements

(2) Any cell, drawn to include a central valid element, the four nearest neighbor valid elements, and the two valid elements immediately above and below, can be seen to include all seven unique digits from 0 to 6. Several examples of such cells are highlighted in Figure 3

4.2. Cellular Network Matrix-Based Data Hiding. According to the features of the cellular network matrix, we are able to use one cover image pixel pair to form the coordinates of a matrix element and hide one base-7 digit within them. It is noted that the proposed scheme is suitable for images of any bit depths as long as we extend the matrix size to adapt

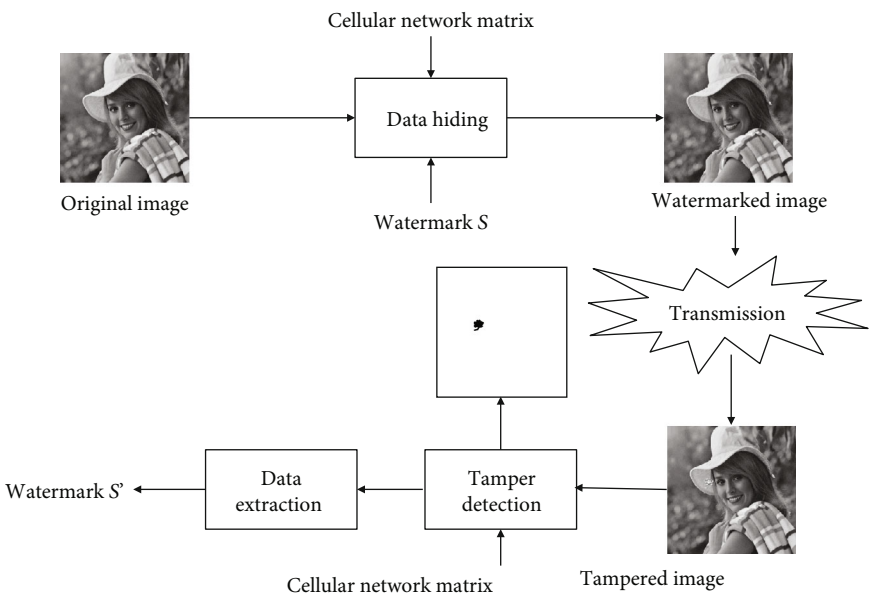


Figure 2: Flowchart of proposed scheme.

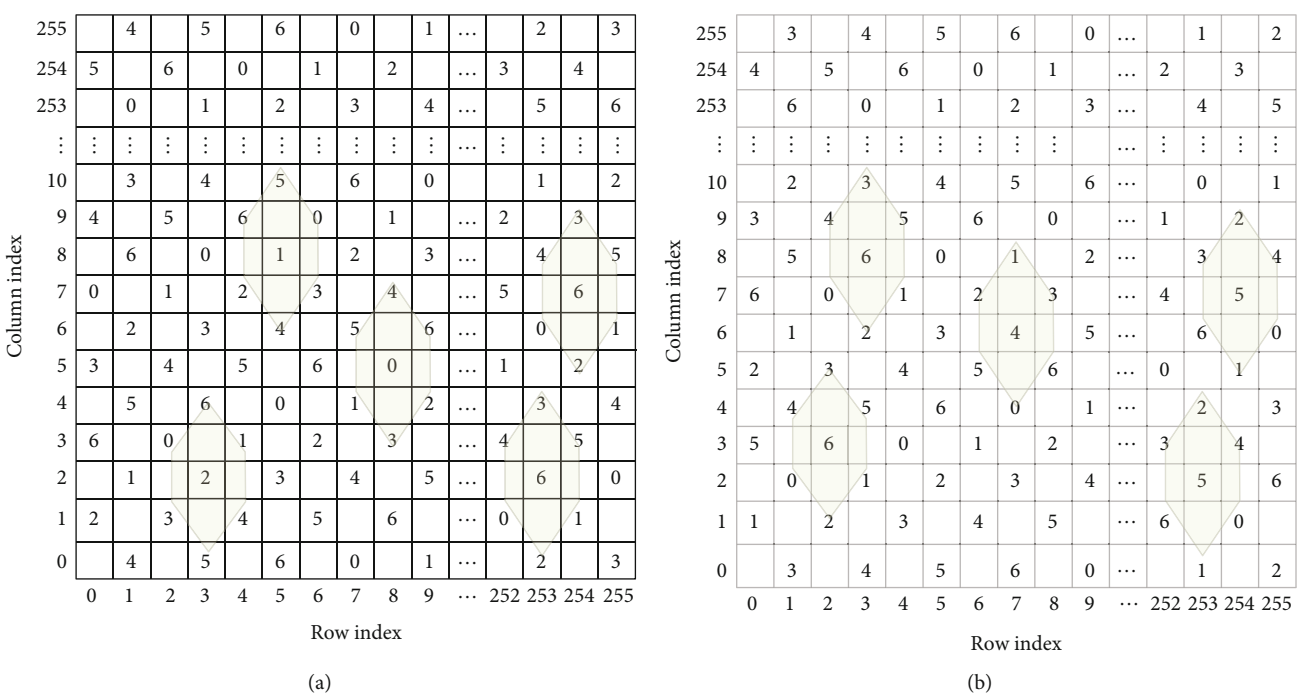


Figure 3: Examples of the cellular network matrix.

the bit depth. For simplicity, we take the gray image as the host image to depict the algorithm. The detailed processes are as follows:

(i) Inputs: the cover image, the binary watermark (generated by a random key), and the cellular network matrix

(ii) Output: a watermarked image

Step 1. Divide the cover image into nonoverlapping pixel pairs

Step 2. Convert the binary watermark into a base-7 digit stream

Step 3. Sequentially embed one digit w_i into each pixel pair (p_i, p_{i+1})

The guiding rule is to consider (p_i, p_{i+1}) as the coordinates of the cellular network matrix element and find the closest element (p'_i, p'_{i+1}) to (p_i, p_{i+1}) that satisfies

$$m\left(p'_i, p'_{i+1}\right) = w_i, \tag{2}$$

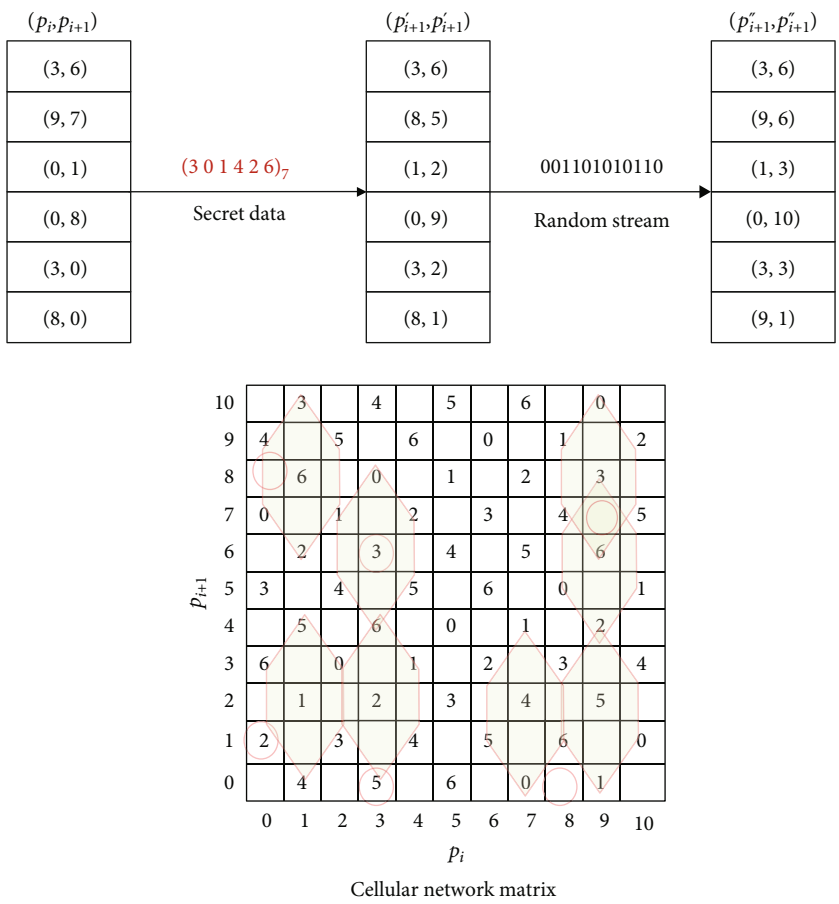


Figure 4: Examples of data hiding.

then replace (p_i, p_{i+1}) with (p'_i, p'_{i+1}) in the cover image. When the embedding is finished, we obtain the watermarked image.

The algorithm to find (p'_i, p'_{i+1}) satisfying Equation (2) can be categorized into four cases:

Case 1. $1 \le p_i \le 254$ and $2 \le p_{i+1} \le 253$ and $m(p_i, p_{i+1})$ is a valid element, for instance, (3,4). Then, draw a cell with center $m(p_i, p_{i+1})$ and find the element satisfying Equation (2) in the cell.

Case 2. $1 \le p_i \le 254$ and $2 \le p_{i+1} \le 253$ and $m(p_i, p_{i+1})$ is a null element, for instance, (5,5). Then, draw two cells with center $m(p_i, p_{i+1} - 1)$ and $m(p_i, p_{i+1} + 1)$, respectively. Note that one of the drawn cells may be not intact; it does not affect the result. Find the element satisfying Equation (2) in both cells.

Case 3. $m(p_i, p_{i+1})$ is a valid element and locates at the edge of the matrix, i.e., $p_i = 0$ or $p_i = 255$ or $p_{i+1} < 2$ or $p_{i+1} > 254$. Then, draw one cell using $m(p_i, p_{i+1})$ as the left bottom or the right bottom or the bottom or the top corner, respectively. Find the element satisfying Equation (2) in the cell.

Case 4. $m(p_i, p_{i+1})$ is a null element and locates at the edge of the matrix, i.e., $p_i = 0$ or $p_i = 255$ or $p_{i+1} < 2$ or $p_{i+1} > 254$. Then, draw one cell with corresponding center of $m(p_i + 1, p_{i+1})$, $m(p_i - 1, p_{i+1})$, $m(p_i, p_{i+1} + 1)$, or $m(p_i, p_{i+1} - 1)$, respectively. Find the element satisfying Equation (2) in the cell.

According to the embedding algorithm described above, when we divide the watermarked image into pixel pairs, the sum of these must be odd. In order to break this condition, which cannot be guaranteed for a natural image, we process the watermarked image using a pseudorandom binary stream according to

$$p'_i = p'_i + r_i, \tag{3}$$

where r_i is the elements of a pseudorandom binary stream $R = r_1, r_2, \cdots, r_{W*H}$ and W and H represent the width and height of the cover image, respectively.

To make the embedding algorithm easier to understand, we give examples in Figure 4. Assume that a string of base-7 digits, e.g., $(3\ 0\ 1\ 4\ 2\ 6)_7$, are embedded into cover pixel pairs {(3,6), (9,7), (0,1), (0,8), (3,0), (8,0)}. Pick up the first pixel pair (3,6) and the to-be-embedded digit "3," which belongs to Case 1 according to the embedding algorithm depicted above. Then, draw a cell in the matrix with center $m(3, 6)$, and find the element $m(3, 6)$ whose value equals "3" (see Figure 3). In this case, pixel pair (3,6) remains unchanged in the cover image. For the second pixel pair (9,7), which

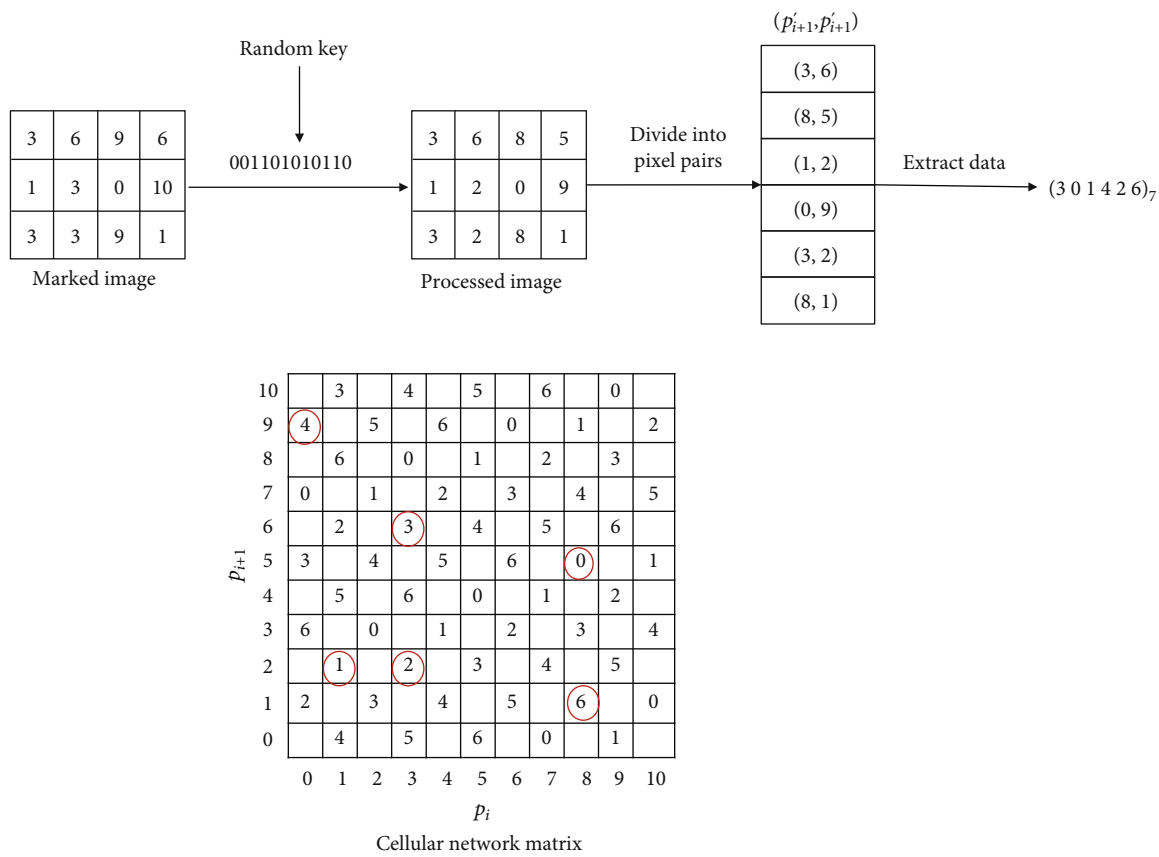


Figure 5: Example of data extraction.

belongs to Case 2, draw two cells with centers $m(9, 8)$ and $m(9, 6)$, respectively. Then, find the closest element whose value equals "0," $m(8, 5)$, and replace pixel pair (9,7) with pixel pair (8,5) in the cover image. The remaining four pixel-pairs belong to Cases 3 and 4. We do not describe these in the text due to the limited available space. After the six digits are embedded, we get the marked pixel pairs {(3,6), (8,5), (1,2), (0,9), (3,2), (8,1)}. In order to disguise the presence of the watermark, a pseudorandom binary stream, e.g., "001101010110" is added to the marked pixels, and the final marked pixel pairs are obtained, i.e., {(3,6), (9,6), (1,3), (0,10), (3,3), (9,1)}.

4.3. Data Extraction. The process of data extraction is the inverse operation of data embedding. The detailed procedures are as follows:

(i) Inputs: the marked image, the key, and the cellular network matrix

(ii) Output: the secret data

Step 1. Use the key to generate a binary stream $R = r_1, r_2, \cdots, r_{W*H}$ and process the watermarked image according to:

$$p'_i = p'_i - r_i, \tag{4}$$

where p'_i and p'_i represent the processed pixel and marked pixel, respectively

Step 2. Separate the processed image into pixel pairs, denoted as (p'_i, p'_{i+1})

Step 3. Consider each pixel pair as coordinates in the cellular network matrix and obtain the value $m(p'_i, p'_{i+1})$ which is precisely the secret digit encoded in base-7

Step 4. After all the secret digits are extracted, convert them into binary form which is the final secret data

An example is given in Figure 5 to illustrate the process of data extraction. As shown in Figure 5, a binary stream "001101010110" is generated using the key, then the processed image is obtained by operating Equation (4) on the marked image and the binary stream. After that, the processed image is separated into pixel pairs (p'_i, p'_{i+1}), and they are considered as coordinates of the cellular network matrix to get the value $m(p'_i, p'_{i+1})$. We now have the secret digits "301426" in base-7, and finally, we convert them into binary form to retrieve the secret data.

4.4. Tampering Detection Phase. If a marked image is tampered with during transmission, the recipient can detect the tampered region without knowing the original watermark. According to the embedding rule, if we select a pixel pair from the marked image and consider it as the coordinates of an element in the cellular network matrix, it must map to a "valid element." Therefore, we can determine that

a pixel pair has been tampered with if it maps to a "null element." Details of tampering detection are described as follows.

(i) Inputs: the marked image, the key, and the cellular network matrix

(ii) Output: the tampered region

Step 1. Follow the same procedures as depicted in steps 1-2 of data extraction

Step 2. Consider each pixel pair as the coordinates of an element in the cellular network matrix. If $m(p'_i, p'_{i+1}) = \text{Null}$, determine that at least one of the two pixels has been tampered with and marks the corresponding position in the tampered map as "1"; otherwise, mark it as "0"

Step 3. Since the tampered pixel pairs do not always map to the "null element," there is some misjudgment of the tampered map. Therefore, a dilation operation is performed on the interim tampered map to generate the final map. The operator is shown in

$$\begin{bmatrix} 0 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 0 \end{bmatrix}. \tag{5}$$

5. Experimental Results

A series of experiments are conducted on the test images shown in Figure 6 to evaluate the performance of the proposed scheme. The test images are grayscale images with size 512×512 sampled from the UCID dataset [18]. Simulation experiments are conducted to evaluate the performance of tampering detection and localization when subjected to attacks, including cropping attack and copy and paste attack. Since we take only two pixels as a unit, tampering localization can be obtained with higher resolution than some state-of-the-art methods.

5.1. Evaluation Metrics. Definitions of the metrics used to evaluate the proposed scheme are described in this section.

(1) ER: embedding ratio

$$\text{ER} = \frac{N}{W \times H}\ (\text{bpp}), \tag{6}$$

where N represents the number of embedded secret bits and W and H denote the width and the height of the cover image, respectively.

(2) PSNR: peak-signal-to-noise-ratio, is calculated using

$$\text{PSNR} = 10 \log_{10}\left(\frac{255^2}{\text{MSE}}\right), \tag{7}$$

where MSE (mean square error) is computed as

$$\text{MSE} = \frac{1}{W \times H}\sum_{i=1}^{W}\sum_{j=1}^{H}\left(I(i,j) - I'(i,j)\right)^2. \tag{8}$$

(3) SSIM: structural similarity index measure is calculated on various windows of an image. The measure between two windows x and y is

$$\text{SSIM}(x,y) = \frac{\left(2\mu_x\mu_y + c_1\right)\left(2\sigma_{xy} + c_2\right)}{\left(\mu_x^2 + \mu_y^2 + c_1\right)\left(\sigma_x^2 + \sigma_y^2 + c_2\right)}, \tag{9}$$

where:

μ_x, μ_y are the average of x and y, respectively

σ_x^2, σ_y^2 are the variance of x and y, respectively

σ_{xy} is the covariance of x and y

c_1, c_2 are two variables to stabilize division with a weak denominator.

The larger the SSIM index is, the higher the similarity is. If SSIM = 1, then the two images are identical.

(4) Tampering detection rate (TDR) [19]: the percentage of tampered pixels that are correctly detected

$$\text{TDR} = \frac{\text{No.of correctly detected pixels}}{\text{No.of actual tampered pixels}} \times 100\%. \tag{10}$$

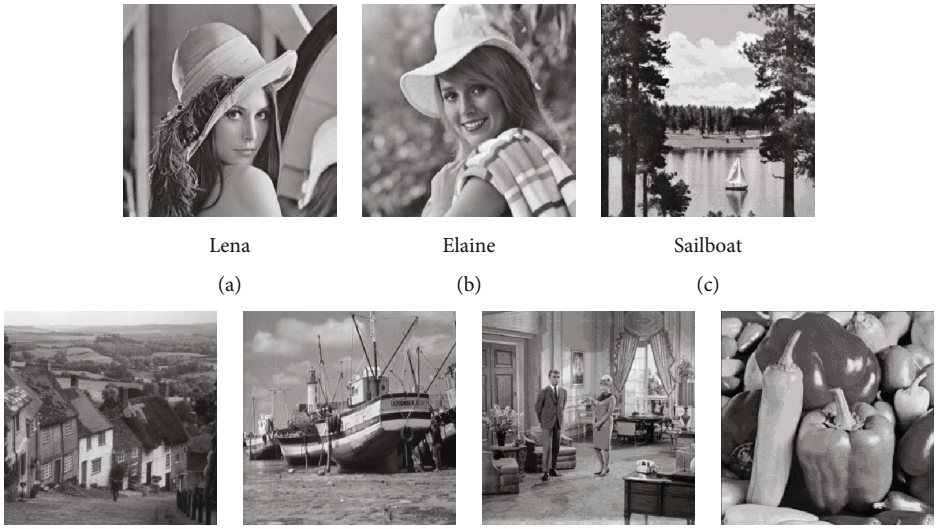
(5) False-positive rate (FPR) [19]: the percentage of nontampered pixels that are incorrectly judged to be tampered pixels

$$\text{FPR} = \frac{\text{No.of misjudged as tampered pixels}}{\text{No.of detected tampered pixels}} \times 100\%. \tag{11}$$

(6) False-negative rate (FNR) [19]: the percentage of tampered pixels that are incorrectly judged to be nontampered

$$\text{FNR} = \frac{\text{No.of misjudged as non} - \text{tampered pixels}}{\text{No.of detected tampered pixels}} \times 100\%. \tag{12}$$

5.2. Executing Efficiency. The execution times for seven test images with size of 512×512 are measured, and experiments are implemented using MATLAB 2017b on a PC with an Intel® Quad-Core i5 CPU @ 1.1GHz with 8 GB RAM. As shown in Table 1, matrix construction costs around 0.0027 s, the watermark embedding and extracting processes



Lena (a) Elaine (b) Sailboat (c)

Goldhill (d) Boat (e) Couple (f) Pepper (g)

Figure 6: Seven test images.

Table 1: Execution times of the proposed scheme (unit: second).

Execution time	Matrix construction	Watermark embedding	Watermark extraction	Tamper detection	Total
Boat	0.0027	0.0226	0.0124	0.0028	0.0405
Couple	0.0027	0.0209	0.0120	0.0030	0.0386
Elaine	0.0027	0.0201	0.0132	0.0034	0.0394
Goldhill	0.0027	0.0197	0.0129	0.0029	0.0382
Lake	0.0027	0.0211	0.0156	0.0033	0.0427
Lena	0.0027	0.0258	0.0141	0.0036	0.0462
Pepper	0.0027	0.0205	0.0158	0.0029	0.0419
Average	0.0027	0.0215	0.0137	0.0031	0.0411

Table 2: Performance of the proposed scheme.

Image	ER (bpp)	PSNR (dB)	SSIM
Boat	1.403	47.17	0.9898
Couple	1.403	47.15	0.9938
Elaine	1.403	47.16	0.9925
Goldhill	1.403	47.16	0.9934
Lake	1.403	47.15	0.9929
Lena	1.403	47.17	0.9903
Pepper	1.403	47.16	0.9907
Average	1.403	47.16	0.9919

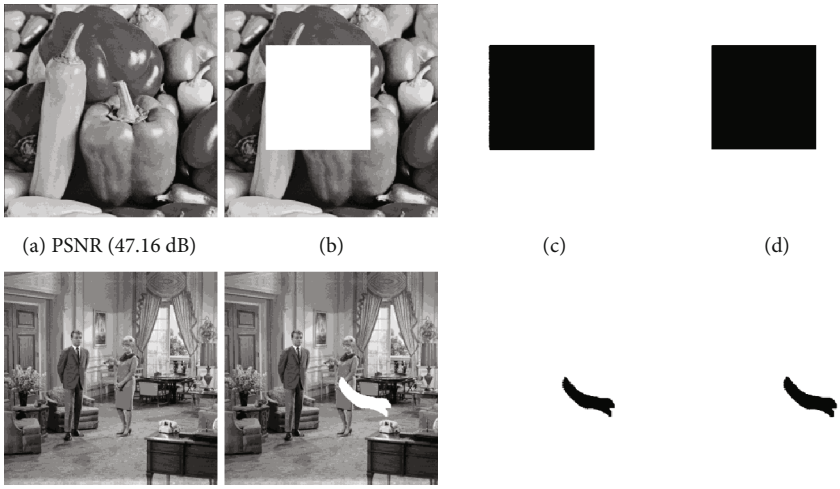
cost 0.0215 s and 0.0137 s, respectively, on average, and tampering detection costs around 0.0031 s. The total time cost is 0.0411 s on average. Therefore, we can declare that the proposed scheme has high efficiency, making it suitable for use in most real-time systems.

5.3. Image Quality and Embedded Data Capacity. In this section, three metrics, ER, PSNR, and SSIM, are used to evaluate the data capacity of the watermark and the imperceptibility of the watermark in watermarked images, with results summarized in Table 2. As described in Section 2.1, one base-7 digit is embedded into each pixel pair, giving a constant ER of 1.403 bpp. Corresponding PSNRs and SSIMs of the watermarked images with maximum payload are given as well. We can see that the average PSNR reaches 47.16 dB and an average SSIM of 0.9919 is achieved, which implies that very good imperceptibility of the watermarked is achieved by the proposed scheme.

5.4. Tampering Detection and Localization

5.4.1. Cropping Attack Detection. First, simulation experiments of cropping attack are conducted on the watermarked images "Pepper" and "Couple," then the proposed scheme is used to detect and locate the tampered region, with the results shown in Figure 7.

Figures 7(a) and 7(e) show the watermarked images, and Figures 7(b) and 7(f) show the tampered images with regular and irregular cropping, respectively. In more detail, "Pepper" is tampered with by cropping a rectangular region, and "Couple" is tampered with by cropping an irregular shape. Since all the tampered pixel values become 255, the sum of two adjacent pixels is even. In other words, pixel pairs from the tampered region inevitably map to a "null



(a) PSNR (47.16 dB) (b) (c) (d)

(e) PSNR (47.15 dB) (f) (g) (h)

Figure 7: Tampering localization of cropping attack: (a), (b), (c), (d) for "Pepper"; (e), (f), (g), (h) for "Couple."

Table 3: Detection results of cropping attack.

Cover images	Results without dilation operation (%)			Results with one dilation operation (%)		
	TDR	FPR	FNR	TDR	FPR	FNR
Peppers	99.80	0.20	0.06	100	1.70	0
Couples	98.78	1.41	0.02	100	8.80	0

element" in the matrix and are detected as tampered pixels. Therefore, the tampered region can be detected accurately except for those pixel pairs on the tampering edge. The detection results without the dilation operation are shown in Figures 7(c) and 7(g), with TDRs of 99.80% and 98.78%, respectively. After one dilation operation, both TDRs are improved to 100% (as shown in Figures 7(d) and 7(h)). Without doubt, the dilation operation would make some nontampered pixels be tampered ones, resulting in an increase of FPR. The statistical results can be found in Table 3, from which we can see that after one dilation operation, the TDR reaches up to 100% with an FNR of 0%, yet the FPR increases slightly. Trade-off between TDR and FPR can be made by conducting the dilation operation.

5.4.2. Copy and Paste Attack Detection. Next, simulation experiments of copy and paste attack are conducted on two watermarked images "Lena" and "Goldhill" generated by the proposed scheme (as shown in Figures 8(a) and 8(e)). Copy and paste attack refers to an attack that replaces one region in the watermarked image with a copy of another region either from within the same image or from an external image. For instance, Figure 8(b) shows an image tampered with by copying a flower onto the shoulder of "Lena" and Figure 8(f) shows an image tampered with by copying a section of white wall and pasting it back on top of a different region within the same image. The detection results without dilation operation can be seen in Figures 8(c) and 8(g), whose TDRs are 55.44% and 49.15%, respectively. FPRs and FNRs are relatively low as well. The statistical results can be found in Table 4, from which we can see that after one dilation operation, TDRs increase shapely; meanwhile, FNRs decrease considerably. Regrettably, FPRs increase accordingly. After two dilation operations, the detection performance improves remarkably (as shown in Figures 8(d) and 8(h)). Both TDRs are close to 100%, and the FNRs are close to 0%. However, the FPRs increase to 16.58% and 7.36%, respectively. The trade-off can be made by the number of dilation operations applied.

5.5. Comparison. Comparisons with some relevant schemes [8, 11, 13, 14] are made in this section. Among them, the scheme in [8] is conducted in the DCT domain, while the remaining schemes are conducted in the spatial domain. Table 5 gives the comparison results. We define unit size to be the dimensions in pixels of the region used for tampering detection and localization. The smaller the unit size, the higher the resolution for detection and localization. We can see that the proposed scheme is significantly superior to schemes [8, 11] in terms of resolution and embedded data capacity. The resolution of the proposed scheme is the same as scheme [13], and it has a similar embedding ratio; however, it performs significantly better in terms of image quality and TDR. Though scheme [14] has a better resolution and a larger embedding rate, the proposed scheme again performs significantly better in terms of image quality and TDR. Overall, the proposed scheme performs well in terms of embedding capacity and tampering detection ability compared to most existing methods and gives significantly improved image quality.

6. Conclusions

We propose a high-performance blind fragile watermarking scheme for image tampering detection. The core contribution of this work is the construction of a cellular network

(a) PSNR (47.17 dB) (b) (c) (d)

(e) PSNR (47.16 dB) (f) (g) (h)

Figure 8: Tampering localization of copy and paste attack: (a), (b), (c), (d) for "Lena"; (e), (f), (g), (h) for "Goldhill."

Table 4: Detection results of copy and paste attacks.

Cover images	Results without dilation operation (%)			Results with one dilation operation (%)			Results with two dilation operations (%)		
	TDR	FPR	FNR	TDR	FPR	FNR	TDR	FPR	FNR
Lena	55.53	3.08	0.57	94.24	9.47	0.08	99.58	16.58	0.01
Goldhill	49.15	2.18	2.04	93.54	4.13	0.26	99.79	7.36	0.01

Table 5: Comparisons between the proposed scheme and others.

Scheme	Domain	Unit size	ER	PSNR (dB)	TDR
[8]	DCT	8×8	4096 bits	42.23	-
[11]	Spatial	32×32	0.25 bpp	≥57	-
[13]	Spatial	1×2	1.5 bpp	42.09	99.50%
[14]	Spatial	1×1	2.0 bpp	44.17	99%
Ours	Spatial	1×2	1.403 bpp	47.16	100%

"-": not present.

reference matrix which is central to a suite of watermarking algorithms that we implement, evaluate, and compare with existing results. Experimental results show that the proposed scheme can achieve a high embedding capacity of 1.403 bpp while maintaining excellent image quality with peak signal-to-noise ratio of 47.16 dB. If the watermarked image is tampered with during transmission, the tampering region can be detected and located with high resolution without knowing the original watermark. For cropping attack and copy and paste attack, the proposed cellular network scheme can provide highly accurate tampering detection, obtaining a tampering detection rate of up to 100%. Furthermore, the algorithm is easy to realize, has high efficiency, and is suitable for use in mobile computing environment.

As the increasing concerns of information integrity, reversible watermarking techniques gain more and more concerns. The technique realizes that the watermarked image can be recovered thoroughly after the watermark is extracted. Chang et al. [20–22] propose several reversible watermarking schemes for protecting privacy in the cloud computing environment. Thus, in the future, we will do research on reversible watermarking for image authentication and tampering localization.

Data Availability

The image data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work is supported by National Natural Science Foundation of China (No. 62172132) and Scientific Research Foundation for the Introduction of Talent at Xiamen University of Technology (YKJ21007R).

References

[1] M. Stoyanova, Y. Nikoloudakis, S. Panagiotakis, E. Pallis, and E. K. Markakis, "A survey on the internet of things (IoT) forensics: challenges, approaches, and open issues," *IEEE Communications Surveys & Tutorials*, vol. 22, no. 2, pp. 1191–1221, 2020.

[2] M. M. Yeung and F. Mintzer, "An invisible watermarking technique for image verification," *Proceedings of international conference on image processing*, vol. 2, pp. 680–683, 1997.

[3] A. K. Kamran and S. A. Malik, "A high capacity reversible watermarking approach for authenticating images: exploiting down-sampling, histogram processing, and block selection," *Information Sciences*, vol. 256, pp. 162–183, 2014.

[4] R. Eswaraiah and E. Sreenivasa Reddy, "Robust medical image watermarking technique for accurate detection of tampers inside region of interest and recovering original region of interest," *IET Image Processing*, vol. 9, no. 8, pp. 615–625, 2015.

[5] X.-L. Liu, C.-C. Lin, and S.-M. Yuan, "Blind dual watermarking for color images' authentication and copyright protection," *IEEE Transaction on Circuits and Systems for Video Technology*, vol. 28, no. 5, pp. 1047–1055, 2018.

[6] A. K. Abdulrahman and S. Ozturk, "A novel hybrid DCT and DWT based robust watermarking algorithm for color images," *Multimedia Tools and Applications*, vol. 78, no. 12, pp. 17027–17049, 2019.

[7] S. P. Singh and G. Bhatnagar, "A new robust watermarking system in integer DCT domain," *Journal of Visual Communication and Image Representation*, vol. 53, pp. 86–101, 2018.

[8] F. Yi, Y. Kim, and I. Moon, "Secure image-authentication schemes with hidden double random-phase encoding," *IEEE Access*, vol. 6, no. 1, pp. 70113–70121, 2018.

[9] C. Qin, P. Ji, X. Zhang, J. Dong, and J. Wang, "Fragile image watermarking with pixel-wise recovery based on overlapping embedding strategy," *Signal Processing*, vol. 138, pp. 280–293, 2017.

[10] H. Zhang, C. Wang, and X. Zhou, "Fragile watermarking based on LBP for blind tamper detection in images," *Journal of Information Processing Systems*, vol. 13, no. 2, pp. 385–399, 2017.

[11] E. Gul and S. Ozturk, "A novel hash function based fragile watermarking method for image integrity," *Multimedia Tools and Applications*, vol. 78, no. 13, pp. 17701–17718, 2019.

[12] S. Bhalerao, I. A. Ansari, and A. Kumar, "A secure image watermarking for tamper detection and localization," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, no. 1, pp. 1057–1068, 2021.

[13] S. Prasad and A. K. Pal, "A tamper detection suitable fragile watermarking scheme based on novel payload embedding strategy," *Multimedia Tools and Applications*, vol. 79, no. 3-4, pp. 1673–1705, 2020.

[14] X. Gong, L. Chen, F. Yu, X. Zhao, and S. Wang, "A secure image authentication scheme based on dual fragile watermark," *Multimedia Tools and Applications*, vol. 79, no. 25-26, pp. 18071–18088, 2020.

[15] N. Memon and A. Alzahrani, "Prediction-based reversible watermarking of CT scan images for content authentication and copyright protection," *IEEE Access*, vol. 8, pp. 75448–75462, 2020.

[16] C.-C. Chang, Y. Liu, and T. S. Nguyen, "A novel turtle shell based scheme for data hiding," in *Proceedings of 2014 International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, pp. 89–93, Kita Kyushu, Japan, 2014.

[17] X. Z. Xie, C. C. Chang, C. C. Lin, and J. L. Lin, "A turtle shell based RDH scheme with two-dimensional histogram shifting," *Multimedia Tools and Applications*, vol. 78, no. 14, pp. 19413–19436, 2019.

[18] *USC-SIPI Image Database*http://sipi.usc.edu/database/.

[19] G. D. Su, C. C. Chang, and C. C. Lin, "High-precision authentication scheme based on matrix encoding for AMBTC-compressed images," *Symmetry*, vol. 11, no. 8, p. 996, 2019.

[20] C.-C. Chang, C.-T. Li, and Y. Q. Shi, "Privacy-aware reversible watermarking in cloud computing environments," *IEEE Access*, vol. 6, pp. 70720–70733, 2018.

[21] C.-C. Chang, C.-T. Li, and K. Chen, "Privacy-preserving reversible information hiding based on arithmetic of quadratic residues," *IEEE Access*, vol. 7, pp. 54117–54132, 2019.

[22] C.-C. Chang, "Neural reversible steganography with long short-term memory," *Security and Communication Networks*, vol. 2021, Article ID 5580272, 14 pages, 2021.

Hindawi
Wireless Communications and Mobile Computing
Volume 2021, Article ID 9310558, 11 pages
https://doi.org/10.1155/2021/9310558

Research Article

Joint Channel Pruning and Quantization-Based CNN Network Learning with Mobile Computing-Based Image Recognition

Huanyu Liu,[1] Qing Luo,[2,3] Mingmei Shao,[1] Jeng-Shyang Pan,[4] and Junbao Li[1]

[1]*School of Electronics and Information Engineering, Harbin Institute of Technology, Harbin 150001, China*
[2]*College of Aerospace Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China*
[3]*Shenyang Aircraft Design & Research Institute, Shenyang 110035, China*
[4]*College of Computer Science and Engineering, Shandong University of Science and Technology, Qingdao 266590, China*

Correspondence should be addressed to Junbao Li; lijunbao@hit.edu.cn

Received 11 August 2021; Accepted 4 September 2021; Published 8 October 2021

Academic Editor: Chi-Hua Chen

The development of the Internet and communication technology has ushered in a new era of the Internet of Things (IoT). Moreover, with the rapid development of artificial intelligence, objects are endowed with intelligence, such as home automation and smart healthcare, which are typical applications of artificial intelligence technology in IoT. With the rise of convolutional neural network (CNN) in the field of computer vision, more and more practical applications need to deploy CNN on mobile devices. However, due to the large amount of CNN computing operations and the large number of parameters, it is difficult to deploy on ordinary edge devices. The neural network model compression method has become a popular technology to reduce the computational cost and has attracted more and more attention. We specifically design a small target detection network for hardware platforms with limited computing resources, use pruning and quantization methods to compress, and demonstrate in VOC dataset and RSOD dataset on the actual hardware platform. Experiments show that the proposed method can maintain a fairly accurate rate while greatly speeding up the inference speed.

1. Introduction

In recent years, with the rapid development of the mobile infrastructure of IoT and the increasing popularity of the application of IoT, the complexity and operability of various mobile applications have been continuously increasing, and the requirements for the intelligence of mobile applications are getting higher and higher. At the same time, artificial intelligence has been gradually applied to all aspects of IoT, such as home automation [1], smart healthcare [2], smart security [3], autopilot [4], and other fields.

In recent years, convolutional neural network (CNN) is regarded as one of the best techniques for understanding image content and has shown great performance in image classification [5], segmentation [6], and detection [7] tasks. The features of CNN such as local connection, weight sharing, and pooling operations can effectively reduce the complexity of the network and reduce the number of training parameters. Since the advent of AlexNet [8], most frontrunners in image processing competitions have adopted CNN-based methods.

However, current state-of-the-art CNN cannot adapt well to today's smart mobile devices. Especially, in tasks with higher complexity like target detection, computing power and on-device memory has become the two major bottlenecks for the CNN model deployment on the mobile side. Therefore, many recent research is aimed at reducing the model computing operations and model parameters with minimal accuracy losses. However, most model compression methods are performed on classification models. Common classification networks usually have small input resolution and simple model architecture, while detection network usually has a larger input resolution and more complicated architecture. Detection network may also include multiple inputs, as well as some postprocessing, which put forward higher requirements for the design of the network compression algorithm. Common compression methods include pruning [9], quantization [10], and distillation. Model

pruning methods can usually be divided into two types: structured and unstructured. Unstructured pruning methods usually only work on specific hardware due to the sparsity on neuron level [11]. Existing target detection datasets have prominent target features and clear details. However, in practical applications, due to the high shooting height, the target size is too small compared to the image, and the target features are incomplete, and the target occurs a certain degree of deformation affected by the shooting angle; the relative motion between the target and the drone causes the target background to change significantly, etc., making the task of drone image target detection challenging. In order to meet the above needs and solve the technical difficulties of UAV target detection, in recent years, researchers have carried out a series of related research. Traditional UAV image target detection methods include frame difference method, background subtraction method, sliding window-based feature extraction algorithm [12], mean-shift algorithm, and edge detection algorithm, and recently, deep learning methods are proposed, for examples, fast deep neural networks with knowledge guided training and predicted regions of interests [13], small unmanned aerial vehicle [14], object-based hierarchical change detection [15], application of unmanned aerial vehicles [16], and real-time implementation using GPUs [17]. Traditional sliding window-based features are usually artificially designed Histogram of Oriented gradient features (HoG) [18], Scale-invariant feature transform features (SIFT) [19], Haar-like wavelet features [20], etc. This method for implementing feature has high computational complexity and cannot be detected in real time. In 2012, S. Janand others combined the multiscale mean-shift algorithm with the edge information of the target to solve the saliency object detection of images taken by the drone. In addition, there are various real-time moving object detections in aerial surveillance algorithms such as local null space pursuit [21]. These methods are slow in calculation and weak in robustness and cannot meet the actual application requirements of real-time detection. In 2016, researchers used neural networks to detect rice field weeds from aerial images of drones flying at a height of 50 meters [22]. Zhang and others searched regions of interest (ROIs) based on the characteristics of adjacent parallel lines [23] and determined the final airport area through transfer learning on the AlexNet network. Xiao and others used the new GoogleNet-LF model to extract multiscale deep integrated feature combination SVM for detection and recognition [24]. In 2018, Wang and others used CNN target detectors with RetinaNet [25] as the backbone network to perform pedestrian detection on the Stanford drone data set [26] and verified the targets of the CNN-based target detector on the drone image advantages in detection. In the highly complex tasks such as target detection, computing power and memory space become the two bottlenecks in the deployment of the model on the mobile terminal. Therefore, many recent research results are aimed at reducing the amount of model calculation and model parameters and affecting the model accuracy as low as possible. However, most of the model compression methods are based on the classification model. Common classification models usually have small input resolution and simple model structure, while target detection models usually have larger input resolution and more complex network structure. The target detection model may also include multiple inputs and some postprocessing processes, which put forward higher requirements for the design of model compression algorithm. Xiao et al. [27] introduced comparative learning into model distillation. The main idea of contrastive learning is to learn a representation. In some metric spaces, the positive sample pairs are close together, and the negative sample pairs are separated as much as possible. The author models the contrast loss between each layer of teacher network and student network, so that the teacher network and student network are not only consistent in the prediction probability of output but also similar in internal representation. The introduction of comparative learning further improves the effect of model distillation. Tian et al. [28] proposed the variational student method. It combines the compressibility of knowledge distillation framework and the sparsity and guiding ability of variational reasoning technology. The author creates a sparse student network. The sparsity of the network is introduced by optimizing the variational parameters of the loss function based on variational reasoning. This is achieved by using the knowledge learned from the teacher network. The author considers Bayesian neural network in a general knowledge distillation framework, in which students adopt a variational penalty least squares objective function. Model compression is a very important part of neural network model training and deployment. Most neural network deployment tools integrate the model compression technology of quantification after training. However, a widely applicable model compression tool is still very rare. Distiller [29] is a toolkit that supports multiple model compression methods. It supports structural and unstructured pruning, quantification after and during training, and knowledge distillation methods. However, these functions can only be used in the classification model. For the more complex target detection model, it only supports unstructured pruning. And it does not provide the interface related to embedded deployment, which hinders the deployment of the final model. The situation of paddleslim is similar to that of distiller, and it also lacks support for target detection network. Recently, TensorFlow [30] supports unstructured model pruning and quantization algorithms during training. Relying on TensorFlow Lite, TensorFlow's model can be easily deployed in the embedded segment. The latest version of PyTorch [31] also supports some basic quantitative tools but less support for the deployment of embedded platforms.

As the previous works, the popular object detection algorithms are usually based on convolutional neural networks, which are difficult to be deployed on platforms with limited computational resources such as embedded platforms due to the limitation of computational amount and large number of parameters. However, with the increasing demand for object detection task in the industry, the detection speed of the common object detection model is not up to the standard. Although the computing power of embedded platform is developing rapidly in recent years, the large amount of

computation and the large number of parameters of object detection algorithm are still the main factors that restrict the practical application of object detection algorithm. Therefore, it is of great significance to study the model compression algorithm of object detection algorithm for the application of object detection algorithm in industry. In this paper, an object detection network suitable for embedded platform deployment is optimized and designed through in-depth study of object detection algorithm. On the basis of optimized network, the object detection network compression algorithm based on pruning and quantization is proposed, which can greatly compress the model size and improve the reasoning speed of the model while maintaining the accuracy of the model. The structured pruning method usually prunes on the channel level, which can be accelerated on common hardware. Quantization methods can generally be divided into posttraining quantization and quantization during training. Quantization converts floating-point numbers to integers symmetrically or asymmetrically according to the quantization bit depth, zero point, and scale parameters. Posttraining quantization usually calculates the quantization parameters of each layer on the calibration set.

The channel pruning algorithm can be divided into three steps: sparse training, channel parameter pruning, and constant channel elimination. The purpose of sparsity training is to make the parameters of the model produce structured sparsity, which is convenient for subsequent screening of channels that have little impact on the accuracy of the network. Structured sparsity means that sparsity occurs not only on a single parameter but also on the whole channel. The second step is to cut out the channels that have little impact on the accuracy. Generally, the closer the parameter to 0, the less the impact on the accuracy of the network. When cutting off the channel, the influence of residual structure and other structures should be considered to prevent the structure of the network from being damaged. Finally, constant channel elimination can eliminate the influence of constant channel and restore the accuracy of the model. In the process of model quantization, we first need to insert a pseudo quantization module into the model. The purpose of this step is to reduce the quantization loss caused by the model quantization process, and the quantization parameters can also be introduced into the calculation diagram. Then, quantitative training, in which the weight and quantitative parameters of the learning model, will be optimized at the same time. Finally, the trained model is transformed into a quantized model. In this step, the learned quantization parameters are applied to the model weight, and the model weight is transformed into a low-precision representation. A pseudo quantization operation corresponds to a tensor, which corresponds to the quantization parameters of the tensor. The pseudo quantization operation has two functions: one is to simulate the quantization loss, and the other is to introduce the quantization parameters into the calculation diagram, so that the back propagation can optimize the quantization parameters. Therefore, we need to add pseudo quantization operation after all tensors to let them learn the quantization parameters of the tensor. These tensors include each operation weight tensor, input tensor, and output tensor. In convolutional networks, most operations can be reduced to three categories: product, addition, and connection. These three operations will change the value domain distribution of the tensor and affect the quantization parameters of the tensor. In addition, the nonlinear activation function will also affect the value range distribution of the tensor. However, the activation functions used in this paper are ReLU activation functions, which can be regarded as piecewise linear. One difficulty of pseudo quantization is how to reverse broadcast the quantized tensor. There are many undefined points in the function curve of pseudo quantization operation. In this paper, we choose to estimate the derivative directly.

In this paper, we apply quantization during training, aka quantization aware training (QAT), that learns and observes weights and quantization parameters at the same time, which can achieve accuracy closer to the floating-point model. And then, there are different computation graphs during training and inference. In training graph, we simulate quantized convolution operation through the fake-quantization module and fuse the BN layer parameters into the convolution weights to better fit the actual value. The distillation method achieves model compression by transferring the knowledge of the trained large model to a smaller model. So, the higher performance is achieved for this method. Common compression methods usually only demonstrate on large models. However, in order to be able to deploy the CNN on platforms with limited computing resources, choose a small network designed specifically for mobile devices, and then, use the compression technique to achieve better results. We use channel pruning and quantization methods to compress the small detection network and test inference time on actual hardware.

2. Proposed Method

In this section, we discuss the details of the methods and principles used to compress the object detection network. First, we describe the architecture of the proposed detection network and then discuss the algorithm of pruning and quantization methods, respectively.

2.1. Network Architecture. We optimized the recognition network as follows. The target recognition network based on deep learning with good generalization is used to complete the target recognition of airports, bridges, and ports under low resolution. The following introduces the identification of the backbone structure of the network, candidate frame generation in the network, calculation of the network loss function, and training strategies.

First, the basic structure of the remote sensing image target recognition network under low resolution is introduced. The basic structure of the remote sensing target recognition network used in this subject is shown in Figure 1; the basic network structure of the VGG16 is continued on the network backbone structure. The first five layers still use the five convolutional layers of the VGG16 network, discarding the fully connected layers of the sixth and seventh layers of the VGG16 network, while using the astrous

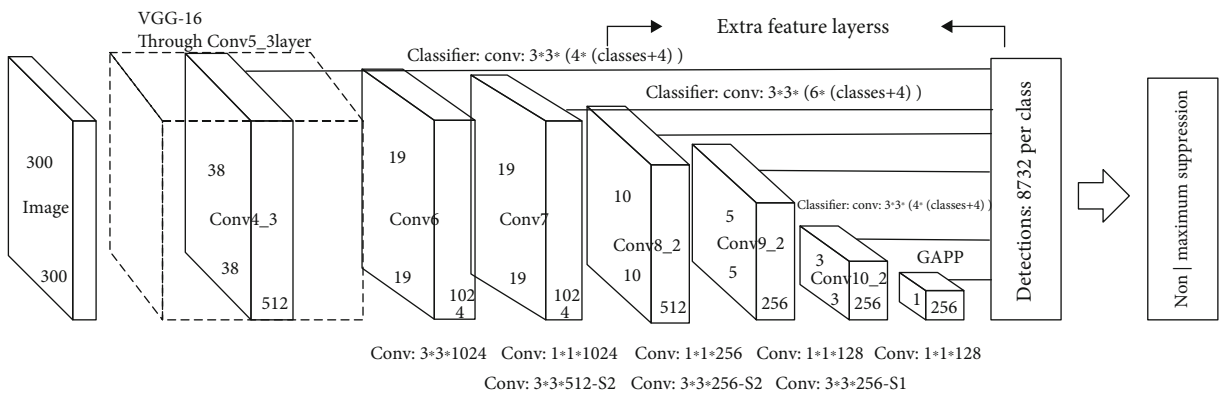


Figure 1: Basic structure of low-resolution remote sensing image recognition network.

algorithm, using the convolutional method to construct two new convolution floor.

The conventional pooling layer in a deep neural network causes a decrease in resolution while increasing the receptive field, and the decrease in resolution loses some feature information. The advantage of this hole convolution is to avoid the decrease in resolution caused by pooling. The comparison between hole convolution and ordinary convolution is shown in Figure 2. It can be seen from the figure that under the same calculation parameters, a larger receptive field can be obtained by using hole convolution instead of ordinary convolution.

After the newly added sixth and seventh convolutional layers, three more convolutional layers (conv8, conv9, and conv10) are added, and layer is added to the network at the end to convert the output feature MAP of the previous layer into a one-dimensional vector. For the remote sensing targets studied in this subject, there is a large intraclass gap for the same type of target, and there is still a problem of scale gap for the same type of target. Therefore, multiscale recognition is particularly important. Considering the scale change of the target object, the network outputs feature MAPs of different scales at different layers and sends them to the detector to predict the degree of confidence and position coordinate offset of each category. As shown in Figure 3, the front-most feature MAP is output after the Conv4_3 layer. The feature MAPs of the first few layers in the network describe the shallower features in the input image, and their receptive fields are relatively small. In contrast, the deeper feature MAPs are responsible for describing the more advanced composite features. Their lower-level feature MAPs of receptive fields are larger, and it also has stronger advanced semantic information. At the end of the network, in order to avoid the result that the same target is detected by the multilayer feature detector at the same time, a nonmaximum suppression process is added, as shown in Figure 3. From this, the final test result is obtained. The network backbone structure does not use a fully connected layer. On the one hand, the output of each layer can only feel the characteristics of the area near the target, not the global information. On the other hand, it also reduces the number of computing parameters in the network.

The architecture of proposed object detection network is shown in Figure 4. We choose MobileNetv2 as our feature extractor. MobileNetv2 is a small network designed specifically for mobile devices, which maintains high accuracy while having a small amount of operations. MobileNetv2 greatly reduces computing operations through the depthwise convolutions and inverse residual structure and is more friendly to CPU. It achieved 72.0% Top-1 accuracy on the ImageNet dataset.

The postprocessing of the network is mainly composed of decoding and nonmaximum suppression (NMS). Decoding converts the outputs of the convolutional layer into predictive value of the bounding boxes. The shape of the tensor is $B \times H \times W \times A \times (C+5)$, where B is the batch size, H and W are the height and width of the feature MAP, A is the number of anchors, and C is the number of categories. The last 5 values of the last dimension are the offsets from the center of the grid point and one foreground probability, denoted as$(\hat{x}_{\min}, \hat{y}_{\min}, \hat{x}_{\max}, \hat{y}_{\max}, p_F)$. Let the grid points of the feature MAP be$P_{i,j} = (x_i, y_j) = (i, j) \quad 0 \le i \le W, 0 \le j \le H$. The relationship between the decoded coordinates and the outputs of the convolution layer will be:

$$\begin{aligned}(x_{\min}, y_{\min}) &= P + \exp\left((\hat{x}_{\min}, \hat{y}_{\min})\right) + 0.5,\\ (x_{\max}, y_{\max}) &= P + \exp\left((\hat{x}_{\max}, \hat{y}_{\max})\right) + 0.5.\end{aligned} \tag{1}$$

$x_{\min}, y_{\min}$ are the elements of the images, and P is the grid points of the feature MAP. We apply NMS on the decoded outputs of detection heads to get the finally bounding boxes. In detection head, we use BCE loss and GIOU loss [22] for classification and box regression.

2.2. Pruning. The channel pruning algorithm can be divided into three steps: sparse training, pruning, and fine-tuning, which is shown in Figure 5. The pruning process can be carried out once or multiple times.

For sparse training, in the training process, L1 norm penalty is applied to γ of the BN layer after the convolutional layer that needs to be sparse. The specific steps of sparse training are as follows. Schematic diagram of channel pruning is shown in Figure 6.

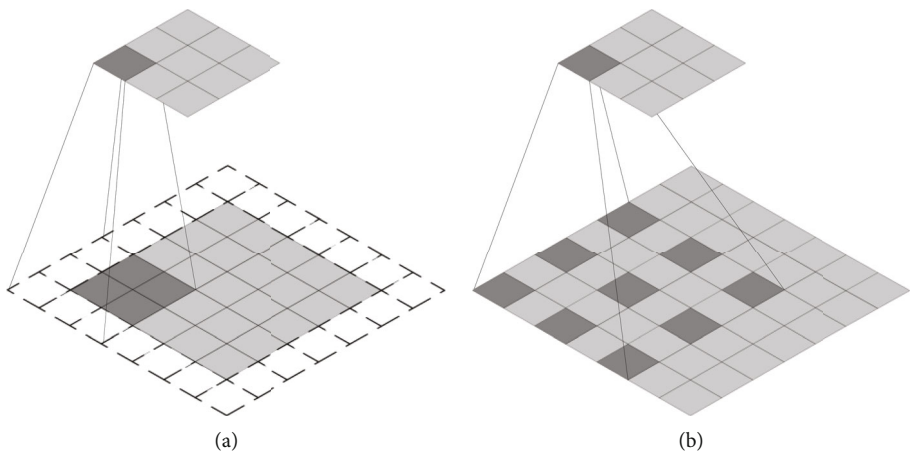


Figure 2: Comparison of ordinary convolution (a) and hole convolution (b).

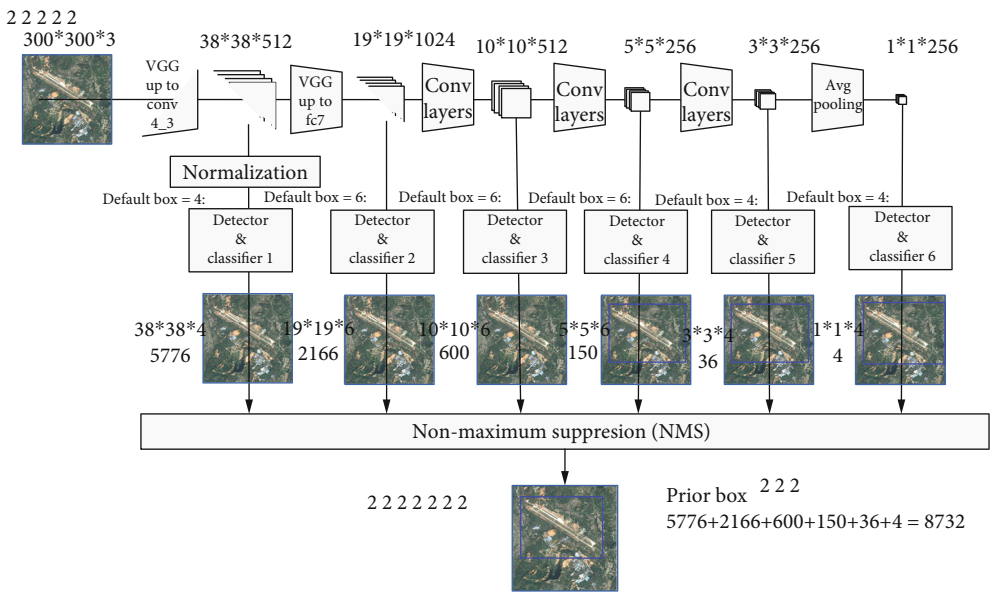


Figure 3: Multiscale detection in the network.

Let the shape of the current layer convolution kernel be $(N_{\text{in}}, N_{\text{out}}, K_h, K_w)$. Then, the convolution kernel after pruning is shown in Equation (2).

$$F_x' = F_x[M_{\text{in}}, M_x, :, :], \tag{2}$$

where $N_{\text{in}}, N_{\text{out}}, K_h, K_w$ represent the number of input channels, the number of output channels, and the convolution kernel height and width, respectively. $A[M_1, M_2]$ means filtering out and restructuring A by M_1, M_2dimension-wise. Colon indicates a mask of all ones.

Finally, set current layer's pruning mask M_x as the next layer's input pruning mask M_{in}.

For depth-wise convolution, the depth-wise convolution can be seen as a convolution with the same number of groups as the number of input channels and output channels. For this type of convolution, it is only necessary to remove the corresponding convolution kernel according to the input pruning mask M_{in}, then set the number of output channels and the number of groups to the new number of input channels.

2.3. Quantization. We have different computation graph during training and inference. The training graph and the inference graph are shown in Figure 7.

In training graph, we simulate quantized convolution operation through the fake-quantization module and fuse the BN layer parameters into the convolution weights to better fit the actual value. We obverse the range of values after each convolution and activation to calculate the quantization parameters. We use 8-bit signed integers to quantize the 32-bit floating-point weights and 8-bit unsigned integers for the activations. Both network input and convolution weights are quantized to 8-bit integers. However, since the biases of convolution are sensitive to errors, 32-bit unsigned integers are used for quantization.

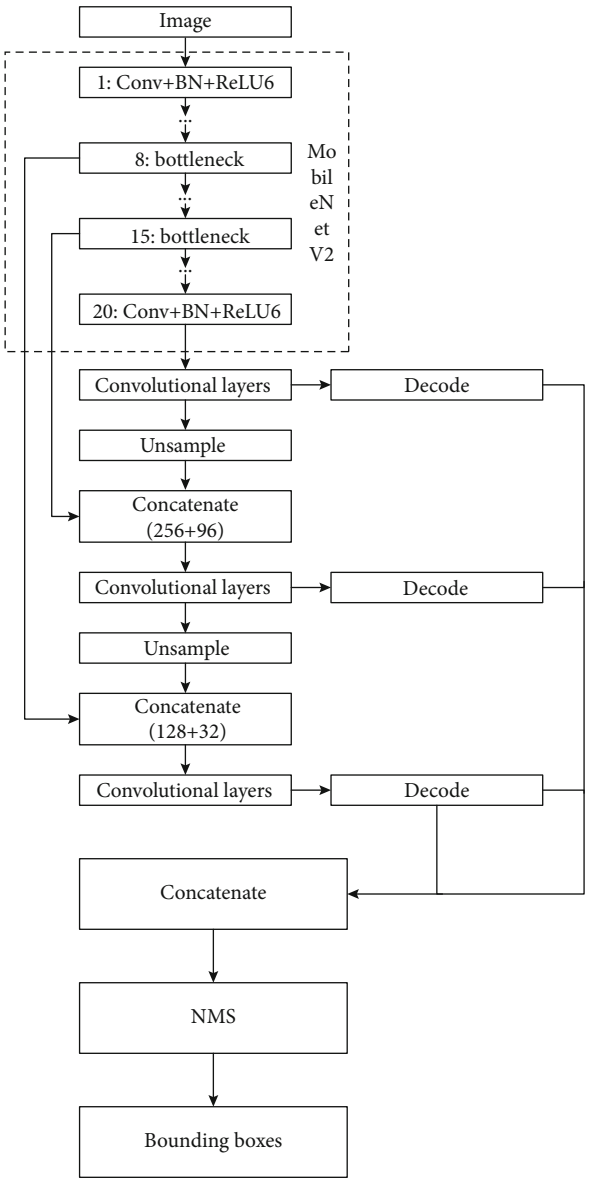


Figure 4: Architecture of the proposed object detection network.

The inference graph only contains quantized convolution operations, since batch-norm layers have been fused into pervious convolution weights.

3. Experiments

This section gives a comparison between proposed network and other state-of-art lightweight detection networks. Then, we show the acceleration of pruning and quantization.

We train our model on the VOC dataset. We set training epoch to 80 for all cases, using standard data augment methods, including random cropping and horizontal flipping, and additionally using mixup. Adopting Adam optimization algorithm, cosine annealing learning rate strategy with warming up, the initial learning rate is 4*e*-3, and the batch size is 16.

The target detection model compression algorithm is evaluated, so the main technical indicators of the algorithm also focus on the performance of the model before and after compression. The technical index shall reflect the compression effect of the compression algorithm on the model performance in terms of computation and memory use. On the server side, the original target detection model is obtained through the target detection training algorithm. Then, the original model is compressed by pruning and quantization through the target detection model compression algorithm. Then, with the help of model deployment tools. The original model and the compressed model are compared in multiple indexes, and then, a number of technical index results are obtained.

According to the needs of the subject, four technical indexes are proposed, such as compression ratio, model calculation, reasoning delay, and accuracy loss, as follows. (1) Compression ratio refers to the ratio of the original model storage space (in bytes) to the compressed model storage space. Compression ratio greater than 6; (2) model calculation quantity refers to the quantity of all multiply accumulate (MAC) operations during forward reasoning of target detection model. The calculation amount of the compressed model is less than 15 MACs; (3) reasoning delay refers to the time consumed from reading the input image to returning the result of forward reasoning in the process of predicting the image by the target detection model. The reasoning delay of the compressed model is less than 50 ms; (4) accuracy loss refers to the difference between the original model and the compressed model in evaluating the mean of average precision (MAP) index on the validation set data set and between the original model and the compressed model MAP. Accuracy loss is less than 5%.

As a result, proposed model got 78.46% of the test set MAP under the input image size of 512×512. FLOPs of our model is 4.25G MACs, and the number of parameters is 6.775M. See Table 1 for comparison with other network in terms of accuracy, MACs, and number of parameters.

The materials and methods section should contain sufficient detail so that all procedures can be repeated. It may be divided into headed subsections if several methods are described.

We have evaluated the versions of Yolo algorithms, and the most advanced Yolo algorithms are YOLOv3 and YOLOv4. YOLOv3 divides the input image into multiple grids on a two-dimensional plane. If the IOU between the detection frame and the predefined anchor frame of a detection target is greater than the threshold, the feature map output vector is responsible for predicting the target. YOLOv3 clusters the detection frames on the data set and divides them into 9 groups according to IOU distance. Using the predefined anchor box, YOLOv3 can assign different targets to different detection heads according to scale and aspect ratio. YOLOv3 uses darknet-53 as a feature extractor. Darknet-53 has the same amount of computation as resnet-50, but its speed and performance exceed resnet-50. YOLOv3 also introduces the idea of feature pyramid network (FPN). After darknet-53, the feature maps of different scales are fused and divided into three detection heads to predict large-, medium-, and small-scale targets, respectively.

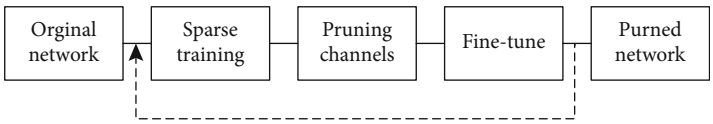


Figure 5: Flow chart of channel pruning.

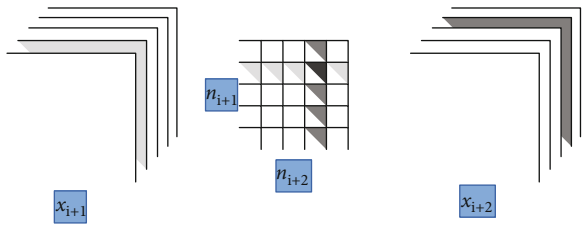


Figure 6: Pruning strategy in general situation.

All model in Table 1 was evaluated on the VOC 2007 test set at the input image size of 512×512. By comparison, the proposed model is close to the YOLOv3 in accuracy, but both MACs and the number of parameters are reduced a lot. Compared with Tiny YOLOv3 and YOLO Nano, we still have great advantages in accuracy and MACs.

The hyperparameter settings of sparse training are the same as previous. In sparse training, we set the sparse rate to 0.01 and reached 75.65% MAP on VOC 2007 test set. After pruning 40% of channels and fine-tuning for 20 epochs, it finally reached 75.44% test set MAP, which decreased by 3.0% compared with the unpruned model. MACs were reduced to 2.606G, and the number of parameters was reduced to 2.31M. Compared with the unpruned model, it was reduced by 38.6% and 65.9%, respectively. The number of channels of each layer before and after pruning is shown in Figure 8. See Table 2 for detailed comparison.

We then apply quantization on 40% pruned model. Training settings are the same as above. We choose 8-bit quantization, freezing BN layer parameters after 10 epochs, and freezing quantization parameters after 15 epochs. The quantized model achieved 76.74% MAP on test set, a decrease of 1.7% from the original model. We benchmark all models on E5-2630 v4 CPU, and the network input is 512×512. The benchmark is performed on the first 100 test set images of VOC dataset. See Table 3 for details.

After pruning and quantization on original model, we achieved 4.76 times acceleration on CPU, and there is only a small loss of accuracy.

We also experiment on RSOD dataset [32]. The RSOD dataset is an open dataset for object detection in remote sensing images. The dataset includes aircraft, oil tank, playground, and overpass. The image sizes range from 500×500 to 2000×1000. We randomly select 80% of the original images as the training set and 20% as the test set. As a result, the training and test sets include 752 images and 187 images, respectively.

During training, we randomly crop a 512×512 patch on original images and then scale it to the size of 320~608. We apply similar augmentation as before but add horizontal flipping. We pad the test images to nearest multiple of 32 and evaluate on original size. The model is trained by 20k iterations in total, and other training hyperparameter settings are as same as before.

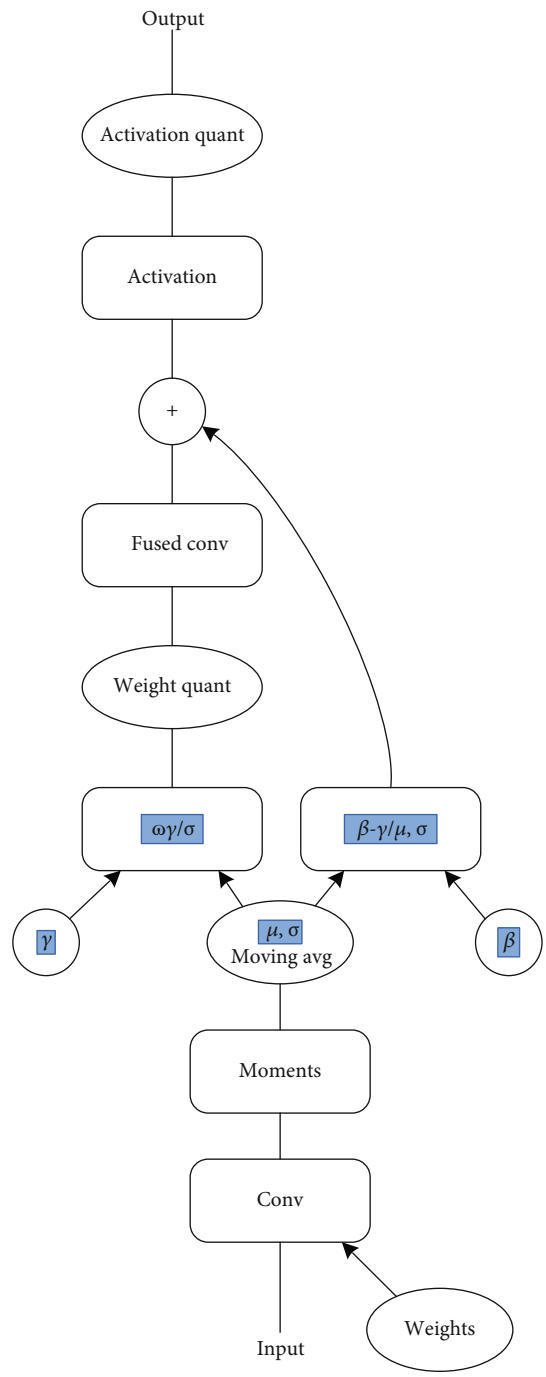


Figure 7: Computation graph of quantization-aware training.

In sparse training, we set sparse ratio to 0.1 and round remaining number of channels to multiple of 8. To avoid

Table 1: Comparison of the accuracy, MACs, and number of parameters.

Name	MAP	MACs	Params
YOLOv3 [20]	79.2%	49.8G	59.25M
Tiny YOLOv3 [20]	61.3%	8.36G	8.5M
YOLO Nano [26]	71.7%	6.92G	1.1M
Ours	78.46%	4.25G	6.775M

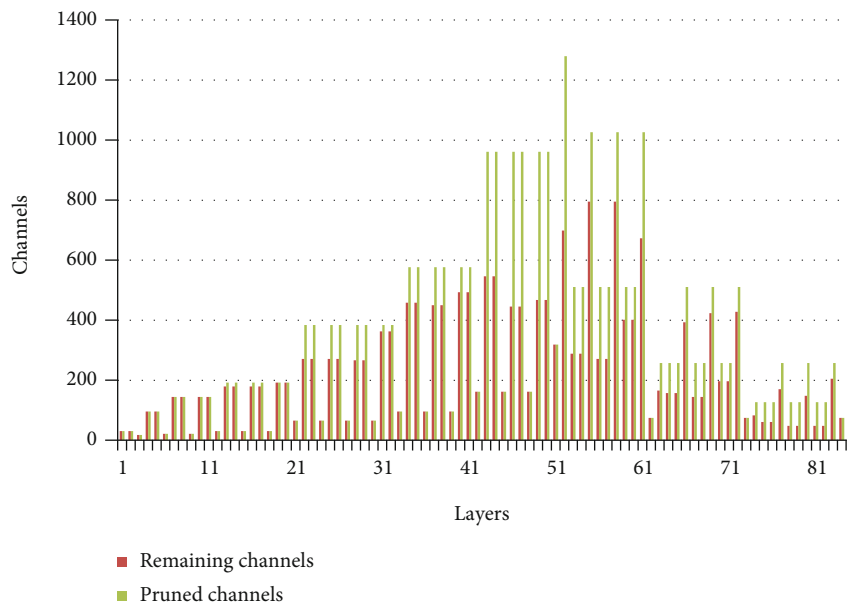


Figure 8: Comparison of the number of channels per layer before and after pruning on VOC dataset.

Table 2: Comparison between original and pruned model on VOC dataset.

Name	MAP	MACs	Params
Original	78.46%	4.25G	6.775M
30% pruned	76.30%	3.195G	3.463M
40% pruned	75.44%	2.606G	2.308M

Table 3: Benchmark for original, pruned, and quantized model on VOC dataset.

Name	MAP	MACs	Params	Inference time	Speedup
Original	78.46%	4.25G	6.775M	126.6 ms	1
40% pruned	75.44%	2.606G	2.308M	102.7 ms	1.23
40% pruned + quantized	76.74%	—	—	26.6 ms	4.76

pruning out all channels, we set the minimum number of channels to 16.

As showed in Table 4, our model reached 92.59% MAP on test set. We found that 85% of pruning on model greatly reduce FLOPs to 1.335G MACs, and the number of parameters is 180.5K, nearly 37x smaller than unpruned model with trivial accuracy drop.

We found that if pruning ratio higher than 85%, MAP drops sharply. The pruning limit is around 0.876 in this case, perhaps that is why.

For comparison, we also draw a figure shown the number of channels with and without pruning (see Figure 9). Note that the layers as main stem of residual blocks make up most of FLOPs.

Finally, we quantize the 85% pruned model with QAT. We benchmark 100 images on CPU with 512×512 input. Table 5 shows comparison between original, pruned, and quantized model. Our final model is 6.31 times faster than original one. The size of final model is only 253 KB.

Table 4: Comparison between original and pruned model on RSOD dataset.

Name	MAP	MACs	Params
Original	92.59%	4.248G	6.689M
30% pruned	91.76%	3.274G	3.506M
60% pruned	91.38%	2.334G	1.085G
80% pruned	90.78%	1.650G	326.3K
85% pruned	90.81%	1.335G	180K
90% pruned	63.53%	1.113G	84.9K

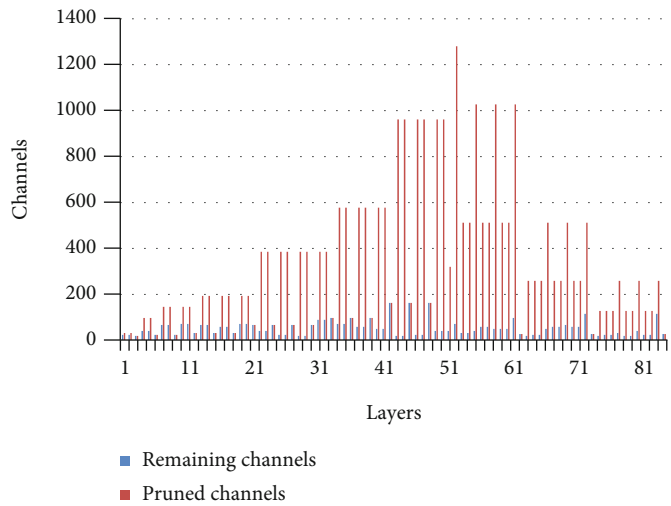


Figure 9: Comparison of the number of channels per layer before and after pruning on RSOD dataset.

Table 5: Benchmark for original, pruned, and quantized model on RSOD dataset.

Name	MAP	MACs	Params	Inference time	Speedup
Original	92.59%	4.25G	6.689M	131.3 ms	1
85% pruned	90.81%	1.335G	180.5M	180.5 ms	2.25
85% pruned + quantized	91.15%	—	—	20.8 ms	6.31

In the future practical application, the compressed algorithm needs to be deployed to the embedded platform for the convolutional neural network target detection system on the embedded platform. Its core technology is the target detection algorithm and network model compression algorithm. It is required to realize target detection based on convolutional neural network at the embedded end. Control the calculation amount of the model below 25 FLOPs. Because the task requires that the target detection algorithm needs to be deployed on the embedded end, it is necessary to select a target detection algorithm that can balance speed and accuracy and is easy to deploy at the same time. Therefore, in future practical applications, attention should be paid to the power consumption of the platform required in practical applications.

4. Conclusions

We propose a lightweight object detection model for hardware platforms with limited computing resources. We achieved this by designing a small network, channel pruning, and quantization. After pruning and quantization, we benchmark proposed model on VOC dataset and RSOD dataset, which achieved 4.76 and 6.31 times acceleration on common CPU with minor accuracy loss. In the future work, the prospects of work and research are listed as follows. In terms of channel pruning, pruning methods can be designed for special network structures such as residual structure, packet convolution, deep separable convolution, and Se module, including sparsity method and pruning channel selection strategy, so as to support target detection networks with more updated structures and achieve higher compression ratio. In the combination of compression methods, more model compression methods can be added, such as distillation and layer pruning. Combining more compression methods can further improve the compression ratio of the compression algorithm.

Data Availability

We have not used specific data from other sources for the simulations of the results. The two popular datasets in this

paper, VOC dataset and RSOD dataset, are free download with the website: https://pjreddie.com/projects/pascal-voc-dataset-mirror/ and https://github.com/RSIA-LIESMARS-WHU/RSOD-Dataset-.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work is supported by stable support project funding Science and Technology on Near-Surface Detection Laboratory with grand no. TCGZ2020C005, the Science and Technology Foundation of National Defense Key Laboratory of Science and Technology on Parallel and Distributed Processing Laboratory (PDL) under Grant No. 6142110180406, the Aeronautical Science Foundation of China under Grant Nos. 2019ZC077006 and 2019-JCJQ-ZD-080-02-00, and the Foundation of Science and Technology on Electro-Optical Information Security Control Laboratory China under Grant No. 6142107200209.

References

[1] G. Evans, "Solving home automation problems using artificial intelligence techniques," *IEEE Transactions on Consumer Electronics*, vol. 37, no. 3, pp. 395–400, 1991.

[2] R. F. Mansour, A. El Amraoui, I. Nouaouri, V. G. Diaz, D. Gupta, and S. Kumar, "Artificial intelligence and Internet of Things enabled disease diagnosis model for smart healthcare systems," *IEEE Access*, vol. 9, pp. 45137–45146, 2021.

[3] Z. Lv, L. Qiao, A. K. Singh, and Q. Wang, "AI-empowered IoT security for smart cities," *ACM Transactions on Internet Technology*, vol. 21, no. 4, pp. 1–21, 2021.

[4] A. K. Yadav and P. Gaur, "AI-based adaptive control and design of autopilot system for nonlinear UAV," *Sadhana*, vol. 39, no. 4, pp. 765–783, 2014.

[5] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, https://arxiv.org/abs/1409.1556.

[6] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3431–3440, 2015.

[7] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 580–587, 2014.

[8] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, 2017.

[9] Z. Liu, J. Li, Z. Shen, H. Gao, S. Yan, and C. Zhang, "Learning efficient convolutional networks through network slimming," in *The IEEE International Conference on Computer Vision (ICCV)*, pp. 2736–2744, 2017.

[10] R. Krishnamoorthi, "Quantizing deep convolutional networks for efficient inference: a whitepaper," 2018, https://arxiv.org/abs/1806.08342.

[11] S. Han, J. Pool, J. Tran, and W. Dally, "Learning both weights and connections for efficient neural network," in *In NIPS*, pp. 1135–1143, 2015.

[12] Z. Chen, Y. Liu, and S. Zhai, *Performance Evaluation of Visual Object Detection for Moving Vehicle*, International CCF Conference on Artificial Intelligence, Springer Singapore, 2019.

[13] W. Cao, J. Yuan, Z. He, Z. Zhang, and Z. He, "Fast deep neural networks with knowledge guided training and predicted regions of interests for real-time video object detection," *IEEE Access*, vol. 6, pp. 8990–8999, 2018.

[14] M. F. B. Ramli, A. Legowo, and S. S. Shamsudin, "Object detection technique for small unmanned aerial vehicle," in *IOP Conference Series: Materials Science and Engineering. IOP Publishing*, 2017.

[15] R. Qin, "An object-based hierarchical method for change detection using unmanned aerial vehicle images," *Remote Sensing*, vol. 6, no. 9, pp. 7911–7932, 2014.

[16] S. Kamate and N. Yilmazer, "Application of object detection and tracking techniques for unmanned aerial vehicles," *Procedia Computer Science*, vol. 61, pp. 436–441, 2015.

[17] D. Jaiswal and P. Kumar, "Real-time implementation of moving object detection in UAV videos using GPUs," *Journal of Real-Time Image Processing*, vol. 17, no. 5, pp. 1301–1317, 2020.

[18] F. Han, Y. Shan, and R. Cekander, "A two-stage approach to people and vehicle detection with HOG-based SVM," *Performance Metrics for Intelligent Systems 2006 Workshop*, pp. 133–140, 2006.

[19] J. Y. Choi, K. S. Sung, and Y. K. Yang, "Multiple vehicles detection and tracking based on scale-invariant feature transform," in *2007 IEEE Intelligent Transportation Systems Conference*, pp. 528–533, IEEE, 2007.

[20] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, 2001.

[21] J. Sokalski, T. P. Breckon, and I. Cowling, "Automatic salient object detection in UAV imagery," in *Proc. 25th International Unmanned Air Vehicle Systems*, 2010.

[22] A. ElTantawy and M. S. Shehata, *Local Null Space Pursuit for Real-Time Moving Object Detection in Aerial Surveillance*, Signal, Image and Video Processing, 2019.

[23] G. V. Konoplich, E. O. Putin, and A. A. Filchenkov, "Application of deep learning to the problem of vehicle detection in UAV images," in *2016 XIX IEEE International Conference on Soft Computing and Measurements (SCM)*, pp. 4–6, IEEE, 2016.

[24] O. Barrero, D. Rojas, and C. Gonzalez, "Weed detection in rice fields using aerial images and neural networks," in *2016 XXI Symposium on Signal Processing, Images and Artificial Vision (STSIVA)*, pp. 1–4, IEEE, 2016.

[25] X. Wang, P. Cheng, and X. Liu, "Fast and accurate, convolutional neural network based approach for object detection from UAV," in *IECON 2018-44th Annual Conference of the IEEE Industrial Electronics Society*, pp. 3171–3175, IEEE, 2018.

[26] T. Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, "Focal loss for dense object detection," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2980–2988, Venice, Italy, October 2017.

[27] Z. Xiao, Q. Liu, G. Tang, and X. Zhai, "Elliptic Fourier transformation-based histograms of oriented gradients for

rotationally invariant object detection in remote-sensing images," *International Journal of Remote Sensing*, vol. 36, no. 2, pp. 618–644, 2015.

[28] Y. Tian, D. Krishnan, and P. Isola, "Contrastive representation distillation," 2019, https://arxiv.org/abs/1910.10699.

[29] N. Zmora, G. Jacob, and L. Zlotnik, "Neural network distiller: a python package for DNN compression research," 2019, https://arxiv.org/abs/1910.12232.

[30] M. Abadi, A. Agarwal, and P. Barham, "TensorFlow: large-scale machine learning on heterogeneous distributed systems," 2016, https://arxiv.org/abs/1603.04467.

[31] A. Paszke, S. Gross, and F. Massa, "Pytorch: an imperative style, high-performance deep learning library," 2019, https://arxiv.org/abs/1912.01703.

[32] A. Robicquet, A. Sadeghian, and A. Alahi, *Learning Social Etiquette: Human Trajectory Understanding in Crowded Scenes*, European conference on computer vision, Springer Cham, 2016.

Hindawi
Wireless Communications and Mobile Computing
Volume 2021, Article ID 4935108, 11 pages
https://doi.org/10.1155/2021/4935108

Research Article

Classification of Markov Encrypted Traffic on Gaussian Mixture Model Constrained Clustering

Junkai Yi,[1] Guanglin Gong,[1] Zeyu Liu,[1] and Yacong Zhang[2]

[1]*College of Automation, Beijing Information Science and Technology University, Beijing 100192, China*
[2]*College of Information Science and Technology, Beijing University of Chemical Technology, Beijing 100015, China*

Correspondence should be addressed to Guanglin Gong; gguanglin@126.com

Received 19 August 2021; Revised 8 September 2021; Accepted 16 September 2021; Published 7 October 2021

Academic Editor: Chi-Hua Chen

In order to solve the problem that traditional analysis approaches of encrypted traffic in encryption transmission of network application only consider the traffic classification in the complete communication process with ignoring traffic classification in the simplified communication process, and there are a lot of duplication problems in application fingerprints during state transition, a new classification approach of encrypted traffic is proposed. The article applies the Gaussian mixture model (GMM) to analyze the length of the message, and the model is established to solve the problem of application fingerprint duplication. The fingerprints with similar lengths of the same application are divided into as few clusters as possible by constrained clustering approach, which speeds up convergence speed and improves the clustering effect. The experimental results show that compared with the other encryption traffic classification approaches, the proposed approach has 11.7%, 19.8%, 6.86%, and 5.36% improvement in TPR, FPR, Precision, and Recall, respectively, and the classification effect of encrypted traffic is significantly improved.

1. Introduction

Network traffic is the data transmitted in the network. Analyzing and monitoring network traffic can allow managers to clearly know the situation of data transmission in the network [1]. In addition to providing references for network control and services, traffic behavior analysis is the basic premise of network security analysis. Therefore, traffic behavior analysis is also called "network visualization." For the needs of transmission protection and user privacy, the encryption ratio of data packets has increased sharply. This change makes traffic analysis and identification very difficult. Before further analysis of network traffic, determining the type of traffic is the basis [2].

Traffic classification is a problem that has been studied very early. In the early reference, most of the traffic was roughly divided into secure shell (SSH), virtual private network (VPN), secure socket layer (SSL), encrypted peer to peer (P2P), voice over internet protocol (VoIP), and other categories [3]. In fact, this level of classification does not have much meaning for further analysis and research in the future. Although traffic classification has been studied very early, due to the characteristics of the traffic itself and the use of encryption technology, traffic classification also faces many problems. First of all, in the network environment, the message exchange between the two parties in the traffic classification feature extraction is a continuous process. The data generated by this process does not have a typical feature description. Moreover, the type, length, and IP address of the message cannot be used directly. Traditional machine learning-based methods cannot analyze network traffic. The second is the scale of network traffic. The specific form of network traffic in the network is data packets. A data packet has a limited size and the limited data it carries. The number of data packets in the network will be very large. Therefore, analyzing and marking network traffic faces a huge challenge.

Encrypted traffic uses a special algorithm to change the original data. Even if it is intercepted during transmission, the content cannot be obtained. Before realizing data transmission in secure socket layer/transport layer security (TLS/SSL), both parties in communication need to go

through cipher and certificate exchange and user data transmission. This process is more complicated, so in the actual process, the communication process is usually simplified to improve communication efficiency. After experiments, the process is divided into three categories:

(1) The first complete communication process generally occurs when the application first applies for communication with the server. The server needs to transmit ciphers and certificates. This process is the most complete and is also the main goal of previous reference behavior analysis

(2) The simplified communication process is used in the session ID reuse phase. This process is suitable for applications to connect again in a short time. The server will reserve resources. When the application requests data, the server will directly obtain the relevant parameters from the reserved resources and will not perform the first type of complete communication. In order to ensure the reliability of the transmission, it will use a ChangeCipherSpec message to transmit the new cipher

(3) The data transmission process is mainly based on the application transmitting user data. Most applications only use transmission control protocol (TCP) message that is used to maintain the connection and Application Data message to complete the task

The existing literatures only consider the first type of situation in the communication process analysis stage and take the message type as the main analysis target. A small amount of literatures consider the relationship between message lengths. The selection of the appropriate analysis object affects the classification efficiency of the classification method.

Encrypted traffic classification methods are constantly changing, and new research results are constantly being produced. Early traffic classification mainly used protocol ports to identify traffic, such as port 21 for file transfer protocol (FTP) and port 80 for hypertext transfer protocol (HTTP). However, many applications currently use the dynamic port negotiation mechanism, which makes the port method no longer applicable. The method based on deep packet inspection (DPI) is considered to be effective and reliable for unencrypted traffic. However, with the widespread use of encryption technology, the amount of encrypted traffic has also increased significantly. Many applications are using protocol encapsulation or obfuscation techniques to circumvent network monitoring. Therefore, DPI-based methods are no longer applicable [4].

At this stage, the neural network is a research hotspot. The combination of encrypted traffic classification and the neural network has become a relatively common method. For example, Liu et al. proposed a novel traffic classification method named High Entropy DistinGuishEr (HEDGE) to distinguish between compressed and encrypted traffic [5]. Ren et al. proposed a tree-structured recurrent neural network (tree-RNN) to classify encrypted traffic, using the tree structure to divide large categories into small categories [6]. Aceto et al. proposed a novel multimodal multitask deep learning approach for traffic classification based on deep learning [7]. However, the above methods are either complicated in feature extraction or difficult in model training.

In 2014, Korczynski and Duda introduced the concept of Markov chain fingerprint recognition for the first time, using hidden Markov models for traffic classification and recognition [8]. They think that the message sequence of communication is a Markov random process, and the current state depends on the previous state. They take advantage of a sequence of message types in the SSL/TLS headers of a given application, which appears in a single-direction flow from a server to a client, to build the first-order Markov chain as a statistical fingerprint for that application. Information embedded in SSL/TLS sessions naturally forms a sequence with time-varying message types, which is analogous to the state transitions in the Markov chain. Therefore, it is reasonable to apply the Markov chain to the construction of application fingerprints. Through Markov modeling the SSL/TLS message sequence, the whole process is divided into three parts: enter probability, transition probability matrix, and exit probability.

Shen et al. [9] believed that there were some shortcomings in the research of Korczynski and Duda [8]. The first is the problem of session ID reuse. The communication between the client and the server will not reestablish the connection within a certain period of time, and there is no complete cipher and certificate exchange process. The second is the problem of the Application Data message. The existing literature only analyzes the encryption process with ignoring the transmission of the Application Data message in the network. From these two points, Shen et al. proposed the classification method of encrypted traffic with second-order Markov chains and application attribute bigrams [9]. However, there are still major limitations in their method: (1) The types of applications are gradually increasing, most of those carry out cipher and certificate interactions in accordance with the protocol, and feature fingerprints have extremely high repetitions. (2) The Certificate message occupies only a small part of the whole message, and most of the message length information is not considered. The length of information of these messages is also an important feature. (3) In the session ID multiplexing stage, the Certificate message appears rarely. Most communication is to keep the connection through TC messages and then send a large number of Application Data messages.

Chen et al. considered that the application had the problem of message duplication that affected the classification effect and proposed a multiattribute-based encrypted traffic classification system named multiattribute associated fingerprint (MAAF) [10]. Liu et al. introduced the concept of length block sequence and proposed a method named multiattribute Markov probability fingerprints (MaMPFs) [11]. However, this method is still based on statistics and belongs to the category of machine learning. It relies heavily on feature selection and usually cannot find exact features. Chen et al. considered the differences among encryption network protocol stacks and proposed a method of encrypted

traffic service classification combining with capsule neural network in a multiprotocol environment by using multiprotocol data unit (PDU) lengths as the features, making full use of Markov property between PDU length sequences [12]. Although the feature extraction time is improved compared with the literature [11], it requires a lot of observation and analysis work in the early stage. Therefore, Yao et al. introduced the GMM and proposed a new traffic classification model based on GMM and hidden Markov models (MGHMMs) [4], which only required fewer features of interpacket time (IPT) and packet size (PS) and calculations to classify traffic.

In order to solve some of the above problems (a summary is presented in Table 1), this paper proposes the method of Classification of Markov Encrypted Traffic on Gaussian Mixture Model Constrained Clustering (MET-GCC). First, the Markov model is established by calculating the initial probability, the completion probability, and the state transition matrix to take fingerprints as a feature of traffic classification. Aiming at the problem of fingerprint duplication and neglect of a large number of messages, an N-gram model [13] based on message length is established. On the basis of the considered Certificate message, the length of other related messages is also taken as an important feature, and the GMM [4] is used to model the packet length. Finally, the method of constrained clustering [14] is proposed, and constrained conditions are added to the clustering parameters to divide the application fingerprints with similar lengths in the same application into a cluster as much as possible and calculate the distribution probability of the application fingerprints.

The MET-GCC method digs out fine-grained features from encrypted traffic, that is, the length of other related messages is also an important feature. The method solves the problem that the existing Markov network traffic classification method only analyzes the traffic classification in the complete stage of communication establishment and ignores the traffic classification in the simplified stage of communication maintenance, and there is a lot of duplication of fingerprints in the network. Through the analysis of the three types of communication processes, the message length is used as the analysis object. The GMM and the constrained clustering method are used to establish the message length model to improve classification efficiency through the distribution probability of fingerprints and realize the classification of traffic in different states.

We briefly summarize our main contributions as follows:

(1) We propose a new type of encrypted traffic classification approach—MET-GCC. By constrained clustering by Gaussian mixture model of message length, the traffic classification of different states is realized. Add constrained conditions to clustering and analyze the probability distribution of packets of different lengths in the same application to improve accuracy. Realize the analysis of the message length through constrained clustering, and improve the classification efficiency by analyzing the distribution of the length of each message

(2) MET-GCC solves not only the traditional Markov model-based network traffic classification method only analyzes the traffic classification in the complete stage of communication establishment with ignoring the traffic classification in the simplified stage of communication maintenance but also duplication problems in application fingerprints during state transition. GMM and clustering methods are introduced on the existing basis to realize the classification of traffic in different states, which effectively improves the classification effect

(3) We compared the classification performance of MET-GCC and related algorithms MCF and SOM. We also compared the classification performance of MET-GCC and the latest Markov and GMM-based MGHMM algorithm on average Precision and Recall. The experimental results show the superiority of the algorithm

The rest of the paper is organized as follows. Section 2 introduces the Markov process and message state transition. Section 3 introduces the derivation of the MET-GCC algorithm in detail. Section 4 verifies the effectiveness of MET-GCC in encrypted traffic classification through a large number of experiments and compares it with traditional and latest classification algorithms. Section 5 gives the conclusion of this article.

2. Markov Process and Message State Transition

There is a certain probability of data packet conversion in the network. This probability is related to the application. Such a random process can be described by the Markov process. The core principle of the Markov process is to include a collection of multiple states. A state transition matrix can describe the transition process. The current state is only related to the previous state and has nothing to do with other states. This is the Markov property. This process of state transition is called the Markov process.

The Markov process is an extremely ideal process, which is a high abstraction of reality, but there are two basic conditions that need to be met:

(1) The state at the current time t is only related to the previous state X_{t-1}, not related to the earlier state

(2) Markov contains at least three parts: state set, transition matrix, and initial state distribution

A set of states of the Markov process is $X_1, X_2, \cdots, X_t$, and the probability of the next moment $(t+1)$ is shown in Equation (1):

$$P(X_{t+1} = x \mid X_1, X_2, \cdots X_t) = P(X_{t+1} = x \mid X_t) = p_{t+1}. \quad (1)$$

The mobile terminal contains a large number of applications and sends a large number of data packets at all times. The acquired features are limited, and these features are not typical features. The classification algorithm in

Table 1: The comparison of existing recent traffic classification algorithms.

Algorithm	Advantages	Disadvantages
HEDGE [5]	High accuracy	Difficult in feature extraction
Tree-RNN [6]	Small classification	Difficult in feature extraction and model training
DISTILLER [7]	Overcome performance limitations of single-modality DL-based TC proposals	Difficult in feature extraction
SOM [9]	Enrich the state transitions in the Markov chain and construct more distinctive application fingerprints	Ignore the state transition of network communication and have duplicate fingerprints
MAAF [10]	Can accurately classify the applications of the same developer	Poor classification to applications with different developers but similar certificates
MaMPF [11]	High accuracy in real networks	Difficult in feature extraction
LS-CapsNet [12]	Solve difficult in feature extraction	High computation overhead
MGHMM [4]	Need fewer features and computational overhead	Constrained by encryption methods

traditional machine learning cannot be used to achieve classification, so the Markov process is a feasible method.

Different types of messages in the network can be defined as the state space in the Markov model. The current state is X_t, and the probability of the next state X_{t+1} is defined as follows:

$$P(X_{t+1} \mid X_t) = p_{t\sim(t+1)}, \tag{2}$$

$$P = \begin{bmatrix} p_{1\sim1} & p_{1\sim2} & \cdots & p_{1\sim n} \\ p_{2\sim1} & p_{2\sim2} & \cdots & p_{2\sim n} \\ \vdots & \vdots & \ddots & \vdots \\ p_{m\sim1} & p_{m\sim2} & \vdots & p_{m\sim n} \end{bmatrix}, \tag{3}$$

where $p_{i\sim j}(i, j \in T)$ is the transition probability.

In order to improve the calculation accuracy, the second-order Markov model can be used:

$$P(X_{t+1} = x \mid X_1, X_2, \cdots, X_t) = P(X_{t+1} = \mid X_t, X_{t-1}) = p_{(t-1)\sim t\sim(t+1)}, \tag{4}$$

$$P = \begin{bmatrix} p_{1\sim1\sim1} & p_{1\sim1\sim2} & \cdots & p_{1\sim1\sim n} \\ p_{1\sim2\sim1} & p_{1\sim2\sim2} & \cdots & p_{1\sim2\sim n} \\ \vdots & \vdots & \ddots & \vdots \\ p_{m\sim m\sim1} & p_{m\sim m\sim2} & \cdots & p_{m\sim m\sim n} \end{bmatrix}. \tag{5}$$

The initial probability is the probability of occurrence of an unknown sequence, so the initial probability (INIP) is defined as shown in Equation (6):

$$\mathrm{INIP} = [\mathrm{ip}_1, \mathrm{ip}_2, \cdots, \mathrm{ip}_m]. \tag{6}$$

INIP represents the probability of a message appearing, and then, an exit probability (EXTP) needs to be defined, as shown in Equation (7).

$$\mathrm{EXTP} = [\mathrm{ep}_1, \mathrm{ep}_2, \cdots, \mathrm{ep}_n]. \tag{7}$$

Assuming that a data packet sequence $\mathrm{seq}_M = <\mathrm{msg}_1, \mathrm{msg}_2, \cdots, \mathrm{msg}_M>$ is captured, the probability that it belongs to a certain application is calculated as shown in Equation (8):

$$P(<\mathrm{msg}_1, \mathrm{msg}_2, \cdots, \mathrm{msg}_M>) = \mathrm{INIP}_{\mathrm{msg}} \times \prod_{i=2}^{l} p(i-1) \sim i \sim (i+1) \times \mathrm{EXTP}_{\mathrm{msg}}, \tag{8}$$

where $\mathrm{INIP}_{\mathrm{msg}}$ is the probability that state msg_1 is the initial state and $\mathrm{EXTP}_{\mathrm{msg}}$ is the probability that state msg_M is the exit state. The transition from the initial state to the exit state is the transition process of the message state.

3. MET-GCC Encrypted Traffic Classification

3.1. Utility Calculations of Application Fingerprints. Obtaining the most typical cipher exchange process in the communication process can more accurately describe the communication process. After a long period of experimental observation, it is found that the number of each type of message in the network is very different. The TCP three-way handshake protocol is widely available in the network, accounting for more than 80% of the total number of messages. Therefore, in addition to the difference in length, it is difficult to conduct behavior analysis through the messages generated by the TCP three times handshake protocol. Although the number of TLS-related messages is about 10%, each time an encrypted communication process is established, the communication between client and server requires a complete cipher and certificate exchange process, which has a greater impact on the analysis of communication behavior. That is, a small number of message types have strong behavior analysis capabilities, so the ability of this message to distinguish data streams is defined as utility, which can be understood as the value of the message itself. It is similar to natural language processing. There are some very frequently used words in natural language, which are used in almost every text, such as "you," "me," and "yes." Although these words have a high frequency of occurrence, they are not helpful for the next step of the algorithm,

because such words are used too often. Therefore, combine with the tf-idf algorithm [15] of evaluating the importance of words in natural language processing to define utility.

Definition 1 (Utility). In the sequence representation $s = <p_l, p_m, \cdots, p_n>$ of an N-gram model, the utility of an item p_m is defined as follows:

$$u(p_m) = \phi\left(n_{p_m}, l\right) \times \varphi\left(n_d, n_d^{p_m}\right), \tag{9}$$

where the number of messages p_m in the sequence s is n_{p_m}, the length of s is l, and the proportion of p_m in the sequence s is $\varphi(n_{p_m}, l)$, which is defined as follows:

$$\phi\left(n_{p_m}, n\right) = \frac{n_{p_m}}{n}. \tag{10}$$

The total number of fingerprints isn_d, $n_d^{p_m}$ is the fingerprint containing message p_m, and $\varphi(n_d, n_d^{p_m})$ is defined as follows:

$$\varphi\left(n_d, n_d^{p_m}\right) = \ln \frac{n_d}{n_d^{p_m} + 1}. \tag{11}$$

Suppose a fingerprint sequence is <11 : 01,11 : 03,11 : 02,23 : 2,11 : 02,23 : 3,23 : 5,23 : 8>, where the utility of the message 23 : 2:

$$u(23 : 2) = \phi(n_{23:2}, n) \times \varphi\left(n_d, n_d^{23:2}\right) = \frac{1}{8} \times \ln \frac{8}{5} = 0.575. \tag{12}$$

Definition 2 (Average utility). The average utility of the sequence $s = <p_l, p_m, \cdots, p_n>$ calculates the average utility of the entire fingerprint based on the utility of the message and is defined as follows:

$$au(s) = \frac{\sum X \in (l, m, \cdots, n) u(p_X)}{l(p_X)}. \tag{13}$$

3.2. Communication State Transition Process. The transmission of network data is a very complicated process that has a large amount of data, many redundancies, and few attributes. The analysis of encrypted traffic has always been a difficult point in research. Literature [9] analyzes the length of the Certificate and Application Data in the message to improve the classification effect and proposes bigram clustering to describe the relationship between them. Figure 1 shows the process of communication state transition described by it.

Figure 1 shows that state X represents the interaction process before Certificate, and state Y represents other types of messages between Certificate and Application Data. The Application Data message only has a difference in length. The probability distribution of this message length is determined by its bigram clustering, and the probability of the Application Data message should also be considered when calculating the probability. However, according to long-term experimental observations, there

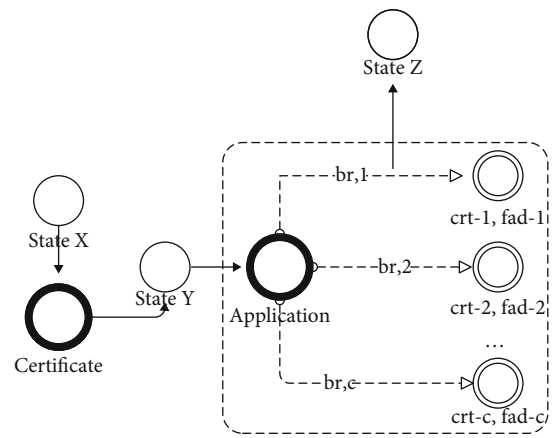


Figure 1: Communication status transition.

are some situations that are not considered in the reference. Figure 2 shows the distribution of the same fingerprint status in network traffic.

Figure 2 shows the conversion between the various states of communication and the distribution of the same fingerprint. state α is the complete process of the above three types of processes, which is juxtaposed with the simplified process of state β, and state γ is the data transmission process. And both state α and state β will convert to state γ. The Certificate message in Figure 2 only appears in state α, and after a large amount of data is transmitted, state α will not occur again for a long time, so Certificate as the core of bigram clustering is difficult to analyze state β and state γ. The other case is to find the fingerprint duplication problem in various states by calculating the average utility of the application fingerprints. As the main object of traffic behavior analysis, state α contains a large number of duplicate fingerprints, which has a greater impact on the classification results.

According to the above problems, this article takes the length of the message as the analysis target, further distinguishes the duplication case between the messages, establishes the N-gram model between the state α, state β, and state γ, and then proposes a message probability distribution model based on Gaussian mixture model constrained clustering to determine the distribution of fingerprints in each state.

3.3. Message N-Gram Model Based on Gaussian Mixture Model Constrained Clustering. The encrypted transmission message has no other valuable features except for the different message length. The length becomes the main feature for analyzing encrypted data. Because of the relative stability of developers and standards, an application generally transmits data in a relatively fixed format. This section analyzes the length-based message N-gram model, which can establish a probability distribution model for messages of different applications and different lengths. Figure 3 is a schematic diagram of the probability distribution model.

In Figure 3, in each state in the communication process, the identical application fingerprint will also have multiple situations, forming the N-gram multivariate model, and each situation will occur with a certain probability. Suppose that in a state, the length of a certain fingerprint contains n

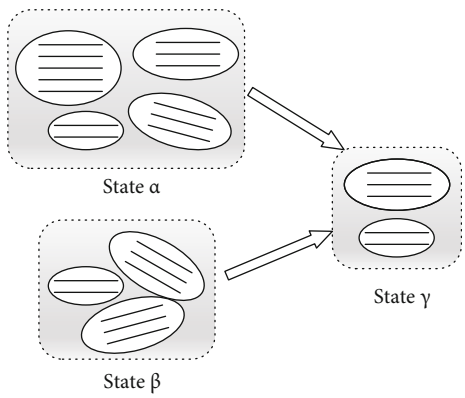


Figure 2: Fingerprint distribution.

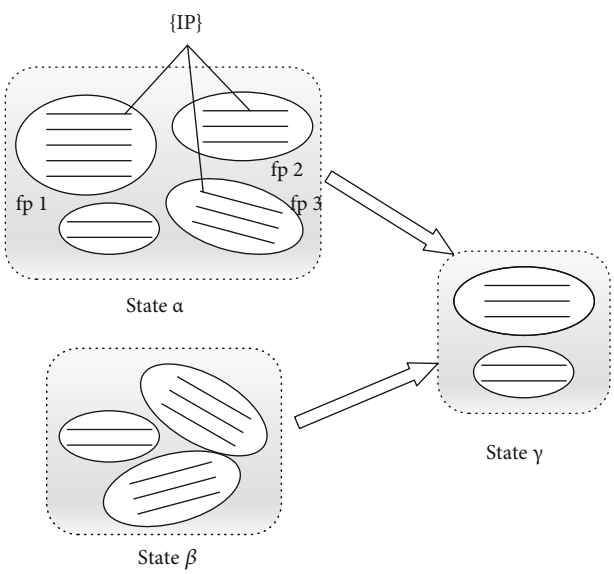


Figure 3: Schematic diagram of probability distribution model.

situations, represented by a vector $\mathrm{fp}^i = (x_1^i, x_2^i, \cdots, x_m^i, \cdots, x_n^i)$, where x_m^i is the length of a message in a fingerprint, so the characteristic fingerprint of an application is an N-gram model and a state has many kinds of N-gram fingerprints $\mathrm{fp}_{\alpha,\beta,\gamma} = (\mathrm{fp}^1, \mathrm{fp}^2, \cdots, \mathrm{fp}^i)$.

$\mathrm{fp}_{\alpha,\beta,\gamma}$ is a combination of multiple distributions whose specific distribution types and parameters are unknown. How to describe $\mathrm{fp}_{\alpha,\beta,\gamma}$ is the key to establishing a model. The GMM is a linear combination of multiple normal distributions. In theory, any kind of unknown distribution can be represented by a linear combination of multiple normal distributions. This is the GMM. Assuming that a certain application fp^i of state $\mathrm{fp}_{\alpha,\beta,\gamma}$ conforms to a normal distribution, then a Gaussian model can be used to describe fp.

$$p\left(x_{\mathrm{fp}} \mid \theta\right) = \sum_{i=1}^{n} a_i p_i\left(x_{\mathrm{fp}}^i \mid \theta_i\right), \tag{14}$$

where a_i is the mixed parameter $a_1 + a_2 + \cdots + a_n = 1$ of each distribution and θ_i is the parameter (μ_i, σ_i) of the normal distribution. Equation (9) can be transformed into the following:

$$p_i\left(x_{\mathrm{fp}}^i \mid \theta_i\right) = f\left(x_{\mathrm{fp}}^i \mid \mu_i, \sigma_i\right) = \frac{\sqrt{\sigma_i}}{\sqrt{2\pi}} e^{-1/2\left(x_{nf}^i - \mu_i\right)^T \sigma_i^{-1}\left(x_{nf}^i - \mu_i\right)}. \tag{15}$$

All the parameters in Equation (15) are $\theta = (\theta_1, \theta_2, \cdots, \theta_n, a_1, a_2, \cdots, a_n)$. Use maximum likelihood estimation to find the parameter θ. Assuming that the collected sample set is $X = (x_1, x_2, \cdots, x_n)$,

$$\log p(X \mid \theta) = \sum_{i=1}^{n} \log p(X_i \mid \theta). \tag{16}$$

Assuming that the estimated parameter value of θ is $\widehat{\theta} = (\widehat{\theta}_1, \widehat{\theta}_2, \cdots, \widehat{\theta}_n, \widehat{a}_1, \widehat{a}_2, \cdots, \widehat{a}_n)$, in order to find $\widehat{\theta}$, $\log p(X \mid \theta)$ needs to be maximized, which is given by Equation (17):

$$\frac{\partial \log p(X \mid \theta)}{\partial \theta} = 0. \tag{17}$$

Obtaining the estimated value $\widehat{\theta} = (\widehat{\theta}_1, \widehat{\theta}_2, \cdots, \widehat{\theta}_n, \widehat{a}_1, \widehat{a}_2, \cdots, \widehat{a}_n)$ of θ, the probability distribution of the same fingerprint type in the state can be obtained:

$$p_{\mathrm{fp}}(X) = \widehat{a}_1 p_1\left(x_{\mathrm{fp}}^1 \mid \widehat{\theta}_1\right) + \widehat{a}_2 p_2\left(x_{\mathrm{fp}}^2 \mid \widehat{\theta}_2\right) + \cdots + \widehat{a}_n p_n\left(x_{\mathrm{fp}}^n \mid \widehat{\theta}_n\right) + \cdots. \tag{18}$$

There are multiple fingerprints in the same state, and each fingerprint is an N-gram model, and the parameters are obtained through maximum likelihood estimation. The calculation of the parameter estimation method shown in Equation (18) is extremely complicated, and the amount of calculation is very large, which is difficult to obtain in the actual process of finding the parameters. In addition, the sample $\mathrm{fp}^i = (x_1^i, x_2^i, \cdots, x_m^i, \cdots, x_n^i)$ describes the length of the message. The same length of messages in the same fingerprint may belong to different applications. The algorithm proposed in this paper requires the probability distribution of fingerprints with messages of different lengths in a fingerprint. The clustering method divides the fingerprints with small differences into a cluster, which is used to calculate the probability distribution of the N-gram model.

The main role of clustering is to separate unlabelled data into discrete sets [16]. The traditional clustering method [17] is to randomly divide several samples into several groups, calculate the distance between each sample and the center of the class by iterative method, and redivide each class. Taking into account the characteristics of network data itself, an IP address will have multiple different fingerprint types, so the fingerprint itself can be divided into different clusters according to the IP address when collecting data. By adding restriction conditions in the parameter calculation, the fingerprints of the same application can be divided into one cluster as much as possible to speed up the

convergence speed and obtain a more accurate state transition matrix. In addition, the number of IP addresses is much larger than the number of classifications, and the obtained data packets are divided into $Y = (y_1, y_2, y_3, \cdots, y_n)$ equivalence sets through clustering.

Assuming that constrained condition (Φ_c) is given to the sample $X = (x_1, x_2, x_3, \cdots, x_n)$ in the process of finding the parameters, there is a class of division $Y = (y_1, y_2, y_3, \cdots, y_n)$, $(x_i \in y_i)$. The sample X can be divided into $X = (x_1, x_2, x_3, \cdots, x_n)$, where $X_i = (x_1^i, x_2^i, x_3^i, \cdots, x_n^i)$ is a subset of X, and the constrained condition is $\Phi_c = \{Y \mid (y_1^i = y_2^i = y_3^i = \cdots = y_n^i)\}$. Therefore, the expected function (θ, θ_y) of the parameters (θ, θ_y) is as follows:

$$F(\theta, \theta_y) = \sum_y \log p(X, Y \mid y \in \Phi_c, \theta) P(y \mid X, y \in \Phi_c, \theta_y), \quad (19)$$

where θ_y is the parameter which is added Φ_c.

Equation (19) is the expected function of the parameter (θ, θ_y), and then, the maximum likelihood is used to estimate the parameter (θ, θ_y). After expanding Equation (19), Equation (20) can be obtained:

$$F(\theta, \theta_y) = \sum_{s=1}^{M} \sum_{l=1}^{K} P(l \mid X_s, y \in \Phi_c, \theta_y) \cdot \left(N_s \log a_l + \sum_{n=1}^{N_s} \log p_l(x_n^s \mid \theta_y) \right) - \sum_{s=1}^{M} \log \sum_{l=1}^{K} (a_l)^{N_s} \Bigg), \quad (20)$$

where $P(l \mid X_s, y \in \Phi_c, \theta_y)$ is the posterior probability of X_s.

Then, it can calculate sequentially the estimated value of the parameter (μ_i, σ_i) of the Gaussian model:

$$\widehat{\mu}_i = \frac{\sum_{s=1}^{M} P(l \mid X_l, y \in \Phi_c, \theta_y) \left(x_1^s + \cdots + x_1^{N_s} \right)}{P(l \mid)X_l, y \in \Phi_c, \theta_y)(N_1 + \cdots + N_M)}, \quad (21)$$

$$\widehat{\sigma}_i = \frac{\sum_{s=1}^{M} P(l \mid X_l, y \in \Phi_c, \theta_y) \left[(x_1^s - \widehat{\mu}_l)(x_1^s - \mu \wedge_l)^T + \cdots + \left(x_{N_S}^s - \widehat{\mu}_l \right) \left(x_{N_S}^s - \mu \wedge_l \right)^T \right]}{P(l \mid X_l, y \in \Phi_c, \theta_y)(N_1 + \cdots + N_M)}. \quad (22)$$

According to $(\widehat{\mu}_l, \widehat{\sigma}_l)$, the probability can be calculated:

$$P(l \mid X_l, y \in \varphi_c, \theta_y) = \frac{|\sigma_l^y|^{N_S/2} e^F}{\sum_{j=1}^{K} |\sigma_l^y|^{N_S/2} e^F}, \quad (23)$$

$$F = \sum_{n=1}^{N_S} \left[-\frac{1}{2} (x_n^s - \mu_l^y) (\sigma_l^y)^{-1} (x_n^s - \mu_l^y) \right] (a_l^y)^{N_l}. \quad (24)$$

Equation (19) requires a very large amount of calculation, where $(x_n^s - \mu_l^y)(\sigma_l^y)^{-1}(x_n^s - \mu_l^y)$ is the Mahalanobis distance [18], the more commonly used Euclidean distance [19] is selected instead, and the cluster with the smallest distance must be selected during the clustering process. Therefore, Equation (23) is simplified to the following:

$$P(l \mid X_s, y \in \Phi_c, \theta_y) = \begin{cases} 1, & \text{if } l = \arg\min \sum_{n=1}^{N_s} \left\| x_n^s - \mu_l^y \right\|^2, \\ 0, & \text{otherwise.} \end{cases} \quad (25)$$

According to Equation (24), the probability distribution of the identical fingerprint can be easily calculated:

$$P_{\text{fp}} = \begin{bmatrix} p_{1,1} & \cdots & p_{1,J} \\ \vdots & \ddots & \vdots \\ p_{K,1} & \cdots & p_{K,J} \end{bmatrix}, \quad (26)$$

where $P_{K,J} = \|X_J\| / \|X_1\| + \|X_2\| + \cdots + \|X_K\|$ is the probability of each application in the identical fingerprint in a cluster.

After the above detailed derivation, we propose the GCC-ETC algorithm. The specific steps are as follows:

According to the algorithm shown, Equation (8) is improved to Equation (27):

$$P(<\text{msg}_1, \text{msg}_2, \cdots, \text{msg}_M>) = \text{INIP}_{\text{msg}} \times \prod_{i=2}^{l} p_{(i-1)\sim i\sim(i+1)} \times \text{EXTP}_{\text{msg}} \times P_{\text{fp}}. \quad (27)$$

4. Experiment and Result Analysis

The dataset required for the experimental test is the network data captured by Wireshark [20] in the actual environment. The device for capturing the experimental data is five

Input: Apply fingerprint.
Step1. Set $\{\mathrm{fp}^1, \mathrm{fp}^2, \cdots, \mathrm{fp}^n\}\{\mathrm{fp}^1, \mathrm{fp}^2, \cdots, \mathrm{fp}^n\}$; Φ_c; K to the equivalent set $X_1, X_2, \cdots, X_K$ according to the restriction Φ_c set by the IP address of the data packet;
Step2. Calculate the average value $\mu_1, \mu_2, \cdots, \mu_K$ of $X_1, X_2, \cdots, X_K$;
Step3. Calculate each sample in $X_1, X_2, \cdots, X_K$ according to $l = \arg\min \sum_{n=1}^{N_s} \|x_n^s - \mu_l^y\|^2$, and generate a new partition $X_1, X_2, \cdots, X_K$;
Step4. Recalculate the mean $\mu_1, \mu_2, \cdots, \mu_K$;
Step5. Repeat step 3 until the sample mean change does not exceed the threshold;
Step6. Recalculate P_{fp}.
Output: Division of application fingerprints $X_1, X_2, \cdots, X_K$.

Algorithm 1

smartphones equipped with Android systems in the lab, and common software is installed on them. According to the classification of each application market, this article installs the 4 most commonly used types of software on smartphones, including video, news, communication, and life, such as QQ, WeChat, email clients, and news clients. After 100 data capture, the average value obtained is used as the experimental dataset which can effectively verify the clustering effect and encrypted traffic classification effect of the algorithm proposed in this paper. The specific dataset is shown in Table 2.

Table 2: Dataset.

Name	Nature category	Time (min)	Number of data packets
Dataset1	Background traffic	15	911
Dataset2	Video	15	335725
Dataset3	News	15	405436
Dataset4	Communication	15	231516
Dataset5	Life	15	241128
Dataset6	Mixed traffic	5	66429
Dataset7	Mixed traffic	10	120447
Dataset8	Mixed traffic	15	181975

Dataset1 is the situation where the smartphone obtains data packets when the user does not run any installed programs. The smartphone sends very few data packets when the smartphone standstill, mainly the push of some messages and the data sent by the operating system itself. Such data is for observing the data used by nonusers and measuring the impact on the data used by users. From the overall view of the above dataset, there are only 911 data packets in 15 minutes, which is very small compared with the mixed traffic collected in 15 minutes. This part of the data situation can be ignored; Dataset2, Dataset3, Dataset4, and Dataset5 are the traffic collection situations that only run related categories of software, and Dataset6, Dataset7, and Dataset8 are mixed traffic, which is the traffic collected by running all applications. The time is 5 min, 10 min, and 15 min.

The experiment verifies the effect of the MET-GCC algorithm proposed above, which is divided into two parts. The first part analyzes the clustering effect of constrained clustering on the length of the data packet. This article adopts the traffic classification based on the Markov process. When calculating the category to which a segment of the data stream belongs, if the calculated probability of each category is similar, the classification effect will be poor, and the opposite is better. Now analyzing how to measure the clustering effect of constrained clustering according to length, the experiment considers the two extreme situations. The first extreme situation is the most ideal situation, and all packets of an application are similar in length so that all messages will be clustered in the same cluster during clustering. In this way, getting a message again of the same length can quickly determine its category. The second extreme situation is the worst situation. The message length of an application is relatively scattered. Messages of various lengths will be grouped into different clusters during clustering so that the probabilities of various categories are similar in the calculation. It is difficult to make a judgment. Therefore, from the analysis of these two extreme situations, the evaluation standard is that the same application message should be in one cluster as much as possible, and only one application should be included in one cluster. Of course, this is the most ideal situation, and it cannot exist in practice, so the experiment defines this evaluation criterion as clustering coefficient (CL-CO) [21], as shown in Equation (28).

$$\text{CL-CO} = \frac{1}{k}\left(\sum_i \prod_k \frac{N^i}{N_k^i} + \sum_k \prod_i \frac{N^i}{N_k^i}\right), \tag{28}$$

where $\sum_i \prod_k N^i/N_k^i$ represents the distribution of the same application in a cluster and $\sum_k \prod_i N^i/N_k^i$ represents the distribution of the message length of each application in the same cluster.

The message length is fixed, and other settings are the same; the CL-CO is related to the distribution of the message length, and the CL-CO is also related to the clustering parameter selection K. The number of clusters K is their input parameter, and then, generate the vector of all cluster centers as the output. In fact, however, we are unable to determine an appropriate value of K in advance, because different applications have a high degree of overlap in packet length [9]. Therefore, enumerate K from 1 to a relatively large number (i.e., the largest K) and use the clustering method to calculate the cluster center vector for each specific

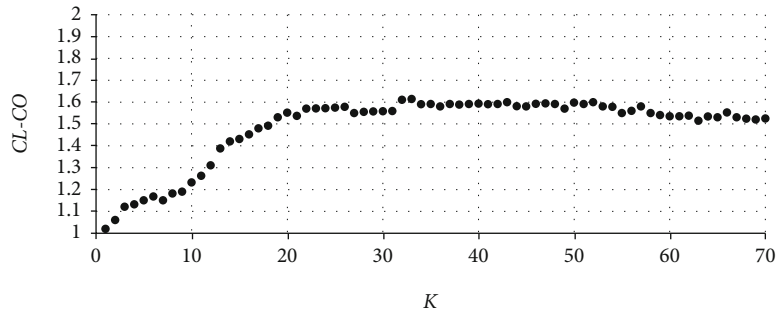


Figure 4: Clustering coefficient (CL-CO).

K. In order to compare all candidate values of K, the aggregation accuracy metric criterion CL-CO is used, and the optimal K value is determined by the value of CL-CO. The mixed traffic Dataset8 is used as the dataset for this analysis because the dataset has a large number of data packets and a long collection time. The distribution of network data packets is relatively stable, and the specific results are shown in Figure 4.

In the most ideal and worst conditions, the CL-CO is infinitely close to 0. Different collected network traffic determines the optimal K value. Therefore, the choice of K value is related to the classification effect. When the value of K is about 33, the CL-CO reaches the best. If the experiment continues to increase the K value, the CL-CO will slowly decrease.

The next experiment analyzes the effect of encrypted traffic classification. It is compared with the MCF algorithm of literature [8] and the SOM algorithm of literature [9]. These two methods are based on the traffic classification method proposed by the Markov model. The difference is that the MCF algorithm only considers the communication establishment phase. SOM is an improvement on the MCF algorithm. On this basis, the impact of different lengths of the Certificate message on the classification effect is considered. Two commonly used values are used as the evaluation criteria for the classification effect, as shown in Equations (29) and (30). TPR represents the current flow is classified into the positive sample category, and the practical positive sample accounts for the proportion of all positive samples; FPR represents the current traffic is incorrectly classified into the positive sample category and the proportion of practical negative samples to the total number of all negative samples.

$$\mathrm{TPR} = \frac{\mathrm{TP}}{(\mathrm{TP} + \mathrm{FN})}, \tag{29}$$

$$\mathrm{FPR} = \frac{\mathrm{FP}}{(\mathrm{FP} + \mathrm{TN})}, \tag{30}$$

where TP is the true positive, which means that traffic belonging to the positive sample category is classified as a positive sample category; FN is the false negative, which means that traffic belonging to the positive sample category is classified as a negative sample category; FP is the false positive, which means negative. The traffic of the sample category is classified as a positive sample category; TN is the true negative, which means that the traffic of the negative sample category is classified into a negative sample category.

Table 3: MCF, SOM, and MET-GCC algorithm comparison.

Application	MCF		SOM		MET-GCC	
	TPR	FPR	TPR	FPR	TPR	FPR
Video	0.59	0.36	0.79	0.29	0.85	0.21
News	0.44	0.49	0.71	0.21	0.80	0.22
Communication	0.71	0.30	0.73	0.28	0.79	0.17
Life	0.55	0.32	0.76	0.33	0.90	0.29

Table 4: MGHMM and MET-GCC algorithm comparison.

Application	MGHMM		MET-GCC	
	Precision	Recall	Precision	Recall
Video	0.79	0.82	0.96	0.85
News	0.99	0.98	0.89	0.80
Communication	0.87	0.75	0.91	0.79
Life	0.85	0.62	0.98	0.90

The experiment here also uses Dataset8 as the tested dataset and selects the relatively better coefficient $K = 33$ for the CL-CO. First, calculate the classification situation of each category application, and the results are shown in the following Table 3.

From the analysis of the classification results in Table 3, it can be seen that the classification effect of the MCF algorithm is not very good, because the MCF algorithm does not consider the communication behavior in the communication maintenance phase. The classification performance of the SOM algorithm has been improved to a certain extent because the SOM algorithm considers the Certificate type of message in the communication maintenance phase. The MET-GCC algorithm is obviously better than these two algorithms because it also analyzes the length of other messages as an important feature in addition to Certificate messages. From the perspective of various classification situations, the more single a category of application traffic behavior is, the better the classification effect is, and the effect of video traffic classification is generally better. The reason is that it mainly sends videos, and the size and format of the video are relatively stable, while the news category contains data of multiple categories such as text and video, which has a certain impact on the classification.

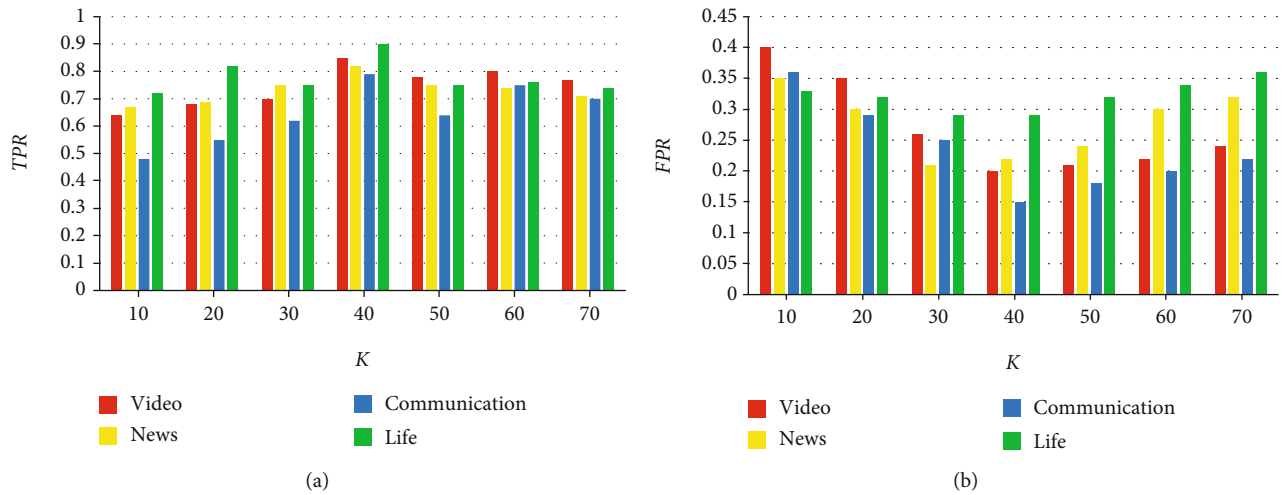


Figure 5: Comparison of TPR and FPR of different K.

Then, compared with the latest MGHMM algorithm based on Markov and GMM of literature [4], the HGHMM algorithm and MET-GCC algorithm both use Markov and GMM traffic classification methods. The difference is that the MGHMM algorithm is based on the analysis of the two traffic characteristics of IPT and PS, and the MET-GCC algorithm is based on the analysis of the message length. At the same time, add restrictions when clustering. It also uses two common values as the evaluation criteria for the classification effect, as shown in Equations (31) and (32). Precision and Recall, respectively, represent among all traffic classified as positive samples, the proportion of real positive sample category and the current traffic is classified into the positive sample category, and the practical positive sample accounts for the proportion of all positive samples.

$$\text{Precision} = \frac{\text{TP}}{(\text{TP} + \text{FP})}, \tag{31}$$

$$\text{Recall} = \frac{\text{TP}}{(\text{TP} + \text{FN})}, \tag{32}$$

where the meaning of TP, FP, and FN is the same as Equations (29) and (30).

Select the same dataset and aggregation coefficient to compare the classification conditions, as shown in Table 4.

Through comparison, it is found that the proposed MET-GCC algorithm has obvious advantages in traffic classification. The average Precision and Recall of classification have increased by 6.86% and 5.36%, respectively. It can be seen from Table 4 that the MGHMM algorithm has outstanding effects in classifying communication and news traffic, especially news. This is because the MGHMM algorithm is based on the analysis of the two traffic characteristics of IPT and PS. Communication and news traffic adopt methods like transmission interval interference and message filling, which have a greater impact on IPT and PS, and it is easy to extract features and facilitate classification. However, when a regularization method that has a small impact on IPT and PS is used to shape the traffic content, the classification effect of the algorithm is significantly reduced. The MET-GCC algorithm uses the analysis of the length of the message, and the classification will not be affected by the difference in the method. As a result, it can be seen that the MGHMM algorithm has its advantages, but the disadvantages are also obvious. The classification effect fluctuates greatly and is restricted by the method. The MET-GCC algorithm has a stable classification effect for various types of traffic, and the average classification effect is also better than the MGHMM algorithm. MET-GCC algorithm is significantly better.

In addition to the above analysis of the classification of each algorithm, the following experiment analyzes the influence of the coefficient K on the classification. From the above experiment, it can be seen that different K values have different CL-CO, and the clustering effect is also different. The experiment still uses Dataset8 to analyze the TPR and FPR for different K values. The results of the analysis are shown in Figure 5.

From Figure 5, the choice of K value in the clustering algorithm has a greater impact on the results of the experiment. Choosing an appropriate K value can increase TPR and reduce FPR. Choosing a K value that is too small or too large will affect the results of the classification experiment. Therefore, it is necessary to select a K value that guarantees a large TPR and a small FPR. It can be seen from the results that when the CL-CO is relatively high, the relative classification effect will also be improved. When $K = 40$, TPR and FPR have the best effects.

5. Conclusions

The paper proposes a new type of encrypted traffic classification algorithm—Classification of Markov Encrypted Traffic on Gaussian Mixture Model Constrained Clustering (MET-GCC). Based on the Markov model, it is through constraining clustering for the GMM of the message length to realize the traffic classification of different states. MET-GCC solves the problem that the traditional algorithm does not consider the state transition of network communication, and the fingerprint is a duplicate. The experiment proves the

effectiveness of the MET-GCC algorithm and reveals that the MET-GCC algorithm performs better than the latest encryption traffic classification algorithm based on Markov and GMM. In the future work, we plan to solve the impact of traffic complexity on the classification effect, so as to improve the classification accuracy of the MET-GCC algorithm for various types of traffic.

Data Availability

The data used to support the finding of this study are included in the article.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (U1636208).

References

[1] F. Xiao, L. Chen, H. Zhu, R. Hong, and R. Wang, "Anomaly-tolerant network traffic estimation via noise-immune temporal matrix completion model," *IEEE Journal on Selected Areas in Communications*, vol. 37, no. 6, pp. 1192–1204, 2019.

[2] A. D'Alconzo, I. Drago, A. Morichetta, M. Mellia, and P. Casas, "A survey on big data for network traffic monitoring and analysis," *IEEE Transactions on Network and Service Management*, vol. 16, no. 3, pp. 800–813, 2019.

[3] Z. Cao, G. Xiong, Y. Zhao, Z. Li, and L. Guo, "A Survey on Encrypted Traffic Classification," in *2014 International Conference on Applications and Techniques in Information Security*, pp. 73–81, Springer, Berlin, Heidelberg, 2014.

[4] Z. J. Yao, J. G. Ge, Y. L. Wu, X. Lin, R. He, and Y. Ma, "Encrypted traffic classification based on Gaussian mixture models and hidden Markov models," *Journal of Network and Computer Applications*, vol. 166, article 102711, 2020.

[5] C. Liu, L. He, G. Xiong, Z. Cao, and Z. Li, "Fs-net: a flow sequence network for encrypted traffic classification," in *IEEE INFOCOM 2019 - IEEE Conference on Computer Communications*, pp. 1171–1179, Paris, France, April 2019.

[6] X. M. Ren, H. Gu, and W. Wei, "Tree-RNN: tree structural recurrent neural network for network traffic classification," *Expert Systems with Applications*, vol. 167, article 114363, 2021.

[7] G. Aceto, D. Ciuonzo, A. Montieri, and A. Pescapé, "DISTILLER: encrypted traffic classification via multimodal multi-task deep learning," *Journal of Network and Computer Applications*, vol. 183-184, article 102985, 2021.

[8] M. Korczynski and A. Duda, "Markov chain fingerprinting to classify encrypted traffic," in *IEEE INFOCOM 2014 - IEEE Conference on Computer Communications*, pp. 781–789, Toronto, Canada, April 2014.

[9] M. Shen, M. Wei, L. Zhu, and M. Wang, "Classification of encrypted traffic with second-order Markov chains and application attribute bigrams," *IEEE transactions on Information Forensics and Security*, vol. 12, no. 8, pp. 1830–1843, 2017.

[10] Y. Chen, T. Zang, Y. Zhang, Y. Zhou, and Y. Wang, "Rethinking encrypted traffic classification: a multi-attribute associated fingerprint approach," in *2019 IEEE 27th International Conference on Network Protocols (ICNP)*, pp. 1–11, Chicago, IL, USA, October 2019.

[11] C. Liu, Z. Cao, G. Xiong, G. Gou, S.-M. Yiu, and L. He, "Mampf: encrypted traffic classification based on multi-attribute Markov probability fingerprints," in *2018 IEEE/ACM 26th International Symposium on Quality of Service (IWQoS)*, pp. 1–10, Banff, AB, Canada, June 2018.

[12] Z. Chen, G. Cheng, B. Jiang, S. Tang, S. Guo, and Y. Zhou, "Length matters: fast internet encrypted traffic service classification based on multi-PDU lengths," in *2020 16th International Conference on Mobility, Sensing and Networking (MSN)*, pp. 531–538, Tokyo, Japan, December 2020.

[13] I. Lopez-Gazpio, M. Maritxalar, M. Lapata, and E. Agirre, "Word n-gram attention models for sentence similarity and inference," *Expert Systems with Applications*, vol. 132, pp. 1–11, 2019.

[14] X. Li, R. Zhang, Q. Wang, and H. Zhang, "Autoencoder constrained clustering with adaptive neighbors," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 1, pp. 443–449, 2021.

[15] I. Yahav, O. Shehory, and D. Schwartz, "Comments mining with TF-IDF: the inherent bias and its removal," *IEEE Transactions on Knowledge and Data Engineering*, vol. 31, no. 3, pp. 437–450, 2019.

[16] C. L. Chowdhary and D. P. Acharjya, "Segmentation of mammograms using a novel intuitionistic possibilistic fuzzy c-mean clustering algorithm," in *Nature Inspired Computing*, pp. 75–82, Springer, Singapore, 2018.

[17] M. J. Gómez-Silva, A. de la Escalera, and J. M. Armingol, "Back-propagation of the Mahalanobis istance through a deep triplet learning model for person re-identification," *Integrated Computer-Aided Engineering*, vol. 28, no. 3, pp. 277–294, 2021.

[18] R. M. Alguliyev, R. M. Aliguliyev, and L. V. Sukhostat, "Efficient algorithm for big data clustering on single machine," *CAAI Transactions on Intelligence Technology*, vol. 5, no. 1, pp. 9–14, 2020.

[19] L. Morin, P. Gilormini, and K. Derrien, "Generalized Euclidean distances for elasticity tensors," *Journal of Elasticity*, vol. 138, no. 2, pp. 221–232, 2020.

[20] P. Narwal, D. Kumar, and S. N. Singh, "A hidden Markov model combined with Markov games for intrusion detection in cloud," *Journal of Cases on Information Technology*, vol. 21, no. 4, pp. 14–26, 2019.

[21] P. Gracar, A. Grauer, L. Lüchtrath, and P. Mörters, "The age-dependent random connection model," *Queueing Systems*, vol. 93, no. 3-4, pp. 309–331, 2020.

Hindawi
Wireless Communications and Mobile Computing
Volume 2023, Article ID 9794634, 1 page
https://doi.org/10.1155/2023/9794634

Retraction

Retracted: Research on Sports Class Load Monitoring System Based on Threshold Classification Algorithm

Wireless Communications and Mobile Computing

Received 11 July 2023; Accepted 11 July 2023; Published 12 July 2023

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. Zhao, "Research on Sports Class Load Monitoring System Based on Threshold Classification Algorithm," *Wireless Communications and Mobile Computing*, vol. 2021, Article ID 3891453, 10 pages, 2021.

Hindawi
Wireless Communications and Mobile Computing
Volume 2021, Article ID 3891453, 10 pages
https://doi.org/10.1155/2021/3891453

Research Article

Research on Sports Class Load Monitoring System Based on Threshold Classification Algorithm

Lin Zhao

Luohe Vocational Technology College, Luohe 462002, China

Correspondence should be addressed to Lin Zhao; zhaolin123@m.fafu.edu.cn

Received 29 June 2021; Revised 24 July 2021; Accepted 2 September 2021; Published 16 September 2021

Academic Editor: Chi-Hua Chen

In order to reduce the sports injury caused by high intensity sports classes, it is necessary to monitor the state of the sports load. Therefore, the sport's load monitoring system based on a threshold classification algorithm is proposed. In this paper, we design the hardware and software structures of the sports load monitoring systems in a physical education class. In this system, the state parameters of the sports load are collected by wireless sensor network nodes, and the feature parameters are fused and clustered by the integrated information fusion method. After that, we establish the movement target image acquisition model, which unifies the ZigBee networking realization to the high intensity sports classroom movement load monitoring. Simulation results show that the designed PE classroom sports load monitoring system based on the threshold classification algorithm has high performance for sports parameter monitoring and can effectively avoid sports injury caused by overload.

1. Introduction

As people pay more and more attention to their physical health, they begin to take intensive physical exercise gradually to improve their physical fitness [1]. But the nonprofessional athlete sometimes cannot grasp well the intensity of the movement load. Therefore, it is necessary to study an effective monitoring system of the sports class exercise load [2, 3]. It is of great significance to study the design method of the monitoring system of the sports load state in a PE class for promoting sports training and health monitoring.

Reference [4] proposes to monitor the athletic performance of male athletes and to measure the change of athletic performance in 1 year after arthroscopic treatment of femoral and acetabular impact injuries. Methods male athletes with the arthroscopic treatment of FAI were tested in preoperative and postoperative 1 year, including acceleration, steering speed, squatting depth, and reaction intensity index. Compared with the healthy athletes, the FAI with symptoms makes the athletes' performance drop significantly and puts the athletes in an obvious sports disadvantage. Reference [5] proposes a method for monitoring adolescent athletic professionalism in a hospital-based outpatient department of pediatric sports medicine. Patients or other participants or athletes injured between the ages of 12 and 17 who visited clinics between 2015 and 2017 and who completed a sports participation survey were invited to participate. The methods mentioned above have the problems of poor anti-interference and low output stability in the monitoring of the sports class load. In order to improve the stability and reliability of the monitoring output, this paper puts forward the monitoring system of the sports class load status based on the threshold classification algorithm, which is based on the collection and information extraction of the parameters of the sports class load status, and combines the optimized information processing technology and the big data information processing technology to monitor the sports class load status and improve the stability and reliability of the monitoring output.

In Reference [6], the authors proposed a multilabel classification based on a random forest algorithm for a nonintrusive load monitoring system. This paper proposes a multilabel classification method using random forest (RF) as a learning algorithm for nonintrusive load identification. Multilabel classification can be used to determine which categories data belong to. This classification can help to identify

the operation states of independent loads from mixed signals without disaggregation. The experiments are conducted in a real environment and public data set. Several basic electrical features are selected as the classification feature to build the classification model. These features are also compared to select the most suitable features for classification by feature importance parameters. The classification accuracy and F-score of the proposed method can reach 0.97 and 0.98, respectively. In Reference [7], the authors proposed a load identification and classification in a nonintrusive load monitoring system based on the data stream. In order to enhance the classification accuracy and reliability of load detection in a nonintrusive load monitoring systems, novel approaches for general load switch detection were proposed. Based on the improved CUSUM algorithm which can determine the start and ending point of the transient process, the transient energy algorithm and correlation method was applied. Further, the comparison results of different ending points show the different differentiation degrees. The information fusion as a criterion was added, and the data fusion error rate threshold was determined to enhance reliability. After that, the correlation method was applied to verify the results. In the case study simulation, the effectiveness of this novel system was proven. In Reference [8], the authors have given the research on sports retrieval recognition of action based on feature extraction and the SVM classification algorithm. The feature extraction speed of the traditional athlete motion retrieval algorithm is slow, and it often takes dozens of minutes or even hours to analyze a video. The speed of this feature extraction obviously cannot meet the needs of big data video analysis. In response to these two problems exposed by Action Bank under large-scale data, this paper proposes to apply the template learning method based on spectral clustering to Action Bank, which replaces the cumbersome manual selection template step and is easy to generalize to different databases. Moreover, in view of the disadvantage of the slow speed of extracting Action Bank features, this paper proposes a fast algorithm for accumulating Action Bank. In addition, this study uses the lookup table method instead of the time-consuming steps of the correlation distance calculation in template matching, which greatly accelerates the time of feature extraction. Finally, this study designed experiments to analyze the performance of the algorithm.

In this paper, we use the threshold classification algorithm to study the sports class load monitoring system. The threshold classification algorithm classifies according to the preset threshold. Those within the set threshold range belong to one category, while those not within the threshold range are divided into another category. The main idea of this method is to divide the individuals within a certain threshold into a class, and the individuals outside that threshold into another class. Compared with other classification algorithms, this algorithm has the advantages of low time complexity and computational complexity.

The main contribution of this paper can be described as follows:

(1) The author studies a new sports class load monitoring system. In recent years, motion detection has been paid more and more attention by researchers, but at present, the research in this field is still in its infancy, and there are little research results

(2) In this paper, a new motion state classification method is proposed, which is called the threshold classification algorithm. Threshold classification algorithm has low computational complexity and better classification effect than other algorithms, so this algorithm can be well applied to motion state monitoring

The structure of the rest of this paper is as follows: Section 2 gives the introduction of the design of a sports load monitoring system for a physical education class. Section 3 gives the object image monitoring model of the sports load state in a physical education class and fusion processing of monitoring sensing information. Section 4 gives the experimental analysis. The conclusion and prospect are given in Section 5.

2. Design of Sports Load Monitoring System for Physical Education Class

2.1. System Hardware Architecture. The hardware of the sports load monitoring system is mainly composed of two parts: the first part is the positioning label and photoelectric heart rate meter. The positioning tag is powered by the battery and can transmit an ultrawideband signal with a frequency of 10 Hz. It is worn as a wristband [9, 10]. The photoelectric heart rate meter must be worn in contact with the skin, in the form of an arm band, and separately worn from the positioning label. The second part is the base station, which is fixed with known coordinates and can receive the UWB signal sent from the positioning tag. Base stations are connected by wire to form a LAN [11]. The parameters of each piece of hardware are shown in Table 1, and the appearance of each piece of hardware is shown in Figure 1.

2.2. System Software Development. The sports load monitoring system software of the physical education classroom is mainly composed of four layers: perception layer, transport layer, service layer, and display layer. The sensing layer is composed of a positioning base station, a positioning tag, and an optoelectronic heart rate meter and is mainly used for automatic data acquisition. When the positioning label enters the coverage of the base station, it automatically registers into the network, determines the configuration parameters of the positioning label, and uploads the relevant information in real time. The transport layer is the transmission channel between the location label, base station, server, and external network, which is transmitted wirelessly. The service layer is composed of a threshold classification algorithm positioning engine, interface software, and system computing software. The display layer is displayed by the service layer through the calculation of a real-time display location label and location information. The display layer displays primarily on the mobile or PC side [12].

Based on the threshold classification algorithm, the monitoring system software of the sports class load includes

Table 1: System hardware parameter list.

Label	Describe	Base station	Describe	Heart rate meter	Describe
Temperature range	-40°C-70°C	Working temperature	-40°C-70°C	Transmission mode	Low power consumption of Bluetooth
Renewal rate	0.01-200 Hz	Antenna gain	12.5 dBi	Software	Wireless upgrade
Frequency range	6.35-6.75 GHz	Frequency range	6.35-6.75 GHz	Frequency range	1 Hz
Power supply	Battery powered	Power supply	External power supply	Power supply	Battery powered
Protection level	IP67	Protection level	IP40	Protection level	30 m underwater
Transmission distance	200 cm	Transmission distance	100 m	Usage time	12 h
Size	7 × 4 × 1 cm	Size	35 × 25 × 8 cm	Size	2 × 2 × 1 cm
Weight	18 g	Weight	1.64 kg	Weight	15 g

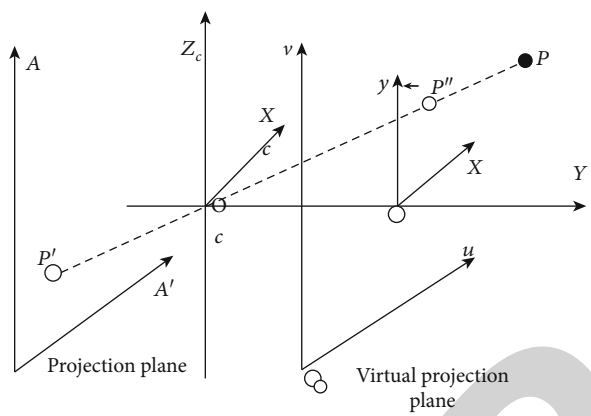


Figure 1: Image monitoring model of sports load target in physical education class.

the main control interface and positioning software. The main control interface is made up of a toolbar, menu bar, label bar, monitor interface, and status bar. The localization software includes four modules: data source setting, system configuration, database management, and network communication. The software development environment mainly uses Microsoft Visual Studio.net version 10 and uses MySQL as the system database and VB.NET as the software development language to complete the positioning software system design.

When the UWB signal is received by the base station whose coordinate position is known, the real-time position of the label is calculated according to the time information, angle information, or intensity information of the label. There are three main methods for a base station to obtain UWB signals: (1) by using the signal flight time (TOF), the signal arrival time (TOA) sent by the base station receiving the positioning label, and the difference between the arrival time of two base stations receiving the signal (TDOA) to obtain the time information; (2) by using the base station receiving signal arrival strength (RSSI) to obtain the signal strength information [13, 14]; and (3) by using the base station receiving signal arrival angle (AOA) to obtain the angle information. Two-dimensional positioning needs at least 2 base stations, and three-dimensional positioning needs at least 3 base stations.

3. Object Image Monitoring Model of Sports Load State in Physical Education Class and Fusion Processing of Monitoring Sensing Information

3.1. Establishment of Sports Load Target Image Monitoring Model in Physical Education Class. The camera that gets the sports load target image is taken as the center of the model O_c; O_c and Y_c are the optical axes generated when getting the sports target. The optical axes are perpendicular to the plane A of the plane projection of the optical axes, and a virtual projection plane A' is created on the right side of the camera plane that gets the sports target image [15]. An image monitoring model of the PE classroom load target is established, which is based on O_c as the origin, Y_c as the Y-axis, and Z_c as the X-axis. The basic principle of the image monitoring model of the sports load target in the physical education class is shown in Figure 1.

In the camera moving target image acquisition model, P is any point in the process of image capture by the robot, the projection point of its imaging on plane A is P', the projection point on plane A' is P", X_c is the X-axis projection, and the plane coordinates and pixel coordinates u, o, and v of the X, O, and y images are set in the model in Figure 1. According to Figure 1, the relationship between the threshold classification algorithms is established. If the focal length of camera for obtaining the moving target image is f, then we can get the position of x and y as follows:

$$\begin{cases} x = X_c \dfrac{f}{Z_c}, \\ y = Y_c \dfrac{f}{Z_c}. \end{cases} \tag{1}$$

If (1) is brought in, then we can rewrite (1) as

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} \dfrac{f}{Z_c} & 0 & 0 & 0 \\ 0 & \dfrac{f}{Z_c} & 0 & 0 \\ 0 & 0 & \dfrac{f}{Z_c} & 0 \end{bmatrix} \begin{bmatrix} X_c \\ X_c \\ Z_c \\ 1 \end{bmatrix}. \tag{2}$$

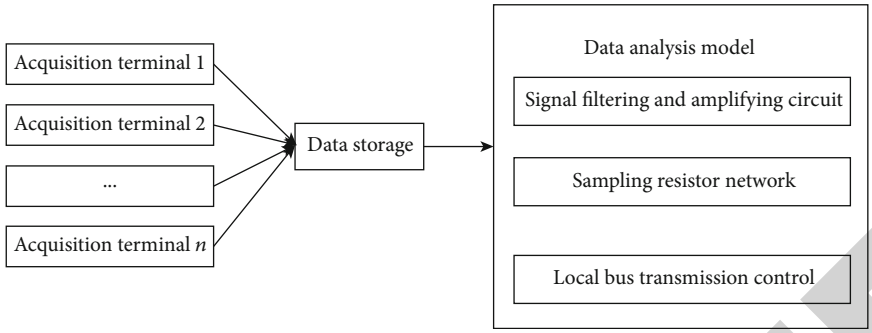


Figure 2: Data acquisition module of exercise load monitoring in physical education class.

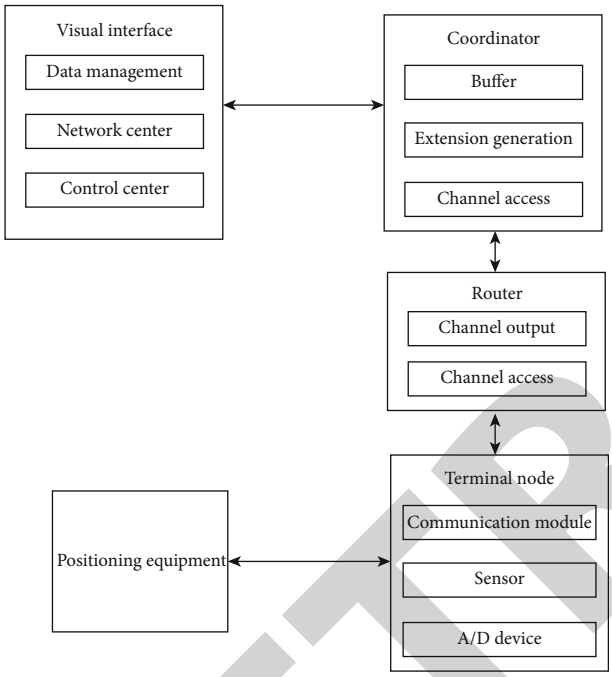


Figure 3: Overall structure module design of the system.

The formula is the mathematical matrix model of moving target image acquisition. In the competition system, the robot searches for the moving target robot which obtains the moving target image through the formula of Figure 1 or (2), and this image belongs to the moving target range graph. Therefore, after obtaining the target image of the sports load in the physical education classroom, it is necessary to calibrate the moving target. After calibrating the moving target image, the obtained image is segmented and processed, and then, the moving target features are extracted [16].

3.2. Monitoring and Sensing Information Collection of Sports Load State in Physical Education Classroom. Combining the design of the ZigBee sensor network, the paper constructs the data acquisition module of the sports classroom load condition monitoring system based on the threshold classification algorithm, as shown in Figure 2.

In the data acquisition module of the sports class load state monitoring shown in Figure 2, the local bus transmission control technology is adopted to realize the information fusion of the sports class load state threshold classification algorithm, and the remote transmission control model of the sports class load state threshold classification algorithm is constructed; the TMS320C50 DSP chip is used as the core processing chip of the sports class load state feature monitoring system to realize the integrated information processing of the sports class load state feature [17]. The FIFO RAM buffer of the sports class load state feature monitoring system is designed, and the synchronous/trigger mode is used to realize the control instruction loading and information control. The overall structure of the system is obtained, and the module design is shown in Figure 3.

In the overall structure module design shown in Figure 3, the collected parameters of the sports class load state mainly include threshold classification algorithms such as VO_2max, VE, o2p, and HR, and the characteristic information display and optimization processing of the sports class load state are realized in the LED display interface [18].

3.3. Acquisition and Fusion Processing of Monitoring Sensor Information

3.3.1. Image Data Information Fusion Processing. When the window function k of the frequency spectrum signal of the threshold classification algorithm of the sports load state in the physical education classroom is determined, the threshold classification algorithm of the sports load state in the physical education classroom is fused and clustered, then

$$X = \sqrt{x(n-k)^2+, \cdots ,+x(n)^2+, \cdots ,+x(n+k)^2}. \quad (3)$$

Then, we can obtain the detection output of the spectrum signal of the classification algorithm of the sports load state threshold:

$$y(n) = \mathrm{Mid}\{x(n-k),\cdots,x(n),\cdots,x(n+k)\}. \quad (4)$$

When the variance $\sigma^2 = 1$, the critical characteristic point of the transition from aerobic metabolism to anaerobic metabolism was taken as the threshold value, and the broadening of the recognition of the state information of the exercise load was obtained as follows:

$$d_n = -2eX_v. \quad (5)$$

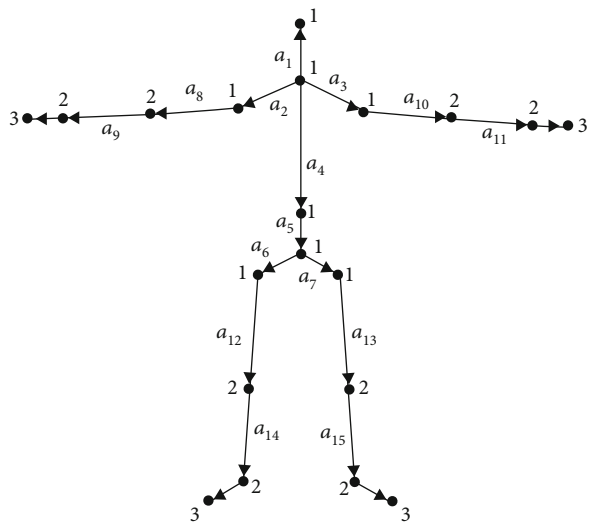


Figure 4: Joint point stratification and bone vector definition.

By using the network protocol model, the threshold classification algorithm of the sports load state in the physical education classroom is segmented. The characteristic parameter of the segmented spectrum is X_v, and V_b is the reference signal collected by the system. When $\tilde{X}$ is the minimum, the fusion component of the sports load state in the physical education classroom is obtained as follows:

$$E = [x(n) - s(n)]^2 - \sum_{v=1}^{V} b_v x_v. \quad (6)$$

By using the method of three terminal linear estimation, this paper constructs the state clustering model of the sports load in the physical education classroom and obtains the threshold classification algorithm function of the sports load state in the physical education classroom:

$$W(n+1) = W(n) - b_v \times S(n). \quad (7)$$

According to the rotational characteristics of sports bones and the topological structure of athletes' bones, 20 athletes' joints collected by Kinect sensors are processed in layers.

(i) *The first layer*: body trunk junctions [19]. The joint of the trunk is the support of the whole body, which is composed of the head, the left and right shoulders, and the spine.

(ii) *The second layer*: limb joint. Most movements of athletes are expressed by limbs. Limb joints contain a lot of characteristic information related to athletes' posture, which is mainly composed of the left and right elbows, wrists, and knees and ankle joints [20, 21].

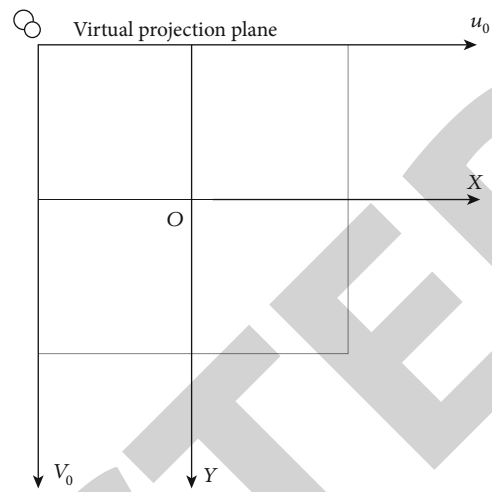


Figure 5: Pixel coordinate system and image plane coordinate system.

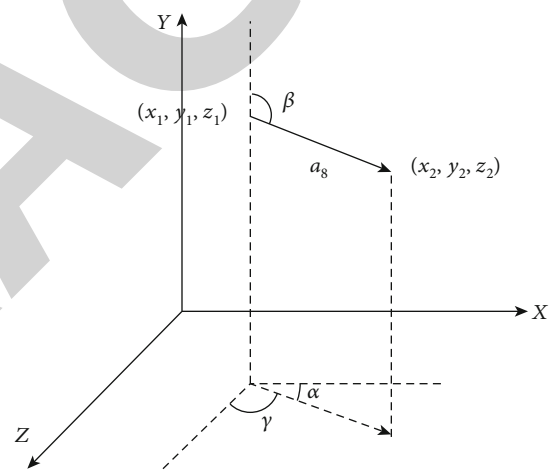


Figure 6: Calculation of direction cosine of bone vector.

(iii) *The third layer*: hands and feet. Left and right hands and left and right foot joints are divided into the third layer.

Joint point stratification and bone vector definition is shown in Figure 4.

3.3.2. Moving Target Image Acquisition. In order to obtain the accurate position in the sports system of the physical education classroom, the geometric position and attribute parameters in Figure 5 need to be one-to-one corresponding to those in the coordinate system, and then, the image is segmented, the features of the moving object are extracted, and the image position of the moving object is determined. Therefore, the pixel coordinate system u, O, V and the image coordinate system X, O, y in the target image monitoring model of the sports load in physical education classroom are extracted, as shown in Figure 5.

In Figure 5, the optical axis coincides with the center of the image plane, so the coincidence point is taken as the

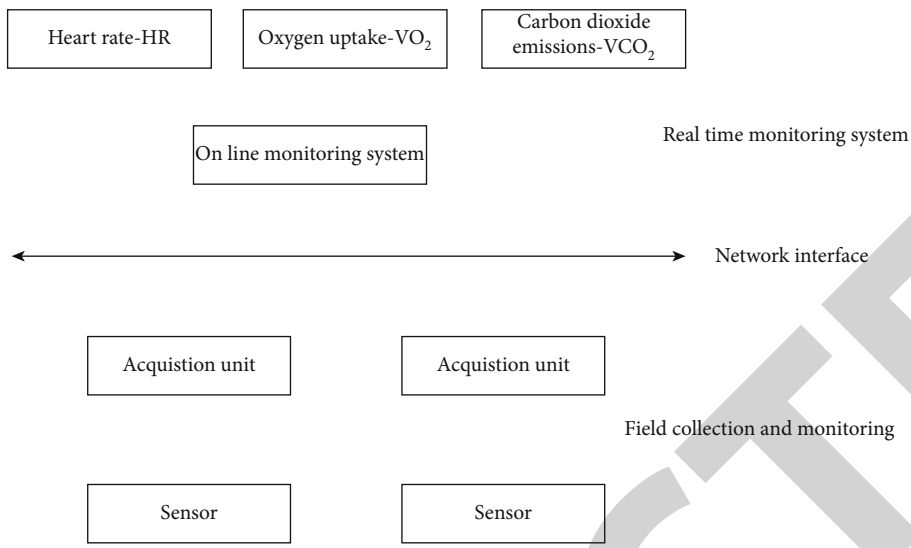


Figure 7: Exercise load monitoring and information fusion in physical education class.

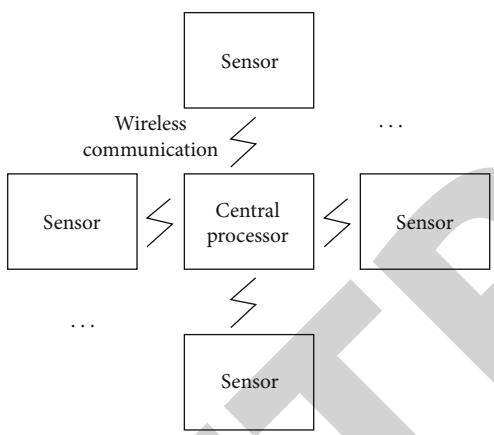


Figure 8: Wireless data acquisition system.

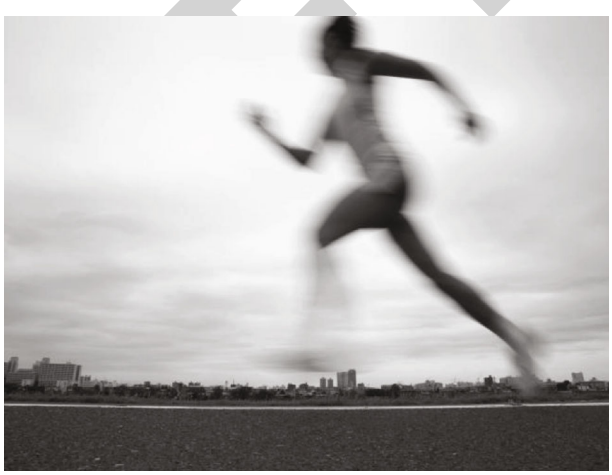


Figure 9: Original moving image.

center point O of the pixel coordinate system and the image coordinate system, and the pixel coordinate system is extracted as u_0, o_0, and v_0. Therefore, if the pixel values in

Figure 10: Results of image stabilization for motion amplitude feature monitoring.

the unit distance of the image plane coordinate system are d_x and d_y, then there are

$$\begin{cases} u = \dfrac{x}{d_x} + u_0, \\ v = \dfrac{y}{d_y} + v_0. \end{cases} \tag{8}$$

According to the calibrated moving target, the moving target image can be acquired.

3.3.3. Moving Target Direction Cosine Feature Extraction. In order to reduce the dimension of the athlete's movement and speed up the calculation, the interested nodes need to be extracted and the ones with less information should be filtered out. Based on the above hierarchical analysis, only the key nodes of the first and second levels are studied [22].

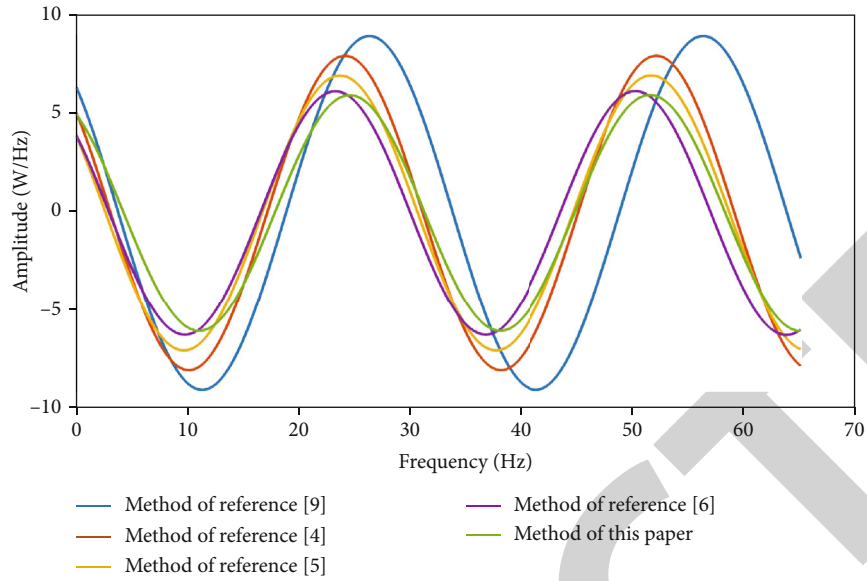


Figure 11: Comparison of double peaks.

Figure 12: Reference image.

In the case of making different movements, the position and angle information of each segment of the athletes' bones will be different. In this section, the direction cosine feature of the skeleton vector is defined to describe some kind of action.

Taking the right shoulder to elbow bone vector (a_8 vector) as an example, the direction cosine feature extraction process is analyzed.

Suppose that the three-dimensional coordinates of the shoulder joint obtained by Kinect are described by (x_1, y_1, z_1) and the three-dimensional coordinates of the elbow joint are described by (x_2, y_2, z_2), as shown in Figure 6. Then, the skeleton vector can be described as

$$a_8 = \{x_2 - x_1, y_2 - y_1, z_2 - z_1\}. \tag{9}$$

Assuming that the three directions of the a_8 vector and the Kinect coordinate system are α, β, and γ in turn, the cosine values of the three directions of the skeleton vector can be obtained by the following formula:

$$\begin{aligned} \cos\alpha &= \frac{x_2 - x_1}{\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2}}, \\ \cos\beta &= \frac{y_2 - y_1}{\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2}}, \\ \cos\gamma &= \frac{z_2 - z_1}{\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2}}. \end{aligned} \tag{10}$$

Through the above process, the direction cosine values of 15 defined bone vectors are regarded as features in order to monitor them. To sum up, the integrated information processing model of the threshold classification algorithm monitoring of the sports load state in the physical education classroom is constructed [23]. According to VO_2max, HR, and other index parameters, the sports load state monitoring and information fusion processing in the physical education classroom are realized, as shown in Figure 7.

3.3.4. Wireless Data Acquisition. In this paper, the data of the senor is sent by wireless communication. As mentioned above, the system needs to collect the data of human movement, which cannot be transmitted through the wired network. The schematic diagram of the wireless transmission of motion data is shown in Figure 8.

4. Experimental Analysis

In order to test the performance and efficiency of the sport's load monitoring system based on the threshold classification algorithm, the simulation experiment was carried out. The

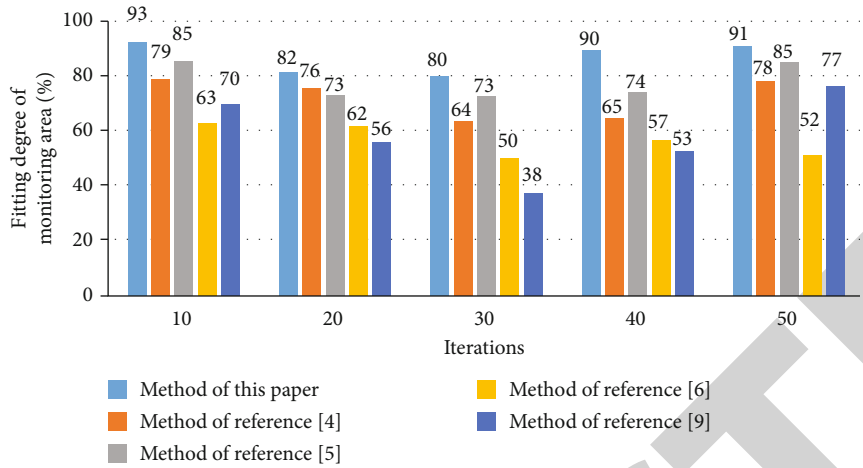


Figure 13: Comparison of monitoring effect of fixed point area of moving target.

hardware environment of the experiment platform was an ordinary desktop computer (CPU 3.30 GHz for Intel Core i3 processor, memory 4 G DDR3). The original image of the experiment is taken by CCD, through which a series of functions such as photoelectric conversion signal storage, transfer (transmission), output, processing, and electronic shutter can be realized, and the CCD data rate can be adjusted. Therefore, CCD can be used to measure the dimension under dynamic, static, and other conditions. The resolution of the image is 320×240. Meanwhile, in the experiment, 100 test samples of each mode in a group of simulation data are 100×8 to constitute the test set, and 20, 50, or all 100 of 100 training samples of each mode are randomly selected to constitute the training set of 20×8, 50×8, and 100×8, respectively. Based on the edge detection results, the image segmentation threshold of the IKONOS and WorldView images is set at 50 and 80, respectively, and the original motion is shown in Figure 9.

Using this method, the original image is preprocessed, and the single frame vision difference analysis method is used to compensate the motion. The output image of motion amplitude detection is shown in Figure 10.

As can be seen from Figure 9, the proposed method can effectively extract the motion amplitude features of moving images and has good performance. Comparing this method with the methods proposed in References [4–6, 9], the paper verifies the position of the bimodal phenomenon. The higher the position is, the worse the monitoring effect is. The verification result is shown in Figure 11. According to this figure, we can draw that the method proposed in this paper has the best performance.

As can be seen from Figure 10, this method has great advantages in comparison with other methods, and its bimodal curve fluctuation range is smaller and lower than other methods, which shows that it is suitable for sports load monitoring in the PE class and has strong practical application.

Figure 11 is a standard MPEC image, which can be used as a reference image. When playing table tennis in the PE class, the scale and shape of the person being tracked have some changes, but do not affect the statistical distribution. The test results are shown in Figure 12.

But in Figure 13, the fitting degree of the monitoring area of the proposed method is up to 98%, which is higher than that of other four methods. According to the other four methods, especially for Reference [9], the original data used for prediction must have correlation; that is, the autocorrelation coefficient must be greater than 0.5, and the prediction effect of the data will be better; otherwise, the prediction result will be extremely inaccurate. Therefore, for these methods, the prediction effect of the method cannot be fully guaranteed. Not only can the monitoring method accurately monitor the new targets in the monitoring area but it can also embody the effect of the monitoring method on the nonrigid object scale transformation and the deformed area, which has good robustness. That is to say, once the objects have big shape difference, the target location will not be unclear or deformed, which can prove that the method has a good effect.

5. Conclusion and Prospect

5.1. Conclusion. The proposed method can effectively extract the motion amplitude feature of the motion image, the monitoring performance is good, and the practical application is strong. The motion target tracking system constructed in this paper can achieve up to 98% of the fitting degree of the tracking region, and the excessive bimodal phenomenon is suppressed, and finally, the tracking image is clear, and the probability of deformation is very small, which can effectively avoid the sports injury caused by overload movement.

The main disadvantage of the threshold classification algorithm is that the classification accuracy is not high. This is mainly because the threshold setting is too simple and the number is too small. If we want to improve the classification accuracy, we must set more thresholds, but this will also increase the computational complexity of classification.

5.2. Prospect

(1) In the future, the monitoring rate of the sports load in the physical education classroom can be improved from the perspectives of a smart phone

sensor, raw data preprocessing, and movement recognition methods, so as to better serve the exercise monitoring

(2) In future studies, the exercise volume can be monitored more accurately based on the association of the daily exercise volume with weight, height, and physique of the study users

(3) During the implementation of the system, in order to reduce the randomness of the system, the data in the open data set can be used to train the classifier. But in fact, these data are optimal for a single user. In the future research, we can construct a personalized movement monitoring model by selecting training data online to improve the accuracy of exercise monitoring

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The author declares that he has no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] F. F. Shiromaru, V. de Salles Painelli, C. Silva-Batista et al., "Differential muscle hypertrophy and edema responses between high-load and low-load exercise with blood flow restriction," *Scandinavian Journal of Medicine & ence in Sports*, vol. 29, no. 11, pp. 1713–1726, 2019.

[2] D. Vaamonde, A. C. Hackney, J. M. Garcia Manso, E. Arriaza Ardiles, and M. Vaquero, "Birth sex ratio in the offspring of professional male soccer players: influence of exercise training load," *Human Reproduction*, vol. 35, no. 11, pp. 2613–2618, 2020.

[3] D. R. Samek, I. J. Elkins, M. A. Keyes, W. G. Iacono, and M. McGue, "High school sports involvement diminishes the association between childhood conduct disorder and adult antisocial behavior," *The Journal of Adolescent Health*, vol. 57, no. 1, pp. 107–112, 2015.

[4] K. Mullins, M. Hanlon, and P. Carton, "Correction to: arthroscopic correction of femoroacetabular impingement improves athletic performance in male athletes," *Knee Surgery Sports Traumatology Arthroscopy*, vol. 13, pp. 96–106, 2019.

[5] V. Sundaresan, G. Zamboni, C. Le Heron et al., "Automated lesion segmentation with BIANCA: impact of population-level features, classification algorithm and locally adaptive thresholding," *NeuroImage*, vol. 202, article 116056, 2019.

[6] X. Wu, Y. Gao, and D. Jiao, "Multi-label classification based on random forest algorithm for non-intrusive load monitoring system," *Processes*, vol. 7, no. 6, p. 337, 2019.

[7] X. F. Tan, T. T. Xuan, and P. C. Zhang, "Load identification and classification in non-intrusive load monitoring system based on data stream," *Chinese Journal of Power Sources*, vol. 40, no. 5, pp. 1110–1112, 2016.

[8] X. Li and S. Geng, "Research on sports retrieval recognition of action based on feature extraction and SVM classification algorithm," *Journal of Intelligent and Fuzzy Systems*, vol. 39, no. 4, pp. 5797–5808, 2020.

[9] S. Banerjee, G. K. Vishwakarma, and A. Bhattacharjee, "Classification algorithm for high-dimensional protein markers in time-course data," *Statistics in Medicine*, vol. 39, no. 10, pp. 89–100, 2020.

[10] S. J. Elliott, "Antismoking adverts and sports sponsorship," *The Lancet*, vol. 340, no. 8831, pp. 1356–1362, 1992.

[11] M. A. Myers, S. Hall, A. Wright et al., "Spinal fractures incurred by sports-related injuries," *World Neurosurgery*, vol. 151, pp. e747–e752, 2021.

[12] J. Nyland, J. Greene, S. Carter, J. Brey, R. Krupp, and D. Caborn, "Return to sports bridge program improves outcomes, decreases ipsilateral knee re-injury and contralateral knee injury rates post-acl reconstruction," *Knee Surgery, Sports Traumatology, Arthroscopy*, vol. 28, no. 11, pp. 3676–3685, 2020.

[13] J. A. Silverman, C. M. Grilo, and J. A. Lydecker, "111. When children's exercise becomes compulsive: associations with youth sports, parenting, and disordered eating," *Journal of Adolescent Health*, vol. 66, no. 2, pp. S57–S58, 2020.

[14] K. E. Ackerman, T. Stellingwerff, K. J. Elliott-Sale et al., "#REDS (relative energy deficiency in sport): time for a revolution in sports culture and systems to improve athlete health and performance," *British Journal of Sports Medicine*, vol. 54, no. 7, pp. 369-370, 2020.

[15] J. Guilherme and A. H. Lancha, "Total genotype score and athletic status: an exploratory cross-sectional study of a Brazilian athlete cohort," *Annals of Human Genetics*, vol. 84, no. 2, pp. 141–150, 2020.

[16] L. Leite, H. P. Santiago, D. C. de Lima, W. Pires, and C. C. Coimbra, "Central losartan administration increases cardiac workload during aerobic exercise," *Neuropeptides*, vol. 77, article 101960, 2019.

[17] E. B. Wasserman, E. L. Sauers, J. K. Register-Mihalik et al., "The first decade of web-based sports injury surveillance: descriptive epidemiology of injuries in US high school boys' baseball (2005–2006 through 2013–2014) and National Collegiate Athletic Association men's baseball (2004–2005 through 2013–2014)," *Journal of Athletic Training*, vol. 54, no. 2, pp. 198–211, 2019.

[18] M. Miller, S. Malekian, J. Burgess, and C. Labella, "Evaluating a commonly used tool for measuring sport specialization in young athletes," *Journal of Athletic Training*, vol. 54, no. 10, pp. 1083–1088, 2019.

[19] D. P. Longman, J. Wells, and J. T. Stock, "Human athletic paleobiology; using sport as a model to investigate human evolutionary adaptation," *American Journal of Physical Anthropology*, vol. 171, pp. 42–59, 2020.

[20] D. A. Dobrosielski, L. Sweeney, and P. J. Lisman, "The association between poor sleep and the incidence of sport and physical training-related injuries in adult athletic populations: a systematic review," *Sports Medicine*, vol. 51, no. 4, pp. 777–793, 2021.

[21] C. McGrew, D. S. MacCallum, D. Narducci et al., "AMSSM position statement update: blood-borne pathogens in the context of sports participation," *British Journal of Sports Medicine*, vol. 54, no. 4, pp. 200–207, 2020.

[22] H. Chaabene, O. Prieske, J. Moran, Y. Negra, A. Attia, and U. Granacher, "Effects of resistance training on change-of-

Hindawi
Wireless Communications and Mobile Computing
Volume 2023, Article ID 9785461, 1 page
https://doi.org/10.1155/2023/9785461

Retraction

Retracted: Remote Judgment Method of Painting Image Style Plagiarism Based on Wireless Network Multitask Learning

Wireless Communications and Mobile Computing

Received 10 October 2023; Accepted 10 October 2023; Published 11 October 2023

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. Wang, "Remote Judgment Method of Painting Image Style Plagiarism Based on Wireless Network Multitask Learning," *Wireless Communications and Mobile Computing*, vol. 2021, Article ID 1345974, 8 pages, 2021.

Hindawi
Wireless Communications and Mobile Computing
Volume 2021, Article ID 1345974, 8 pages
https://doi.org/10.1155/2021/1345974

Research Article

Remote Judgment Method of Painting Image Style Plagiarism Based on Wireless Network Multitask Learning

Zhijun Wang

Academy of Fine Arts, Shanxi University, Taiyuan Shanxi 030006, China

Correspondence should be addressed to Zhijun Wang; wzj777@sxu.edu.cn

Received 21 July 2021; Revised 7 August 2021; Accepted 24 August 2021; Published 9 September 2021

Academic Editor: Chi-Hua Chen

Since the artistry of the work cannot be accurately described, the identification of reproducible plagiarism is more difficult. The identification of reproducible plagiarism of digital image works requires in-depth research on the artistry of artistic works. In this paper, a remote judgment method for plagiarism of painting image style based on wireless network multitask learning is proposed. According to this new method, the uncertainty of painting image samples is removed based on multitask learning algorithm edge sampling. The deep-level details of the painting image are extracted through the multitask classification kernel function, and most of the pixels in the image are eliminated. When the clustering density is greater than the judgment threshold, it can be considered that the two images have spatial consistency. It can also be judged based on this that the two images are similar, that is, there is plagiarism in the painting. The experimental results show that the discrimination rate is always close to 100%, the misjudgment rate of plagiarism of painting images has been reduced, and the various indicators in the discrimination process are the lowest, which fully shows that a very satisfactory discrimination result can be obtained.

1. Introduction

In the commercial society, the competition in the cultural industry has increased fiercely, and commercial imitation and plagiarism may have a direct impact on the economic interests of both parties [1]. In order to avoid infringement disputes that may arise in the future, consciously avoid copying directly as it is, and adopting methods of reference and imitation to intercept fragments or expressions from others' works and absorb them into their works has become a creative norm, which frequently triggers plagiarism disputes [2]. Whether the similarity of works constitutes infringement must consider the balance between personal interests and social interests, and neither can we turn a blind eye to the obvious plagiarism of other people's works, nor can it be regarded as infringement without consideration for any acts involving plagiarism. In the plagiarism of painting works, plagiarism often refers to the content and style of the existing works and consciously changes the color, angle, and relative position of the main content in the painting. Change the size, shape, and details of the object, and rebuild other objects in the painting [3].

The intellectual property rights of digital image works are mainly embodied in the use of image color, composition, and artistry of expression. The artistic plagiarism of digital image works is regenerative [4]. Like other forms of digital works, the identification of regenerative plagiarism is obviously difficult, and the identification of regenerative plagiarism is even more difficult because the artistry of the work cannot be accurately described. The identification of reproducible plagiarism of digital image works requires in-depth research on the artistry of artistic works and refines the artistry into quantifiable artistic indicators. Then, based on these indicators, the two works are tested for similarity. The reproducible plagiarism identification of digital image works will be an important direction of plagiarism identification research in the future.

Reference [5] discusses the painting art style conversion technology in NPR, uses image representation derived from convolutional neural networks, and designs an art style algorithm. The algorithm can separate and recombine the content and style of the image and can produce new images with high quality. Taking the painting style of Xin'an School as an example, this algorithm is used to convert the sample

image into a new image of Xin'an School. Reference [6] proposed the Cycle GANSN algorithm. After the Cycle GAN algorithm identifies each convolutional layer of the network, a spectral normalization layer is added, the spectral norm of the convolutional layer parameter matrix is estimated by the power iteration method, and the stochastic gradient descent method is used to update the convolutional layer parameters. Since the change of parameters in each update is very small, only one iteration can quickly estimate the maximum singular value of the matrix. According to the obtained maximum singular value, the convolutional layer parameters are normalized, so that the entire discriminant network meets 1-Lipschitz continuity. Reference [7] proposes to recognize similar parts from multiple execution trajectories of multithreaded programs under the same input and abstract behavior motifs that are not easily affected by thread interleaving to realize plagiarism detection for multithreaded programs. This method captures the dynamic execution trajectory of the program. After trajectory pruning, gram matching, expansion, and abstraction, motif birthmarks are extracted to model the behavior of multithreaded programs. Finally, the potential plagiarism between programs is determined by measuring the similarity of motif birthmarks. In Reference [8], it investigates the usefulness of the normalized compression distance (NCD) for image similarity detection. Instead of the direct NCD between images, the paper considers the correlation between NCD-based feature vectors extracted for each image. The vectors are derived by computing the NCD between the original image and sequences of translated (rotated) versions. Feature vectors for simple transforms (circular translations on horizontal, vertical, and diagonal directions and rotations around image center) and several standard compressors are generated and tested in a very simple experiment of similarity detection between the original image and two filtered versions (median and moving average). The promising vector configurations (geometric transform and lossless compressor) are further tested for similarity detection on the 24 images of the Kodak set subject to some common image processing. While the direct computation of NCD fails to detect image similarity even in the case of simple median and moving average filtering in 3×3 windows, for certain transforms and compressors, the proposed approach appears to provide robustness at similarity detection against smoothing, lossy compression, contrast enhancement, and noise addition and some robustness against geometrical transforms (scaling, cropping, and rotation). In Reference [9], the authors declare that binary code similarity detection (BCSD) plays an important role in malware analysis and vulnerability discovery. Existing methods mainly rely on the expert's knowledge for the BCSD, which may not be reliable in some cases. More importantly, the detection accuracy (or performance) of these methods is not so satisfactory. To address these issues, we propose BinDeep, a deep learning approach for binary code similarity detection. This method firstly extracts the instruction sequence from the binary function and then uses the instruction embedding model to vectorize the instruction features. Next, BinDeep applies a Recurrent Neural Network (RNN) deep learning model to identify the specific types of two functions for later comparison. According to the type information, BinDeep selects the corresponding deep learning model for similarity comparison. Specifically, BinDeep uses the Siamese neural networks, which combine the LSTM and CNN to measure the similarities of the two target functions. Different from the traditional deep learning model, our hybrid model takes advantage of CNN spatial structure learning and LSTM sequence learning. The evaluation shows that our approach can achieve good BCSD between cross-architecture, cross-compiler, cross-optimization, and cross-version binary code.

For two images that are exactly the same, the pixel values of the corresponding positions are the same. At present, the common method of image consistency detection software on the Internet is to use a cryptographic hash algorithm for judgment. When the image pixel values are exactly the same, the result calculated by the cryptographic hash algorithm is also exactly the same. Therefore, as long as the calculation results are verified to be the same, it can be judged whether the image content is exactly the same. However, the limitation of this method is that it requires a large amount of calculation for point-by-point discrimination and requires high image consistency. If you modify the details of the image, such as toning, cropping, and zooming, the calculated results will be completely different.

This paper proposes a remote judgment method of painting image style plagiarism based on wireless network multitask learning. Firstly, the edge sampling is based on the multitask learning algorithm to remove the uncertainty of painting image samples. Secondly, the multitask classification kernel function is used to extract the deep-level details of the painting image, and most of the pixels in the image are excluded. Lastly, when the cluster density is greater than the threshold value, the two images can be considered to have spatial consistency. It can also be judged based on this that the two images are similar, that is, there is plagiarism in the painting.

2. Wireless Network Multitask Learning

2.1. Remove the Uncertainty of Painting Image Samples. The selection of painting image samples based on the multitask learning algorithm can better avoid the influence of uncertainty conditions on the Internet storage capacity, and all the instruction behaviors in the entire storage processing process maintain the multitasking ability. Therefore, regardless of the increase or decrease of the painting image sample space, the multitask learning algorithm can better stabilize the transmission and application capabilities of valuable information parameters. It is stipulated that 1 represents the information measurement value in the painting image sample space, and 2 represents the storage amount of the painting image sample when the measurement value is 3.

The painting image sample is a description of the storage capacity of the network painting image. With the support of the multitask learning algorithm, the larger the painting image sample space, the stronger the storage capacity of the matching network painting image [10]. Uncertainty of painting image samples is a generalized application ability.

In the space of painting image samples, two adjacent painting image data have the ability to be infinitely close. But it will never be completely relieved, so the larger the sample space of the painted image, the weaker the sample uncertainty ability [11]. The selection of painting image samples based on the multitask learning algorithm can better avoid the influence of uncertainty conditions on the Internet storage capacity, and all the instruction behaviors in the entire storage processing process maintain the multitasking ability. Therefore, regardless of the increase or decrease of the painting image sample space, the multitask learning algorithm can better stabilize the transmission and application capabilities of valuable information parameters. It is stipulated that χ represents the information measurement value in the painting image sample space, and P_X represents the storage amount of the painting image sample when the measurement value is χ. Combining the above physical quantities, the uncertainty condition of the painting image sample based on the multitask learning algorithm can be defined as follows:

$$K = \frac{\operatorname{argmax}|1 - P_x|}{\hat{y} \times \dot{I}}, \tag{1}$$

where $\hat{y}$ is the valuable information feature value in the painting image sample space and $\dot{I}$ is the uncertain search condition of the painting image sample.

Edge sampling is a common processing method in multitask learning algorithms. Due to the existence of uncertain conditions of painting image samples, this processing behavior can directly determine the final execution result of painting image classification. When the information space of the painting image is large, the real-valued result of the edge sampling coefficient will also increase; when the information space of the painting image is small, the real-valued result of the edge sampling coefficient will also decrease [12]. In other words, the edge sampling coefficient is a flexible index parameter that can interfere with the actual processing behavior of painting image classification. And as the uncertainty of painting image samples weakens, the amount of valuable information of painting images affected by multitask learning algorithms will gradually increase. This is also the main reason why the edge sampling coefficient can determine the direction of painting image classification processing to a certain extent [13]. α represents the sampling parameter value of valuable painting image information, and ε represents the sampling parameter value of the multitask learning algorithm. Simultaneous formula (1) can express the edge sampling coefficient of painting image classification as follows:

$$A = \exp\left\{\frac{f.\|u_1 - u_2\|}{(\alpha - \varepsilon)K^2}\right\}. \tag{2}$$

In this formula, u_1 and u_2 represent two different valuable painting image information respectively, and f represents the information extraction coefficient.

2.2. Extraction of Deep-Level Detail Feature of Painting Image. The multitask classification kernel function determines the practical application ability of the multitask learning algorithm. When the uncertainty condition of the painting image sample is allowed, the network host often needs to store a large amount of painting image sample data at the same time. With the increase of the edge sampling coefficient, the range of the kernel function will continue to expand until the multitask learning algorithm can fully meet the actual storage requirements of the painting image sample data. If the accurate calculated value of the edge sampling coefficient is known, the space where the sample data of the painting image has been stored can be arranged according to the specific requirements of the multitask learning algorithm. On the one hand, it provides information parameter support for subsequent image classification and collaborative processing, and on the other hand, it can also avoid the obvious accumulation of stored image data [14]. i_1 represents the first image information classification condition, and i_n represent the n image information classification condition. The simultaneous formula (2) can define the classification kernel function based on the multitask learning algorithm as follows:

$$F = \sum_{i_1}^{i_n} A.\frac{x_1 + x_2 + \cdots + x_n}{n \times |\bar{x}|^2}, \tag{3}$$

where x_1, x_2, and x_n represent n different image information classification coefficient items, respectively, and $\bar{x}$ represents the average value of n coefficient items.

For the painting image reconstruction method, it is first assumed that the painting image and the painting image are sparse expressions for their respective dictionaries, and then, the image sample features are obtained through the PCA Net deep network, and a pair of over-complete dictionaries D_h and D_1 are obtained by combining the dictionaries. In the painting image reconstruction stage, the painting image is processed in the same way, deep-level feature extraction is performed through PCA Net, and the coefficients expressed by the sparseness of the painting image feature D_1 are directly applied to D_h to obtain the corresponding painting feature image. Realize the painting reconstruction of painting images [15]. In the mining of image samples through PCA Net deep network, better image features than nondeep networks can be obtained. The deep feature dictionary established can also improve its description ability and significantly improve the quality of image reconstruction.

It is necessary to sample K fuzzy images in the training data set, adjust them to the same size, obtain painting images corresponding to the painting images, and combine them into a model sample pair: $T = \{X_h, X_1\}$, where $X_h = \{x_h^i\}_{i=1}^k$ and $X_1 = \{x_h^i\}_{i=1}^k$ represent the painting feature. Calculate matrix blocks for all samples in the data set, and select a sliding window with a size of $k_1 \times k_2$ (normally, a square window of image pixels of 3, 5, or 7 is used). After feature extraction of all images through the above-set sliding window, a new $N \times m \times n$-column data matrix X can be

obtained. Each column of this matrix represents an image block, with a total of $k_1 \times k_2$ elements.

The formula for obtaining the i painting image $\bar{x}_{hi}$ training sample is as follows:

$$\bar{X}_{hi} = [\bar{x}_{hi,1}, \bar{x}_{hi,2}, \cdots, \bar{x}_{hi,mn}]. \quad (4)$$

The features are extracted using the data matrix X obtained above, and the extracted features are regarded as feature samples in the ScSR model and substituted into the feature dictionary of PCA Net.

Suppose T_{hi}^{L} represents the painting image result after sparse coding, the sparse coding result is quantized, the histogram coding is completed, and the deep-level detail features of the painting image are extracted, namely,

$$F_{hi} = \left[\mathrm{Bhist}\left(T_{hi}^{1}\right), \cdots, \mathrm{Bhist}\left(T_{hi}^{L_1}\right)\right]^T \in R^{\left(2^{L1}\right)L_1 B}. \quad (5)$$

Through the same processing process as the painting image, the result of extracting the deep details of the painting image is given, namely,

$$F_{li} = \left[\mathrm{Bhist}\left(T_{hi}^{1}\right), \cdots, \mathrm{Bhist}\left(T_{hi}^{L_1}\right)\right]^T \in R^{\left(2^{L1}\right)L_1 B}. \quad (6)$$

In the formula, F_{hi} and F_{li}, respectively, represent the feature extraction results of the painting image; Bhist represents the encoding process of the straight image; B represents the number of divided image sample blocks.

If the pixel point P is used as the detection point, a discrete circle with a radius of 3 is made with it as the center, and 16 pixel points formed on the circumference are obtained to surround the central pixel P. Set a threshold suitable for feature point t. When there are n consecutive pixels, the gray value satisfies formula (1). That is, when the gray values of n consecutive pixels are larger than the threshold value t or smaller than the threshold value t of the pixel of the P point, it is determined that the pixel point P is a corner point. Feature extraction is shown in Figure 1.

$$V = |I_X - I_o| = p > t. \quad (7)$$

In the formula, V is the scoring function, which represents the score of the detection point, I_X represents the pixel value of n pixels on the circle and represents the pixel value of P points, and I_p takes the point with the largest score within the radius as the distinctive feature. In order to increase the calculation speed, $n = 12$ is generally selected for calculation. At the same time, the pixels on the 16 circles are numbered in turn clockwise. In corner detection, the 4 pixels numbered 1, 5, 9, and 13 can be judged first. Only when at least 3 of these 4 pixels satisfy the above formula can it be confirmed as a corner point. Therefore, the FAST feature extraction algorithm can quickly eliminate most of the pixels in the image, improve the calculation efficiency of corner detection, and save time.

3. Remote Judgment of Plagiarism of Painting Image Style

If two pictures have spatial consistency in a painting, then the two works can be considered similar. According to the principle of spatial consistency, if two pictures are similar, the lines corresponding to the matching feature points of the two pictures are parallel [16]. Therefore, most of the angle values obtained after Hough transform fall within the same angle clustering range.

Assuming that there are images A and B with pixels of $m \times n$ and $p \times q$, the image copy detection process is as follows:

(1) *Image Expansion*. First, expand the image A into an $(m+p) \times (n+q)$ image G and use A to tile to the right and down in order, as shown in Figure 2.

(2) *Scan Comparison*. First, align the upper edge of B with the upper edge of G, align the left edge of B with the left edge of G, and cover B on G. Compare the part B and G are covered by B, as shown in Figure 3(a). Then, shift B to the right by one pixel and then compare B and the part of G covered by B. Continue like this until the right edge of B is aligned with the right edge of G, as shown in Figures 3(b) and 3(c).

Move B to the left of G again so that the left edge is aligned with the left edge of G, but the top edge is moved down one pixel. Repeat the horizontal scan comparison until the right edge of B is aligned with the right edge of G, as shown in Figures 3(d)–3(f).

This continues until the lower edge of B is aligned with the lower edge of G, and the right edge of B is aligned with the right edge of G, as shown in Figures 3(g)–3(i).

(3) *Comparison Result*. In the scan comparison, a copy amount, copy pattern, and copy method are obtained for each comparison. The maximum value of the copy amount is regarded as the copy amount of A and B, and the copy pattern with the largest copy amount is regarded as the copy pattern of A and B.

(4) *Copy Judgment*. According to the set copy amount threshold, it is judged whether there is a possibility of copying between A and B, and finally, it will be further judged by human work.

When there is any x that makes the calculation result satisfy formula (2), it can be judged that the image has spatial consistency, that is, the two pictures are similar, where x is any value in the interval, FW is the threshold that controls the clustering range, TH is the judgment threshold, and n is the total number of connections. $\theta(x)$ represents the number of angles equal to x after transformation and represents the total number of $\theta(x)$ within the interval clustering range

(a) Original image (b) Feature extraction

Figure 1: Feature extraction.

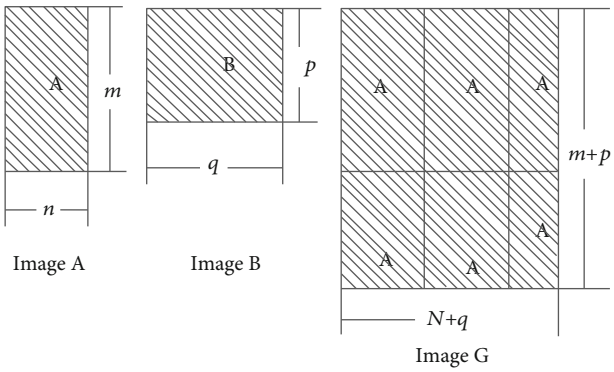


Figure 2: The expansion of image A.

[17].That is, when there is a cluster density greater than the determination threshold in any interval range, it can be considered that the two images have spatial consistency. It can also be judged based on this that the two images are similar, that is, there is plagiarism in the painting.

$$N(x) = \sum_{i=-\mathrm{FW}}^{\mathrm{FW}} \theta(x-t) > \mathrm{TH}^{*} n. \tag{8}$$

When different cluster width values are used, the obtained judgment results are different, and the final judgment results will also be different when different judgment thresholds are taken. Therefore, the cluster width and the judgment threshold are important parameters that need to be adjusted. Image clustering density curve is shown in Figure 4.

The blue line represents the similar image cluster density curve, and the orange line represents the dissimilar image cluster density curve. It can be seen that as the cluster width FW increases, the probability of meeting the specified threshold TH also increases, and the probability that the input image is judged as a similar image increases. The opposite is the same. In addition, in the case of the same cluster width, the threshold TH decreases, and the probability of the input image being judged as a similar image will increase and vice versa.

4. Experimental Study

According to the experiment, 1620 pairs of typical images were selected as the training set from the works of the National Science and Technology Innovation Competition for Children's Science Fantasy Painting Competition, and the parameter optimization experiment was carried out. In addition, a total of 30,870 pairs of images were selected as the test set to verify the reliability of the algorithm. The main performance indicators of the effect of the painting duplicate check algorithm are the precision rate and the recall rate.

Discrimination rate: it indicates the ratio of the number of identified incidents to the total incidents of plagiarism of painting images within a certain time range when applying the discrimination method. The specific calculation method is as follows:

$$\mathrm{IR} = \frac{\mathrm{DR}}{S} \times 100\%. \tag{9}$$

In the formula, IR represents the discrimination rate, DR represents the number of plagiarism incidents of the identified painting image, and S represents the number of actual painting image incidents. In order to verify the performance of the method proposed in this paper, we do the experiment as well as with the other five methods proposed in References [5–9], and the results of IR-FIR are shown in Figure 5.

It can be seen from the above figure that when the false alarm rate is the same, the discrimination rate of this method is always better than the other five methods proposed in References [5–9]. When the false alarm rate reaches 1%, the discrimination rate of the method in this paper is close to 82%, the discrimination rate of the method in Reference [7] is 43%, the discrimination rate of the method in Reference [8] is 57%, and the discrimination rate of the method in Reference [9] is 70% (63% and 61% for References [5, 6], respectively). Therefore, the method proposed in this paper has the best discrimination effect performance. When the judgment rate is at the same level, it can be found that the misjudgment rate of the method proposed in this paper is lower than the other three methods. When the judgment rate is 60%, the misjudgment rate of the methods proposed in References [7–9] is 2.0%, 1.2%, and 0.7%, respectively, and the misjudgment rate of the method proposed in this paper is 04%. The reason that the method proposed in this paper has the best performance might be that this method can make full use of the advantages of multitask learning, and it can make use of the acquired knowledge to get better classification effect. According to this figure, we can draw that the effectiveness of the method in this paper is significantly better than the other three methods.

False alarm rate: it indicates the ratio of the number of false alarm events to the total number of detected events within a certain time range. The specific calculation method is as follows:

$$\mathrm{FIR} = \frac{\mathrm{FN}}{\mathrm{DR}} \times 100\%. \tag{10}$$

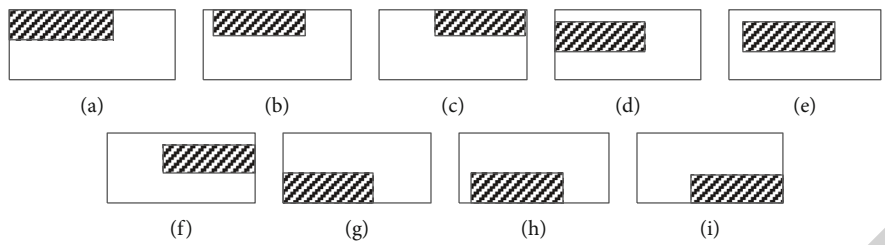
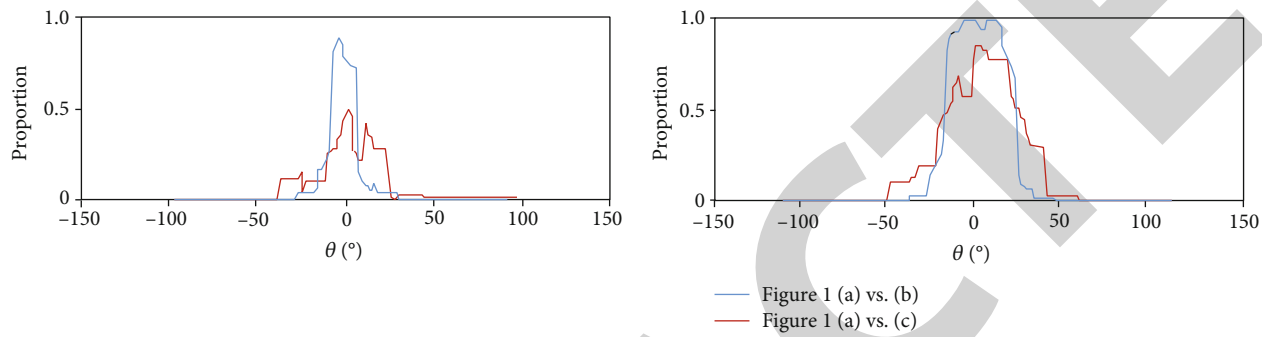


Figure 3: Scanning comparison.



(a) FW = 7 cluster density

(b) FW = 20 cluster density

Figure 4: Image clustering density curve.

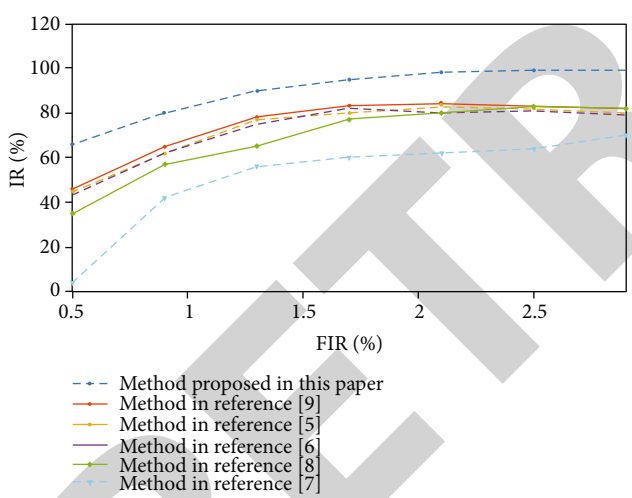


Figure 5: IR-FIR relationship diagram.

In the formula, FIR represents the false alarm rate, FN represents the number of false alarms, and DR represents the number of detected events.

Average discrimination time: it means the average value of the time difference from when an event occurs within a certain time interval to the end of the algorithm discrimination. Due to the difference between the detector and the painting image discrimination method, the average detection time will also change, and the range of change will remain within 4-10 min.

$$\mathrm{MTTI} = \frac{1}{n}\sum_{i=1}^{n}[\mathrm{TI}(i) - \mathrm{AT}(i)]. \tag{11}$$

In the formula, $\mathrm{TI}(i)$ represents the actual time of the event i, $\mathrm{AT}(i)$ represents the estimated time of the event i, and n represents the number of traffic incidents.

Result analysis of false alarm rate and average discrimination time is shown in Table 1.

It can be seen from the above table that the method in this paper meets the requirements of three-level alarm-related indicators, and the false judgment rate of plagiarism of painting images is reduced. On the basis of meeting the relevant requirements of the first-level alarm, the relevant requirements for the third-level alarm indicators can also be met, ensuring efficient management of the identification of plagiarism in painting images.

In order to verify the accuracy of the judgment result, the mean square percentage error index is selected as the evaluation index. The specific calculation formula is as follows:

$$\mathrm{MSPE} = \frac{1}{N}\sqrt{\sum_{t=1}^{N}\left(\frac{x_{\mathrm{real}} - x_{\mathrm{pre}}}{x_{\mathrm{real}}}\right)^2} \times 100\%. \tag{12}$$

In the above formula, MSPE also represents the distribution of errors, reflecting the degree of deviation between the predicted value and the true value, and the two are inversely proportional. In this paper, we use the average of MSPEs as the forecast index which is also used to compare the performance of the four methods. The forecast index is defined as formula (13).

$$\mathrm{FI} = \frac{M}{N}. \tag{13}$$

Table 1: Result analysis of false alarm rate and average discrimination time.

	Threshold		First-level alarm			Secondary alarm			Three-level alarm		
	Upper reaches	Downstream	IR (%)	MTTI (min)	FIR (%)	IR (%)	MTTI (min)	FIR (%)	IR (%)	MTTI (min)	FIR (%)
1	0.24	0.54	98.66	2.05	0.41	98.32	2.26	0.05	97.66	2.25	0.01
2	0.24	0.34	99.66	2.22	0.52	99.66	1.88	0.10	99.32	1.88	0.02
3	0.14	0.54	99.66	1.66	0.51	99.66	1.90	0.04	99.32	2.01	0.01
4	0.24	0/34	97.0	2.27	0.20	96.0	2.40	0.05	96.0	2.44	0.01
5	0.44	0.14	95.66	2.38	0.28	94.32	2.38	0.08	93.0	2.38	0.05
Average			98.06	2.12	0.40	97.7	2.16	0.06	97.46	2.1	0.02

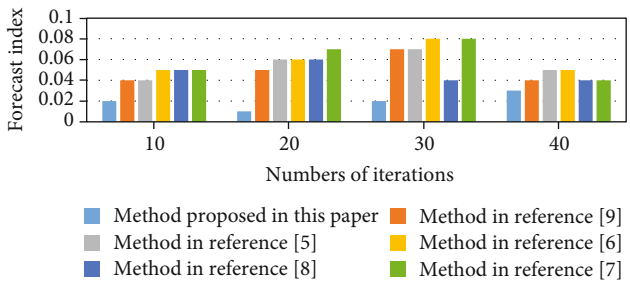


Figure 6: Comparison of prediction results of different discrimination methods.

where FI is forecast index, M is the sum of MSPE of N times, and N is the number of experiment. Figure 6 gives the comparison of the forecast index value of different methods.

A comprehensive analysis of the experimental data in Figure 6 shows the forecast index of the six prediction methods. In order to verify the performance, we have given the performance of FI as well as compare the method proposed in this paper with methods proposed in References [5–9]. Among the six methods for identifying plagiarism in painting images, the indicator of the proposed method is the lowest, which fully shows that the proposed method can achieve very satisfactory results. It should be noted that, when the number of experiments is 40 in this figure, the FI of the proposed method is much higher, and the main reason is that the method in this paper is based on the edge sampling of the multitask learning algorithm to remove the uncertainty of painting image samples. Meanwhile, there may be negative transfer in multitask learning. Therefore, its performance may be poor at some times. However, with the accumulation of knowledge and continuous training, the network performance will eventually be very good. According to the method proposed in this paper, it can take advantage of "experience." The multitask classification kernel function can extract the deep-level details of the painting image, eliminate most of the pixels in the image, and improve the calculation efficiency of corner detection. According to References [8, 9], the FI of the method in Reference [9] is higher than that of the method in Reference [8] when the number of experiments is 30, and the reason might be that the performance of deep learning can be affected by the change of parameters in the learning process. Meanwhile, we can see that the performance of the method proposed in Reference [7] is higher than the other four methods. Besides, the FI of the method in Reference [7] is lower than that of the method in Reference [6] in other cases, and the main reason is the change of parameters in the learning process too. References [6, 9] are two methods based on neural network, so these two methods have the same drawbacks. Since the advantage of the multitask learning, this paper can take the advantage of transfer learning, so it can use the "knowledge," which can further improve the performance of the method proposed in this paper.

5. Conclusion

This paper proposes a remote judgment method of painting image style plagiarism based on wireless network multitask learning strategy. Firstly, the proposed method uses edge sampling based on multitask learning algorithm to remove the uncertainty of painting image samples. Secondly, the deep-level detail features of painting images are extracted through multitask classification kernel function. Excluding most of the pixels in the image, when the clustering density is greater than the determination threshold, it can be considered that the two images have spatial consistency. It can also be judged based on this that the two images are similar, that is, there is plagiarism in the painting.

Under this circumstance, it can make the "substantial similarity" more reasonable and the judgment result more objective and fair. To further enhance the predictability of the trial results to the public, it is necessary to conduct a deeper discussion and research on the judgment of "substantial similarity" in copyright infringement.

The method proposed in this paper can use the basic features of the painting image for similarity detection. According to this method, it can remove the uncertainty of painting image samples by the edge sampling, and then, the deep-level detail features of painting images can be extracted according to the multitask classification kernel function. To sum up, the method proposed in this paper can extract the features of painting image from multiple angles, so as to help judge whether there is plagiarism. However, the solution of the multitask learning method, which is a nonconvex, is generally difficult to obtain. Therefore, this method is difficult to be applied to the actual system. In the future, we need to introduce more efficient feature extraction algorithms.

Hindawi
Wireless Communications and Mobile Computing
Volume 2023, Article ID 9764052, 1 page
https://doi.org/10.1155/2023/9764052

Retraction

Retracted: Research on Acquisition and Tracking Algorithm of Global Satellite Positioning Receiver Based on UWB

Wireless Communications and Mobile Computing

Received 10 October 2023; Accepted 10 October 2023; Published 11 October 2023

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. Yang, D. Sun, and H. Ruan, "Research on Acquisition and Tracking Algorithm of Global Satellite Positioning Receiver Based on UWB," *Wireless Communications and Mobile Computing*, vol. 2021, Article ID 9025086, 9 pages, 2021.

Hindawi
Wireless Communications and Mobile Computing
Volume 2021, Article ID 9025086, 9 pages
https://doi.org/10.1155/2021/9025086

Research Article

Research on Acquisition and Tracking Algorithm of Global Satellite Positioning Receiver Based on UWB

Li Yang,[1] **Danshi Sun**,[2] **and Haote Ruan**[1]

[1]*College of Geography and Environmental Science, Henan University, Kaifeng, 475000 Henan, China*
[2]*School of Geodesy and Geomatics, Wuhan University, Wuhan 430079, China*

Correspondence should be addressed to Li Yang; 10130147@henu.edu.cn

Received 24 July 2021; Revised 4 August 2021; Accepted 11 August 2021; Published 27 August 2021

Academic Editor: Chi-Hua Chen

In order to overcome the problems of the traditional algorithm, such as the time-consuming execution of acquisition instructions, low signal tracking accuracy, and low signal capture accuracy, a global satellite positioning receiver acquisition and tracking algorithm based on UWB technology is designed in this study. On the basis of expounding the pulse generation method and working principle in UWB technology, this paper analyzes in detail the characteristics of UWB technology, such as antimultipath, low power consumption, and strong penetration. Then, on the basis of window function filtering, in the process of three-dimensional search of global satellite positioning signal, firstly, the satellite signal entering the GPS software receiver is processed by RF front-end mixing and AD sampling, and then, the signal tracking and navigation message solving are completed according to the relationship between the influence factor and Doppler frequency offset. The experimental results show that the execution time of the acquisition instruction of the proposed algorithm varies between 1129 ms and 1617 ms; the signal tracking accuracy ranges between 0.931 and 0.951, and the signal capture accuracy ranges between 93.3% and 95.6%, which proves that the proposed algorithm has achieved the design expectation.

1. Introduction

Global positioning system (GPS) has realized real-time, reliable, and accurate positioning services anywhere on the earth, and GPS receiver is the key to realize accurate and fast positioning [1]. Compared with the hardware receiver, the software receiver transplants all the signal processing process to the programmable microprocessor for implementation and has significant advantages such as high flexibility, strong expansibility, and convenient upgrade [2].

In general, GPS software receiver includes RF front-end module, ADC module, capture module, tracking module, and navigation message solving module. The capture module is the foundation of the following tracking module and the message solving module, which directly affects the performance of GPS software receiver [3]. The function of the acquisition module is to determine the visible satellite and roughly estimate the C/A code phase and Doppler frequency offset of the visible satellite GPS signal, which provides support for the follow-up signal tracking and position calculation. With the continuous expansion of the application range of GPS technology, GPS positioning technology under weak signal conditions (tunnel, forest, indoor, etc.) has received extensive attention, and GPS signal acquisition with high sensitivity has become a hot research issue.

For a specific visible satellite, GPS signal acquisition is a two-dimensional search problem in the signal Doppler frequency domain and the phase domain of C/A code [4]. Early GPS receivers usually adopt sliding correlation acquisition algorithm, which has a large amount of computation and a long acquisition time. To improve the capture speed, in reference [5], a new parallel fast acquisition algorithm is proposed for C/A code acquisition of GPS software receiver under weak signal conditions. In this method, the average correlation technique is used to reduce the autocorrelation loss of C/A codes, and the cumulative error of C/A codes is reduced based on the carrier frequency error compensation technique. The superposition correlation method is used to reduce the complexity of coherent accumulation operation, and the time consumption of frequency stepping

search is reduced by cyclic shift. Experimental results show that the algorithm can effectively achieve GPS signal acquisition. The comparative simulation results show that this algorithm can capture weak GPS signals with moderate computation and is suitable for high sensitivity software receivers. In the research of reference [6], in order to make full use of the GNSS receiver in the processing process, the prior information of the navigation filter is used to assist the acquisition of the baseband signal, which improves the acquisition speed and sensitivity. In this method, the concept of "virtual path" is introduced to calculate the phase of the pseudo code, and the algorithm of directly calculating the phase of the received pseudo code by navigation filter is derived in detail. Then, the error source and propagation characteristics of the pseudo code phase and Doppler frequency are analyzed in detail. However, the above two methods have the drawback of long capture instruction execution time.

In reference [7], a fast satellite signal acquisition algorithm based on Sparse Fast Fourier Transform (SFFT) is presented. Combined with the strong autocorrelation of satellite signal pseudo random code, the original estimation of SFFT amplitude is removed. By using the time-domain serial acquisition method, the local pseudo random code corresponding to the large coordinate point output in SFFT algorithm is correlated with the received satellite signal, and then, the satellite signal is captured. In order to solve the problem that the Fourier transform of the traditional parallel frequency acquisition algorithm in the Beidou satellite receiver needs to process a large amount of data and thus affects the acquisition speed of satellite signals, a fast Beidou signal acquisition algorithm based on coherent drop sampling is proposed. By using FPGA+DSP (high speed digital signal processor + field programmable logic gate array), the coherent drop sampling module is added into the traditional parallel frequency acquisition algorithm. After carrier stripping and pseudo code stripping, the amount of data to be processed by Fourier transform is reduced by reducing the sampling frequency, and then, the satellite signal is searched in three dimensions. However, in the practical application, it is found that the above two methods have the problems of low tracking accuracy and poor capture accuracy.

Although there have been many research results in this field, however, the current research results still have some shortcomings, such as high algorithm computational complexity and time complexity, and the research results can not be applied to the actual system. In order to solve these problems of the traditional methods, this paper designs a global satellite positioning receiver acquisition and tracking algorithm based on UWB technology. On the basis of expounding the pulse generation method and working principle in UWB technology, this paper analyzes in detail the characteristics of UWB technology, such as anti-multipath, low power consumption, and strong penetration. Then, on the basis of window function filtering, in the process of three-dimensional search of global satellite positioning signal, firstly, the satellite signal entering the GPS software receiver is processed by RF front-end mixing and AD sampling, and then, the signal tracking and navigation message solving are completed according to the relationship between the influence factor and Doppler frequency offset.

The main work of this paper can be concluded as follows:

(1) We study the problem of acquisition and tracking algorithm of global satellite positioning receiver based on UWB, which is now a very hot research point

(2) We proposed a new acquisition and tracking algorithm of global satellite positioning receiver. This new method is based on the UWB, and we analyze in detail the characteristics of UWB technology, such as anti-multipath, low power consumption, and strong penetration

The rest of this paper is organized as follows: Section 1 gives the relevant concepts. Section 3 gives the acquisition, and tracking algorithm of GPS receiver is designed based on UWB technology. Section 4 is the experiment and the result analysis. The conclusion is given in Section 5.

2. Relevant Concepts

2.1. GPS Technical Analysis. GPS, also known as the global positioning system (GPS), is a midrange circular orbit satellite navigation system that combines the technology of satellite and communication development and uses navigation satellites for time measurement and ranging [8]. GPS is a new generation of satellite navigation and positioning system developed by the United States from the 1970s, which lasted more than 20 years and cost 20 billion US dollars. It was fully completed in 1994 and has the ability to fully implement three-dimensional navigation and positioning in the sea, land, and air. After nearly ten years of use by surveying and mapping departments in China, the global positioning system has won the trust of surveying and mapping workers with the characteristics of all-weather, high precision, automation, and high efficiency and successfully applied to geodesy, engineering survey, aerial photogrammetry, vehicle navigation and control, crustal movement monitoring, engineering deformation monitoring, resource investigation, geodynamics, and other disciplines, so as to bring a profound technical revolution in the field of surveying and mapping [9].

The global positioning system is composed of a monitoring center and a mobile terminal, and the monitoring center is composed of a communication server and a monitoring terminal. The communication server is composed of the main controller and the GSM/GPRS receiving and sending module [10]. Among them, the mobile terminal is composed of GPS receiver, GSM transceiver module, main control module and external probe, etc. In fact, the GPS positioning system is a "3G" system composed of GSM, GPS, and GIS with high new technology. The structure of GPS receiver is divided into antenna unit and receiving unit. The global positioning system is shown in Figure 1.

The basic positioning principle of GPS is that the satellite continuously sends its own ephephis parameters and time

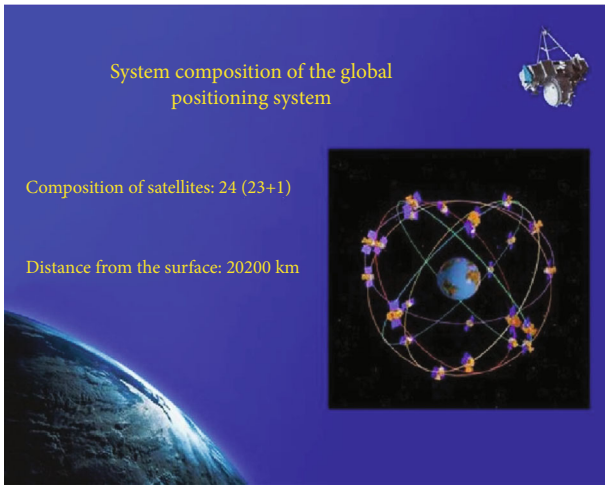


Figure 1: Composition of GPS.

information. After receiving these information, the user calculates the three-dimensional position, three-dimensional direction, and motion speed and time information of the receiver [11].

In recent years, China has built permanent GPS tracking stations in Beijing, Wuhan, Shanghai, Xi'an, Lhasa, Urumqi, etc., to carry out precise orbit determination of GPS satellites, provide observation data and precise ephephis services for high-precision GPS positioning and measurement, and commit to the establishment of China's independent wide-area differential GPS (WADGPS) scheme. Participate in the preparation of the Global Navigation Satellite System (GNSS) and GPS Enhancement System (WAAS). At the same time, China has started to build its own satellite navigation system (binary satellite positioning system), capable of producing navigation GPS receivers. The application of GPS technology is developing at a deeper level [12].

2.2. UWB Technical Analysis. Ultrawide band (UWB) is a wireless carrier communication technology that does not use a sinusoidal carrier but uses narrow pulses of nonsinusoidal waves at the nanosecond level to transmit data, thus covering a wide spectrum. UWB technology has the advantages of low system complexity, low power spectral density of transmitted signals, insensitivity to channel fading, low ability of interception, and high positioning accuracy, which is especially suitable for high-speed wireless access in dense multipath places such as indoor [13].

UWB technology began with the rise of pulse communication technology in the 1960s. UWB technology uses extremely wide spectrum of ultrawide baseband pulse to communicate, so it is also called baseband communication technology, wireless carrier communication technology, mainly used in military radar, and positioning and communication systems with low interception rate/low detection rate. Instead of using a sinusoidal carrier, UWB uses narrow pulses of nonsinusoidal waves at the nanosecond level to transmit data, so it covers a wide spectrum and, despite wireless communication, can transmit data at rates of several hundred megabits per second or more [14]. UWB technology can be used to transmit signals over very wide bandwidth, the United States Federal Communications Commission (FCC) regulations for UWB technology: in the 3.1 to 10.6 GHz band of 500 MHz or more bandwidth.

In essence, UWB is a carrier-free spectrum spreading technology using impulse pulse with very low duty cycle as information carrier. It directly modulates the impulse pulse with very steep rise and fall time. A typical UWB directly transmits an impulse train, which no longer has the traditional concepts of IF and RF. The transmitted signal can be regarded as either baseband signal (in terms of conventional radio) or RF signal (in terms of the spectral component of the transmitted signal).

The impulse pulse usually adopts a single period Gaussian pulse, and one bit of information can be mapped into hundreds of such pulses. Single-period pulses are nanosecond in width and have a wide spectrum. UWB has developed a new wireless channel with gigahertz capacity and the highest space capacity [15]. UWB pulse wireless transceiver based on CDMA generates a pulse sequence with a certain repetition period in the clock generator at the sending end. The information to be transmitted by the user and the pseudo random code representing the user's address are, respectively, or combined to modulate the above periodic pulse sequence in a certain way, and the modulated pulse sequence drives the pulse generation circuit. A pulse sequence of certain pulse shape and regularity is formed and then amplified to the required power and then coupled to the UWB antenna to transmit. At the receiving end, the signal received by the UWB antenna is amplified by a low noise amplifier and sent to one of the input ends of the correlator. At the other input end of the correlator, a locally generated pulse sequence modulated by the user pseudo random code synchronizes with the initiator is added. The receiving signal and the pulse train modulated by the local synchronous pseudo random code go through the multiplication, integration, and sampling and holding operations in the correlator together to generate a signal separated from the user's address information, which only contains the user's transmission information and other interferences, and then demodulate the signal [16].

3. The Acquisition and Tracking Algorithm of GPS Receiver Is Designed Based on UWB Technology

Based on the above analysis, this paper designs a GPS receiver acquisition and tracking algorithm based on UWB technology.

3.1. Application of UWB Technology. UWB communicates by sending an instantaneous spike wave or pulse wave. These pulses are limited to a width of less than 1 nm and emit pulses representing data in a very short period of time, enabling transmission rates of up to hundreds of megabytes.

There are two ways to generate a pulse source:

(a) *Photoelectric Method.* The basic principle is to obtain pulse signal by using the steep rising edge of the

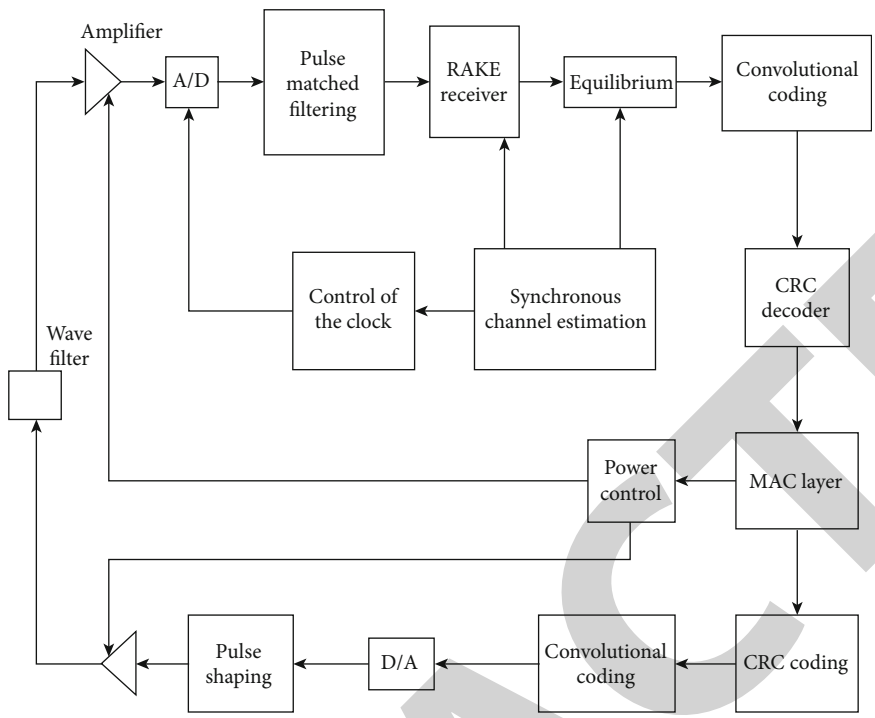


Figure 2: Structure diagram of UWB transceiver.

photoconductive switch. The pulse width obtained from the laser pulse signal can reach the order of picosecond (10-12), which is the most promising method [17]

(b) *Electronic Methods*. The basic principle is to reverse charge the semiconductor PN junction, so that it reaches the avalanche state, and at the moment of conduction, take the steep rising edge as a pulse signal. The very short pulse obtained after shaping is the most widely used scheme at present, but due to the limitation of transistor voltage resistance characteristics, this method can only produce tens to hundreds of volts of pulse; pulse width can reach 1 ns below

There are a variety of UWB pulses, among which the typical one is the Gaussian mono-periodic pulse, whose time-domain and frequency-domain formulas can be expressed as follows:

$$v(t) = 6A\sqrt{\frac{e\pi}{3\tau}} \times \frac{t}{\tau} \exp\left(-6\pi\left(\frac{t}{\tau}\right)^2\right), \quad (1)$$

$$v(f) = -j\left(\frac{2f^2}{3}\right)\sqrt{\frac{e\pi}{2}} \exp\left(\frac{\pi}{6f^2\tau^2}\right), \quad (2)$$

where A is pulse amplitude, τ is time delay length, t is time, and f is frequency. From the formula, we can know, center frequency $f = 1/\tau$, and half power bandwidth $W = f \times 1.16$. Thus, a single cycle pulse with a pulse width of 0.5 ns can be obtained. Its center frequency is 2 GHz, and half power bandwidth is about 2 GHz.

As shown in Figure 2, the structure of the transceiver is divided into baseband part and RF part. Because there is no complex RF and IF transformation, the analog signal processing of the UWB transceiver can be simplified. When the data is sent, the signal sent is coded and interwoven by the digital signal processor. After the digital signal is converted by D/A, the pulse is formed, amplified, and then transmitted.

The received pulse signal is directly converted to A/D through antenna matching, filtering, and amplification, and the complex baseband processing such as demodulation, matched filtering, diversity receiving, channel estimation, and equalization is handed over to the digital signal processor for processing. The whole system can be done on a chip. Of course, the design of UWB system also faces some challenges, such as the UWB system used in the ultrawide band antenna, ultranarrow pulse generation, low energy narrow pulse detection, A/D and D/A transformation accuracy and speed, and accurate timing of pulse reception. All need to be further studied.

Global positioning system is a dense multipath system. Therefore, combined with the characteristics of UWB technology, its application advantages in GPS network are as follows:

(a) *Multipath Resistance*. Multipath fading is a major obstacle to wireless communication. Traditional wireless technology is vulnerable to multipath interference, which makes wireless transmission characteristics worse, and it is difficult to accurately track the location with multipath dry disturbance. UWB signal has very high resolution due to its very high

frequency and very low duty cycle, and the resolution of different paths can be reduced to the order of ns, which is insensitive to channel multipath fading

For conventional wireless signal multipath fading depth of 10-30 dB, multipath environment for UWB signal fading less than 5 dB has an excellent antimultipath performance.

(b) The low power consumption of the system is conducive to intrinsically safe design. The UWB system does not use carrier and has a very low signal duty cycle (as low as 0.5%). The system consumes only a few hundred μW to a few tens of mW during high-speed communication, which is only 1/10-1/100 of the power consumption of existing radio stations

(c) *Strong Penetration*. The attenuation characteristics of UWB signal in the form of baseband narrow pulse are different from that of general wireless communication signal which carries information by sinusoidal carrier. The sinusoidal electromagnetic wave emitted by an antenna is a spherical wave whose attenuation in free space is inversely proportional to the square of the distance. However, the UWB transient pulse with appropriate waveform has strong orientation, and its attenuation is inversely proportional to or less than the distance. Therefore, at the same power, the UWB signal in the baseband narrow-pulse form can travel a longer distance than the ordinary modulated carrier signal

(d) The transmission distance can be increased by increasing the transmitting power. According to the characteristics of UWB technology, the application of UWB technology in GPS network can reduce the transmission rate to achieve long distance wireless access service in many occasions

3.2. Window Function Filtering. The main mathematical tool in digital signal processing is Fourier transform. The Fourier transform deals with the whole time domain and the frequency domain. However, when engineering test signal processing is realized by computer, it is impossible to measure and calculate the infinite length of signal but to analyze the finite time segments. The method is to intercept a time segment from the signal and then use the intercepted signal time segment for periodic extension processing, to get a virtual infinite long signal, and then, the signal can be Fourier transform, correlation analysis, and other mathematical processing. When a signal of infinite length is truncated, its spectrum is distorted, and the energy that was concentrated at $f(0)$ is spread over two wider bands (a phenomenon known as spectral energy leakage).

Window function is to realize the operation time reach the key steps of the linear, but using the traditional window function of truncation will cause the spectrum leakage, so in this paper, the window function by the Gaussian window function and the convolution of the rectangular window function make it in time domain and frequency domain energy can be concentrated, thus use of window function to the UWB signal filter processing.

The flat window function constructed in this study is expressed by g in the time domain and G in the frequency domain, and its parameter is $1/B, 1/2B, \delta, \omega$. K is signal sparsity, window width is $B = O(\sqrt{NK/\varepsilon \log(N/\delta)})$, and $\omega = O(B \log(N/\delta))$.

Sequence y can be obtained by multiplying $P_{\sigma,\tau}x$ and g. The process is as follows:

$$y = P_{\sigma,\tau}x \cdot g. \tag{3}$$

Different window functions have different effects on the signal spectrum, which is mainly because different window functions have different leakage sizes and different frequency resolution abilities. Signal truncation produces energy leakage, while FFT algorithm is used to calculate the spectrum and produces fence effect. In principle, these two errors cannot be eliminated, but we can suppress their effects by selecting different window functions. The main lobe of the rectangular window is narrow, the side lobe is large, the frequency identification accuracy is the highest, and the amplitude identification accuracy is the lowest. The main lobe of Blackman window is wide, the side lobe is small, and the frequency identification accuracy is the lowest, but the amplitude identification accuracy is the highest.

In the process of designing the filter, the given amplitude characteristic is usually piecewise constant, and there are discontinuity points at the frequency band boundary. The infinite impulse response (IIR) digital filter is needed to approximate this characteristic. In order to realize the finite impulse response (FIR), an appropriate and finite time wide window sequence can be selected to weight the infinite impulse response sequence, so as to constitute the finite impulse response digital filter to realize the given characteristics. Because the multiplication of two sequences in the time domain is equivalent to the convolution of the discrete Fourier transform of two sequences in the frequency domain, therefore, the convolution of the window function with a given filter property can achieve the goal of approximating the given property with a finite impulse response.

In order to improve the resolution, it is required to increase the length of recorded data when using the classical periodic graph to estimate the power spectrum, but the result will lead to the increase of the fluctuation of the periodic graph, that is, the increase of the estimation variance. In order to smooth the spectrum estimation and reduce the variance of the estimation, the data can be windowed, and then, the period graph estimation can be made to achieve the tradeoff between reducing the variance and improving the resolution.

3.3. Tracking Capture. The three-dimensional search process of global positioning signals is shown in Figure 3. According to this figure, we can get the search process of global positioning signals. In the capture range, we can use the code phase estimate and carrier frequency estimation to search the target.

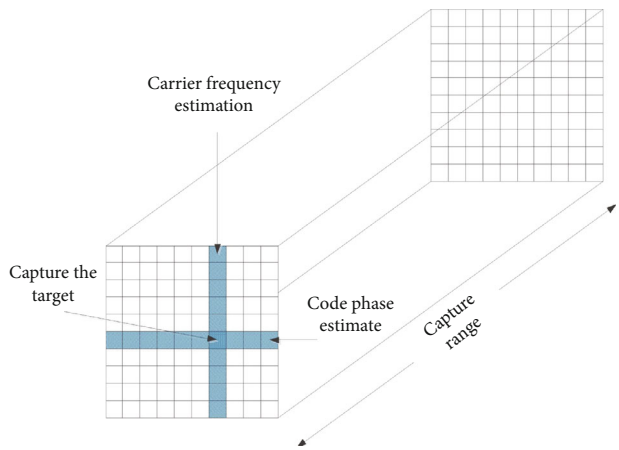


Figure 3: 3D search schematic diagram of global positioning signals.

After the RF front-end mixing processing and AD sampling processing of the satellite signal entering the GPS software receiver, the digital baseband signal can be expressed as

$$x(t_n) = qD_k(t_n)C_k(1+\eta)(t_n\text{-}t_\tau)\cos((\omega_{\mathrm{IF}} - \omega_D)t_n + \phi_0) + v(t_n). \tag{4}$$

Among them, q represents the signal amplitude, t_n represents the time series after sampling, $D_k(t_n)$ represents the GPS navigation message data sequence with an interval of 20 ms, and $C_k(t_n)$ represents the pseudo random coding of the satellite, namely, the C/A code to be captured. t_τ represents the initial phase of C/A code, η represents the influence factor of the phase of Doppler frequency offset C/A code, ω_{IF} represents the IF frequency output after mixing, ω_D represents the Doppler frequency offset generated by the relative motion between GPS receiver and satellite, ϕ_0 represents the initial phase, and $v(t_n)$ represents the additive white Gaussian noise.

Ignoring the influence of the ionosphere, there is the following relationship between the influence factor and Doppler frequency offset:

$$\eta = \frac{\omega_D}{2\pi \times 1575.42 \times 10^6}. \tag{5}$$

C/A code capture is to estimate the initial code phase and Doppler frequency offset ω_D according to the sampled digital base band signal, which supports subsequent signal tracking and navigation message solution; η is the influence factor.

For example, to capture the signal of the j-th satellite, the codirectional and orthogonal components of the IF signal of the satellite's C/A code are, respectively, copied in the software receiver:

$$I(t_n) = C_k(1+\widehat{\eta})\left(t_n - \widehat{t}_\tau\right)\cos(\omega_{\mathrm{IF}} - \widehat{\omega}_D)t_n, \tag{6}$$

$$Q(t_n) = \text{-}C_k(1+\widehat{\eta})(t_n - t_\tau)\sin(\omega_{\mathrm{IF}} - \widehat{\omega}_D)t_n. \tag{7}$$

Table 1: Experimental parameter table.

Parameter	The numerical
Intermediate frequency	1.40 MHz
Sampling frequency	5.71 MHz
Doppler frequency offset	1000 Hz
Code phase	900
Coherent integration time	5 ms
Window function parameter B	144
The sparse degree	4

In the formula, $C_k(1+\widehat{\eta})(t_n - t_\tau)$ represents the C/A code of local replication, $\cos(\omega_{\mathrm{IF}} - \widehat{\omega}_D)t_n$ and $\sin(\omega_{\mathrm{IF}} - \widehat{\omega}_D)t_n$ represent the codirectional carrier and orthogonal carrier of local replication, respectively, and $\widehat{\eta}$, $\widehat{t}_\tau$, and $\widehat{\omega}_D$ represent the estimated value of η, t_τ, and ω_D, respectively. Because of the coupling relationship between η and U, the ω_D code capture problem is a two-dimensional search problem about.

After correlating the locally replicated codirectional carrier and orthogonal carrier with the IF signal of GPS software receiver, the following can be obtained:

$$z_I(t_n) = \sum_{n=0}^{N-1} x(t_n)I(t_n), \tag{8}$$

$$z_Q(t_n) = \sum_{n=0}^{N-1} x(t_n)Q(t_n). \tag{9}$$

The two signals after correlation are composed of complex signals, and the process is as follows:

$$Z(t_n) = z_I(t_n) + jz_Q(t_n). \tag{10}$$

When $\widehat{t}_\tau = t_\tau$ and $\widehat{\omega}_D = \omega_D$, $Z(t_n)$ gets the maximum value. Therefore, GPS signal capture and tracking can be achieved by traversing t_τ and ω_D and comparing the maximum value of $Z(t_n)$ with the threshold.

4. Experiment and Result Analysis

4.1. Experimental Design. In order to verify the actual application performance of the acquisition and tracking algorithm for the global positioning receiver based on UWB technology designed in this study, the following simulation experiments are designed.

Experimental parameters are shown in Table 1.

In order to effectively solve the unilateral and one-sided problems of the experimental results, the paper compares the traditional GPS software receiver signal fast acquisition algorithm (algorithm of reference [5]) and the satellite signal acquisition algorithm based on the relevant SFFT (algorithm of reference [7]) and the fast acquisition algorithm of Beidou signal based on FPGA+DSP (algorithm of reference [8]).

Table 2: Statistical results of capture instruction execution time of different algorithms (ms).

Number of experiments/time	Algorithm of reference [5]	Algorithm of reference [7]	Algorithm of reference [18]	Algorithm of this paper
100	4240	3040	3799	1617
200	4421	3319	3842	1129
300	4325	3407	3695	1130
400	4021	4011	4005	1534
500	4217	4485	4016	1437

Table 3: Comparison of signal tracking accuracy of different algorithms.

Number of experiments/time	Algorithm of reference [5]	Algorithm of reference [7]	Algorithm of reference [18]	Algorithm of this paper
100	0.831	0.827	0.725	0.931
200	0.842	0.831	0.726	0.939
300	0.820	0.841	0.759	0.940
400	0.832	0.881	0.772	0.944
500	0.855	0.910	0.786	0.951

4.2. The Experiment Indicators

(1) *Capture Instruction Execution Time*. This indicator can reflect the working efficiency of different algorithms. The less time it takes to execute the capture command, the higher the efficiency of the algorithm, and the GPS receiver can capture the positioning signal in a shorter time. On the contrary, the more time it takes to execute the capture instruction, the lower the efficiency of the algorithm

(2) *Signal Tracking Accuracy*. This index can reflect the reliability of different algorithms. The value range of signal tracking accuracy is [0,1]. The closer the signal tracking accuracy is to 1, the higher the reliability of the algorithm is

(3) *Accuracy of Signal Capture*. this index can directly reflect the effectiveness of different algorithms. The higher the accuracy of signal capture, the higher the effectiveness of the algorithm, and the more effective the satellite positioning signal can be obtained. On the contrary, the lower the accuracy of signal capture, the lower the effectiveness of the algorithm

4.3. Experimental Results and Analysis

4.3.1. Capture Instruction Execution Time. First, the capture instruction execution time of different algorithms was counted, and the results are shown in Table 2.

By comparing the results in Table 2, it can be seen that after the application of algorithm of reference [5], with the increase in the number of experiments, the execution time of capture instruction varies between 4021 ms and 4421 ms. With the application of algorithm of reference [7], the execution time of capture instruction varies between 3040 ms and 4485 ms with the increase of the number of experiments. After the application of algorithm of reference [18], with the increase of the number of experiments, the execution time of capture instruction varies between 3695 ms and 4016 ms. However, with the application of algorithm of this paper, the execution time of capture instruction varies between 1129 ms and 1617 ms with the increase of the number of experiments.

Through the comparison of the above experimental results, it can be seen that the algorithm of this paper takes less time to execute the capture instruction, indicating that the higher the efficiency of the algorithm, the shorter the time the global satellite positioning receiver can capture the positioning signal.

4.3.2. Signal Tracking Accuracy. By comparing the signal tracking accuracy of different algorithms, the comparison results are shown in Table 3.

By analyzing the experimental results shown in Table 3, it can be seen that the signal tracking accuracy of the algorithm of reference [5] is between 0.831 and 0.855 with the increase of the number of experiments. The signal tracking accuracy of the algorithm of reference [7] is between 0.827 and 0.910. Algorithm of reference [18], whose signal tracking accuracy is between 0.725 and 0.786. The signal tracking accuracy of the algorithm of this paper is between 0.931 and 0.951.

Compared with the three traditional algorithms, the signal tracking accuracy of the algorithm of this paper is closer to 1. Therefore, it can be seen that the acquisition and tracking algorithm of global positioning receiver based on UWB technology designed in this paper is more reliable.

4.3.3. Accuracy of Signal Capture. The accuracy of signal capture of different algorithms was counted, and the comparison results are shown in Figure 4.

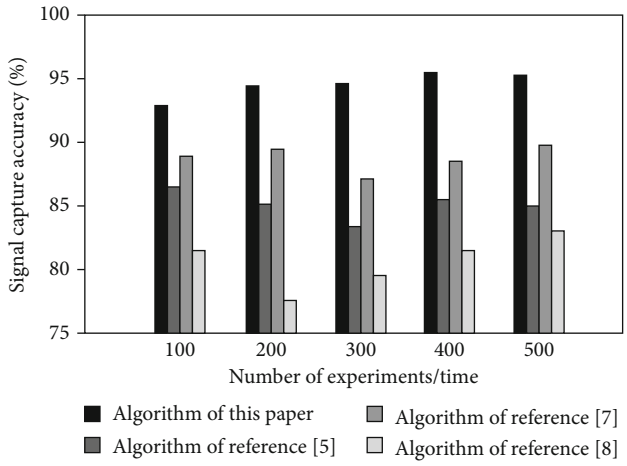


Figure 4: Comparison of signal capture accuracy of different algorithms (%).

By analyzing the experimental results shown in Figure 4, it can be seen that the accuracy of defect feature extraction by different methods keeps increasing. The accuracy of the signal capture for the algorithm of reference [5] ranges from 82.5% to 86.8%, and the accuracy of the algorithm of reference [7] ranges from 84.7% to 90.1%. The signal capture accuracy of the algorithm of reference [18] ranges from 77.5% to 83.0%, while that of the algorithm of this paper ranges from 93.3% to 95.6%.

Compared with the three traditional algorithms, the algorithm of this paper has higher signal capture accuracy. Therefore, it can be concluded that the acquisition and tracking algorithm of global positioning receiver based on UWB technology designed in this study can obtain more effective satellite positioning signals.

To sum up, the acquisition and tracking algorithm for global satellite positioning receiver based on UWB technology designed in this study has the advantages of less time consuming in the execution of capture instructions, high signal tracking accuracy, and high signal capture accuracy and is suitable for promotion.

5. Conclusion

In order to solve the problems in the traditional signal capture and tracking algorithm, this paper proposed a global satellite positioning receiver acquisition and tracking algorithm based on UWB technology. The main work of this paper is to design the window function filtering process after analyzing the multipath resistance, low power consumption, and strong penetration characteristics of UWB technology in detail. In the process of 3D search of GPS positioning signals, the satellite signals that enter the GPS software receiver are firstly processed by RF front-end mixing and AD sampling, and then, the signal tracking and navigation message solving are completed according to the relationship between influence factors and Doppler frequency offset.

After completing the algorithm design, this study also carried out experiments. The experimental results show that the execution time of the acquisition instruction of the proposed algorithm varies between 1129 ms and 1617 ms, the signal tracking accuracy ranges between 0.931 and 0.951, and the signal capture accuracy ranges between 93.3% and 95.6%, which fully proves the effectiveness of the proposed algorithm.

In the future, we will continue to study other acquisition and tracking algorithm of global satellite positioning receiver and develop practical application systems as soon as possible.

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 conflict of interest.

References

[1] F. Farzan and L. René, "Multi-constellation software-defined receiver for Doppler positioning with LEO satellites," *Sensors*, vol. 20, no. 20, pp. 5866–5884, 2020.

[2] L. I. U. Xu, Y. A. O. Zheng, L. Y. U. Hongli, and L. U. Mingquan, "SIC-based anti near-far effect acquisition method for pseudolites systems," *Gnss World of China*, vol. 45, no. 1, pp. 12–18, 2020.

[3] M. Li, H. Yu, and P. Liu, "An automated safety risk recognition mechanism for underground construction at the pre-construction stage based on BIM," *Automation in Construction*, vol. 91, no. 28, pp. 284–292, 2018.

[4] K. Rouabah, S. Atia, M. Flissi, M. S. Bouhlel, and S. Mezaache, "Efficient technique for DLL S -curve side zero-crossings cancellation in global positioning system/Galileo receiver," *IET Signal Processing*, vol. 13, no. 3, pp. 338–347, 2019.

[5] G. U. Jianhua, Y. A. N. Guojun, and Y. A. N. G. Jiudong, "High sensitivity fast acquisition algorithm for GPS software receiver signal," *Fire Control & Command Control*, vol. 44, no. 6, pp. 67–71, 2019.

[6] Z. H. A. O. Lin, L. U. O. Zhibin, D. I. N. G. Jicheng, and W. U. Mouyan, "Signal acquisition technique aided by navigation filter in GNSS receiver," *Journal of Harbin Institute of Technology*, vol. 52, no. 3, pp. 165–172, 2020.

[7] D. Lu and L. Yali, "Satellite signal acquisition algorithm based on correlated SFFT," *Journal of Signal Processing*, vol. 36, no. 8, pp. 1227–1233, 2020.

[8] S. P. Rana, M. Dey, and M. Ghavami, "ITERATOR: a 3D gait identification from IR-UWB technology," in *The 41st IEEE Engineering in Medicine & Biology Society (EMBC) (EMBC*, IEEE, 2019.

[9] J. I. Kaiyuan, Z. H. A. N. G. Boya, and J. I. A. N. G. Changhui, "A GNSS carrier tracking algorithm utilizing support vector machine," *Electronics Optics & Control*, vol. 26, no. 9, 2019.

[10] L. U. O. Wenjie, Z. H. A. N. G. Xiyue, and L. I. U. Xinglong, "Simulation of multi-path suppression algorithm for civil aviation satellite navigation jamming signal," *Computer Simulation*, vol. 37, no. 3, pp. 47–51, 2020.

[11] C.-H. Chen, F. Song, F.-J. Hwang, and L. Wu, "A probability density function generator based on neural networks," *Physica*

Hindawi
Wireless Communications and Mobile Computing
Volume 2023, Article ID 9836176, 1 page
https://doi.org/10.1155/2023/9836176

Retraction

Retracted: Study on Regional Control of Tourism Flow Based on Fuzzy Theory

Wireless Communications and Mobile Computing

Received 11 July 2023; Accepted 11 July 2023; Published 12 July 2023

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] W. Shang and G. Chuangle, "Study on Regional Control of Tourism Flow Based on Fuzzy Theory," *Wireless Communications and Mobile Computing*, vol. 2021, Article ID 9648879, 7 pages, 2021.

Hindawi
Wireless Communications and Mobile Computing
Volume 2021, Article ID 9648879, 7 pages
https://doi.org/10.1155/2021/9648879

Research Article

Study on Regional Control of Tourism Flow Based on Fuzzy Theory

Wei Shang and Guo Chuangle

Chengdu University of Information Technology, Chengdu 610200, China

Correspondence should be addressed to Guo Chuangle; guochuangle001@chd.edu.cn

Received 12 July 2021; Revised 26 July 2021; Accepted 12 August 2021; Published 19 August 2021

Academic Editor: Chi-Hua Chen

In order to solve the problems of poor regional control effect and high control difficulty coefficient of a traditional tourism flow, this paper puts forward the research of a regional control of tourism flow based on fuzzy theory. The capacity of regional tourism is determined by analyzing the factors that influence the regional control of the tourism flow. The regional tourism flow is divided into different time series by automatic clustering algorithm, the same sample data is fused, and the Euclidean distance between traffic is obtained. The regional tourism flow prediction model is constructed according to fuzzy theory. On this basis, the real-time capacity of regional scenic spot flow is calculated, and the regional tourist flow control model is constructed to realize the regional tourist capacity control. The experimental results show that the regional control error of tourism flow is always lower than 0.40, and the difficulty coefficient of control is low, which has certain advantages.

1. Introduction

Global tourism has developed rapidly in recent years. More and more countries will speed up the development of tourism as a strategic decision, and our country will position tourism as a strategic pillar industry and modern service industry to cultivate it and issued the Tourism Law to ensure and promote the sustainable and healthy development of tourism [1]. Tourism's share of GDP continues to rise. The focus of the development of world tourism is gradually moving eastward, and the tourism market in Asia, especially in China, accounts for more and more of the global tourism market. Worldwide, the external environment of tourism in China is further optimized, and the development trend of popular tourism is becoming more and more obvious. At the same time, tourism has become a hot field of industrial investment. Tourism is called "the promotion catalyst of national economy," " forever sunrise industry," and the demand for tourism service of our people is gradually increasing [2]. On the one hand, the development of tourism can meet the increasing material and cultural needs of people; on the other hand, it can directly or indirectly promote the development of national economy. At the same time, tourism is a very related industry. In the process of its operation, it can not only promote the development of six tourism characteristic industries, such as travel, housing, purchase, entertainment, food, and travel, but also promote the development of a series of industries, such as business and service industry. Regional tourism demand has become the foundation of tourism development [3].

At present, the regional tourism demand is rising. If the forecast result of tourism demand is not accurate, it will lead to waste of resources and repeated construction. The phenomenon of low tourism service quality and excessive reception load caused by tourism demand is [4]. The accurate forecast of tourism demand is the basic basis and reference of tourism policy making and investment development, and also the fundamental of regional control of tourism flow [5]. Therefore, the related research has carried on a lot of research and has obtained certain results.

A tourism flow prediction model based on a gradient lifting regression tree is proposed in [6]. The accurate

prediction of tourist flow is the key problem in tourism economic analysis and development planning. Based on the idea of integrated learning, a tourism flow prediction model based on gradient lifting regression tree is proposed. In this model, the tree generation algorithm of the original model is optimized to minimize the nonanalytical solution of the objective function, and the person correlation coefficient is used to analyze the correlation of each influencing factor to construct the feature vector to predict the tourist flow accurately. Taking Guilin tourist flow from 2015 to 2018 as an example, the prediction accuracy of the exponential smoothing algorithm and support vector machine algorithm is analyzed by comparing the average error, mean square error, and error. This method has higher prediction accuracy and better application value in tourist flow prediction, but little consideration is given to regional tourist flow control. Literature [7] puts forward the real-time tracking and prediction method of tourist flow data in scenic spots under cloud computing. This method can accurately track and predict the tourist flow data of scenic spots in real time. It is necessary to consider the continuous passenger flow state of scenic spots under cloud computing and complete the real-time tracking of tourist flow data through the state equation of scattered passenger flow data. The traditional real-time tracking method of tourist flow data in scenic spots does not consider the state of continuous passenger flow in scenic spots, which leads to poor tracking accuracy. The real-time tracking and prediction method of tourist flow data in scenic spots under cloud computing is put forward. The state parameters of continuous passenger flow under cloud computing are modeled, the unidirectional passenger flow and the aggregation passenger flow are estimated effectively, and the state equation and observation equation of dispersed passenger flow data are obtained. Based on the adaptive Kalman filter algorithm to monitor the dynamic change of scattered passenger flow data in cloud computing, and to correct the system state noise and observation noise variance, finally, on the basis of data prediction sorting and similar clustering, the effective tracking of tourist traffic data in scenic spots under cloud computing is realized. The method has higher tracking accuracy and stable and reliable performance, but the control and prediction are time-consuming and have some limitations.

Mobile computing is a new technology emerging with the development of mobile communication, Internet, database, distributed computing, and other technologies. Mobile computing technology will enable computers or other information intelligent terminal devices to realize data transmission and resource sharing in the wireless environment. Its role is to provide useful, accurate, and timely information to any customer anytime and anywhere. This will greatly change the way people live and work.

Based on the shortcomings of the above methods, this paper proposes a regional control method of tourism flow based on fuzzy theory. By analyzing the factors that affect the regional control of tourism flow, the capacity of regional tourism is determined, the regional tourism flow is divided into different time series by automatic clustering algorithm, the same sample data is fused, and the Euclidean distance between traffic is obtained. The technical route of this method is as follows:

(1) To determine the capacity of regional tourism by analyzing the factors affecting the regional control of tourism flows

(2) The regional tourism traffic is divided into different time series by automatic clustering algorithm, the same sample data is fused, and the Euclidean distance between traffic is obtained, and the regional tourism traffic prediction model is constructed according to fuzzy theory

(3) On this basis, the real-time traffic capacity of regional scenic spots is calculated, and the regional tourist flow control model is constructed to realize the regional tourist capacity control

(4) Experimental analysis

(5) Concluding remarks

2. Regional Tourism Demand and Capacity Analysis

2.1. Analysis of Factors Influencing Regional Tourism. Due to time constraints, regional tourism has become the most popular choice in current tourism. Before controlling the regional tourism flow, it is necessary to analyze the influencing factors of regional tourism. On this basis, the regional tourism capacity is divided to lay the foundation for the subsequent regional control of tourism flow [8]. Regional tourism impact factors include the following:

(1) Seasonal factors

Seasonality is a prominent feature of the tourism industry. The phenomenon of seasonal characteristics in tourism industry is not only limited to individual countries or regions to fully understand this important feature of tourism industry but also the premise of rational development and utilization of tourism resources. Seasonal characteristics are the part that needs to be attached great importance to in the process of tourist volume prediction. For a long time, how to grasp the seasonal volatility of tourist volume in the process of tourist volume analysis has been an important and complex issue. In the literature of tourist volume analysis, there are two ways to deal with the seasonal variation of tourist volume, that is, it is regarded as random or definite. As to which of the two methods will produce more accurate prediction results in the process of tourist volume prediction, there is no conclusion at present [9]. There are many complicated reasons behind the phenomenon that the tourist volume fluctuates with the season, including the reason of the tourist destination itself. It also includes the reasons from tourists and the comprehensive role of the whole tourism industry chain. Therefore, in the process of tourist volume prediction, we may need to consider the seasonal characteristics of tourism in different regions.

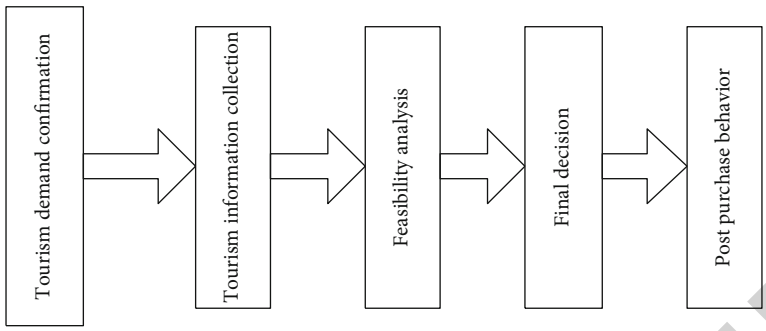


Figure 1: Traveller's regional tourism decision-making process.

(2) Sudden events

The impact of various unexpected special events on local tourism is very significant. For example, during natural disasters such as earthquakes and tsunamis, during severe outbreaks or during terrorist attacks and people's riots, tourism will suffer heavy losses due to these unpredictable emergencies. These natural or perceived events have caused huge losses to the number of visitors. The method used to measure the specific losses caused by these sudden events is to model and analyze the data before the event occurs, and to predict the local tourist volume under normal conditions. Then, compared with the actual observations at the time of occurrence, the difference between them is regarded as the loss caused by the event to the local tourist volume.

(3) Policy change

With the development of global international trade, economic and political exchanges between countries are frequent. In this context, tourism is also ushered in more foreign tourists. Countries' economic, political, and other related policies have a real-time impact on domestic tourism development. Therefore, this influence factor also becomes the important influence factor which affects the tourism industry progress and the development.

2.2. Regional Tourism Capacity Analysis. The so-called tourism market usually refers to the tourism demand market or the tourist source market, that is, the frequent buyers and potential buyers of tourism products: tourism consumption refers to people in the process of playing, the act and activity of meeting the needs of individual enjoyment and development by purchasing one or some tourism products. Tourism products are not stored as tangible products that are common in everyday life. They are essentially "perishable," [10] for example, hotel rooms, flight seats, and ticket tickets at a particular time and place will lose their current product value if they are not sold in time. Therefore, for the tourism market, it is very important to study and analyze the consumer's purchase behavior and make accurate tourism demand forecast.

A consumer purchase behavior model is an important theoretical reference for marketing activities. Through this model, we can determine the factors that affect consumers' decisions before and during the purchase of goods or services. In order to help marketers better carry out sales work, the decision-making process of consumer purchase is one of the important components of this behavior model and the decision-making process of tourists' regional tourism as shown in Figure 1.

When tourists make regional tourism decisions, they first "confirm the demand," and the purchase behavior will not occur for no reason. It must be because of some reason or stimulation that the demand for a certain commodity or service will trigger the desire to buy. The inducement may be accumulated for a long time, or may be temporarily stimulated by the behavior of acquaintances and friends, the effect of advertising, etc., such as seeing the photos of friends playing on social networks, suddenly producing the desire to travel outside. Then, we "collect information "; most people have to collect information to help make decisions before making decisions, information is the basis for decision-making, and purchase decisions are the same. Especially when it comes to larger amounts of money, consumers want to have a comprehensive understanding of information about goods or services, or to consult their own information, or to ask for advice from family and friends, or to study and analyze different package preferences, and so on. The larger the amount, the more detailed the information collection is. The characteristics of tourism services (inseparability, quality differences, family names, nonstorage, and nontransferability of ownership) determine that tourism consumption decisions are usually accompanied by higher risks, so consumers are more likely to conduct extensive information searches to reduce risk; again," evaluate goods ", which is based on the various types of information collected in the previous link, aggregate and further compare them, thus forming a consumer perspective and evaluation of a commodity or service, deciding whether to enter the next link or to give up buying. The final "decision to buy," that is, after confirming the availability of goods or services, occurs naturally, but this does not mean that the purchase is terminated [11].

When the regional tourism consumption of the above tourists is terminated, the regional tourist attractions need to plan the capacity to realize the control of the regional tourism capacity. The capacity of regional tourist attractions analyzed in this paper is shown in Figure 2.

The tourism capacity is the total capacity of the regional tourist attractions, and the composition of the capacity set to Z includes three parts of the influence factors, which are set

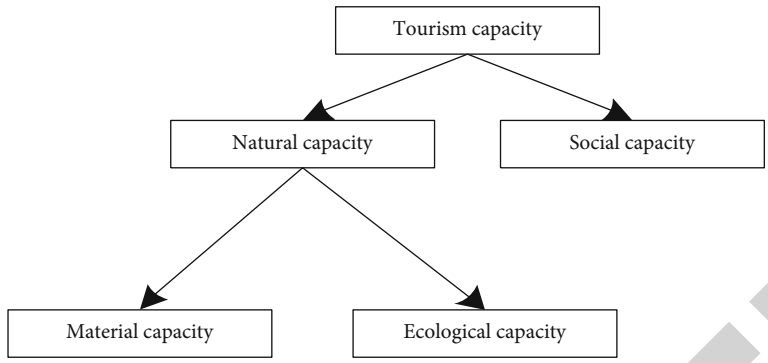


Figure 2: The capacity of regional tourist attractions analyzed.

to P, E, and S, respectively. The calculation formula can be expressed as

$$Z = P + E + S. \tag{1}$$

The content of formula (1) is not a simple addition, indicating that there is a vector relationship between the three, which is an inclusive relationship. Natural capacity is a supply-oriented capacity calculation, which mainly includes basic material capacity and natural ecological environment capacity. Among them, material capacity is a kind of facility capacity, according to the facility capacity of regional tourism to regional tourism. Social capacity is a kind of demand-oriented capacity, which is constructed on the psychological capacity of tourists and residents in tourist areas [12].

3. Regional Control of Tourism Flows

3.1. Construction of Regional Prediction Model of Tourism Flow Based on Fuzzy Theory. In order to realize the regional control of tourism flow, a regional prediction model of tourism flow is constructed to determine the flow of the tourism area.

First, the regional tourism traffic is divided into different time series by automatic clustering algorithm, and the historical sample is set as

$$y_i = (y_1, y_2, \cdots y_n). \tag{2}$$

By reordering the samples in formula (2), the same sample data is fused and [13], and the sorted sample data are expressed as follows:

$$y_i^{'} = \left(y_1^{'}, y_2^{'}, \cdots y_n^{'}\right). \tag{3}$$

On this basis, the mean value of the sample data is obtained, namely,

$$\text{arv} = \frac{\sum_{i=1}^{n}\left(y_{i+1} - y_i^{'}\right)}{n-1}. \tag{4}$$

Then, the Euclidean distance between different sample data is calculated, that is,

$$d\left(x_i, x_j\right) = \sqrt{\left(x_i - x_j\right)^2\left(x_i - x_j\right)}. \tag{5}$$

According to formula (5), the clustering center can be obtained as

$$c_j = \frac{1}{n}\sum_{i=1} d. \tag{6}$$

Finally, the regional prediction model of tourism flow is constructed according to fuzzy theory. A logical relation existing in fuzzy theory sets the interval between sample data as a set of fuzzy theories; fuzzy processing of sample travel flow data [14] obtained is as follows:

$$\begin{cases} A_1 = \dfrac{f_{11}}{u_1} + \dfrac{f_{12}}{u_2} + \cdots \dfrac{f_{1n}}{u_n}, \\ A_2 = \dfrac{f_{21}}{u_2} + \dfrac{f_{22}}{u_2} + \cdots \dfrac{f_{2n}}{u_n}, \\ \cdots. \\ A_i = \dfrac{f_{i1}}{u_1} + \dfrac{f_{i2}}{u_2} + \cdots \dfrac{f_{\text{in}}}{u_n}. \end{cases} \tag{7}$$

After the above fuzzy processing, the logical relationship in the sample data of tourism flow is determined, and on this basis, the prediction model is constructed, that is,

$$A_j = \frac{p + M[A_i]}{s+1}. \tag{8}$$

Construction of the regional prediction model of tourism flow based on fuzzy theory is shown in Figure 3

3.2. Regional Tourist Capacity Control. According to the prediction results of the regional tourism flow prediction model, the control method of regional tourist capacity is designed to achieve the expected goal. For regional tourist attractions, the real-time control of capacity should be achieved, which requires scenic spots to predict the daily capacity. Setting up the entrance capacity of regional tourist attractions, that is, the main factor limiting ticket sales, is calculated by formula (9) and obtained [15]:

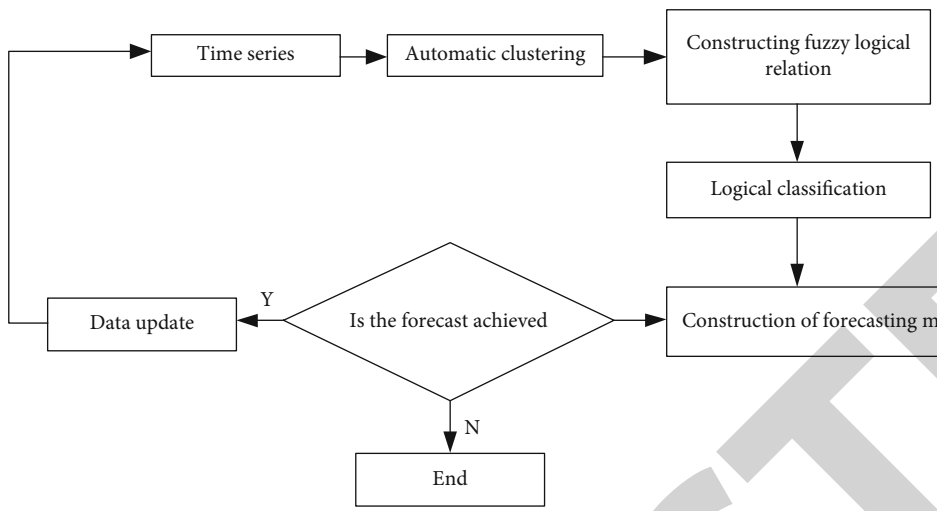


Figure 3: Flow of regional prediction model of tourism flow based on fuzzy theory.

$$T_i = \frac{(t - t_1)}{t_t \times F}. \quad (9)$$

In the formula, T_i represents the total entrance capacity of regional scenic spots, F represents the total number of tickets sold, and t_t represents the per capita flow of a single entry.

In the regional tourist capacity control, the key factor determining its capacity is the main play items in the scenic area, and the mathematical formula is as follows:

$$R_i = \frac{s_1}{s} \times \frac{t_{t-1}}{R_1 + R_2}. \quad (10)$$

In the formula, R_i represent the tourism facility application market, s_1 represents the landscape area, and R_1 represents the length of the tour.

According to the above analysis, [16] of regional tourist flow control model is constructed, and the following is obtained:

$$U = \frac{t_t - R_i}{T_i}. \quad (11)$$

In the regional tourist capacity control, according to the flow prediction model constructed by fuzzy theory, the real-time capacity of regional scenic spot flow is calculated, and the regional tourist flow control model is constructed to realize the regional tourist capacity control [17].

4. Experimental Analysis

4.1. Design of Experimental Scheme. In order to verify the scientific effectiveness of the proposed method, an experimental analysis was carried out. On the basis of the correlation analysis of real-time passenger flow data in a regional scenic spot, the data values with relatively high correlation degree are selected as the target research objects. Four data parameters in representative group 3 were selected as the main research objects. The experiment is carried out on the MATLAB platform. The experimental operating system is a WINDOWS XP system, and the system runs 16 GB of memory.

Table 1: Experimental parameters.

Parameter	Short-cut process
Regional passenger flow (h)	500
Sampling interval (min)	5
Data statistics software	SPSS 7.0
Sample data volume	3000
Number of iterations (times)	10

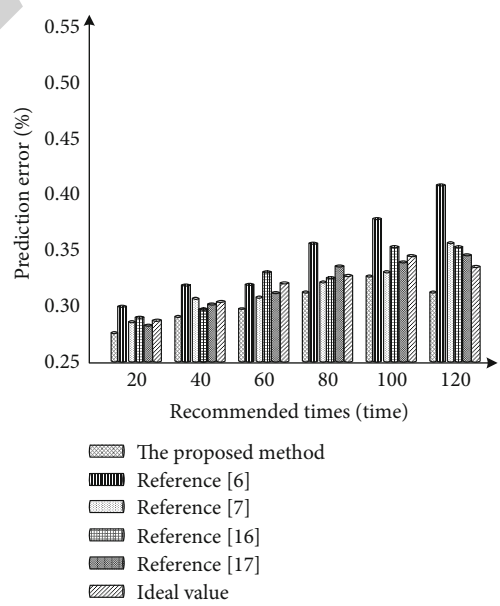


Figure 4: Comparison of regional traffic data prediction error in different methods.

4.2. Design of Experimental Parameters. In order to verify the effectiveness of the proposed method, it is shown in Table 1.

Table 2: Comparison of the regional control coefficient of tourism flow with different methods.

Number of controls (times)	Methods of this paper	Document [6] methodology	Document [7] methodology	Document [18] methodology	Document [19] methodology
20	0.11	0.52	0.63	0.45	0.39
40	0.13	0.42	0.68	0.49	0.48
60	0.14	0.53	0.71	0.58	0.51
80	0.13	0.54	0.65	0.62	0.57
100	0.12	0.57	0.62	0.59	0.59

Under the above experimental environment and experimental parameter setting, the proposed method, literature [6] method, and literature [7] method are used to control the sample data, and the accuracy of regional flow prediction and the difficulty coefficient of control are taken as the experimental indexes. The prediction error is calculated by the following formula:

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{c=1}^{N}(y_i - y\wedge_i)^2}. \tag{12}$$

In the formula, $\widehat{y}_t$ and $\widehat{y}_i$ represent two different observations of passenger flow and $\bar{y}_t$represents the average measured passenger flow.

4.3. Result. To verify the validity of the proposed method, the error of the proposed method, the literature [6] method, and the literature [7] method in predicting the sample flow data is analyzed experimentally as shown in Figure 4:

The analysis in Figure 4 shows that the error of sample data prediction by using the proposed method, literature [6] method, and literature [7] method. When the number of iterations is 40, the prediction error of the proposed method is about 0.29, which is lower than the ideal error value. The prediction error of the literature [6] method is about 0.33, and the prediction error of the literature [7] method is about 0.32. When the number of iterations is 80, the prediction error of the proposed method is about 0.31, which is lower than the ideal error value. The prediction error of the literature [18] method is about 0.38, and the prediction error of the literature [19] method is about 0.33. When the number of iterations is 100, the prediction error of the proposed method is about 0.30, which is lower than the ideal error value. The prediction error of the literature [18] method is about 0.45, and the prediction error of the literature [19] method is about 0.34. In contrast, the prediction error of the proposed method is low, which is due to the fact that the proposed method determines the capacity of regional tourism by analyzing the factors that affect the regional control of tourism flow. The regional tourism flow is divided into different time series by automatic clustering algorithm, the same sample data is fused, and the Euclidean distance between traffic is obtained. The regional tourism flow prediction model is constructed according to fuzzy theory, in order to improve the accuracy of the proposed method.

Table 3: Comparison of RMSE with different methods.

Methods	RMSE
Methods of this paper	0.56
Document [6] methodology	0.72
Document [7] methodology	0.65
Document [18] methodology	0.71
Document [19] methodology	0.61

To further verify the effectiveness of the proposed method, the experimental analysis of the proposed method, the literature [6] method, and the literature [7] method on the tourism flow regional control coefficient, the value of the control coefficient range between 0 and 1, where the lower value represents the better control effect; the results are shown in Table 2:

The analysis of the data in Table 2 shows that under the same experimental environment, there are some differences in the difficulty coefficient of sample data control using the proposed method, literature [6] method, and literature [7] method. When the control times are 40 times, the control difficulty coefficient of the proposed method is 0.13, the control difficulty coefficient of the literature [6] method is 0.42, and the control difficulty coefficient of the literature [7] method is 0.68. When the number of iterations is 80, the control difficulty coefficient of the proposed method is 0.13, the control difficulty coefficient of the literature [18] method is 0.62, and the control difficulty coefficient of the literature [19] method is 0.57. When the number of iterations is 100, the control difficulty coefficient of the proposed method is 0.12, the control difficulty coefficient of the literature [18] method is 0.59, and the control difficulty coefficient of the literature [19] method is 0.59. It has some advantages. In addition, the RMSE of each comparison method is given in Table 3.

5. Conclusion

In view of the shortcomings of traditional methods, this paper puts forward a regional control study of tourism flow based on fuzzy theory. The capacity of regional tourism is determined by analyzing the factors that influence the regional control of tourism flow. The regional tourism flow is divided into different time series by automatic clustering algorithm, the same sample data is fused, and the Euclidean distance between traffic is obtained. The regional tourism flow prediction model is constructed according to fuzzy

Hindawi
Wireless Communications and Mobile Computing
Volume 2023, Article ID 9838651, 1 page
https://doi.org/10.1155/2023/9838651

Retraction

Retracted: Analysis of Feasibility Modeling of Multimode Physical Fitness Training for Long-Distance Runners

Wireless Communications and Mobile Computing

Received 11 July 2023; Accepted 11 July 2023; Published 12 July 2023

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] X. Xia, "Analysis of Feasibility Modeling of Multimode Physical Fitness Training for Long-Distance Runners," *Wireless Communications and Mobile Computing*, vol. 2021, Article ID 5817396, 8 pages, 2021.

Hindawi
Wireless Communications and Mobile Computing
Volume 2021, Article ID 5817396, 8 pages
https://doi.org/10.1155/2021/5817396

Research Article

Analysis of Feasibility Modeling of Multimode Physical Fitness Training for Long-Distance Runners

Xu Xia

Hunan University of Humanities, Science and Technology, Hunan 417000, China

Correspondence should be addressed to Xu Xia; 164104134@stu.cuz.edu.cn

Received 22 June 2021; Revised 24 July 2021; Accepted 4 August 2021; Published 15 August 2021

Academic Editor: Chi-Hua Chen

Because the traditional multimode feasibility modeling analysis method of physical fitness training for long-distance runners has the problems of long modeling time and low modeling accuracy, a new multimode feasibility modeling analysis method for physical fitness training of long-distance runners is proposed. The improved LLE (local linear embedding) method is used to reduce the dimensionality of the training data for the physical fitness of long-distance runners. According to the processing results, the information theory is used to analyze the information content of the physical fitness training features of the long-distance runners, the information entropy of each feature is calculated, and the long-distance runners are extracted. Athlete's physical fitness enhancement training characteristics, combined with quantitative regression analysis method to carry out the information regression analysis of the long-distance runners' physical fitness training multimode statistical sequence, construct the feasibility evaluation model of the long-distance runners' physical fitness training multimode and complete the feasibility of the long-distance runners' physical fitness training multimode feasibility study Mode analysis. The simulation experiment results show that the proposed method has higher accuracy and shorter modeling time for multimode feasibility modeling of physical fitness training for long-distance runners.

1. Introduction

In long-distance running, aerobic oxidation is the main body energy metabolism of athletes, supplemented by anaerobic metabolism. The training period is long, and the load and intensity of training and competition are relatively large. Long-distance running has higher requirements for athletes' physical level, and compared with ball or other collective events, the requirements for skills and tactics are relatively low [1]. Therefore, whether for professional or amateur long-distance running training, athletes' fitness enhancement training and fitness evaluation have become an important part of training. Only when athletes have a good physical fitness level can they ensure good results in training and competitions. Long-distance runners not only need to have a good endurance level and certain technical and tactical skills but also need to have good physical fitness. In recent years, with the improvement and improvement of physical fitness training models, new requirements and challenges have been put forward for the physical fitness training of long-distance runners. For long-distance runners, due to their own conditions and exercise choices, they will affect the improvement of physical training level in long-distance running training [2]. In addition, due to the training system, the training time and training load of Chinese long-distance runners are usually higher than those of athletes from other countries. However, due to the insufficient number of physical coaches and team doctors, and the deviation of their responsibilities, according to the athletes' own characteristics and requirements, the conditions for individual specific physical training of athletes are insufficient. In the training of injured athletes, it is often passive treatment after the injury occurs, and there is a lack of injury risk assessment and injury prevention training before the injury. The treatment of injuries is often limited to the location of the injury itself, without considering the overall situation as a whole, and not analyzing the mechanism of injury from a specific perspective, resulting in the long-term treatment of injuries or repeated attacks after healing. Such problems have led to a high incidence of long-distance runners' injuries, especially during the concentrated

preparation period before the big competition. A large number of long-distance runners suffered injuries and even could not participate in the competition, which seriously affected China's long-distance running preparations [3].

To sum up, first, the increasingly fierce long-distance running competitions nowadays have raised the athletes' specific physical fitness requirements to a higher level; second, the special physical training of long-distance runners in my country is still at the initial stage, and no systematic training mode and methods have been formed; third, the rapid development of current competitive sports theory and practice promotes the renewal and improvement of traditional training methods; fourth, my country's long-distance runners have a high incidence of injuries and show common special characteristics and injuries. Therefore, combining the new training model with Chinese long-distance running practice, sorting out the special characteristics of long-distance running events, analyzing the positioning of special physical fitness in long-distance running training, and constructing a special physical training model for outstanding long-distance runners with Chinese characteristics are the job responsibility. It is also an exploration and attempt on the practice of special physical fitness enhancement training for outstanding athletes [4].

In literature [5], in order to scientifically and systematically study the optimization and innovation of the physical training methods of the Shandong middle- and long-distance running team, this article uses the observation method, the literature review method, the expert interview method, and the logical analysis method to compare the Shandong middle- and long-distance running team. Comparing the traditional physical training methods with the traditional physical training methods, the following conclusions are drawn: (1) the Shandong middle- and long-distance running team has optimized the strength training methods; (2) the Shandong middle- and long-distance running team has carried out the joint maintenance training methods Innovation and optimization; (3) Shandong Province Middle- and Long-Distance Running Team optimized the restorative training methods; (4) Shandong Province Middle- and Long-Distance Running Team carried out innovative and optimized training methods for sensitivity and coordination; (5) Shandong Province Middle- and Long-Distance Running Team's endurance quality and speed quality are not optimized. Literature [6] constructs a special physical training model for excellent fencers in China and explores effective ways and methods of special physical training for excellent fencers in China, so as to improve the special physical fitness level of excellent fencers in China. The research suggests that the special physical training for excellent fencers is supported by the special characteristics of fencing and physical fitness, with index testing, initial evaluation, model construction, practice operation, and result evaluation as the main links, feedback adjustment as the core, and improvement of athletic ability as the goal, summarizing training practice problem solutions. There are certain logical clues and a relatively stable reference style. However, the above two methods both have the problems of low modeling accuracy and long modeling time.

Mobile computing is a new technology emerging with the development of mobile communication, Internet, database, distributed computing, and other technologies. Mobile computing technology will enable computers or other information intelligent terminal devices to realize data transmission and resource sharing in the wireless environment. The multimode feasibility modeling analysis method in this paper has a very close relationship with Wireless Communications and Mobile Computing. The sensors are directly dependent on Wireless Communications and Mobile Computing technology.

Based on this, this paper proposes a multimode feasibility modeling analysis method for physical fitness training for long-distance runners and verifies the effectiveness of this method through simulation experiments. Lay the foundation for my country's long-distance running performance.

Our contribution is threefold:

(1) To the problems of long modeling time and low modeling accuracy, a new multimode feasibility modeling analysis method for physical fitness training of long-distance runners is proposed

(2) According to the processing results, the information theory is used to analyze the information content of the physical fitness training features of the long-distance runners and complete the feasibility of the long-distance runners' physical fitness training multimode feasibility study Mode analysis

(3) The simulation experiment results show that the proposed method has higher accuracy and shorter modeling time for multimode feasibility modeling of physical fitness training for long-distance runners

The remainder of this paper is organized as follows. Section 2 introduces the analysis of feasibility modeling of multimode physical fitness training for long-distance runners. Section 3 discusses the experiment and analysis. Section 4 presents the conclusions of the study.

2. Analysis of Feasibility Modeling of Multimode Physical Fitness Training for Long-Distance Runners

2.1. Long-Distance Runners' Physical Fitness Enhancement Training Data Dimensionality Reduction Processing. First, collect the physical fitness training data for long-distance runners. However, due to the large amount of physical fitness training data and uneven data dimensions for long-distance runners, the physical fitness training data dimension of long-distance runners is too high, which reduces the accuracy of multimodal evaluation of physical fitness training for long-distance runners. Before the subsequent processing, the training data for physical fitness enhancement of long-distance runners needs to be processed for dimensionality reduction. At present, data dimensionality reduction methods are mainly divided into two categories: linear mapping and nonlinear mapping methods. Among them, the principal component analysis algorithm and the linear

discriminant analysis method are linear mapping methods, and the nonlinear mapping methods mainly include the kernel method (kernel+linear), two-dimensional and Zhang quantization (two-dimensional+linear), and manifold learning methods. Here, an LLE (Locally Linear Embedding) method in the nonlinear mapping method is used for dimensionality reduction. The algorithm is roughly divided into three steps:

Step 1. Find the k-nearest neighbor points of each long-distance runner's physical fitness training data sample point.

Step 2. Calculate the neighboring points of each long-distance runner's physical fitness training data sample and obtain the local reconstruction weight matrix of the sample point.

Step 3. Obtain the low-dimensional embedding output value of the training data sample point of the long-distance runner according to the calculation result of Step 2 [7].

Although the LLE method can solve the problem of data dimensionality, the dimensionality reduction effect is easily affected by the number of neighbors. For this reason, this chapter constrains the reconstruction error by constructing approximate reconstruction coefficients to achieve the improvement of the LLE method. The improved LLE method dimensionality reduction process is as follows.

For a given high-dimensional data set $X = \{x_1, x_2, \cdots, x_N\}$, $x_i \in R^D$, where R^D is the D-dimensional data, sampled from the d-dimensional manifold, find the low-dimensional coordinates $Y = \{y_1, y_2, \cdots, y_N\}$.

$$d_{ij} = \left[\sum_{k=1}^{D} \left\|x_{ik} - x_{ij}\right\|^2\right]^{1/2}. \tag{1}$$

$$\begin{cases} \min e(W) = d_{ij} \sum_{i=1}^{N} \left\| x_i - \sum_{j=1}^{N} w_{ij} x_j \right\|_2^2 \\ s.t. \sum_{j=1}^{k} w_{ij} = 1 \end{cases}. \tag{2}$$

Step 1. Find the $k(k < N)$ neighboring points of the training data sample point x_i of the physical fitness enhancement training of the long-distance runners and realize it by using the Euclidean distance formula [8].

Step 2. Calculate the local reconstruction weight matrix (w_{ij}) through the following formula (2).

Among them, w_{ij} is the weight between x_i and x_{ij}. $x_{ij}(j = 1, 2, \cdots, k)$ is the k-nearest neighbors of x_i [9].

Combining with the restrictive conditions, equation (2) can be rewritten as follows:

$$\begin{aligned} \min e(W) &= \sum_{i=1}^{N} \left\| \sum_{j=1}^{N} w_{ij} (x_i - x_{ij}) \right\|^2 \\ &= \sum_{i=1}^{N} \left\| (x_i - x_{ij}) w_i \right\|^2 \\ &= \sum_{i=1}^{N} (w_i)^T Z_i w_i. \end{aligned} \tag{3}$$

Among them,

$$Z_i = (x_i - x_{ij})^T (x_i - x_{ij}), \tag{4}$$

$$w_i = [w_{i1}, w_{i2}, \cdots, w_{ik}]^T. \tag{5}$$

In the formula, Z_i is the local covariance matrix of the ith sample point; w_i is the local reconstruction weight of the ith sample point [10].

Introducing Lagrange multiplier to solve formula (3), then

$$L(W) = \sum_{i=1}^{N} (w_i)^T Z_i w_i + \left(\sum_{j=1}^{N} w_{ij} - 1\right). \tag{6}$$

Usually, a simple solution method is used, let $Z_i w_i = 1$, to find w_i.

$$s_i^j = \frac{d_w^j L(W)}{d_b^j \min e(W)}. \tag{7}$$

Step 3. Obtain the approximate reconstruction coefficient s_i^j of the reconstruction error of the long-distance runner's physical fitness training data sample point x_i through formula (7) and construct the low-dimensional embedding output value Y according to the approximate reconstruction coefficient, namely,

In the formula, d_w^j is the average distance within the class of the training data sample points of the j-type long-distance runners; d_b^j is the distance between the j-type long-distance runners' physical fitness training data sample and the center of the overall long-distance runners' physical fitness training data sample point.

Using the Lagrange multiplier method, find

$$S = (I - W)^T (I - W) s_i^j. \tag{8}$$

In the formula, I represents a N-dimensional identity matrix; W is a diagonal matrix.

Step 4. Take the eigenvectors corresponding to the $2 \sim (d + 1)$ non-zero eigenvalues of W as the low-dimensional coordinates Y to complete the dimensionality reduction processing

of the training data for physical fitness enhancement of long-distance runners [11].

2.2. Feature Extraction of Physical Fitness Training for Long-Distance Runners Based on Decision Tree. Decision tree is a tree structure and an important technology in data mining. It uses the principle of information theory to analyze the information content of the physical fitness training characteristics of long-distance runners, calculates the information entropy of each feature, and extracts the physical fitness training characteristics of long-distance runners [12]. The quantitative recursive analysis model is used to quantify the physical fitness training feature quantities of the long-distance runners at the ith point $\boldsymbol{x}_i$ and the jth point $\boldsymbol{x}_j$. The cost function of the multimode physical fitness training for long-distance runners is constructed as

$$G\left(U\middle|\mu_k, \sum_k\right) = (2\pi)^{-d/2}\left|\sum_k\right|^{-1/2} \times \exp\left[-\frac{1}{2}(U-u_k)^T\sum_k^{-1}(U-u_k)\right]. \tag{9}$$

In the above formula, $G(U|\mu_k, \sum_k)$ is the constraint feature quantity of the physical fitness training for long-distance runners. Calculate the optimal estimates $\widehat{a}$, $\widehat{b}_i$ of the correlation dimension features a and b_i of the bit sequence of the training sample sequence bit sequence for the long-distance runners' physical fitness enhancement, and the calculation formula is expressed as follows:

$$D'_{t+1} = 1-(1-\lambda)\sum_{n=0}^{\infty}\Omega_{m+n+1}((n+1)b-t), \tag{10}$$

Table 1: Determination of single machine connection code.

Project	Code	Stand-alone type
Height	00	Height measuring instrument
Body weight	01	Weight measuring instrument
Vital capacity	02	Spirometry
Step index	03	Step index test equipment
Long run	04	Long-distance running measuring equipment
500 m run	05	Timing equipment

$$L'_{t+1} = (1-\lambda)\sum_{n=0}^{\infty}\Omega_{k+n}(n+1)b. \tag{11}$$

In the above formula, Ω represents the associated information entropy function, and the obtained distribution feature is defined as follows:

$$x_k = f\{x_{k-1}, u_{k-1}, w_{k-1}\}. \tag{12}$$

Among them, u_k represents the model parameter; w_k represents the fuzzy statistical value of the training characteristics of physical fitness enhancement of long-distance runners.

Using the decision tree method to extract the training features of physical fitness enhancement of long-distance runners, the expression is

$$\begin{cases} x(k+1) = \left(\begin{bmatrix} 1 & 0.6 \\ -0.4 & 0.5 \end{bmatrix} + \begin{bmatrix} 0.02 & 0.01 \\ -0.02 & 0.12 \end{bmatrix}\right)x(k) + \begin{bmatrix} 1 \\ 1 \end{bmatrix}kx(k-\tau_k) + \begin{bmatrix} 0.1 \\ 0.1 \end{bmatrix}w(k) \\ z(k) = [1 \quad 1]x(k) + 0.1u(k) + 0.1w(k) \end{cases}. \tag{13}$$

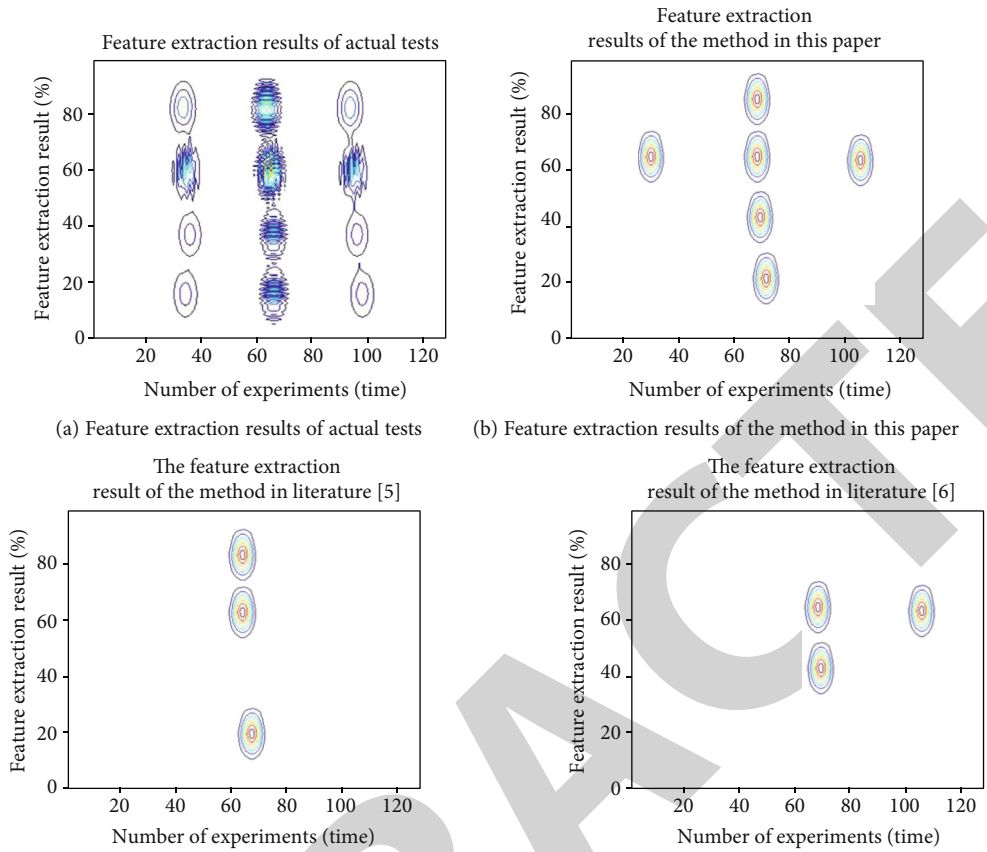
2.3. The Feasibility Evaluation Model of Multimode Physical Fitness Training for Long-Distance Runners. Based on the feature extraction of physical fitness training for long-distance runners, the feasibility evaluation model of multimode physical fitness training for long-distance runners is constructed, and the feasibility analysis of the multimode physical fitness training for long-distance runners is carried out. The evaluation feature quantity distribution model is

$$x_{ij} = [\lg r]_{ij}, y_{ij} = [\lg C_N(r)]_{ij}. \tag{14}$$

In the above formula, j is the characteristic sampling point for multimode feasibility evaluation of physical fitness enhancement training for long-distance runners under a certain embedding dimension i, denoted as follows:

$$y_{ij} = ax_{ij} - b_i. \tag{15}$$

In the above formula, $a = D$, combined with the fuzzy feature analysis method, obtains the joint distribution probability of the multimode feasibility statistical sample sequence $x(t+\tau)$ in the state estimation region of the long-distance runners, which is random sampling. Using the spatial embedding theorem obtains the correlation dimension of



(a) Feature extraction results of actual tests (b) Feature extraction results of the method in this paper

(c) The feature extraction result of the method in literature [5] (d) The feature extraction result of the method in literature [6]

Figure 1: The results of feature extraction of physical strength enhancement training for long-distance runners.

the multimode feasibility evaluation of the long-distance runners' physical enhancement training

$$D = \lim_{r \longrightarrow 0} \frac{\lg C_m(r)}{\lg r}, \tag{16}$$

where $C_m(r)$ is the correlation integral under the sampling delay, and when the Euclidean distance r of the phase space distribution trajectory is small enough, the fuzzy correlation characteristic component of the multimode feasibility evaluation of the long-distance runners' physical fitness enhancement training is calculated [13]. The information characteristics of multimode feasibility evaluation of physical fitness enhancement training for long-distance runners are as follows:

$$I(\tau) = -\sum_{ij} p_{ij}(\tau) \ln \frac{p_{ij}(\tau)}{p_i p_j}. \tag{17}$$

When $I(\tau) = 0$, then $x(t+\tau)$ is the statistical sample sequence of multimode feasibility evaluation of long-distance runners' physical fitness enhancement training. If it meets the convergence solution, it means that it is unpredictable, that is, $x(t), x(t+\tau)$ is independent and completely uncorrelated. Calculate the correlation mutual information $I(\tau)$ to adjust the prediction possibility. When $I(\tau)$ is the minimum, the corresponding sampling time points $x(t)$ and $x(t+\tau)$ are uncorrelated, It indicates that the multimode feasibility statistical sample sequence of physical fitness enhancement training for long-distance runners is

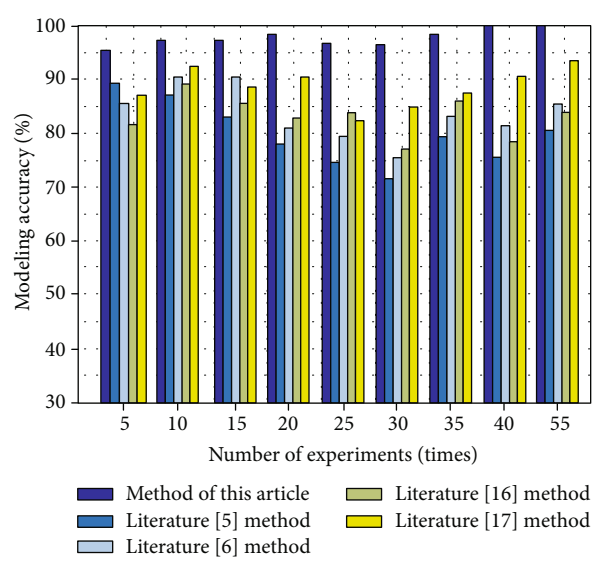


Figure 2: Comparison results of multimode feasibility modeling accuracy.

Table 2: Time comparison results of multimode feasibility modeling.

Number of experiments/times	Modeling time/(s)				
	Method of this article	Literature [5] method	Literature [6] method	Literature [16] method	Literature [17] method
20	7.58	10.58	14.87	10.58	14.87
40	13.74	17.33	21.41	17.33	21.41
60	16.52	23.75	27.63	23.75	27.63
80	24.63	30.96	34.85	30.96	34.85
100	30.22	35.75	40.96	35.75	40.96
120	35.85	41.96	48.69	41.96	48.69
140	40.78	57.85	57.87	57.85	57.87
160	45.63	64.27	66.25	64.27	66.25
180	51.78	71.85	72.54	71.85	72.54
200	60.52	77.54	78.96	77.54	78.96

completely predictable [14]. Using the least square fitting method for descriptive statistical analysis, combined with quantitative regression analysis method for information regression analysis of multimode statistical sequence of long-distance runners' physical fitness enhancement training [15], the multimode feasibility evaluation model of long-distance runners' physical fitness enhancement training is obtained as follows:

$$\Pi_2 = \begin{bmatrix} \bar{A}^T P\bar{A} - P + K^T RK & \bar{A}^T P\bar{B} & \bar{A}^T PF_1 \\ \bar{B}^T P\bar{A} & \bar{B}^T P\bar{B} - R & \bar{B}^T PF_1 \\ F_1^T P\bar{A} & F_1^T P\bar{B} & F_1^T PF_1 \end{bmatrix}. \tag{18}$$

Assuming that the probability random variable of the multimode feasibility time series $x(t)$ appearing in the distribution interval i satisfies the convergence condition, the multimode feasibility model of long-distance runners' physical enhancement training is completed.

3. Simulation Experiment Analysis

In order to verify the performance of the multimode feasibility modeling and analysis method of long-distance runners' physical fitness enhancement training proposed in this paper in practical application, the long-distance runners of a sports school are selected as the experimental objects, and the joint programming design of Visual C + + and MATLAB 7 is adopted. The simulation experiment is carried out in the operating system of Windows XP, server of Tomcat 5.5, 2.0 GHz CPU, 16 GB memory, and 7200 rpm hard disk. C language is a general programming language, which not only has all the characteristics of a high-level language but also can directly carry out structured programming. Among them, c51 is a structured language with high programming efficiency, strong portability, support for floating-point calculation, and strong real-time performance. Using this language to write the object program can not only shorten the algorithm cycle but also facilitate the maintenance. It is a very ideal language program.

To test the long-distance runner's physical strength enhancement, this paper connects different types of test stand-alone through the controller, and the connection code of the stand-alone is shown in Table 1.

Under the above background, the multimode feasibility modeling analysis method, literature [5] method, and literature [6] method proposed in this paper are used to extract the characteristics of long-distance runners' physical enhancement training, and the results are compared with the actual test results. The comparison results are shown in Figure 1.

It can be seen from Figure 1 that the feature extraction results of the multimode feasibility modeling analysis method for long-distance runners' physical enhancement training proposed in this paper are consistent with those of the actual test, while the feature extraction results of long-distance runners' physical enhancement training proposed in literature [5] and literature [6] are quite different from those of the actual test. The results show that the feature extraction accuracy of the multimode feasibility modeling analysis method of long-distance runners' physical enhancement training proposed in this paper is higher than that of the methods in literature [5] and literature [6], because before the feature extraction of long-distance runners' physical enhancement training, the feature extraction accuracy of long-distance runners' physical enhancement training proposed in this paper is higher than that of literature [5] and literature [6]. This paper uses the improved LLE method to reduce the dimension of long-distance runners' physical strength enhancement training data, so as to improve the feature extraction accuracy of long-distance runners' physical strength enhancement training.

In order to verify the effectiveness of this method, this paper makes a comparative analysis on the feasibility modeling accuracy of the multimode of long-distance runners' physical fitness enhancement training proposed in this paper, the method in literature [5], and the method in literature [6]. The comparison results are shown in Figure 2.

According to Figure 2, the modeling accuracy of the multimode feasibility analysis method for long-distance runners' physical enhancement training proposed in this paper is more than 95%, and the highest is 100%. However, the

modeling accuracy of the multimode feasibility analysis method for long-distance runners' physical enhancement training proposed in literature [5, 6] is lower than that of this method. The results show that the multimode feasibility modeling of long-distance runners' physical fitness enhancement training is effective.

In order to further verify the effectiveness of this method, this paper makes a comparative analysis of the multimode feasibility modeling analysis method of long-distance runners' physical fitness enhancement training proposed in this paper, the method of literature [5], and the multimode feasibility modeling time of long-distance runners' physical fitness enhancement training proposed in literature [6]. The comparison results are shown in Table 2.

According to Table 2, the time of multimode feasibility modeling of long-distance runner's physical enhancement training proposed in this paper is only 60.52 s, and the time of multimode feasibility modeling of long-distance runner's physical enhancement training proposed in literature [5] and literature [6] is 77.54 s and 78.96 s, respectively, It shows that the multimode feasibility modeling analysis method of long-distance runners' physical enhancement training proposed in this paper has shorter modeling time and higher modeling efficiency.

4. Conclusion

Long-distance running is a kind of periodic endurance project dominated by physical ability, which is characterized by long distance, low intensity, overcoming self-weight, and repeating the same action for a long time. At present, with the improvement of competitive level and the fierce competition, the requirement of athletes' speed level is higher and higher. The level of physical ability plays a leading and core role in the components of their competitive ability. Whether long-distance runners can win in the fierce competition mainly depends on the size of their physical ability. Physical fitness is based on the energy metabolism activities of the three major energy supply systems of the human body, which are expressed through the skeletal muscle system. Physical fitness includes physical quality, shape, and function. Each of the three factors has its own relatively independent role, and they are closely related, restricting, and influencing each other. The level of one of the factors will affect the overall level of physical fitness. Among them, the change of form and function is the material basis of the change of physical ability, and sports quality is the external performance of physical ability. In sports training, the development of various sports quality is the basic content of physical training.

In order to enhance the physical fitness of long-distance runners, this paper proposes a multimode feasibility modeling analysis method for long-distance runners. On this basis, according to the principle of information theory, this paper analyzes the characteristics of long-distance runners' physical enhancement training, obtains the information entropy of each feature, extracts the characteristics of long-distance runners' physical enhancement training, and adopts the quantitative regression analysis method according to the feature extraction results. This paper makes a quantitative regression analysis on the multimode statistical sequence of long-distance runners' physical enhancement training and constructs a multimode feasibility evaluation model of long-distance runners' physical enhancement training. The simulation results show that the proposed method for multimode feasibility modeling of long-distance runners' physical fitness enhancement training has higher accuracy and faster modeling efficiency.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

Declares that he has no conflict of interest.

References

[1] X. Fei, "Research on physical fitness training of excellent long-distance runner Xue Fei in winter training," *Neijiang Technology*, vol. 279, no. 2, pp. 82–117, 2018.

[2] J. Yin and X.-q. Wang, "Study on safety mode of dragon boat sports physical fitness training based on machine learning," *Safety Science*, vol. 120, pp. 1–5, 2019.

[3] Y. Yang and W. Liu, "The influence of public physical education curriculum on college students' physical health," *Revista Brasileira de Medicina do Esporte*, vol. 27, no. spe, pp. 83–86, 2021.

[4] Lesya Ukrainka Volyn National University, S. Savchuk, N. Zakhozha et al., "Special physical training of high educational institutions athletes in a short running race," *Physical Education Sport and Health Culture in Modern Society*, vol. 4(52), pp. 60–66, 2020.

[5] M. Wang, "Research on optimization of physical training methods for middle and long distance running in Shandong Province [J]," *Stationery and sports goods and technology*, vol. 443, no. 10, pp. 218-219, 2020.

[6] C. Zhen and L. Heng, "Discussion of risks in chinese construction market-contractors perspective [J]," *Journal of Construction Engineering and Management*, vol. 132, no. 3, pp. 327–329, 2006.

[7] W. S. Jung, H. Moon, and H. Y. Park, "the effects of 12-week combined exercise participation on body composition and daily-living physical fitness in active obesity elderly men," *The Korean Journal of Growth and Development*, vol. 26, no. 3, pp. 297–304, 2018.

[8] L. Galamandjuk, A. Siedlaczek-Szwed, G. Iedynak et al., "Evaluation of the physiological characteristics of girls with different handedness using various types of physical training [J]," *Journal of Physical Education and Sport*, vol. 6, pp. 19–30, 2019.

[9] H. Wang, "Discussion on the application of stratified evaluation in middle school football teaching and training," *Education Research Frontier: Chinese and English Edition*, vol. 9, no. 3, pp. 80–84, 2019.

[10] R. Stănciulescu and C. Stănciulescu, "Considerations regarding the possibility of implementing a model for the evaluation of the level of cadets' physical training," *Land Forces Academy Review*, vol. 26, no. 1, pp. 49–54, 2021.