Wireless Communications and Mobile Computing

Learning Methods for Urban Computing and Intelligence

Lead Guest Editor: Wei Wang
Guest Editors: Qingchen Zhang, Ahmedin Mohammed, and Bingxian Lu

Learning Methods for Urban Computing and Intelligence

Wireless Communications and Mobile Computing

Learning Methods for Urban Computing and Intelligence

Lead Guest Editor: Wei Wang
Guest Editors: Qingchen Zhang, Ahmedin Mohammed, and Bingxian Lu

Contents

Hindawi
Wireless Communications and Mobile Computing
Volume 2023, Article ID 9893103, 1 page
https://doi.org/10.1155/2023/9893103

Retraction

Retracted: Analysis on the Construction of Personalized Physical Education Teaching System Based on a Cloud Computing Platform

Wireless Communications and Mobile Computing

Received 8 August 2023; Accepted 8 August 2023; Published 9 August 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This article has been retracted by Hindawi following an investigation undertaken by the publisher [1].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] Z. Zhang and H. Min, "Analysis on the Construction of Personalized Physical Education Teaching System Based on a Cloud Computing Platform," *Wireless Communications and Mobile Computing*, vol. 2020, Article ID 8854811, 8 pages, 2020.

Hindawi
Wireless Communications and Mobile Computing
Volume 2020, Article ID 8885670, 7 pages
https://doi.org/10.1155/2020/8885670

Research Article

Algorithm for Target Detection in Smart City Combined with Depth Learning and Feature Extraction

Feng Wang[1], **Zhiming Xu**[1], **Zemin Qiu**[1], **Weichuan Ni**[2], **Jiaqi Li**[1] and **YiLan Luo**[1]

[1]*Department of Information Science, Xinhua College of Sun Yat-Sen University, Guangzhou, China*
[2]*Department of Equipment and Laboratory Management, Xinhua College of Sun Yat-Sen University, Guangzhou, China*

Correspondence should be addressed to Feng Wang; iswf@xhsysu.edu.cn

Received 10 April 2020; Revised 3 August 2020; Accepted 20 September 2020; Published 5 October 2020

Academic Editor: Wei Wang

The target detection algorithms have the problems of low detection accuracy and susceptibility to occlusion in existing smart cities. In response to this phenomenon, this paper presents an algorithm for target detection in a smart city combined with depth learning and feature extraction. It proposes an adaptive strategy is introduced to optimize the algorithm search windows based on the traditional SSD algorithm, which according to the target operating conditions change, strengthening the algorithm to enhance the accuracy of the objective function which is combined with the weighted correlation feature fusion method, and this method is a combination of appearance depth features and depth features. Experimental results show that this algorithm has a better antiblocking ability and detection accuracy compared with the conventional SSD algorithms. In addition, it has better stability in a changing environment.

1. Introduction

The concept of a smart city originated from the idea of smart earth proposed by IBM in 2008. Among them, the target detection algorithm is one of the key technologies of smart cities. However, existing computer vision algorithms are difficult to deal with target detection problems in complex backgrounds, such as the effects of light, target size changes, and target occlusion. The introduction of deep learning has opened up a new path for target detection. In recent years, more and more researchers have begun to conduct in-depth research on deep learning algorithms in target detection [1, 2]. Deep learning avoids the drawbacks of the traditional method of manually extracting features, because of the characteristics that its deep structure can effectively learn from large amounts of data [3, 4]. Currently, based on the target detection algorithm, the depth study of literature is not much. However, from the perspective of the depth model, it can be broadly classified into a target detection algorithm based on CNN and a target detection algorithm based on SAE [5–7]. The target detection algorithm of the SAE depth model is usually combined with the traditional classical algorithms. It uses hidden layers to learn a representation of data and to preserve and better obtain more efficient information by using a nonlinear feature extraction method that does not use classification tags. This method is not conducive to information classification, but visual tracking itself needs to distinguish the target from the background. Therefore, target tracking is not the strength of the SAE algorithm. The target detection algorithm based on CNN combines artificial neural networks and convolution operations. It can recognize a wide variety of target modes, and a certain degree of distortion and deformation has good robustness. Therefore, we use the target detection algorithm based on CNN for this article. Among them, the SSD algorithm recognition structure is superior to a similar algorithm in the mAP and training speed [8, 9]. By using different sizes and different proportions of anchors at different levels, the algorithm can find the best matching anchor with ground truth for training. However, the recognition effect of the target on the small size is relatively poor and is easily affected by the occluder, which undoubtedly affects the application of the algorithm in practical applications.

Regarding the issue above, this article proposes a target detection method based on the SSD algorithm and feature extraction fusion. It is based on the traditional SSD algorithm [10, 11]. Its search window is dynamically adjusted by using an adaptive strategy changes according to its operating conditions, which can reduce unnecessary calculation accuracy problems during the entire detection target fixed occurring. At the same time, in order to improve the classification ability of the features of the algorithm, we did the following optimization; like in the feature fusion method of weighted correlation, we combined with the appearance depth feature and the motion depth feature to improve the accuracy of the objective function and perform experiments in different complexity image environments. This algorithm is more time-consuming than the traditional SSD algorithm. The accuracy of the target gradually increases with the complexity of the image. The gap between the detection accuracy and success rate and the traditional SSD algorithm gradually widens and remains at about 86%. It is said that the algorithm can be applied to a variety of environments and maintain good stability. It has good practical value in the development of a smart city.

2. Principle of Algorithm

Traditional SSD is based on a forward-propagating CNN network. It produces a series of fixed-size bounding boxes and has the possibility of containing object instances in each box then performs a nonmaximal suppression to get the final predictions. The model network structure is as follows.

From the structure diagram of Figure 1, we can get the SSD network can be divided into two parts: the basic network and additional functional layers; the former is used for the standard network for image classification, but all the layers involved in the classification are eliminated; the latter mainly achieves the following goals [12]:

Multiscale feature maps for detection: the convolutional feature layer is added to obtain feature layers of different scales so as to achieve multiscale target detection.

Convolutional predictors for detection: for each added feature layer, a set of convolution filters is used to obtain a fixed set of target detection predictions.

Each of these convolutions results in a set of scores or coordinate offsets from the default candidate regions. Finally, combining the obtained detection and classification results, the position of each object in the image and the object category in the image can be obtained.

3. The Algorithm of This Paper

3.1. Select the Aspect Ratio of the Default Box. The feature map will be smaller and smaller at deeper layers. This is not only to reduce the computational and memory requirements but also has the advantage that the last extracted feature map will have some degree of translation and scale invariance [13].

In the SSD structure, the default boxes do not necessarily correspond to the receptive fields of each layer. Predictions are made by introducing m feature maps; the size calculation formula of the default box in each feature map satisfies:

$$s_k = s_{\min} + \frac{s_{\max} - s_{\min}}{m-1}(k-1), \tag{1}$$

where k is $[1, m]$, $s_{\min}$ value 0.2, and $s_{\max}$ value 0.95.

But usually, the size of each default box will not be adjusted after this calculation. This is undoubtedly unfavorable for detecting the object whose size will change and then affecting the detection effect. Figure 2 shows the structure of the target of the traditional SSD algorithm.

Among them, the objective loss function is a weighted sum of the localization lossloc and confidence loss in Figure 2:

$$L(x, c, l, g) = \frac{1}{N}\left(l_{\text{conf}}(x, c) + \alpha L_{\text{loc}}(x, l, g)\right). \tag{2}$$

In the formula:

$$\begin{aligned} L_{\text{loc}}(x, l, g) &= \sum_{i\in\text{Pos}}^{N} \sum_{m\in(cx,cy,w,h)} x_{ij}^{k}\text{smooth}\left(l_i^m - \widehat{g}_j^m\right), \\ L_{\text{conf}}(x, c) &= -\sum_{i\in\text{Pos}}^{N} x_{ij}^{p}\log\left(\widehat{c}_i^p\right) - \sum_{i\in\text{Neg}}\log\left(\widehat{c}_i^0\right), \end{aligned} \tag{3}$$

among them, $\widehat{c}_i^p = \exp\left(c_i^p\right)/\sum_p \exp(c_i^p)$, N is the number of matching default boxes, x indicates whether the matched box belongs to category p, value {0,1}, l is a predictive box, and g is the true value of that ground truth box. c is the confidence that the selected target belongs to category p. Weight item α=1.

We need to filter out the boxes; we finally give from these boxes.

The pseudocode is as follows.

for every conv box:
for every class:
if class_prob < theshold:
continue
predictive box = decode(convbox)
nms(predictive box) # Remove very close boxes

In this way, the target coordinates can be found effectively, thereby improving the detection effect of the algorithm on target detection.

3.2. Adaptive Strategy. Usually, the width and height of each default boxes are fixed during the entire target detection process. However, when the behavior of the controlled object changes due to changes in the characteristics of the object, this type of parameter fixation tends to produce undesirable results. Control the effect, so you need to make use of adaptive strategies for dynamic adjustments.

In this paper, the width and height are adaptively adjusted by combining the size of the default box obtained above. The calculation formula is as follows:

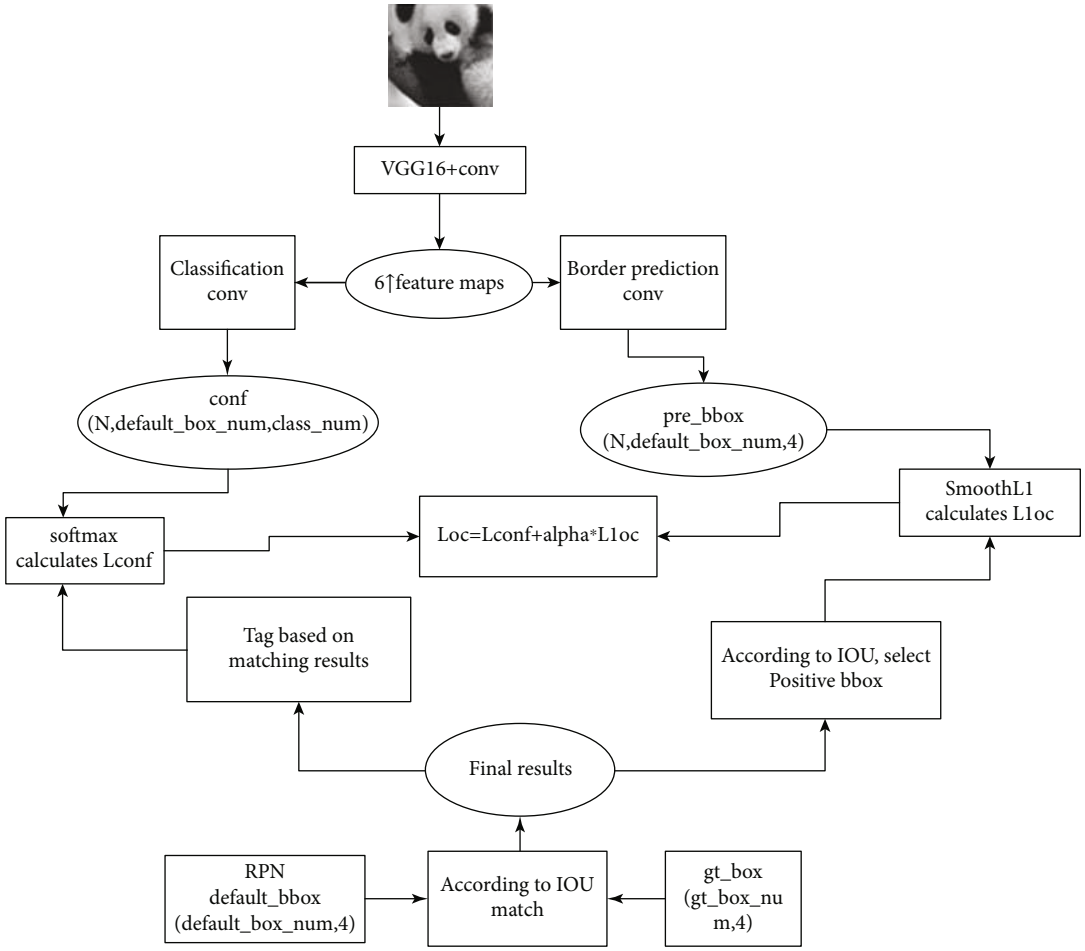


Figure 1: SSD model network structure.

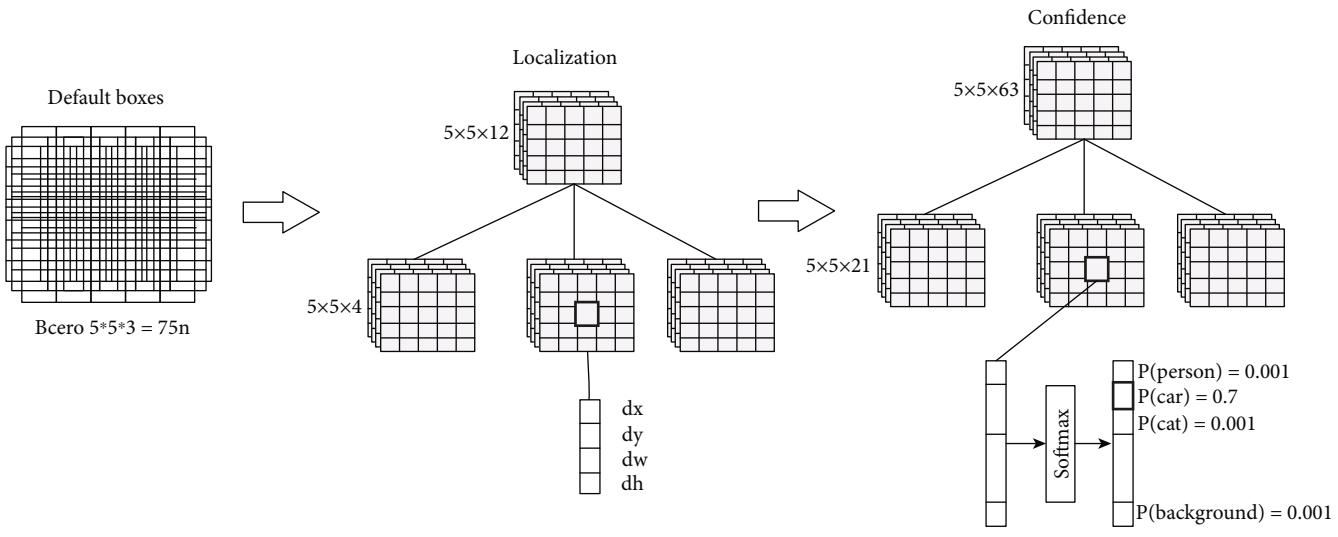


Figure 2: Block diagram of the traditional SSD algorithm target positioning.

$$L_w = \sqrt{\frac{\sum_{i=1}^{n} s_{k_i}^{\ 2}}{n}},$$
$$L_h = \sqrt{\frac{\sum_{j=1}^{n} s_{k_j}^{\ 2}}{n}}, \quad (4)$$
$$L = \max\left(L_w, L_j\right),$$

where n is equal to the number of blocks in the image; s_{ki}, s_{kj} is the component of the horizontal direction i and the vertical direction j of the objective function, respectively; L_w, L_h is newly born into the width and height of the default boxes; and L is the whole frame of motion intensity.

By determining the motion complexity of the block based on the calculated motion vector and the motion vector of the current block, the degree of motion of the current block can be effectively judged by the degree of difference between the horizontal component and the vertical component. The formula is as follows:

$$S_w = \text{Max}\left[\text{horizontal}(s_k) - \text{horizontal}\left(s_{kimiddle}\right)\right],$$
$$S_h = \textbf{\textit{Max}}\left[\text{vertical}(s_k) - \text{vertical}\left(s_{kimiddle}\right)\right], \quad (5)$$
$$S = \text{Max}\ (S_w, S_h).$$

Among them, $s_{kimiddle}$ is the median of the three macroblocks in the left, top, and upper right directions. $s_{ki}$is the objective function of the i-th block obtained above. S_w, S_h separately expressed the horizontal and vertical movement complexity. S represents the complexity of the motion of the current block.

The search window size is as follows:

$$W = \begin{cases} L & S < L \\ S + L & \text{other} \end{cases}. \quad (6)$$

3.3. Feature Extraction of Weighted Correlation. In order to improve the classification ability of features, a feature fusion method based on weighted correlation is used to combine the appearance depth feature and the movement depth feature to form a multidimensional feature vector. For the convenience of presentation, the appearance depth feature and the movement depth feature are denoted by y_1 and y_2, respectively. The merged feature y is:

$$y = (\omega_1 y_1, \omega_2 y_2) \quad (7)$$

Here, ω_i is the weighting factor, and $(\omega_1)^2 + (\omega_2)^2 = 1$. The weighting coefficients are determined according to intraclass consistency and class separability. Intraclass consistency: It is generally expected that the samples in the same class are as close as possible in the feature space. However, there is usually a large variance in the sample characteristics in the same class. Therefore, it is not necessary to require all samples in the same class to be close to each other. A tradeoff is to ensure that the samples in the same neighborhood within the same class are as close as possible.

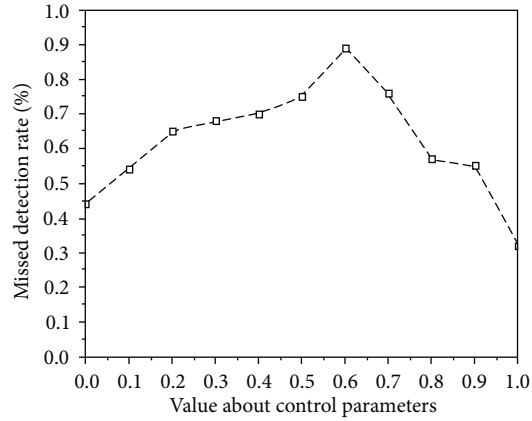


Figure 3: Control parameter λ_s value curve.

Assume $y_i = (\omega_1 y_1^i,\ \omega_2 y_2^i)$ and $y_j = (\omega_1 y_1^j,\ \omega_2 y_2^j)$.

Let us denote the i-th and j-th samples; then, the intraclass consistency is defined as:

$$S_C = \sum_{i=1}^{N} \sum_{j \in N_R(F_i)} \frac{< y_i, y_j >}{\|y_i\| \|y_j\|} = \sum_{i=1}^{N} \sum_{j \in N_R(F_i)} \frac{\sum_{k=1}^{2} w_k^2 y_i^k y_j^k}{\sqrt{\sum_{k=1}^{2} w_k^2 \left(y_i^k\right)^2} \sqrt{\sum_{k=1}^{2} w_k^2 \left(y_j^k\right)^2}}. \quad (8)$$

In the formula, $N_R(F_i)$ means the sample F_i and with F_i which belongs to the index set of the k nearest neighbor samples of the same class.

According to the target characteristics with good intraclass consistency, this paper determines the weighting coefficients by solving the following optimization problems:

$$\max\ \{S_C + \lambda_s \|\omega\|\}, \quad (9)$$

among them, $\omega_k > 0$, $\|\omega\| = 1$, and λ_s is a control parameter.

Combining the above equations, the gradient descent method is used to solve the equation, which can be solved:

$$\omega_k(t+1) = \omega_k(t) + \eta \left.\frac{\partial L}{\partial \omega_k}\right|_{\omega_k = \omega_k(t)}, \quad (10)$$

where t is the number of iterations, η is the iteration step, and $L(S_C, \omega) = S_C + \lambda_s \|\omega\|$ is the objective function.

Among them

$$\frac{\partial L(S_C, \omega)}{\partial \omega_k} = \sum_{i=1}^{N} \sum_{j \in N_r(x_i)} \frac{\partial h_{ij}(\omega)}{\partial \omega_k} + \lambda_s \omega_k, \quad (11)$$

$h_{ij} = \sum_{k=1}^{2} \omega_k^2 y_i^k y_j^k / \sqrt{\sum_{k=1}^{2} \omega_k^2 \left(y_i^k\right)^2} \sqrt{\sum_{k=1}^{2} \omega_k^2 \left(y_j^k\right)^2}$.

Thus, you can get the final objective function:

$$L = L(x, c, l, g) + L(S_C, \omega) \quad (12)$$

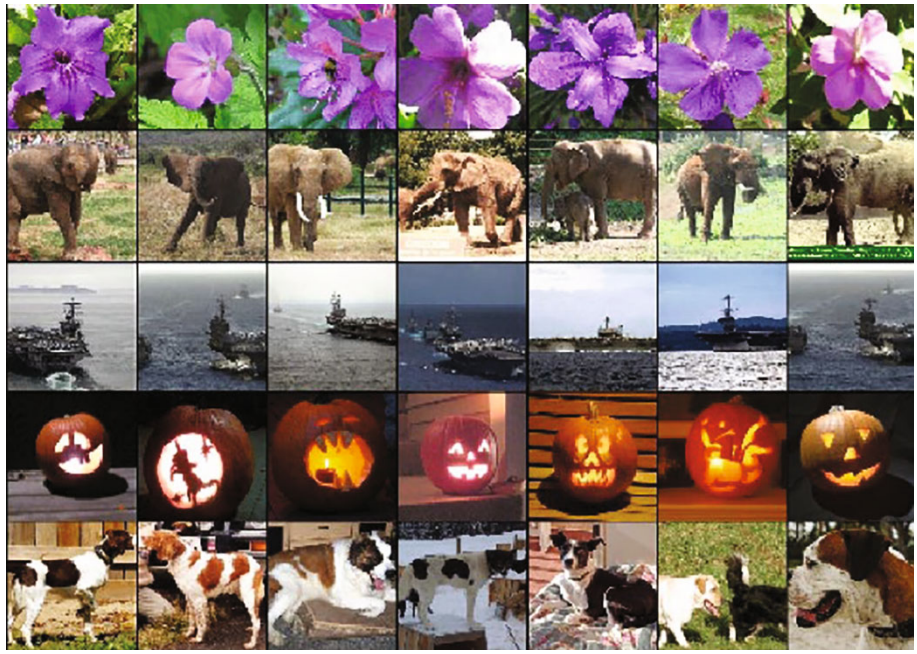


Figure 4: ImageNet datasets.

Table 1: Time spent in training and testing.

Method	Time spent on testing/s		Training period time spent/s	
	Evaluation window time taken	Target retrieval time	Evaluation window time taken	Target retrieval time
Traditional SSD algorithm	0.35	2.11	0.25	3.63
Literature [14] algorithm	0.30	1.92	0.18	2.51
Article algorithm	0.31	2.01	0.20	2.98

After many experiments, the summary data is obtained in Figure 3; we can see the experiments have found that blindly increasing the control parameters does not have any new improvement in the detection accuracy. It is appropriate when the number of key region control parameters is $\lambda_s = 0.6$. This ensures that a unique solution that can converge to a global optimum through gradient descent can be guaranteed.

4. Simulation Experiment

4.1. Data Sets and Test Standards. In order to test the speed and accuracy of the algorithm, the training data from this paper comes from the ImageNet dataset, which contains more than 14 million pictures, covering more than 20,000 categories [15], like Figure 4. The ImageNet dataset is a field that is currently applied in the field of deep learning images. Most research work on image classification, positioning, and detection is based on this dataset. It is a huge image library for image/visual training. It has been widely used in research papers in the field of computer vision and has almost become the "standard" data set for the performance testing of algorithms in the field of deep learning images.

4.2. Experiments and Results. The experiment was simulated using a laptop computer, tested using Python 3.6, TensorFlow v0.12.0, Pickle, OpenCV-Python, and Matplotlib (optional), and the data was analyzed using MATLAB 2014a. In order to detect the target detection effect of the algorithm, this paper compares the traditional SSD algorithm with the literature algorithm [14].

In order to test the retrieval efficiency of the algorithm, the data analysis is performed on the time window of the evaluation of the image and the time required for the detection of the target. Observing Table 1, we can see that in the testing phase, the evaluation window of this algorithm is more time-consuming than the traditional SSD. In algorithm 0.04s, the target detection time is better than the traditional SSD algorithm 0.1s, and in the training phase, the evaluation time of the algorithm in this paper is better than the traditional SSD algorithm 0.05s, and the target detection time is better than the traditional SSD algorithm 0.65s.

From Table 1, it can be seen that the algorithm is time-consuming in the test phase and the training phase is better than the traditional SSD algorithm, but the text is slightly weaker than the literature algorithm [14]. In addition to focusing on the time-consuming, the target detection algorithm needs to analyze and evaluate the accuracy and success rate of the detection target. Therefore, this paper detects the precision and success rate of the target and in different complex scenarios. The test was conducted, in which the selected image material was reconstructed from low to high (food, vegetable, bird, person). The data is shown in Table 2.

Table 2: Algorithm accuracy data tables in different complex scenarios.

Image	Accuracy/%	Traditional SSD algorithm	Literature algorithm [14]	Article algorithm
Food	Accuracy rate	0.848	0.875	0.874
	Success rate	0.854	0.859	0.859
Vegetable	Accuracy rate	0.795	0.804	0.854
	Success rate	0.805	0.814	0.866
Bird	Accuracy rate	0.757	0.750	0.879
	Success rate	0.754	0.771	0.862
Person	Accuracy rate	0.711	0.708	0.852
	Success rate	0.712	0.705	0.871

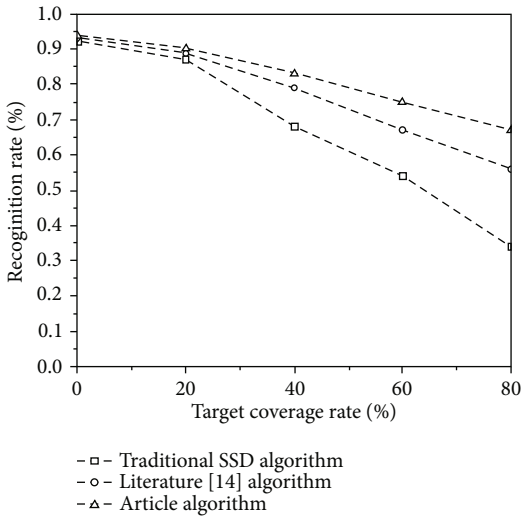


Figure 5: Food cover experiment.

As can be seen from Table 2, it is clear that the target detection accuracy of this algorithm in different environments is better than the traditional SSD algorithm, and as the complexity of the image is gradually increased, the accuracy of detection under the image and the success rate gap gradually increase; the detection accuracy rate is maintained at about 86%, while in the literature algorithm [14] gradually decreases as the complexity increases, and the target detection effect in a variety of complex scenarios embodies the algorithm and has a certain degree of universality.

In order to test whether the algorithm has accuracy in the presence of a shelter, this paper needs to use two sets of experiments to verify. One group adopts a food image with smaller image complexity, and one group employs a person image with the highest complexity; data acquisition is performed using the above algorithms, respectively, and the data image is shown.

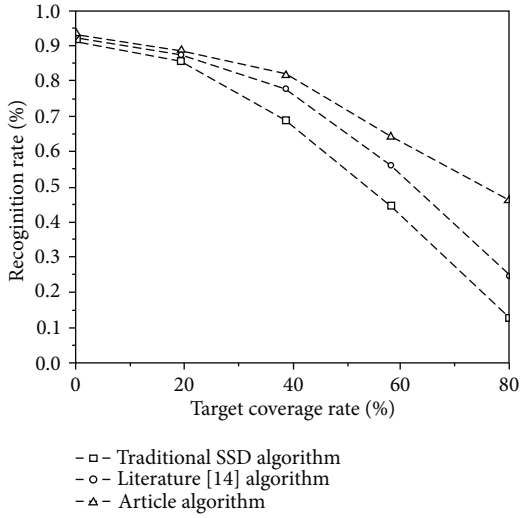


Figure 6: Person cover experiment.

Figure 7: Target extraction image.

By observing Figures 5 and 6, we can see that with the increase of target coverage rate, the setting rate of each algorithm gradually decreases, but it can be seen through observation that the data collected by using this algorithm are kept above the other two algorithms. When the target coverage rate reaches 60% and higher, the setting rates of both the algorithm [14] and the traditional SSD algorithm begin to decline drastically. However, although there is a problem of drop in the article algorithm, the decrease rate of the algorithm is slower than that of the other two methods, which effectively proves the feasibility of the feature extraction.

In order to test the detection effect of the algorithm in the actual environment, we have selected the daily traffic scene for the target detection, real-time extraction of the people, and vehicles appearing in the traffic; the following effect map is obtained.

By observing Figures 7 and 8, we can find that it can be seen from the observation that the algorithm can accurately retrieve the target in the complicated traffic area and track and identify it. Although the process will be blocked by vehicles or pedestrians, there is still no problem that is currently lost, which effectively validates the feasibility of the algorithm.

Figure 8: Target extraction image 2.

5. Conclusion

We propose a target detection method based on the SSD algorithm and feature extraction fusion. The algorithm is based on the traditional SSD algorithm. The algorithm adopts an adaptive strategy to dynamically adjust the search window according to the change of the running status of the image and combines the appearance depth feature and the movement depth feature in combination with the feature fusion method of weighted correlation, and finally improves the precision extraction of the objective function. Through experiments, the algorithm can maintain a high and stable target detection effect under different complexity of the image environment and is more suitable for the environment changeable target detection environment. However, it still cannot be effectively reduced in the time-consuming aspect of the algorithm. This will serve as a research focus in the future and will be further studied.

Data Availability

The data used to support the results of this study needs to be obtained with the consent of the corresponding author.

Conflicts of Interest

The authors declare that they have no competing interests.

Authors' Contributions

The authors have equally contributed to the manuscript. All authors read and approved the final manuscript.

Acknowledgments

This article is supported by Sun Yat-Sen University Xinhua College 2017 School-Level Scientific Research Startup Fund General Project: Research and Design of Target Trajectory Tracking System Based on Depth and Visual Information Fusion (Project Code: 2017YB001).

References

[1] L. Fan, L. Pengyuan, and L. Bing, "Deep learning model design of video target tracking based on tensor flow platform," *Laser&Optoelectronics Progress*, vol. 9, no. 15, pp. 277–285, 2017.

[2] G. Hao, X. Xiang-Yang, and A. Zhi-Yong, "Advances on application of deep learning for video object tracking," *Acta Automatica Sinica*, vol. 42, no. 6, pp. 834–847, 2016.

[3] Y. Zheng, C. Quanqi, and Z. Yujin, "Deep learning and its new progress in object and behavior recognition," *Journal of Image and Graphics*, vol. 19, no. 2, pp. 175–184, 2014.

[4] S. Wu, S. Wang, R. Laganiere, C. Liu, H. S. Wong, and Y. Xu, "Exploiting target data to learn deep convolutional networks for scene-adapted human detection," *Ieee Transactions On Image Processing*, vol. 27, no. 3, pp. 1418–1432, 2018.

[5] C. Shiyu, L. Yuehu, and L. Xinzhao, "Vehicle detection method based on fast R-CNN," *Journal of Image and Graphics*, vol. 22, no. 5, pp. 671–677, 2017.

[6] Z. Guangjun, W. Xuchu, N. Yanmin, T. Liwen, and Z. Shaoxiang, "Deep SAE feature learning based segmentation for digital human brain image," *Journal of Computer-Aided Design & Computer Graphics*, vol. 28, no. 8, pp. 1297–1305, 2016.

[7] X. Y. Qian, L. Han, Y. Wang, and M. Ding, "Deep learning assisted robust visual tracking with adaptive particle filtering," *Signal Processing-Image Communication*, vol. 60, no. 1, pp. 183–192, 2018.

[8] Z. Tang, H. Wu, W. Wang, J. Wei, and T. Huang, "self-adaptive SSD caching system for multiobjective optimization in virtualization environment," *Journal of Software.*, vol. 28, no. 8, pp. 1982–1998, 2017.

[9] W. Jiewen, Z. Yinwei, L. Weilin, and G. Canzhang, "Batch renormalization of real-time object detection algorithm YOLO," *Application Research of Computers*, vol. 35, no. 11, pp. 1–9, 2018.

[10] Y. Han and H. Hahn, "Visual tracking of a moving target using active contour based SSD algorithm," *ROBOTICS AND AUTONOMOUS SYSTEMS.*, vol. 53, no. 3-4, pp. 265–281, 2005.

[11] H. U. Yin and Y. A. N. G. Jing-yu, "Tracking algorithm based on fusion of SSD and MCD robust to partial occlusion," *Journal of System Simulation*, vol. 22, no. 4, pp. 908–911, 2010.

[12] H. Song, X. Zhang, B. Zheng, and T. Yan, "Vehicle detection based on deep learning in complex scene," *Application Research of Computers*, vol. 35, no. 4, pp. 1–5, 2018.

[13] W. Liu, D. Anguelov, D. Erhan et al., *SSD: single shot multi box detector*, Springer International Publishing, 2016.

[14] S. Hong, T. You, S. Kwak, and B. Han, "Online tracking by learning discriminative saliency map with convolutional neural network," in *Proceedings of the 32nd International Confernce on Mackine Learning*, pp. 597–606, 2015.

[15] "Imagenet dataset [DB/OL]," 2020, http://www.image-net.org/download-imageurls.

Hindawi
Wireless Communications and Mobile Computing
Volume 2020, Article ID 8879616, 10 pages
https://doi.org/10.1155/2020/8879616

Research Article

Research and Analysis of Sports Training Real-Time Monitoring System Based on Mobile Artificial Intelligence Terminal

Biao Ma,[1,2] Shangqi Nie,[3] Minghui Ji,[4] and Jeho Song[2]

[1]*Anhui Province Ying Shang County Shibalipu Town Center School, 236200 Fu Yang, China*
[2]*Wonk Wang University, Sports Science, Iksan City, Republic of Korea*
[3]*School of Physical Education, HuangHuai University, Zhumadian, China*
[4]*Anhui Province Ying Shang County Balihe Town Center School, 236200 Fu Yang, China*

Correspondence should be addressed to Jeho Song; sjhao@wku.ac.kr

Received 9 June 2020; Revised 10 August 2020; Accepted 27 August 2020; Published 28 September 2020

Academic Editor: Wei Wang

With the rapid development of artificial intelligence, related technologies and applications come into being, and industries based on artificial intelligence are booming, among which image recognition and target tracking technologies are widely used in various fields, especially in the fields of security monitoring and augmented reality. In this paper, combined with the characteristics of athletes, based on mobile artificial intelligence terminal technology, the C/S mode of athlete training process monitoring system is developed and designed, which uses GPS to obtain the real-time position information of athletes and provide real-time guidance for athletes. In order to reveal the changing rules of various indexes of athletes in training state, the author makes synchronous tracking analysis from the aspects of individual sports function characteristics of athletes, training plan arrangement of coaches, brain function state, routine physiological and biochemical indexes, nutrition regulation, and injury conditions.

1. Introduction

It has long been a good and strong wish of mankind to thoroughly understand the mechanism of human intelligent behavior and to create intelligent machines that can simulate intelligent behavior [1]. During the long-term medical supervision with the sports team and the monitoring of athletes' physical function, a large number of original data have been generated, and the influence of various indexes on athletes' functional state is different in technology and algorithm [2]. Real-time monitoring of athletes' functional status and their positions in the training process and targeted determination of guidance programs and real-time guidance are very important means of scientific training today. A visual target tracking for mobile terminals is a research direction developed in the field of computer vision in recent decades [3]. The visual target tracking system of mobile terminal includes three main functional modules, namely, image acquisition module, image processing module, and image display module. The emergence and rapid development of artificial intelligence has brought hope to the realization of this wish [4]. Its research has prolonged the function of the human brain, deepened and expanded the intelligent labor of human beings, and made the scientific and technological revolution develop at an unprecedented speed. Many latest intelligent technologies, such as telepresence technology, virtual field technology, multi-intelligence technology, artificial neural network, multisensor fusion technology, and so on, have been applied in intelligent mobile robots.

Artificial intelligence is a frontier discipline, which is developed on the basis of computer science, cybernetics, information theory, systems science, philosophy, and other disciplines [5]. In recent years, mobile devices, especially mobile phones and tablet computers, have developed very rapidly. The chip processing speed of mobile devices is getting closer to personal computers. The memory capacity of mobile devices is almost the same as that of ordinary PC, and some even exceed PC [6, 7]. If the athletes' functional information, position, road conditions, and other information can be mastered in real time during training, the best

guidance scheme can be obtained through computer processing and provided to coaches, and the athletes can be communicated and guided in real time through wireless transmission; it will greatly change the current situation of middle-distance and long-distance running training in our country and improve the training quality [8]. Therefore, it is proposed to use computer technology and network technology as auxiliary means for data collection, data calculation, and conclusion analysis, and a real-time monitoring system for athletes' functional status is developed. The combination of sports and artificial intelligence is an inevitable trend of social development to conform to the new trend of the development of artificial intelligence in the world. This article hopes to attract more people to participate in the research of intelligent sports and provide new thinking for solving practical problems in sports.

The introduction part mainly introduces the popularity of mobile devices, the application status, and the wide application prospect of visual target tracking algorithm on mobile terminals. The related work in the second section summarizes the research status of visual target tracking in universities and scientific research institutions at home and abroad and briefly describes the key problems and solutions to be solved in the realization of mobile visual target tracking algorithm. The third part discusses the realization of a target tracking system based on the KCF-LD algorithm on the mobile artificial intelligence terminal, taking the KCF-LD algorithm as the target tracking module of the system, and introduces the design and implementation of the real-time monitoring system for physical training in detail. In the fourth part of this paper, the target tracking system implemented in this paper is used to experiment on real scenes, the hardware configuration of the experiment is listed, and the experimental scenes are classified. Finally, the experimental results are classified and compared, and some factors affecting the real-time performance of the system are analyzed, which proves the practicability of the KCF-LD algorithm in mobile terminals. At last, this paper summarizes the work done in this paper, expounds the innovation of this paper, and summarizes the shortcomings of the real-time monitoring system of sports training based on mobile artificial intelligence terminal, and looks forward to the future work.

2. Related Work

As computers have entered almost all fields more and more widely, it is an important response of computer intelligence to let computers exchange information with people in the way that people are accustomed to [9, 10]. Although sports practical computer science has made rapid development in just over ten years and has effectively promoted the progress of sports, we have to calmly see that sports practical computer technology still lags far behind the development of computer science [11]. Anderson studied the related problems in the field of artificial intelligence [12]. Arkhipov proposes "Explanation of Artificial Intelligence: Insights from Social Science" research theory [13]. Santos proposed artificial intelligence: why we must do it well [14]. Let real objects be more vivid and easier to understand and at the same time enhance user experience. When augmented reality technology is applied to mobile terminals, real objects are displayed in the mobile devices in the form of images through cameras. Due to the movable characteristics of mobile terminals, objects of interest to users in the images will follow the moving positions [15, 16]. However, for the time being, most coaches and researchers in our country focus their monitoring on athletes before or after training, and the monitoring during training only uses rhythm telemetry to monitor training. Limited by the data collection methods, the most basic artificial intelligence analysis can analyze the tracking of a single target and mine valuable information [17]. For example, in ball games, players control the trajectory of the ball to win the score. Therefore, how to design an effective human body detection algorithm to solve these two problems has become the focus of research in computer vision-related fields.

3. Methodology

The physical function of athletes is an important factor that determines their competitive level. It directly affects the performance of their athletic ability and technical level, thus affecting their athletic performance. The whole process monitoring system of athlete training developed and designed in this paper adopts C/S architecture. The database of this architecture is composed of two parts: client application program and database server program, which are called foreground program and background program, respectively. The whole system consists of electromyogram collector, interface card, synchronizer and electromyogram acquisition, and analysis software [18]. For collective events, the tactical cooperation and tactical choice between athletes are important factors that determine the competition. It is necessary to continuously track and analyze the athletes' movements in the competition. The movable intelligent control terminal is used for intelligent control of household appliances, but most household appliances are currently realized by infrared communication. The infrared communication part of the system includes infrared emission and infrared receiving functions, and the infrared receiving function is mainly used for the learning function of the system on other remote controllers [19, 20]. Human motion recovery of target video refers to using multiple cameras to capture human motion at the same time and recovering three-dimensional human posture through fusion processing of multiple data. Then, through computer programming, the computer simulates the reasoning thinking process (the process of selecting knowledge) peculiar to human beings, thus completing the intelligent problems that only human beings can solve.

Sixteen athletes were randomly and evenly divided into experimental group and control group according to male, female, and events. The basic conditions of the experimental group and control group were roughly the same. Before the experiment, the indexes of the experimental group and the control group were statistically tested. The results showed that there was no significant difference, and they were in accordance with statistics. Contrast requirements are shown in Table 1 and Figure 1.

Table 1: Comparison of basic conditions between the experimental group and the control group.

Group	Training years (years)	5000 m score (min)
Experimental group	5.3	16.25
Control group	6.4	15.33

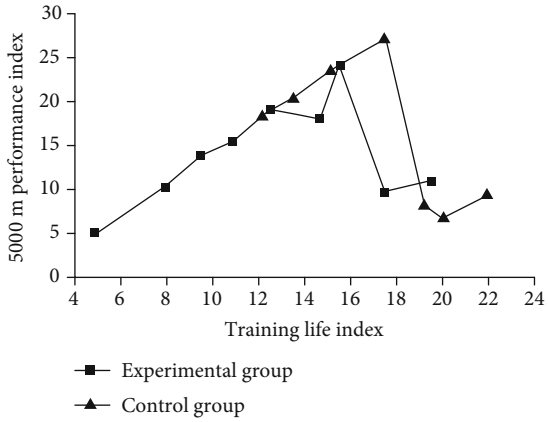


Figure 1: Comparison of basic conditions between the experimental group and the control group.

During the experiment, the GPS system was used to obtain the changes of athletes' basic sports quality in different positions in real time, and the wireless dialogue system was used to guide the athletes' training process and control their control intensity within the target range. This article takes the comparison results of 100 m and 50 m freestyle events as an example for analysis, as shown in Table 2 and Figure 2.

Spatial consistency means that the motion speeds of pixel points in adjacent positions are assumed to be consistent. This is a special assumption of optical flow method, because there is only one basic equation of optical flow method, but to find the velocity in horizontal and vertical directions according to the equation, only multiple equations can be solved simultaneously. According to the above assumptions, we can obtain the optical flow equation, see equation (1).

$$F = K \times \frac{(V - Vt)^2}{V}. \tag{1}$$

Equation (1) indicates that the gray value of the pixel point remains unchanged after V time, K, and t displacement. The first order Taylor series expansion is carried out on the above equation to obtain formula (2).

$$V = \sqrt{\frac{F \times Vt}{K} + \frac{F^2}{4 \times K^2}} + Vt + \frac{F}{2 \times K}. \tag{2}$$

Assuming that the brightness is constant in a small local neighborhood of I, equation (3) is obtained.

$$\beta_{xy} = \frac{I_{xy}}{\sum_{a-1}^{g-1} I_{xa}}. \tag{3}$$

Table 2: Comparison of 100 m and 50 m freestyle performance between the experimental group and the control group after training.

Group	Distance (m)	Time (s)
Experimental group	50	30
Control group	100	53

Assuming R_i is used to represent the motion speed of every pixel point, of course, the speed here is a vector that contains the magnitude and direction information of the motion speed of the pixel point; then, the above formula can be simplified as follows:

$$R_i = M_i \sum_{i=1}^{N_r} \widehat{R}_{i,r}. \tag{4}$$

According to the present knowledge in the sports field, various artificial intelligence technologies and knowledge engineering methods are applied to deeply understand the internal semantics of user questions, mine the answers that users really care about, and correlate and recommend relevant knowledge that users may be interested in. Whether the imaging is fast and the system is smooth is the final decisive factor for the rationality of the architecture and algorithm design for the input reactants. With the continuous movement of the target in the scene, the target and the background near the target are constantly changing, so the template established for the target also needs to be constantly changed, which requires continuous learning of the target template in the tracking process and the establishment of a suitable model for the target. According to the overall design requirements, heart rate monitoring sensors, satellite positioning receivers, and language transmitters are developed and selected to realize real-time monitoring of various indexes of athletes' training process and obtain various information [21]. On the one hand, this evaluation can be used to judge the fatigue degree and recovery degree of athletes caused by sports training and to judge the fitness of athletes to sports load and the training effect during sports training. At the same time, we can know the changes of athletes' physical function, sports quality, and competitive ability at different positions at any time. We can use wireless cluster network to realize the communication between athletes and coaches during the training process. Athletes only need to wear a waterproof earphone to hear the guidance of coaches at any time.

After 3 months of experimental training, the 12 athletes in the experimental group and the control group were tested under the same conditions of time, place, and climate of their respective events. In order to avoid accidental factors, we conducted the same second test on the 10th day after the first test and compared the best results. The results show that the improvement of sports performance of the experimental group trained with real-time monitoring system is significantly better than that of the control group trained with traditional training (see Table 3 and Figure 3).

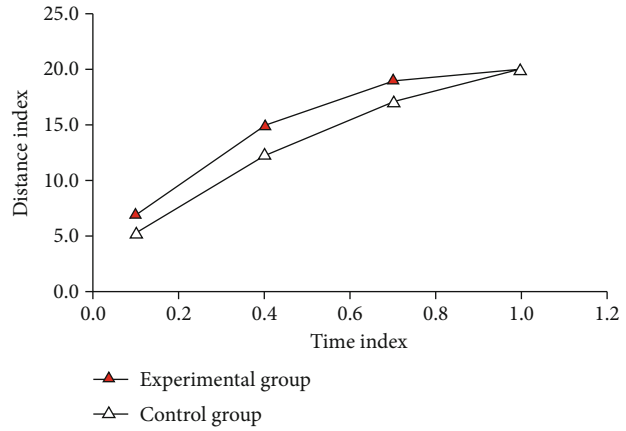


Figure 2: Comparison of 100 m and 50 m freestyle performance between the experimental group and the control group after training.

Table 3: Table comparing the growth of sports scores between the experimental group and the control group after training.

Projects	Original score (min)	Current score (min)
Marathon	180.33	179.66
5 000 m	20.56	18.24
Swimming	2.03	1.65

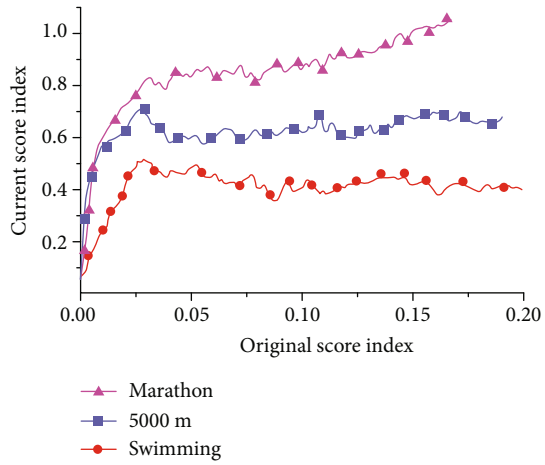


Figure 3: Comparative chart of sports performance growth between the experimental group and the control group after training.

Let the original target window be D, and its coordinates, width, and height are Dx, Dy, D_{width}, and Dheight. The candidate set takes D as the original target and 1.3 as the scaling factor and makes 10 transformations with different scales. See formula (5) for enlargement transformation formula and formula (6) for reduction transformation formula.

$$\left(D_{\text{width}}, D_{\text{height}}\right) * 1.2^{k}. \tag{5}$$

$$\frac{\left(D_{\text{width}}, D_{\text{height}}\right)}{1.2^{k}}. \tag{6}$$

In formulas (5) and (6), k is 1, 2, 3, 4, 5, 6, 7, 8, 9, and 10, respectively, and the candidate window is guaranteed to be no less than 10 pixels in the scaling process.

Given a training sample set and labels f_i and X, the training process of classifier B is the process of minimizing the regularization loss function value by finding the best parameters. The linear classifier is generally expressed as equation (7).

$$B\left(\vec{X}\right) = \prod_{i=1}^{m}\left(f_i\left(\vec{X}\right) - f_i\left(\vec{X}_w\right)\right)^{1/m}, \qquad R_i = M_i \sum_{i=1}^{N_r} \widehat{R}_{i,r}. \tag{7}$$

m in the above formula represents the vector dot product, and minimizing the square error of the classifier for samples and labels is expressed as formula (8).

$$B\left(\vec{X}\right) = \prod_{i=1}^{m}\left(f_i\left(\vec{X}\right) - f_i\left(\vec{X}_w\right)\right)^{w_i}. \tag{8}$$

A support vector machine and a regularization least square method are also based on this framework, in which SVM uses hinge loss function:

$$B\left(\vec{X}\right) = \prod_{j=1}^{n} u_j\left(x_j\right)^{v_j}. \tag{9}$$

Ridge regression in the least square method uses a square loss function:

$$\omega_s = \frac{2n\pi}{l}\sqrt{\frac{H}{m}}(n = 1, 2, 3, \cdots). \tag{10}$$

As can be seen from the graphs in Table 4 and Figure 4, the KCF tracking algorithm lacks precision when dealing with target disappearance and low resolution, among which the TLD algorithm performs better when dealing with these two situations. If the target center error is allowed to exceed 35 pixels, TLD algorithm's precision exceeds KCF.

The interface card directly communicates with the computer and software through wired mode and sends commands and receives data to the electromyography collector and synchronizer in wireless mode at the same time. The collector is worn by athletes to collect electromyography, store data, and send in real time. In addition, the external contour loses image information that may be useful for understanding the target, such as texture and internal edge. For example, walking, trotting, and other actions have very similar human body contours on some image frames. The movable intelligent control terminal needs to accurately control the DC motor. A rotary encoder is introduced to detect the rotation feedback amount of the motor to judge the rotated angle, and the coordinates of the actual position of the motor are calculated through an algorithm and fed back to the control circuit. The synchronizer sends a measurement start and

Table 4: KCF algorithm precision curve disadvantage table.

Algorithm	Target center error	Accuracy
KCF algorithm	20.33	0.37
TLD algorithm	40.15	0.42

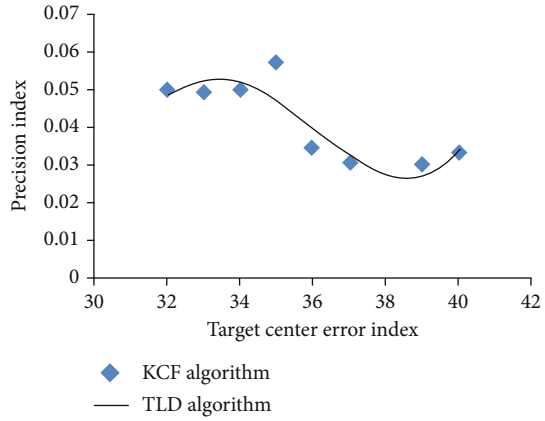


Figure 4: KCF algorithm precision curve inferior position chart.

end synchronization instruction to the camera or other equipment to coordinate the synchronization of each measurement equipment. The acquisition and analysis software performs measurement control, real-time monitoring, data management, and electromyography signal analysis. The technical and tactical analysis methods of different sports are different, but the basic steps are similar: first collect the original data, then extract the effective information, and then conduct in-depth analysis on the data. At the monitoring terminal, desktop geographic information system (GIS), mobile geographic information system, and virtual reality technology are used to present athlete's heart rate information and position information in the form of data. Meanwhile, spatial analysis technology is applied to analyze athlete's speed and heart rate at any position and time period. Through the use of computer network technology and artificial intelligence methods, real-time monitoring of athletes' functional states is realized, thus greatly improving the accuracy and efficiency of evaluation and diagnosis of athletes' functional states.

The following Table 5 and Figure 5 list the processing results of KCF algorithm tracking algorithm. From the figure, it can be concluded that KCF tracking algorithm performs better in dealing with deformation, occlusion, cluttered background, and other aspects; solves most of the key difficulties in target tracking; and has low computational complexity, which is suitable for application to mobile terminals.

The tracker of TLD algorithm uses optical flow method. KCF-LD replaces the tracking algorithm in TLD with the KCF tracking algorithm. The KCF-LD algorithm uses KCF tracker tracking result as reference and uses the nearest neighbor classifier to judge the tracking result. At this time, the detector is used to detect in the whole image. The method is still using the sliding window mode. All candidate windows obtained from the sliding window are input to the cascade classifier for screening. Finally, the remaining candidate windows are clustered to finally obtain the location of the target. The algorithm framework is shown in Figure 6.

Table 5: Advantage table of precision curve of KCF algorithm.

Processing contents	Target center error	Accuracy
Deformation	26.41	0.21
Shelter	30.12	0.35
Cluttered background	38.24	0.29

Table 6 and KCF and DCF in Figure 7, respectively, show that the correlation filter with Gaussian kernel, and linear kernel is used for the kernel function. Table 6 shows that the correlation filter algorithm with Gaussian kernel function and directional gradient histogram feature has the highest accuracy, while the correlation filter algorithm with linear kernel function and directional gradient histogram feature has the fastest speed.

According to the characteristics of athletes' training center, rate information is obtained by heart rate monitoring sensors, real-time athletes' running distance, road conditions and positions obtained by differential global satellite positioning technology, etc. The tracker uses the median optical flow method to calculate the position of the target in the current frame according to the pixel states of the previous frame image and the current frame image. The detector scans the subwindows in the image. All subwindows pass through the cascade classifier and then filter out the background window. Because of the requirement of real-time performance, a simple noise suppression method is adopted in real-time detection, i.e., if the absolute value of the data is less than a predetermined threshold within a certain period of time, it is background noise, and the data is assigned a zero value. The sensor with the function of heart rate detection and sensing and GPS are integrated into a light and thin intelligent bracelet to be worn on athletes' hands. Through wireless cluster mobile communication technology, coaches only need to hold a microcomputer or a mobile PC to monitor various index data information in the training process of athletes. Highlight the individual characteristics of athletes, pay attention to the internal relations of various indicators of athletes, especially the dynamic changes of indicators according to the time sequence with the training process, from the characteristics of individual athletes' sports function. Its working principle is to use photoelectric conversion to replace the mechanical geometric displacement on the output shaft with the corresponding electric pulse or electronic digital quantity. The composition of photoelectric encoder includes grating disk, photoelectric detection, and other devices, wherein the grating disk is provided with fan-shaped light transmission channels uniformly on a circular plate with a certain diameter.

4. Result Analysis and Discussion

Single-feature human body descriptors based on machine learning are sensitive to noise interference, and multifeature fusion can further distinguish target objects that cannot be distinguished by single-feature description. Coaches' training

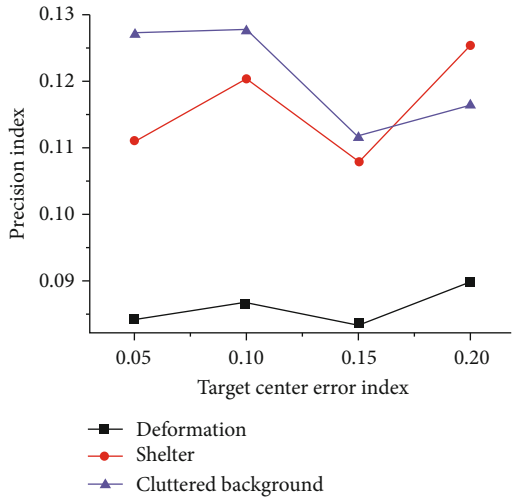


Figure 5: KCF algorithm precision curve dominance chart.

plan arrangement, brain function state, routine physiological and biochemical indexes, nutrition regulation, injury and illness conditions, and other aspects are systematically tracked and measured for a long time to find out the change rule of their function state. At the same time, it is possible to talk back to athletes through waterproof earphones equipped with voice transmitters in the athletes' ears, which will not affect athletes' training results. The noise caused by electromagnetic radiation is weakened by good shielding in hardware. A 50 Hz notch filter is used to filter out 50 Hz pollution, but the EMG signal frequency spectrum is concentrated in 20~200 Hz, which filters out noise and EMG signal at the same time. By collecting the heart rate data of wheelchair racers throughout the season, the relationship between aerobic training, strength training, and athletic performance was studied. The concepts in the real world and the relationships between concepts are abstracted into entities and methods. The accumulation of knowledge instances is completed through entities, and the enrichment of knowledge expression is realized through methods. The construction of knowledge base can be completed more quickly based on customer historical data. Before training the weak classifier, set the same initial weight for each sample, generate a weak classifier for each round of training, and iteratively update the weight for each sample. Especially in the case of extremely high frequency or very steep pulse preface, more attention should be paid; only in this way can the influence of distributed capacitance between wires be reduced.

TLD, KCF, and KCF-LD are used to test the 26 target tracking reference videos. Two methods are used to determine whether an image is correctly tracked. The first method is based on the overlap degree. When the proportion of the overlap area between the tracking result and the real target position exceeds 25%, the tracking is considered correct. See Table 7 and Figure 8 for the statistical results of tracking correct frames and accuracy based on overlap, and see Table 8 and Figure 9 for the statistical results of tracking correct frames and accuracy based on center offset distance.

High-reliability hardware foundation can ensure no hardware failure, but there are a large number of signal noise interference sources in life and production. The training samples are given initial weights, and important features with strong discrimination ability are selected from a large set of candidate features. The weights are updated in each iteration, and the weighted classification errors are used as the cost function to select features and add them to the overall strong classifier. Several target windows are obtained, and the learner integrates the results of the tracker and the detector and extracts the region most similar to the target from the integrated results to train the classifier so that the classifier can learn the latest appearance of the target. At this time, if the dynamic changes of athletes' heart rate are mastered through real-time monitoring and the heart rate information of the body responding to the exercise load is transmitted in real time through appropriate methods, the exercise intensity can be adjusted at any time to achieve ideal training effect. The choice of software system mainly considers its performance, function, price, practicability, compatibility, ease of management, etc. At the same time, it should also consider its future upgrade. Use single interface drive to enhance user experience. At the same time, the system can realize remote login. When coaches cannot supervise and guide athletes on the spot, they can monitor the whole training process and guide them in real time only through remote login.

Assuming that an $N \times 1$ one-dimensional vector is used to represent an image block, denoted as x, x is taken as a reference sample, and the reference sample is shifted to obtain a series of negative samples, thus obtaining the positive and negative samples required for training the classifier. This operation can be represented by the following permutation matrix, see equation (11).

$$H = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}. \tag{11}$$

The set can also be seen as consisting of two parts; half of the elements are obtained by forward cyclic shift of the reference sample, and the other half of the elements are obtained by reverse cyclic shift of the reference sample.

$$HD = \frac{1}{l} \sum_{i=1}^{l} A_i \oplus B_i. \tag{12}$$

If the vector x is the pixel value of the first row in the image, a square matrix of $n \times n$ dimensions can be obtained by performing a cyclic shift operation on x; each row of the square matrix is a vector obtained by a cyclic shift operation of the vector x, and the square matrix can be expressed as formula (13).

$$C \times \left(\sum_{i=1}^{I} q_{i,j} \right) \geq u_j. \tag{13}$$

The cyclic matrix structure is selected because when the classifier evaluates a plurality of different subwindows, the

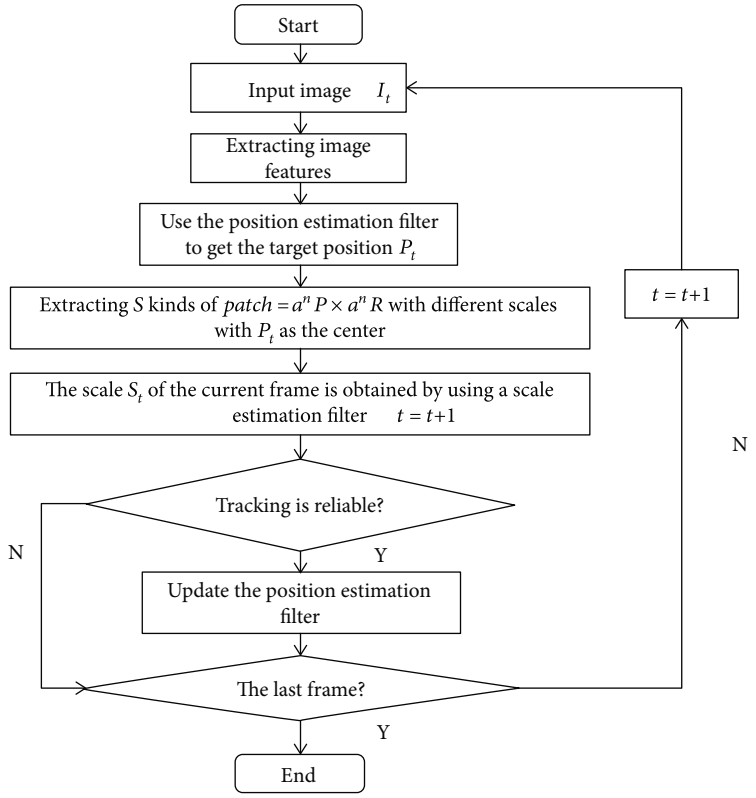


Figure 6: KCF-LD algorithm framework diagram.

Table 6: Comparison of average results between KCF algorithm and DCF algorithm.

Algorithm	Average accuracy (%)	Average frame rate
KCF algorithm	72.3	170
DCF algorithm	74.1	216

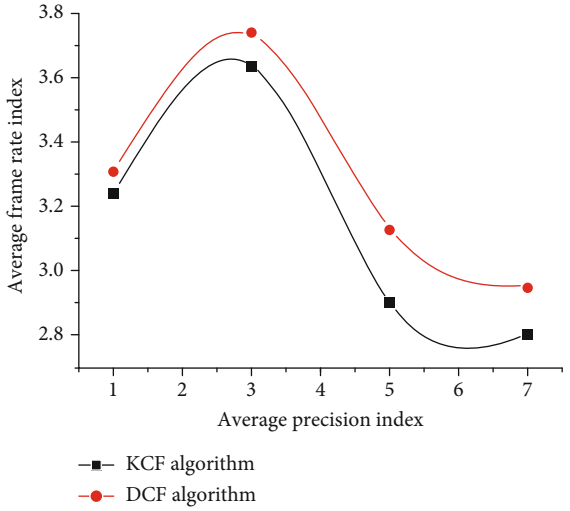


Figure 7: Comparison of average results between KCF algorithm and DCF algorithm.

vector convolution of the cyclic matrix can be used to obtain the calculation result. If the product C represents the convolution of the vector u_j and the vector $q_{i,j}$, the result can be calculated in the Fourier domain:

$$C \times \left(\sum_{i=1}^{I} q_{i,j} \times \left(\frac{f_{i,j}}{d_{i,j} + g} \right) 2 \right) \geq u_j. \tag{14}$$

This cyclic matrix structure can be converted into diagonal matrix by discrete Fourier transform, and this attribute is not related to the generation of reference sample x, so the cyclic matrix structure can be reexpressed as follows in combination with discrete Fourier transform:

$$x(t) = Ax(t) + Bu(t). \tag{15}$$

If the above cyclic matrix structure is applied to linear regression, the training samples in equation (15) are all obtained by cyclic shift of the reference samples. It can be regarded as a noncentral covariance matrix, which is brought into equation (16) to obtain:

$$y(t) = Cx(t). \tag{16}$$

In this experiment, the Volkswagen video where the target disappeared was first tested. In the experiment, the KCF tracker threshold value was set to 0.27, and a smaller value was set because it has certain resistance to low resolution. On the other hand, it also indicates that the KCF-LD

Table 7: Statistical results of correct frames and accuracy of tracking based on overlapping degree.

Algorithm	Correct frame number	Accuracy
TLD algorithm	23	0.3
KCF algorithm	52	0.8
KCF-LD algorithm	29	0.4

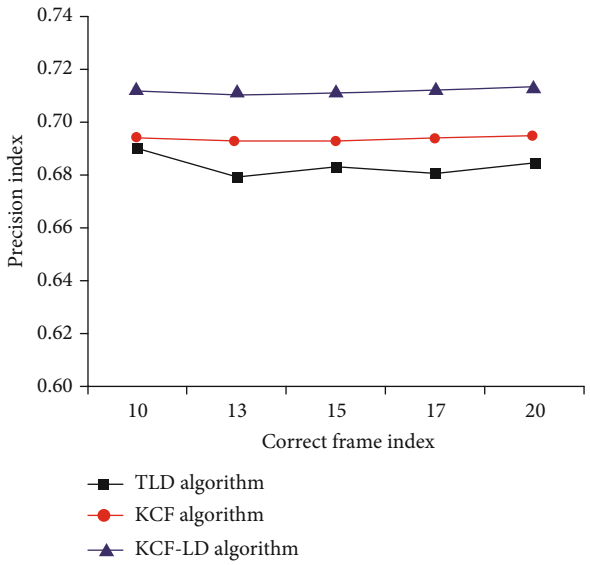


Figure 8: Statistical results of correct frames and accuracy of tracking based on overlapping degree.

Table 8: Statistical results of correct frames and accuracy of tracking based on center offset distance.

Algorithm	Correct frame number	Accuracy
TLD algorithm	32	0.6
KCF algorithm	48	0.2
KCF-LD algorithm	26	0.5

algorithm gives greater trust to the tracker in a low resolution video. Because the detector needs to pass through three-level classifiers when detecting several image blocks, the classifier with larger error will make the result of the detector inaccurate, thus affecting the tracking accuracy of the whole tracking algorithm. Figure 10 shows the test results of the target disappearance video Volkswagen.

Motion artifacts are mainly introduced by the relative sliding between the electrode and the skin and the motion of the connecting lead wire. The main frequency range of motion artifacts is concentrated in 0 ~ 20 Hz, and the frequency spectrum of surface electromyography data is concentrated in 20~ 20~200 Hz. Through the integration of indicators and comprehensive analysis, controlling the relationship between health status and fatigue degree in training not only helps to improve the training level but also forms the basis for evaluating whether players are suitable for playing and playing time. The weights of the training samples are

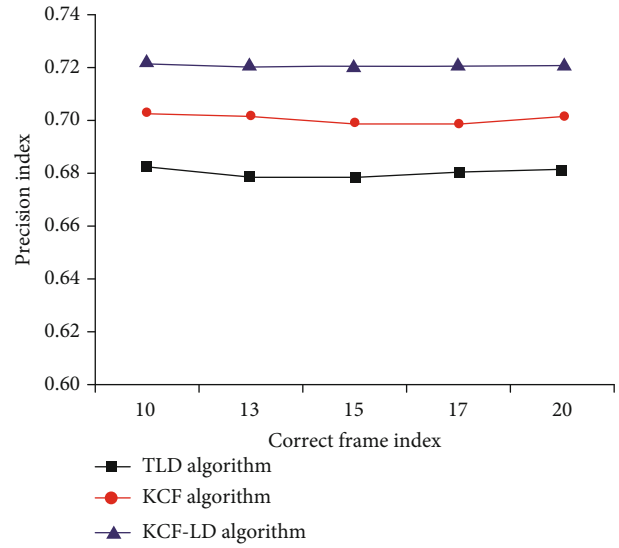


Figure 9: Statistical results of correct frames and accuracy of tracking based on center offset distance.

regarded as the same in each round, so all the features need to be trained once to form a weak classifier, and the classification results are calculated once and stored in the statistical table. Use shielded wire to connect the input terminal with weak signal, and the outer shielding layer of shielded wire must be well connected with common ground wire. When measuring with an oscilloscope, try to cut off the influence brought by the distributed capacitance of the connecting line. The signal connection mode is selected to be connected with the oscilloscope probe. Usually, the resting potential is deducted from the electromyography signal to eliminate the influence on the electromyography, but other factors are considered in this study; only the instrument bias is corrected, and the resting potential is not deducted. Although the team did not disclose the specific details, it is reasonable to believe that it is one of the means to help the team to make continuous progress by collecting physiological indexes through wearable equipment, continuously improving training parameters and optimizing training models through machine learning and using the results to guide the competition.

Using physiological and biochemical data indexes such as heart rate, blood lactic acid, blood urea, endocrine, and the like, the training process of athletes can be monitored in real time in a planned way, the adaptive characteristics of athletes can be mastered, and the scientificity of training implementation can be guaranteed. Due to the versatility of the development tool, it is very convenient to transplant the application program. When replacing the background database management system, only the appropriate database interface needs to be selected, and the application program itself does not need to be modified, which provides great convenience for future system upgrade and background database replacement. This is also the basis for monitoring training intensity by heart rate. According to this principle, we chose a mobile exercise heart rate monitor to carry out wireless heart rate monitoring to provide real-time heart rate during athletes' training for comprehensive integrated information. The task

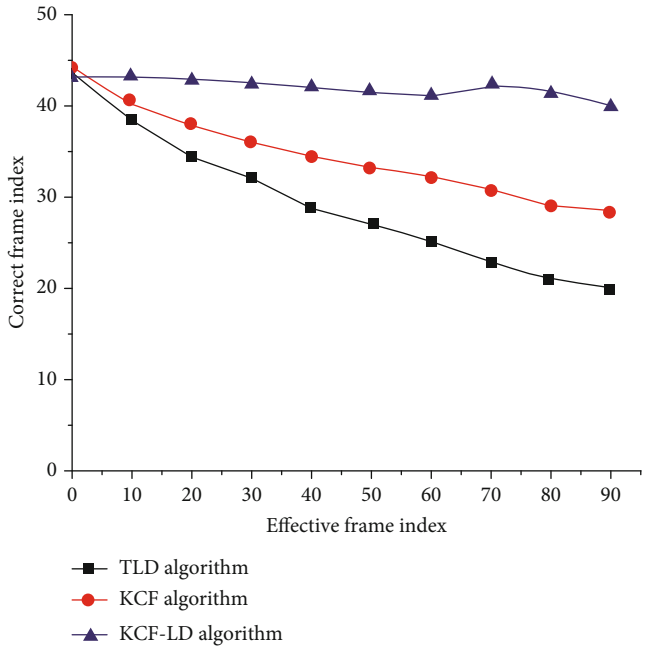


Figure 10: Tracking results of TLD, KCF, and KCF-LD on video Volkswagen.

of the tracking algorithm is to estimate the motion state of the target. The tracker first estimates the motion of some feature points in the rectangular region and estimates the reliability of these motions. The median of some of the most reliable motions is taken as the motion of the final rectangular frame. By testing whether the immune indexes are within the normal range, the body's health condition can be checked, and the possible health risks can be predicted. The change of EMG signal frequency can be regarded as fatigue analysis. Through the change trend of EMG signal frequency with time, whether the muscle is fatigue can be judged according to the physiological characteristics of the tested muscle. In this system, the athlete's position is directly displayed on the background of the electronic map. Taking the athlete as a geographical object, it can measure, analyze, and simulate the spatial information on which the athlete is based, thus indirectly obtaining the athlete's training status.

5. Summary

The establishment of training real-time monitoring system has realized the functions of obtaining athletes' heart rate information and position information in real time, processing and analyzing the information by computer, establishing a real-time scientific guidance scheme, and carrying out real-time information exchange. On the basis of fully considering the system structure, an intelligent mobile and learning system based on a single chip microcomputer is designed. Among them, the intelligent mobile function module integrates temperature and infrared sensors, as well as display and voice modules, which can automatically track and avoid obstacles and indicate the temperature, time, and date of the scene, making the design close to life. The establishment of real-time monitoring system for training combines the laws of human physiology, human movement, and modern science and technology, which has strong theoretical basis, practical application value, and scientific content. This system can monitor the training process of athletes online in real time and analyze and compare the training effect offline, which has guiding significance for the training process.

Because of the limitation of research time, methods, and experimental objects, it is necessary to make further research. At the same time, it is suggested that coaches of relevant departments and high-level sports teams can fully understand and participate in scientific training from thought to action so that the training level can reach a new level.

Data Availability

All the data used in this manuscript are included in the manuscript.

Conflicts of Interest

The author declares that there is no conflict of interest regarding the publication of this paper.

Acknowledgments

This paper was supported by Wonkwang University in 2020.

References

[1] Q. Guoyong, Z. Hucheng, and S. Yafei, "Management and application of high school students' physical fitness monitoring data," *Physical Education*, vol. 39, no. 6, pp. 60-61, 2019.

[2] Z. Hengliang, M. Wanli, and L. Hongyou, "Using data series to infer the analysis of the load data of college football special classroom teaching competitions," *Sports Science Research*, vol. 39, no. 6, pp. 83–90, 2018.

[3] C. Feng, H. Kun, and W. Lu, "Data analysis on physical health of primary school students in Wuhan based on monitoring of third-party organization," *Hubei Sports Science and Technology*, vol. 38, no. 1, pp. 4–7, 2019.

[4] L. Xiaolan, "Data collection system and recognition design of human motion," *Computing Technology and Automation*, vol. 38, no. 2, pp. 146–150, 2019.

[5] Q. Di and Z. Feng, "Research on the development of artificial intelligent sporting products," *Sports Culture Guide*, no. 6, pp. 104–108, 2018.

[6] L. Zhenhua, L. Yu, and W. Wenbo, "Research on the construction of early warning mechanism for college students' physical health," *Liaoning Sports Science and Technology*, vol. 40, no. 6, pp. 79–85, 2018.

[7] J. Wilkins, "Is artificial intelligence a help or hindrance?," *Network Security*, vol. 2018, no. 5, pp. 18-19, 2018.

[8] L. Xuan, "Application and research of LBS position trajectory tracking service platform in marathon movement," *Science Teaching Journal*, no. 8, pp. 249-250, 2018.

[9] T. Miller, "Explanation in artificial intelligence: insights from the social sciences," *Artificial Intelligence*, vol. 267, pp. 1–38, 2019.

[10] A. Curioni, "Artificial intelligence: why we must get it right," *Informatik Spektrum*, vol. 41, no. 1, pp. 7–14, 2018.

[11] M. Milan, M. Branko, M. Dragan et al., "Artificial intelligence in clinical medicine and dentistry," *Vojnosanitetski Pregled*, vol. 74, no. 3, pp. 267–272, 2017.

[12] D. Anderson, "Artificial intelligence and applications in PM&R," *American Journal of Physical Medicine & Rehabilitation*, vol. 98, no. 11, pp. e128–e129, 2019.

[13] V. V. Arkhipov and V. B. Naumov, "Artificial intelligence and autonomous devices in legal context: on development of the first Russian law on robotics," *SPIIRAS Proceedings*, vol. 6, no. 55, pp. 46–62, 2017.

[14] O. C. Santos, "Artificial intelligence in psychomotor learning: modeling human motion from inertial sensor data," *International Journal of Artificial Intelligence Tools*, vol. 28, no. 4, article 1940006, 2019.

[15] G. Liang, X. Lan, J. Wang, J. Wang, and N. Zheng, "A limb-based graphical model for human pose estimation," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 48, no. 7, pp. 1080–1092, 2018.

[16] Y. Wang, "Design and development of sports video analysis system based on sports training," *Automation Technology and Application*, vol. 8, pp. 148–151, 2019.

[17] P. Shapshak, "Artificial intelligence and brain," *Bioinformation*, vol. 14, no. 1, pp. 38–41, 2018.

[18] P. Chemouil, P. Hui, W. Kellerer et al., "Special issue on artificial intelligence and machine learning for networking and communications," *IEEE Journal on Selected Areas in Communications*, vol. 37, no. 6, pp. 1185–1191, 2019.

[19] H. Feng and W. Dingjun, "Research on the solution of intelligent work order analysis system based on text similarity," *Electronic Technology and Software Engineering*, vol. 141, no. 19, pp. 222-223, 2018.

[20] L. Lu, "Frontier hotspot and evolution analysis of international sports artificial intelligence research based on Cite Space," *Journal of Hebei Institute of Physical Education*, vol. 2, pp. 37–44, 2020.

[21] Z. Bin, "Research on the design of sports training system under wireless sensor network," *Computer Knowledge and Technology*, vol. 14, no. 4, pp. 12-13, 2018, 21.

Hindawi
Wireless Communications and Mobile Computing
Volume 2020, Article ID 8814733, 12 pages
https://doi.org/10.1155/2020/8814733

Research Article

The Influence of Demographic Characteristics on Employee Promotion: Research Based on Data Mining and Game Theory

Chang Zhang,[1,2] Ting-jie Lv,[1] Chun-hui Yuan,[1] Yuan-yuan Ren,[2] and Shuo Wang[2]

[1]*School of Economics and Management, Beijing University of Posts and Telecommunications, Beijing 100876, China*
[2]*Shi Jiazhuang Posts and Telecommunications Technical College, Shijiazhuang 050021, China*

Correspondence should be addressed to Chun-hui Yuan; yuanchunhui@139.com

Received 17 June 2020; Revised 28 July 2020; Accepted 24 August 2020; Published 24 September 2020

Academic Editor: Wei Wang

The motivation of the paper is to find out the influence of demographic characteristic factors on employees' career promotion in China's state-owned enterprises and study the relationship between promotion mechanism and enterprise benefits. More than 6,500 pieces of human resource data from China's state-owned enterprises were studied. Logistic regression is adopted to analyze the correlation between demographic characteristics and promotion. Meanwhile, different data mining methods are used to summarize the characteristics of promotion. On this basis, this study uses the principal-agent model of game theory to analyze the profits of employees and enterprises under different promotion mechanisms and demonstrates the conditions for enterprises to obtain the maximum benefit from the employees' promotion. The results provide a theoretical reference for the assessment indicator selection of enterprise promotion and help executives better understand the impact of promotion mechanism on enterprises and employees. The application of results can reduce the information asymmetry in promotion incentive, prevent the emergence of adverse selection, and achieve a win-win situation for enterprises and employees.

1. Introduction

1.1. Research Background. Promotion is an essential organizational process for both management and employees [1]. It mainly refers to the promotion of the staff post, namely, transferring from the old post to another post to take on more responsibilities. The promoted employees should bear greater risk and liability; accordingly, they will have greater power and enjoy better benefits. An excellent promotion system can optimize the distribution of human resources and bring benefits to the enterprise.

At present, some scholars [2] in Japan have carried out empirical research: in many large enterprises in Japan, employees prefer to get position promotion rather than a salary increase, that is, the promotion of position can make an employee get more incentive than a salary increase. Many large enterprises in China are similar to those in Japan. Particularly, some state-owned enterprises are owned by the state, and their executives usually have the dual identities of government officials and entrepreneurs ([3], even they are more like officials than professional managers [4]. At the same time, as the salaries of executives in state-owned enterprises are constantly transparent and standardized, employees in state-owned enterprises will pay more attention to the administrative rights and hidden benefits brought by senior positions, which will also make promotion a very important incentive for employees.

In fact, when considering the reasons for promotion, in addition to the dimension of output or performance, there are some employees' demographic characteristic factors that may affect the promotion. Generally speaking, demographic characteristics are not obtained through employees' hard work in an enterprise but exist in employees objectively. Demographic characteristic factors cannot represent employees' subjective ability improvement after they join the work and are very easy to identify, such as gender, age, length of service, and preemployment education. In many large enterprises, especially some of the large state-owned enterprises in China, the output or performance is not directly proportional to individual promotion; some demographic characteristic factors may play a role in promotion. When studying the promotion of officials in China (the

promotion of state-owned enterprises is similar to the officials), some scholars (Rongrong [5]) analyzed the promotion data of some officials in China from 1990 to 2013 and found that some demographic characteristic factors, such as education level, have a direct impact on promotion. In addition, some studies (Shisong [6]) have also shown that gender is highly correlated with promotion in large state-owned enterprises in China.

Whether demographic characteristic factors have an effect on promotion in large state-owned enterprises is a very important question. Because employees may feel that promotion through demographic characteristics is unfair [7], the fairness of promotion can lead to different incentives for employees. Therefore, in order to improve the objectivity of the research, the factors influencing the promotion can be extended; some demographic characteristic factors can be added to the promotion study as independent variables.

Besides the correlation research between demographic characteristics and promotion, how to use the basic human resource data to make verifiable generalization and prediction of the prospect and characteristics of employee promotion is a difficult problem that many enterprises and consulting institutions have paid extensive attention to. The data mining method [8] can be adopted to find potential knowledge and patterns from a large amount of data, which can enhance the informationization and intelligence of human resource management, so as to effectively solve such problems. Algorithms such as the decision tree, the random forest, and the artificial neural network can be used to discover knowledge and patterns, making intelligent predictions. At present, relevant scholars have used data mining algorithms to study the influence that the joint effect of different factors has on promotion. For example, the C4.5 decision tree algorithm in classification is employed to find the characteristics and rules of employee promotion and to predict the controllable accuracy of employee promotion (Hamidah [9]). In addition, few researches have explored the pattern and rule of employee promotion by using data mining methods, especially the research based on the demographic characteristic factors as input.

Furthermore, game theory also plays an important role in the study of employee promotion. The representative study researches the incentive and cooperation of officials in the promotion game with the promotion championship model (Li'an [10]). However, the promotion game model studied by many scholars at present is still mainly based on the linear relationship between the subjective variables of employees' effort or output and promotion [11], and the game analysis is carried out according to different situations [12]. These game studies rarely consider the role of other objective factors in promotion. In fact, different managers may adopt different assessment forms in the process of employee promotion. For example, the assessment method mainly relies on performance assessment and is supplemented by age, educational background, and other factors. However, such a detailed performance assessment will increase the economic cost of assessment. Otherwise, in order to reduce the economic cost of assessment, the employee promotion is directly based on objective factors such as educational background, age, and gender. Different promotion assessment methods will have different incentive effects on the employees, so that the promoted employee will create different values in the new position, which will affect the total earnings of the enterprise.

1.2. Research Content. Based on a large amount of human resource data of large Chinese state-owned enterprises, this study used the method of logistic regression and the random forest to analyze and find demographic characteristic factors that affect promotion and then summarized the promotion mode of enterprises. Then, based on the results of data analysis, this study establishes a promotion game model based on the information asymmetry principle between enterprise managers and employees, further analyzes the incentive effect of different promotion assessment mechanisms on promoters in different situations, and also elaborates the total return of enterprises, thus triggering thoughts on the promotion mechanism of enterprises and providing a theoretical basis for enterprises to build a promotion system with balanced resource allocation and the incentive function, so as to create a virtuous circle for the development of enterprises and employees. To be more clear, the study structure is shown in Figure 1.

1.3. The Main Points. In this study, data mining and game theory were used for empirical analysis, and the results showed the following:

(1) Apart from the influence on performance, the age, length of service, gender, and educational background of employees have a strong positive correlation with their promotion. That is to say, in Chinese state-owned enterprises, there are still promotion mechanisms that take the demographic characteristic factors of employees as the main factor

(2) Furthermore, different promotion mechanisms require different costs and have different incentive effects on employees, which will make promoters create different benefits in the new positions and ultimately affect the total benefits of the enterprise

(3) In order to maximize the total benefits of the enterprise, the total incomes of the promoted employee in the new positions should be in direct proportion to the difference of the possible benefits created by the promoted employee under different promotion mechanisms

2. Demographic Characteristic Factors Related to Promotion

2.1. Factor Selection of Demographic Characteristics. This study proposes the hypotheses for factors that may affect promotion.

Hypothesis 1. There is a strong correlation between demographic characteristic factors (such as age, length of service, gender, and educational background) and promotion.

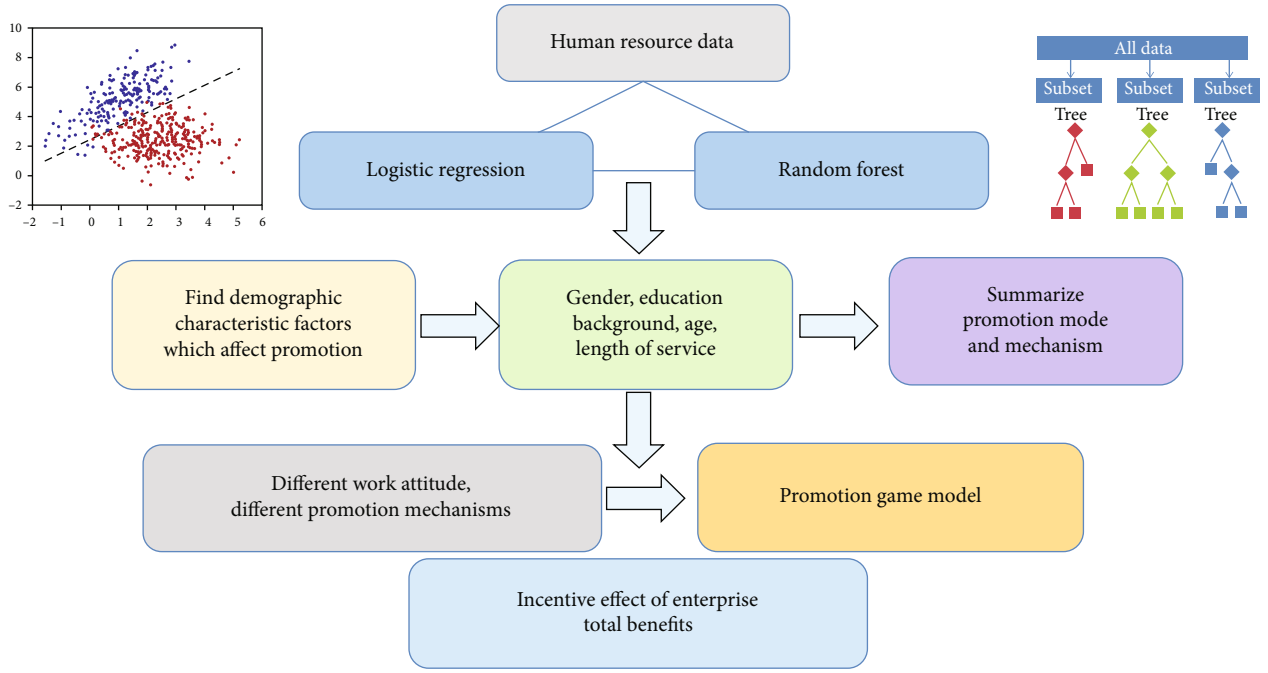


Figure 1: Study structure.

It is generally accepted that people with different levels of education have different abilities and qualities. People with higher educational backgrounds have stronger learning and working abilities and have higher probabilities of cultivating excellent talents. Therefore, many enterprises will pay less information screening cost if they choose among those with higher education levels instead of those with lower education levels. Recruitment is like this and so is promotion. Some scholars have confirmed that the education level sent a positive signal to enterprises, that is, individuals have higher cognitive abilities and stronger sense of responsibilities, which is conducive to the performance improvement of enterprises [13]. Some scholars adopted the education level as a formative indicator of human capital and confirmed that the education level had a positive impact on the welfare level [14]. At the same time, the differences in employees' academic qualifications, including those obtained in postwork continuing education, can also directly reflect employees' personal demands which are reflected in their demands for promotion. Highly educated employees have higher demands for promotion. At the same time, their strong abilities and low training cost have become consideration factors for the promotion of employees.

Age and length of service sometimes can be regarded as the measurement of work experience and social capital in enterprises. With the accumulation of employees' age and working age, they will accumulate more work experience and social capital in the enterprise, so that they are more capable of creating benefits for the enterprise and thus getting more promotion space.

Gender is regarded as a reference factor for promotion in many enterprises, and many studies show that although there is no gender difference in salary return for promotion in many jobs, the promotion probability of women is significantly lower than that of men [15]. According to the theory of human capital, due to the influence of fertility, family division of labor, and other reasons, women tend to invest more energy [16]. Since women do more housework, they passively reduce their efforts to work in the enterprise. In this case, their productivity is reduced, and lower performance output than men appears, thus reducing their chances of promotion. At the same time, studies have shown that one of the possible reasons for the low promotion rate of women is that women are significantly more likely to withdraw from the labor market due to family reasons than men. However, enterprises need specific human capital investment before the promotion of employees, which will reduce the rate of return on investment of women in the workplace and thus lower the value of training women [17].

2.2. Data Analysis. This study takes the human resource data of a large state-owned enterprise in China as an example for analysis. The data directly derived from the employee management system of an enterprise through project cooperation; the employee management system records the employees' information objectively and updates the data in real time, which excludes the interference of many subjective factors. We did data cleansing to the original data, including data integration and construction of useful dimensions, to make the whole data tidy and practical.

In all, more than 6,500 pieces of data were analyzed here. The main data dimensions adopted in this study include the age of employees, length of service in the

enterprise, gender, and education level based on the above analysis; among them, the education level is divided into the highest degree and the first degree. The highest degree is the final degree obtained by employees, which can reflect their knowledge reserve. The first degree generally refers to the degree obtained through the first curricula education accepted after the national college entrance examination. Many Chinese enterprises believe that the first degree can also identify the basic quality and learning ability of an employee. The target dimension is promotion; whether to be promoted is based on the promotion to the middle management level of the enterprise as the threshold, because when taking the middle management position in the enterprise, people will have a larger space to exercise management power independently and give play to personal ability.

In the data sample, the ratio between the nonpromoted and the promoted employees is 1 : 9, that is, there is a very common problem in the study of promotion data: the number of the nonpromoted is much higher than that of the promoted. If promotion is taken as the output, the serious imbalance of dimension value will affect the objectivity and accuracy of data mining results.

In this study, the k-means clustering method [18] is adopted to try to solve this problem. A large number of nonpromoted employees' data were analyzed by clustering that based on the partition method, the highest similarity within each cluster and the lowest similarity between clusters could be ensured. Thus, random selection in proportion in each cluster can effectively regulate the data of the nonpromoted employee and finally make it similar to the total data of the promoted employee.

For the convenience of follow-up research, some discrete dimensions' partitioning and assignment should be set to ensure the accuracy of data analysis, as shown in Table 1.

Then, the age of promoters was analyzed, as shown in Figure 2.

Figure 2 shows most promotions occur between the ages of 40 and 50, and 46-48 is the most likely age to be promoted. In Chinese state-owned enterprises, human capital is quite important, which will become higher and higher with the accumulation of time. A common phenomenon in state-owned enterprises is that midlevel leaders are mostly elder employees. Therefore, according to the principle that employees retire at 60, it is reasonable for employees to break through the threshold of promotion to leadership at around 46.

Then, the promoters' length of service was compared, as shown in Figure 3.

Figure 3 shows most promotions occur almost between the length of service of 20-28 years; like the age of promotion above, this phenomenon is still the cause of human capital accumulation. Then, the relationship between gender and promotion was analyzed, as shown in Figure 4.

From Figure 4, we can find male employees are promoted at a significantly higher rate than female employees. Then, the relationship between first degree and promotion was analyzed, as shown in Figure 5.

From Figure 5, we can find the proportion of employees with a higher first degree is significantly higher than that of employees with a lower first degree. Then, the relationship between highest degree and promotion was analyzed, as shown in Figure 6.

Table 1: Dimension value partitioning and assignment.

Dimension	Dimension value partitioning	Assignment
First degree	Below bachelor degree	0
	Bachelor degree or above	1
Highest degree	Below bachelor degree	0
	Bachelor degree	1
	Master degree or above	2
Gender	Female	0
	Male	1
Promotion or not	Nonpromoted	0
	Promoted	1

From Figure 6, we can find the higher the employee's highest degree, the more likely he or she is to be promoted.

Then, by comparing the relationship synthetically between different dimensions and the target dimension, Figures 7 and 8 can be obtained.

Multidimensional scatter plot (a) and multidimensional scatter plot (b) are, under the condition of first degree = 0 and first degree = 1, the relationship between age (x axis), length of service (y axis), highest degree (big circle = 2, middle circle = 1, and small circle = 0), gender (square = 1, circle = 0), and promotion (red = 1, blue = 0). It can be seen that with the increase of age and length of service, the number of promoted people is increasing. The higher the degree, the more people are promoted, and men are promoted far more than women. In addition, through the comparison of the two figures, it can be seen that the higher the first degree, the higher the possibility of promotion.

After that, in this study, there are only two dimension values in the target dimension, so we can use a logistic regression model to analyze the correlation between promotion and factors including age, length of service in the enterprise, gender, the first degree, and the highest degree. The method of logical regression is suitable for the classification of discrete values in the target dimension, and the calculated regression coefficients can be used for correlation analysis between different variables. The maximum likelihood method is a suitable method to estimate all coefficients in logistic regression. After calculation, the fitting information of logistic regression is shown in Table 2.

In the model fitting information table, the significance level value in the last column is less than 0.05, indicating that the model has statistical significance and has passed the test. In addition, its pseudo-r-square table is shown in Table 3.

In the pseudo-r-square table, the three pseudo-r-square values listed in turn are all low, with a maximum of 0.325, indicating that the model has a general explanatory degree for the variation of the original variable, and some information cannot be explained, so the fitting degree is not very outstanding.

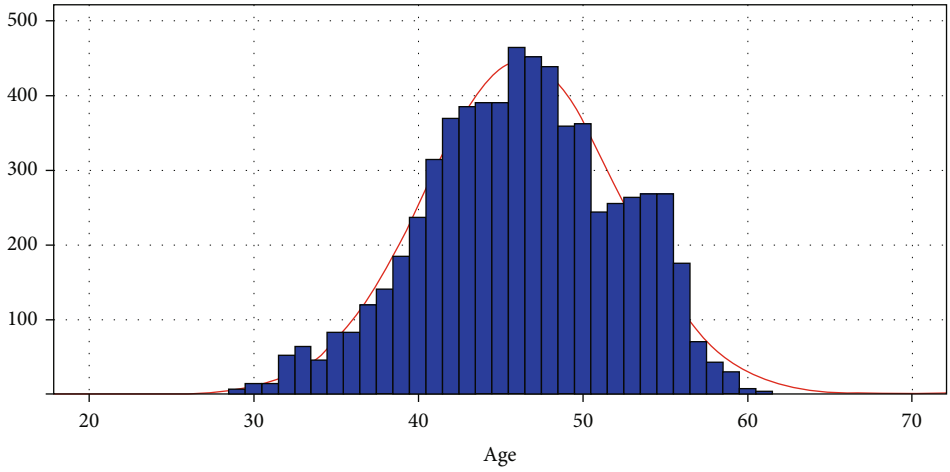


Figure 2: Age distribution of promoters.

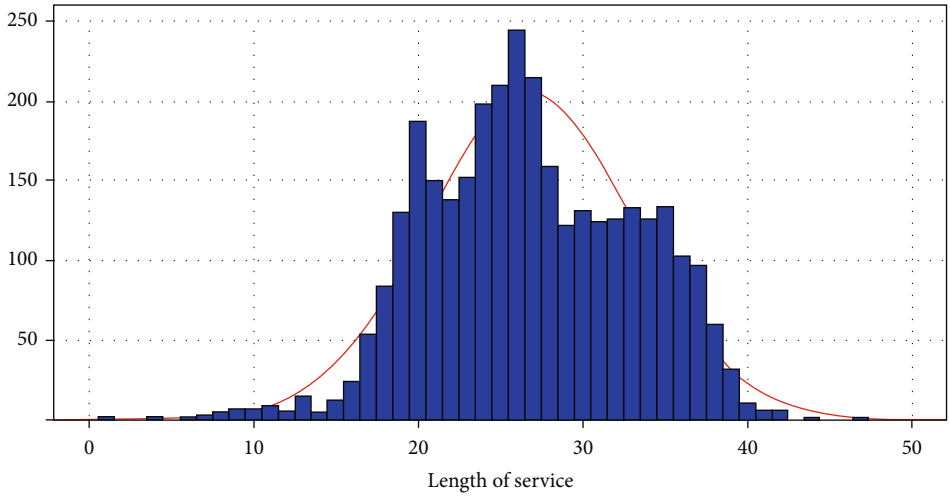


Figure 3: Length of service distribution of promoters.

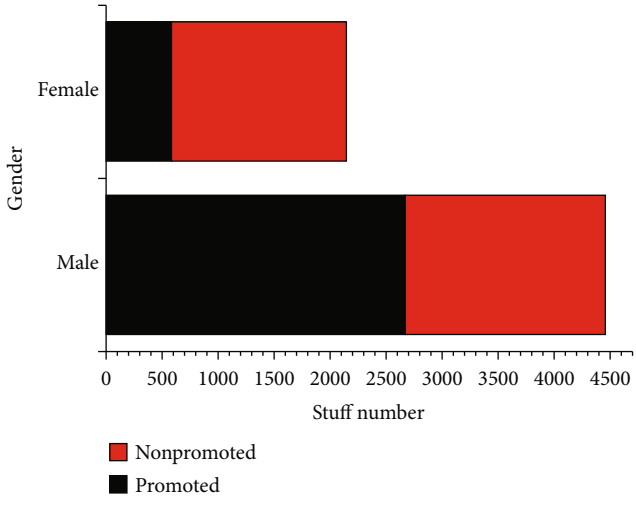


Figure 4: The relationship between gender and promotion.

The parameter estimation of the regression model is shown in Table 4.

According to the significance test in column 7 of the table, it can be seen that the significance levels are all less than 0.05, indicating that the corresponding independent variables have a significant impact on the dependent variables. As can be seen from the B value in column 2, the length of service and age are positively correlated with promotion. If "the first degree = below bachelor degree" is set as the reference group, "the first degree = bachelor degree or above" is positively correlated with promotion. In other words, people with higher first degrees are more likely to get promotion space, while those with lower first degrees are the opposite. If "gender = female" is set as the reference group, "gender = male" is positively correlated with promotion, that is, men are relatively easy to get promotion opportunities, while women are on the contrary. If "the highest degree = below bachelor degree" is set as the reference group, "the highest degree = bachelor degree" and "the highest degree = master degree or above" are both

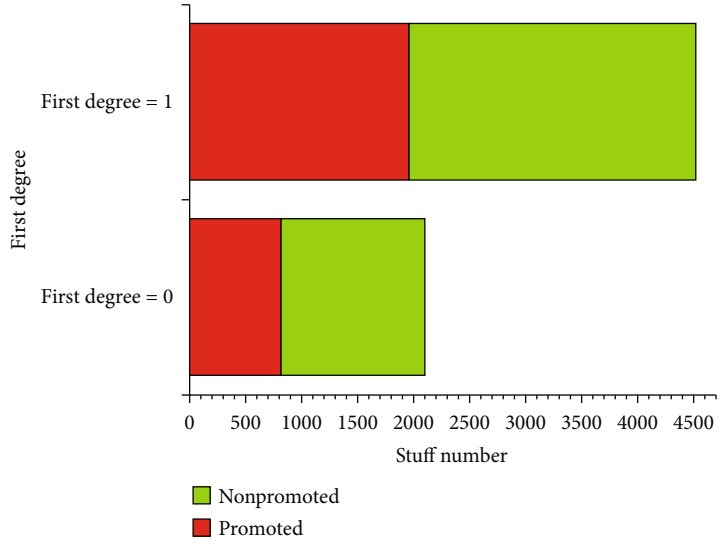


Figure 5: The relationship between first degree and promotion.

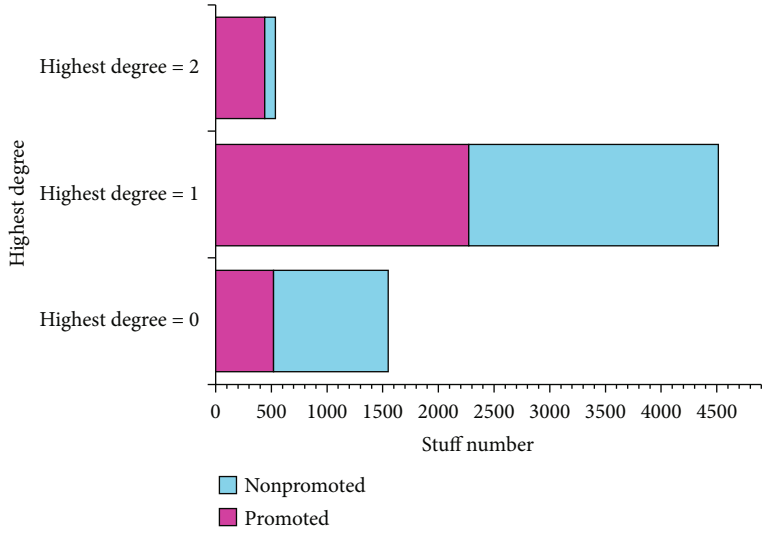


Figure 6: The relationship between highest degree and promotion.

positively correlated with promotion; as it is shown above, the higher the highest degree is, the more likely one is to be promoted. The promotion equation based on the logistic regression model is as follows:

$$\begin{aligned}(\text{Promotion} = 1) = {} & 0.03936 * \text{length of service} \\ & + 0.1024 * \text{age} + 0.4417 \\ & * (\text{first degree} = \text{bachelor degree or above}) \\ & + 1.252 * (\text{gender} = \text{male}) + 2.749 \\ & * (\text{the highest degree} = \text{master degree or above}) \\ & + 1.131 * (\text{the highest degree} = \text{bachelor degree}) \\ & +-7.751. \end{aligned} \tag{1}$$

In this study, the data set is divided into the training set and the test set, and the logistic regression model obtained from the training set above is brought into the test set for precision verification. The accuracy is 78.9% which is within a reasonable range. Therefore, Hypothesis 1 is verified.

Furthermore, in order to study the comprehensive effect of the above factors on promotion and improve the precision of data mining results, this study used the methods of C4.5 decision tree, C5.0 decision tree, the support vector machine (SVM), and random forest (RF) to analyze data. The random forest algorithm has good performance for solving classification problems, its training speed is faster, and the result is more accurate. Meanwhile, because the target dimension value is discrete, we can also use the method of decision tree for classification analysis, while C4.5 and C5.0 are typical decision tree algorithms. At the same time, under the existing amount of data, SVM can also be better applied to classification problems and have a strong learning ability.

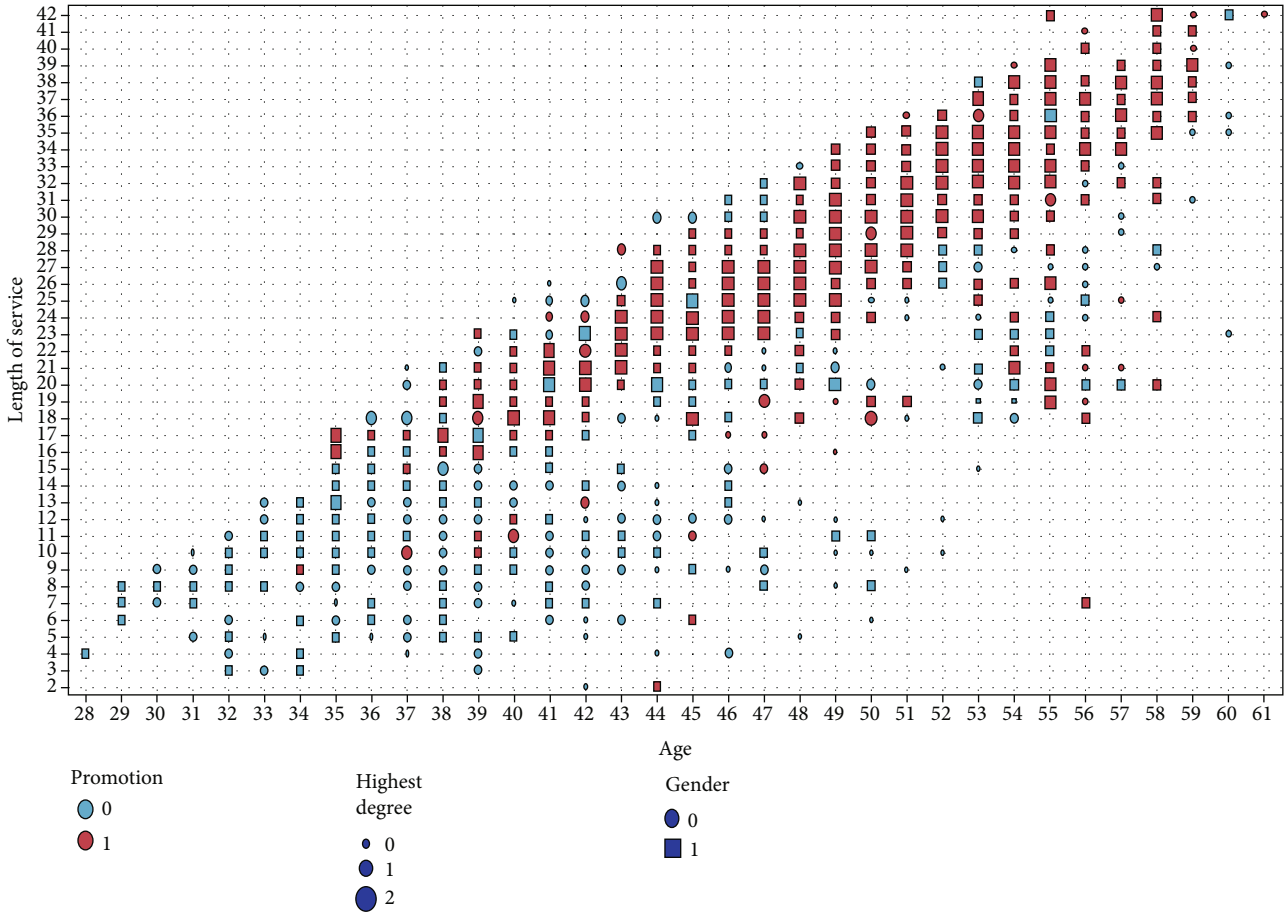


Figure 7: Multidimensional scatter plot (a).

The employee's age, length of service in the enterprise, gender, the first degree, and the highest degree are taken as input, and whether they are promoted is taken as output. In order to increase the objectivity of the conclusion, we selected different sample sizes for algorithm analysis. Through the training of the models, the results are obtained and the accuracy is checked in the test set, as shown in Table 5.

As can be seen from Table 5, the random forest algorithm is the most accurate in this study. Because random forest is a classifier that contains multiple decision trees, and its output category is determined by the majority of the output category of each tree [19]. Compared with the traditional decision tree algorithm, the analysis of the random forest algorithm is more accurate and can prevent overfitting. The algorithm works as follows.

Firstly, the bootstrap sampling technique is used to extract N training sets from the original samples. Secondly, the algorithm constructs the CART decision tree for each training set and generates a forest composed of N CART decision trees. Thirdly, the algorithm randomly selects m($m < M$) feature variables from all M feature variables. According to the principle of the minimum Gini coefficient, the characteristics of the training set are selected and internal node branches are carried out. Finally, the output of N decision trees is set for voting, and the class with the most votes is taken as the decision result of the random forest.

In the random forest algorithm, the Gini coefficient is used to measure the effect of feature segmentation in node segmentation. Suppose that there are K classes in sample D, then its Gini coefficient is

$$\text{Gini}(D) = 1 - \sum_{k=1}^{K} \left(\frac{C_k}{D}\right)^2. \tag{2}$$

In the above formula, C_k is a subset of the k class in D, and K is the number of classes. If after a partition, the sample set D is divided into m parts, $D_1, D_2 \cdots D_m$, the Gini coefficient of this partition is

$$\text{Gini}_{\text{split}}(D) = \frac{|D_1|}{|D|}\text{Gini}(D_1) + \frac{|D_2|}{|D|}\text{Gini}(D_2) \cdots + \frac{|D_m|}{|D|}\text{Gini}(D_m). \tag{3}$$

The Gini coefficient is inversely proportional to the segmentation effect. The value of the Gini coefficient can be calculated to rank the importance of all features. The larger the value is, the more important the feature will be.

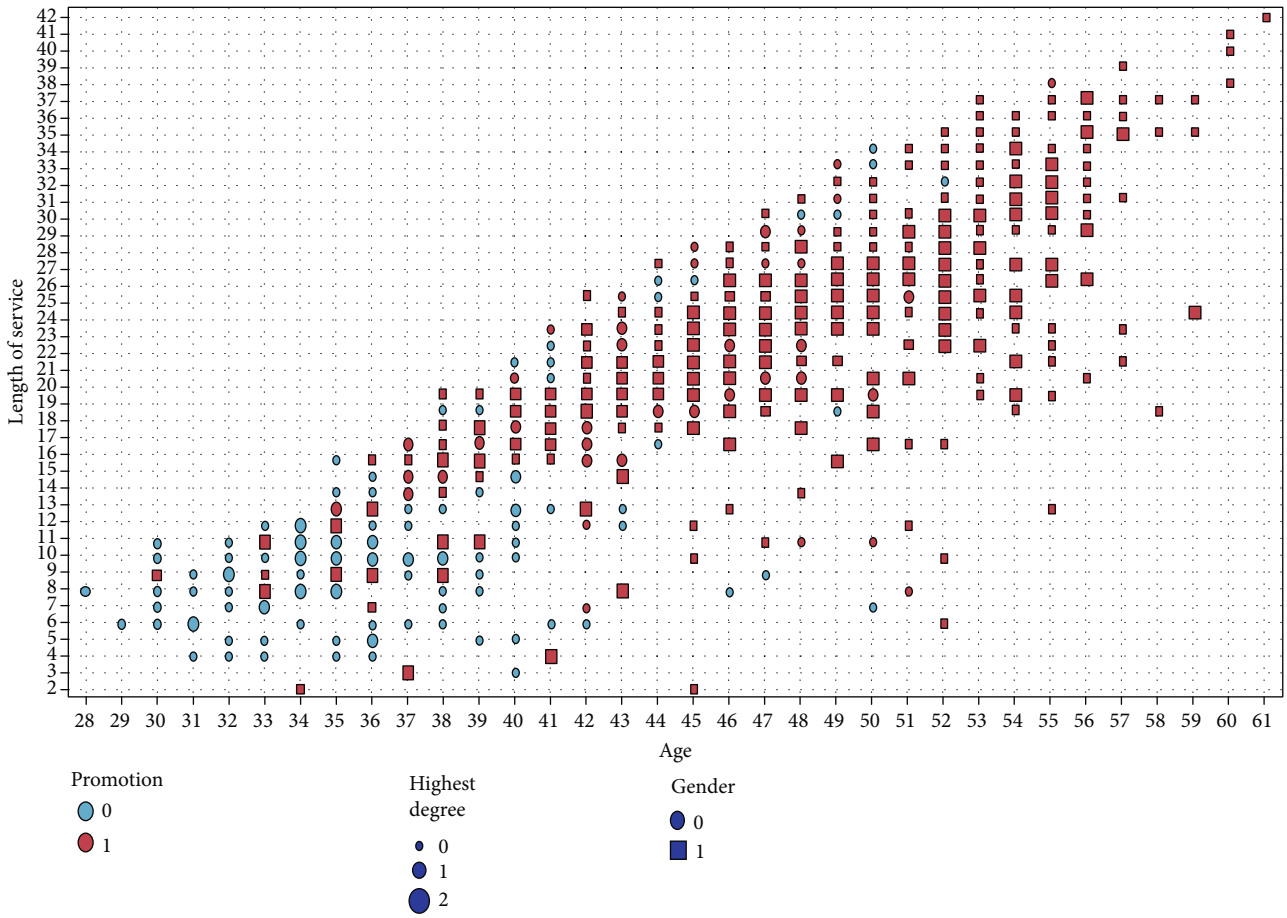


Figure 8: Multidimensional scatter plot (b).

Table 2: Fitting information.

Model	Model fitting criteria -2 log likelihood	Likelihood-ratio test Chi-square	df	Sig.
Intercept only	5019.676			
Final	3187.831	1831.845	6	0.000

Table 3: Pseudo-r-square.

Cox and Snell	0.244
Nagelkerke	0.325
McFadden	0.202

In this study, the promotion model of an employee in an enterprise can be summarized through the random forest model by the steps above.

The male employees with the first degree and the highest degree of bachelor degrees or above and at the age from 41 to 52 are most likely to be promoted to middle-level leadership in the enterprise.

The accuracy of this model is higher than the previous logistic regression model. The result also shows demographic characteristics are helpful for promotion.

3. Research on the Promotion Mechanism Based on Game Theory

3.1. The Influence of Different Promotion Mechanisms on Employees. Based on the above research, it can be seen that except the work performance or the level of effort, demographic characteristic factors may also become an important condition for enterprises to select talents for promotion. Based on the above data analysis and the field survey of enterprises, the study shows that when employees are promoted, large enterprises have different promotion mechanisms, which mainly include the following:

(1) Comprehensively and reasonably investigating employees' contribution, performance, effort, working status, and so on, making the best choice

(2) Promotion mainly based on demographic characteristic factors, such as the educational background, age, and gender

For the first kind, the enterprise may need a high screening cost for promotion, because the enterprise is essentially a teamwork production; the products are produced by team members' work together, but as a result of the existence of information asymmetry, it is difficult to depict each

Table 4: Parameter estimation of the regression model.

Job level		B	Std. error	Wald	df	Sig.	Exp(B)	95% confidence interval for Exp(B)	
								Lower bound	Upper bound
Promotion	Intercept	-7.751	0.353	482.015	1	0.000			
	Length of service	0.039	0.009	20.729	1	0.000	1.040	1.023	1.058
	Age	0.102	0.011	92.501	1	0.000	1.108	1.085	1.131
	First degree = below bachelor degree	0(b)	.	.	0	.	.	.	.
	First degree = bachelor degree or above	0.442	0.068	42.152	1	0.000	1.555	1.361	1.777
	Gender = female	0(b)	.	.	0	.	.	.	.
	Gender = male	1.252	0.064	381.902	1	0.000	3.498	3.085	3.967
	Highest degree = below bachelor degree	0(b)	.	.	0	.	.	.	.
	Highest degree = bachelor degree	1.131	0.078	211.468	1	0.000	3.097	2.660	3.607
	Highest degree = master degree or above	2.749	0.148	343.395	1	0.000	15.627	11.684	20.900

Table 5: Comparison of algorithm accuracy.

Algorithm	C4.5	C5.0	SVM	RF
Accuracy (100% of total sample size)	75.3%	78.2%	77.4%	83.4%
Accuracy (80% of total sample size)	74.4%	77.6%	75.5%	82.8%
Accuracy (60% of total sample size)	74%	75.8%	75.2%	81.5%
Accuracy (40% of total sample size)	72.2%	74.5%	73.7%	81%
Accuracy (20% of total sample size)	71%	73.6%	72.8%	80.3%

member's individual contribution accurately; if the enterprise does so, the cost of supervision and identification of marginal output will increase significantly. Therefore, it becomes very difficult or even impossible to determine promotion candidates based on marginal output and marginal contribution.

However, the first promotion mechanisms will make employees feel fair, which will stimulate the employees' enthusiasm and improve the income of the enterprise.

For the second kind, although the screening cost of the enterprise is low, employees will feel unfair and have negative emotions, which will reduce the work efficiency and the income of the enterprise. Some scholars [7] found that employees would compare their promotion approaches with others and judge the fairness of promotion. Most employees believe that factors such as work performance could serve as the important basis for promotion. Other scholars [20] found through research that the higher the fairness of promotion perceived by employees is, the harder they will work, thus bringing more profits to the enterprise.

Therefore, the selection of different employee promotion mechanisms will have an impact on the overall resource allocation and revenue of enterprises. If there are problems in the promotion mechanism, such as taking the demographic characteristics of employees as the main factor of promotion, employees will think that the enterprise lacks the fairness of promotion, which will have a negative impact on employees' work enthusiasm. In this way, employees who are promoted by demographic characteristics rather than performance will not be inclined to work hard in their new positions even if they are paid more, because they will think such demographic characteristic factors will help them continue to be promoted, and the higher salary is deserved after promotion. Therefore, this promotion mechanism does not generate positive incentives for employees and eventually bring losses to the enterprise. In some enterprises, managers do not have rigid and clear promotion indicators and systems but promote employees through decision-making meetings or subjective decisions of managers; that is to say, the two promotion mechanisms may exist at the same time, or promotion in the two mechanisms left and right sway, such enterprises are most prone to the above problem, resulting in the loss of corporate interests.

3.2. Promotion Game Strategy Based on the Maximization of Enterprise Benefit. In the employee promotion, there is a problem of information asymmetry that refers to the information that some participants have in one aspect but others lack or have no complete information (Weiying [21]). It is an important factor that affects the efficiency of transactions, especially for human capital transactions. Therefore, the game theory of information economics has been introduced into the promotion problem.

Based on the choice of different promotion mechanisms, there is an important game between managers and promoted employees. In large state-owned enterprises, many promotion decisions are based on managers' subjective decision research, so there may be included different promotion factors, such as the employee's performance and the

Table 6: Promotion game matrix.

	Y	N
T	$0, -h$	$w, v_2 - w$
S	$w - g, v_1 - w - h$	$w - g, v_1 - w$

Table 7: Nash equilibrium.

	Y	N
T	$0, -h$	$\underline{w}, \underline{v_2 - w}$
S	$\underline{w - g}, v_1 - w - h$	$w - g, \underline{v_1 - w}$

Table 8: No Nash equilibrium solution.

	Y	N
T	$0, \underline{-h}$	$\underline{w}, v_2 - w$
S	$\underline{w - g}, v_1 - w - h$	$w - g, \underline{v_1 - w}$

demographic characteristic factors; then, different types of incentive are produced in the promotion, so how to use the method of game theory to gain optimal benefits for enterprise is a problem worthy of study.

In this study, we focus on the enterprises where the two different promotion mechanisms exist. Based on the consideration of enterprises' overall benefits, we believe that the benefits provided by the enterprise for promoted employees are not the higher the better; high benefits do not necessarily produce high incentives, so the benefits should be related to those created by the employees in different working attitudes and thus put forward the following assumptions:

Hypothesis 2. Different promotion mechanisms will generate different incentives for employees. If the overall benefits of the enterprise are to be optimal, the benefits provided by the enterprise for promoted employees should be in direct proportion to the difference of the benefits brought to the enterprise by the promoted employees under different incentives.

In order to prove the rationality of this hypothesis, this study is based on the principal-agent model of traditional game research [22]. Here, the client is the enterprise manager and decision maker, and the agent is the employee waiting to be promoted. There are two strategic spaces for enterprise managers as follows:

First, enterprise managers should investigate the performance, effort, and working status of employees comprehensively and reasonably so as to screen the promotion of employees, set as Y.

Second, enterprise managers screen the promotion of employees mainly based on demographic characteristics, set as N.

The promoted employees are also featured with two different strategic spaces before promotion: hard work for performance or perfunctory work, when they are set as S and T, respectively. In addition, we suppose the following:

(1) Compared with strategy N, the cost of enterprise screening by strategy Y is h

(2) The overall welfare of the promoted employees after promotion is w, and $w > h$

(3) Compared with strategy T, all efforts put by employees into promotion using strategy S are set as g, and $w > g$

(4) The income created in the new position by employees who are positively motivated in promotion is v_1. If the enterprise can operate normally, the income is expressed as $v_1 > w$

(5) The income created in the new position by employees who are not motivated in promotion is v_2

It is assumed that the employees waiting for promotion have certain human capital, such as long working years, the gender of males, and advanced education, but the promotion strategy adopted by enterprise managers is not completely known by employees.

If the enterprise adopts strategy Y for screening, and the working state of employees is T, the employees will not be promoted, and their promotion utility is 0. In this case, as the enterprise has lost some screening costs, it gained $-h$.

If the enterprise adopts strategy Y for screening, and the working state of the employees is S, their promotion utility will be $w - g$. Besides, since strategy Y has positive incentives for the employee, the enterprise will gain $v_1 - w - h$ after the promotion.

If the enterprise adopts strategy N for screening, and the working state of the employee is T, the employee will also be promoted by demographic characteristic factors, since the employee fails to make corresponding work effort g, and the promotion utility of the employee is w. However, as employees do not get positive incentives and the company does not suffer from the loss of screening cost h, the enterprise gains $v_2 - w$ after the promotion.

If the enterprise adopts strategy N for screening, and the working state of the employee is S, the promotion utility of employees is $w - g$, since the employees really worked hard and got the promotion, so employees will mistakenly believe that they are promoted through hard work. Therefore, they will still create v_1 benefits in the new position, and furthermore, the income obtained by the enterprise is $v_1 - w$.

To sum up, the following game matrix is shown in Table 6.

When $v_2 - w \geqq -h$, that is, $v_2 \geqq w - h$, the problem can be expressed as the following matrix shown in Table 7, where there is a Nash equilibrium solution of pure strategy (w, $v_2 - w$).

However, there is a situation that can be commonly seen. In other words, v_2 without positive excitation will be extremely small, and then, $v_2 < w - h$ occurs, as shown in Table 8.

There is no Nash equilibrium solution of pure strategy, and therefore, we need to study mixed strategies of the Nash equilibrium. Suppose the probability of employees adopting strategy T is x, then the probability of adopting strategy S

and enterprise adopting strategy Y is $1-x$ and y accordingly. In this case, the probability of adopting strategy N is $1-y$.

Then, the expected profit of the enterprise is

$$\mu_1(\sigma_1,\sigma_2) = y(-x*h+(1-x)(v_1-w-h)) + (1-y)(x(v_2-w)+(1-x)(v_1-w)). \quad (4)$$

According to the definition of the Nash equilibrium, given the employees' mixed strategy $(x, 1-x)$, the value of y is sought to maximize $\mu_1(\sigma_1,\sigma_2)$. Therefore, in the equation mentioned above, take the partial derivative to y and set the equation as 0. Then, we can get $x = h/(w-v_2)$.

Similarly, employees' expected profits after promotion are

$$\mu_2(\sigma_1,\sigma_2) = x(0*y+w(1-y))+(1-x)((w-g)y + (w-g)(1-y)). \quad (5)$$

According to the definition of the Nash equilibrium, given the enterprise's mixed strategy $(y, 1-y)$, the value of x is sought to maximize $\mu_2(\sigma_1,\sigma_2)$. Therefore, in the above equation, take the partial derivative to x and set the equation as 0. Thus, we can get $y = g/w$.

Therefore, the Nash equilibrium solution of the mixed strategy for this promotion problem is

$$\left(\frac{h}{w-v_2}, 1-\frac{h}{w-v_2}\right), \left(\frac{g}{w}, 1-\frac{g}{w}\right). \quad (6)$$

When bringing it into the expected profit of the enterprise, there is

$$\mu_1(\sigma*1,\sigma*2) = h\frac{v_2-v_1}{w-v_2}+v_1-w. \quad (7)$$

Generally speaking, v_1, v_2, and h are relatively fixed values, while the overall welfare given by enterprises to promoted employees is a variable. Here, take the partial derivative to w and set it as $\mu_1(\sigma_1,\sigma_2)=0$. We can obtain the following:

$$(w-v_2)^2 = h(v_1-v_2). \quad (8)$$

That is,

$$w = \sqrt{h(v_1-v_2)}+v_2 \quad (9)$$

or

$$w = v_2-\sqrt{h(v_1-v_2)}. \quad (10)$$

From (10), there is $v_2 = w+\sqrt{h(v_1-v_2)} > w-h$, which contradicts the above condition $v_2 \leqq w-h$ of mixed strategy of the Nash equilibrium. Therefore, in order to achieve the maximization of enterprises' benefits, enterprises offer benefits to the promoted employees, which should refer to formula (9).

In other words, the total welfare provided by the enterprise for the promoted employees should be in direct proportion to the difference of benefits brought to the enterprise by the promoted employees under different incentives and also in direct proportion to the promotion screening cost of the enterprise. Therefore, Hypothesis 2 is verified.

4. Discussions and Conclusions

Overall, this study analyzed the demographic characteristic factors that influence the employee promotion in state-owned enterprises of China based on massive human resource data and conducted empirical research on the factors of gender, age, length of service, and the education level. The logistic regression method was adopted to verify the strong correlation between these factors and the promotion of an employee. At the same time, different classification methods of data mining are used to analyze and summarize the patterns and characteristics of employee promotion in enterprises, which can be used for intelligent and informationized enterprise human resource management.

On this basis, the principal-agent model of game theory is further employed to investigate the impact of different promotion mechanisms on enterprises. The performance-oriented and comprehensive employee promotion mechanism needs high talent screening cost, but it will generate positive incentives for employees, thus creating more benefits for the enterprise. The promotion mechanism based on demographic characteristics needs less screening cost, but it will generate negative incentives for employees, thus reducing the benefits for the enterprise. Therefore, in this study, the conditions for enterprises were analyzed to obtain the maximum benefits when the enterprise chooses different promotion mechanisms and employees choose different work attitudes.

In theory, this paper provides a theoretical basis for whether there are factors other than performance in employee promotion. That is, employees' demographic characteristics have a great influence on promotion in large state-owned enterprises in China, which can provide a theoretical reference for the assessment indicator selection of enterprise promotion. Secondly, this paper helps the executives of state-owned enterprises better understand the impact of employee promotion on enterprises and the significance of improving promotion mechanism.

In terms of managerial implications, the paper finds out the characteristics of employees promoted in state-owned enterprises in China through the method of data mining, which has a certain guiding role for state-owned enterprises to summarize the law of employee promotion, carrying out informationalized talent management and talent development. In addition, our conclusions can provide a practical basis for executives of state-owned enterprises to reduce their labor costs and to enhance the allocation of human resources. The executives of state-owned enterprises can take this study as a reference to provide welfare benefits for promoters, so as to ensure the maximization of profits of enterprises under the influence of different promotion mechanisms. The application of conclusion also can reduce the problem

of information asymmetry in promotion incentive, prevent the emergence of adverse selection, and achieve a win-win situation for enterprises and employees.

However, data mining methods of our study have some limitations. First, the dimension values of some dimensions are not very average, for example, the proportion of men and women is not quite equal in the gender dimension, which will have a certain impact on the analysis results. Second, since data of employee performance or output are difficult to measure and obtain, the disturbance 1of employee performance on promotion is not excluded when analyzing the influence of demographic characteristics on promotion. In other words, it is necessary to analyze the influence of demographic characteristics on promotion under the circumstance of the employees' same or similar performance level.

In the future research, we will further rationalize the data preprocessing and obtain the data resources that can represent employees' performance through enterprise research, analyze the demographic characteristics of employees at the same level of performance, and deeply study the relationship between demographic characteristic factors and promotion. In addition, we will continue to construct more interesting factors and analyze their correlation with promotion, such as the factors like employees' prepromotion positions and the relationship between promoted employees and senior executives.

Data Availability

The xls data used to support the findings of this study were supplied by the Talent Assessment Center of China Post under license and so cannot be made freely available. Requests for access to these data should be made to Hao Kou, 18810751@qq.com.

Conflicts of Interest

We declare that we have no conflicts of interest, financial or otherwise, in this work.

References

[1] J. T. Delaney and M. A. Huselid, "The impact of human resource management practices on perceptions of organizational performance," *Academy of Management Journal*, vol. 39, no. 4, pp. 949–969, 1996.

[2] K. Takahashi, "Effects of wage and promotion incentives on the motivation levels of Japanese employees," *Career Development International*, vol. 11, no. 3, pp. 193–203, 2006.

[3] D. Chen, J. B. Kim, O. Z. Li, and S. Liang, "China's closed pyramidal managerial Labor market and the stock price crash risk," *The Accounting Review*, vol. 93, no. 3, pp. 105–131, 2017.

[4] X. Lu, Y. He, and T. Wu, "Is SOE executives' political promotion incentive a long-term solution," *Economic Management Journal*, vol. 7, pp. 94–106, 2016.

[5] R. Lin, "How does human capital affect the promotion of officials – based on the promotion process of provincial leaders from 1990 to 2013," *Journal of Political Science*, vol. 1, pp. 91–105, 2019.

[6] S. Qing, "Gender discrimination in promotion," *Management World*, vol. 11, pp. 28–38, 2011.

[7] T. A. Beehr, V. N. Nair, D. M. Gudanowski, and M. Such, "Perceptions of reasons for promotion of self and others," *Human Relations*, vol. 57, no. 4, pp. 413–438, 2016.

[8] C. C. Wu, Y. L. Chen, Y. H. Liu, and X. Y. Yang, "Decision tree induction with a constrained number of leaf nodes," *Applied Intelligence*, vol. 45, no. 3, pp. 673–685, 2016.

[9] H. Jantan, "Human talent prediction in HRM using C4.5 classification algorithm," *International Journal on Computer Science and Engineering*, vol. 2, pp. 2526–2534, 2010.

[10] L.'a. Zhou, "The motivation and cooperation of government officials in the promotion game," *Economic Research Journal*, vol. 6, pp. 33–40, 2004.

[11] M. Chen and X. Chen, "Game analysis of traditional promotion system," *East China Economic Management*, vol. 9, pp. 94–97, 2005.

[12] Y. Lei and P. Yongjian, "Game analysis of human resource promotion model," *Journal of Business Research*, vol. 15, pp. 64-65, 2006.

[13] T. W. H. Ng and D. C. Feldman, "Human capital and objective indicators of career success: the mediating effects of cognitive ability and conscientiousness," *Journal of Occupational and Organizational Psychology*, vol. 83, no. 1, pp. 207–235, 2010.

[14] M. Hayek, C. H. Thomas, M. M. Novicevic, and D. Montalvo, "Contextualizing human capital theory in a non-Western setting: testing the pay-for-performance assumption," *Journal of Business Research*, vol. 69, no. 2, pp. 928–935, 2016.

[15] F. D. Blau and J. Devaro, "New evidence on gender differences in promotion rates: an empirical analysis of a sample of new hires," *Industrial Relations*, vol. 46, no. 3, pp. 511–550, 2007.

[16] G. S. Becker, "Human capital, effort, and the sexual division of labor," *Journal of Labor Economics*, vol. 3, 1, Part 2, pp. S33–S58, 1985.

[17] T. Pekkarinen and J. Vartiainen, "Gender differences in promotion on a job ladder: evidence from Finnish metalworkers," *Industrial and Labor Relations Review*, vol. 59, no. 2, pp. 285–301, 2016.

[18] J. Han and M. Kambe, *Data Mining: Concepts and Techniques*, Elsevier, Amsterdam, 2011.

[19] L. Breiman, "Random forest," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[20] S. Foley, D. L. Kidder, and G. N. Powell, "The perceived glass ceiling and justice perceptions: an investigation of Hispanic law associates," *Journal of Management*, vol. 28, no. 4, pp. 471–496, 2016.

[21] W. Zhang, *Game Theory and Information Economics*, Shanghai people's publishing house, Shanghai, 1996.

[22] J. J. Laffont, *The Theory of Incentives the Principal-Agent Model*, Princeton University Press, Princeton, 2002.

Hindawi
Wireless Communications and Mobile Computing
Volume 2023, Article ID 9893103, 1 page
https://doi.org/10.1155/2023/9893103

Retraction

Retracted: Analysis on the Construction of Personalized Physical Education Teaching System Based on a Cloud Computing Platform

Wireless Communications and Mobile Computing

Received 8 August 2023; Accepted 8 August 2023; Published 9 August 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] Z. Zhang and H. Min, "Analysis on the Construction of Personalized Physical Education Teaching System Based on a Cloud Computing Platform," *Wireless Communications and Mobile Computing*, vol. 2020, Article ID 8854811, 8 pages, 2020.

Hindawi
Wireless Communications and Mobile Computing
Volume 2020, Article ID 8854811, 8 pages
https://doi.org/10.1155/2020/8854811

Research Article

Analysis on the Construction of Personalized Physical Education Teaching System Based on a Cloud Computing Platform

Zhifei Zhang and Hyunjoo Min

College of Arts & Physical Education, Gangneung-Wonju National University, 7, Jukheon-gil, Gangneung-si, Gangwon-do, Republic of Korea

Correspondence should be addressed to Hyunjoo Min; hjmin@gwnu.ac.kr

Received 9 May 2020; Revised 10 August 2020; Accepted 27 August 2020; Published 19 September 2020

Academic Editor: Wei Wang

The emergence of cloud computing, the change of education methods, and the requirements of lifelong education make the traditional teaching platform face great challenges. With the rapid development of network technology and computer technology, the speed of updating knowledge is accelerating day by day, and the way of education is gradually changing. Facing the informationization of education, our physical education teaching methods and means are still stuck in the traditional words and deeds, which obviously cannot meet the needs of the development of physical education and health curriculum. In terms of the overall development of sports, school sports is the cornerstone of the country's sports development. This research is based on cloud computing technology, breaks the framework of the traditional sports model, establishes a personalized sports teaching system according to the basic theory of physical education, and designs and discusses the future college sports model. The construction and application of digital teaching resources of physical education courses in colleges and universities can help solve the problems such as shortage of teachers and contradiction between learning and training. The construction of digital teaching resources of physical education courses in colleges and universities based on cloud computing can save costs and improve resource utilization efficiency.

1. Introduction

The education in the new century is mainly to strongly advocate and promote quality education, and the goal of education has also turned to cultivating students' innovative spirit and innovative ability [1]. With the continuous development and construction of the cloud platform and cloud hosting, the current education and teaching reform has ushered in a golden development period. In the development process of teaching in the information age, the learning methods are constantly reformed, and the learning carriers and platforms are constantly changing [2]. In developing China, sports has become a national strategy and an important strategy to promote the realization of the Chinese dream and complete the great rejuvenation of the Chinese nation [3]. Promoting the development of sports for all is the result of the progress of China's times and the inevitable choice for the development of world history. In the face of education informatization, our physical education teaching methods and means still remain in the traditional words and deeds, which obviously cannot meet the needs of the development of physical education and health curriculum [4]. Under cloud hosting, the premise of teaching makes personal space construction continuously improved and also makes the function of cloud hosting more fully explored [5]. The teaching system is the key to cultivating students' talents and the main way to promote students to become pillars. In college teaching, the traditional teaching model has been strongly impacted by informatization, and the reform of the teaching model is imperative. At the same time, the organic combination of computer networks and college teaching is gradually becoming a new development trend.

The cloud space of physical education teaching in the new information age effectively makes up for the lack of teaching in our country and makes physical education keep pace with the development of the times, full of vitality and

vitality [6]. The rapid development of network technology and computer technology has accelerated the update of knowledge. In order to obtain more and more timely knowledge, the way of education has gradually changed [7]. Online education must be a kind of personalized teaching that meets the needs of individualized learning. Online courses must dynamically adjust the page structure according to the learner's interests, frequency of visits, and time to better meet the needs of learners [8]. With the development of information technology and the widespread application of cloud computing technology in the field of education, the construction of cloud computing has become a hot spot in the work of education informationization [8]. In the past, under the influence of traditional concepts in China's college physical education, the emphasis was placed on the standardization of movements as well as training intensity, making physical education in colleges and universities with certain mandatory specifics [9]. In this way, students do not like and pay attention to it, and it also brings difficulties to PE teachers in teaching courses. The traditional learning platform is a website based on a single hardware facility, which provides learning content through streaming media services and other technical means. The main disadvantages of this method are large investment, long cycle, high maintenance cost, poor scalability, and so on. In order to successfully construct a teaching mode that respects the individual differences of students, combines the individual characteristics of students, and earnestly implements the principle of teaching students according to their aptitude, college physical education teachers must completely change the old traditional teaching mode [10].

With the continuous development of information technology and computer networks, information in the network era has begun to grow by leaps and bounds [11]. Haipeng and others put forward that education cloud technology is to deploy innovative educational applications on the cloud platform through the integration of relevant educational information resources, so as to achieve seamless communication, interaction, resource sharing, and processing among education authorities, universities and educational institutions, and teachers and students [12]. Chen believes that cloud computing adopts service-oriented architecture, which can better support the personalized service of basic education resources. In the actual physical education classroom teaching, many physical education teachers cannot accurately grasp the profound connotation of individualized physical education in colleges and universities [13]. As far as the overall development of sports is concerned, school sports is the cornerstone of the national sports development. Jingrui proposed that the combination of cloud computing and curriculum, and the scientific management of it, can effectively combine various informal learning means and finally promote the development of personal lifelong learning in an all-round way [14]. Haoyue and Xiuqi mentioned in their research that with the help of cloud computing technology, different regions can share high-quality teaching resources and allocate educational projects according to regional characteristics, which can not only greatly promote the reform of the teaching mode but also realize the regional equalization of educational resources as soon as possible [15]. This research is based on cloud computing technology, breaking the framework of the traditional sports model and, according to the basic theory of physical education, establishing a personalized physical education teaching system to design and discuss the future college sports model.

2. Feasibility of Cloud Computing-Assisted Physical Education

2.1. Changes Brought about by Cloud Computing-Assisted Physical Education. With the development of information technology and the wide application of cloud computing technology in the field of education, the construction of a sports teaching service platform focusing on cloud computing is a new hot spot in the in-depth development of school sports information. Cloud computing is an Internet-based, public participation supercomputing model [16]. That is, a large amount of information and processor resources stored on personal computers, mobile phones, and other devices are centralized and work together. The existing network teaching platform has insufficient storage compatibility and cannot implement teaching well across platforms. Faced with numerous digital devices and their complicated operating systems, most physical education teachers are afraid. The cloud platform can provide users with corresponding cloud services. The platform will put the written programs into the cloud and then feed back these information and resources when customers use cloud platform services. The storage performance of cloud hosting is very high, and the personal space behind it is a massive information repository based on cloud computing technology. The new personalized comprehensive teaching platform integrates the characteristics of virtualization, cluster teaching, and experiment; provides a unified solution; and has the characteristics of integration, high density, and multiplatform. Cloud computing, with its superscale, high scalability, high reliability, virtualization, on-demand distribution, cheapness, and versatility, enables ordinary users to enjoy the storage and computing capabilities of high-performance computers on computers with simple configuration, bringing great convenience to human life and work [17]. The "school sports cloud computing" model, which emerged from the application of cloud computing technology in sports teaching, will bring many new changes and development opportunities to sports teaching. At the university stage, students' independence and autonomy are more demanding, but students at this stage are often affected by less social experience. Their independence and autonomy often manifest themselves as arbitrary, sometimes contradictory, and often lead to a loss of self. It can be seen that students at this stage need teachers' active and correct guidance.

The biggest characteristic of cloud computing method should be embodied in personalization. Using the information obtained from cloud computing, materials to be learned in teaching resources should be selected and organized in a targeted way dynamically. Two hundred track and field athletes were selected as test samples for this experiment. The basic information of athletes is shown in Table 1.

Table 1: Basic information of athletes.

	Number of cases	Average height (cm)	Average weight (kg)	Average age
Male	100	180 ± 3.25	75.6 ± 4.52	22 ± 3.28
Female	100	165 ± 2.78	53 ± 3.64	22 ± 2.34

If physical education teachers in colleges and universities can personalize their physical education classes based on their own specialty and students' reality and rely on their own practical teaching experience, they will certainly form their own personalized classroom teaching style that students like and have their own characteristics [18]. Cloud computing-assisted physical education is to give full play to the advantages of computers and the Internet and effectively integrate excellent physical education teaching resources from various places and schools into the physical education teaching process to create an information-based teaching environment. The primary factor to promote the development of college students' sports personality lies first in the training objectives set by colleges and universities. The training objectives of personalized education can be divided into two points. First, students should meet the basic requirements of the objectives of college physical education. Second, according to the characteristics of different students' sports literacy, the designated training plan for teaching students is in accordance with their aptitude [19]. Cloud computing can realize the informatization without its infrastructure. Each unit does not need to build any facilities, as long as the network purchases the required sports teaching resources. The personalized online teaching assistant system based on cloud computing is the assistance and extension of classroom teaching, and it is a tool to help students consolidate and reduce the workload of teachers after class review. At present, the education models commonly used in higher education are unified and standardized. Students expect to meet the needs of sports skills and lifelong physical fitness in physical education.

2.2. Positioning, Content, and Technical Realization of Cloud Computing-Assisted Physical Education. The goal of a personalized learning system is to recommend the knowledge and information that students are interested in and want to learn according to their personal characteristics, their learning interests, and learning needs, so as to help learners to swim out of the vast ocean of information. With the continuous development of teaching cloud hosting, teaching curriculum resources have been better decomposed. Through the construction of teaching cloud hosting, the updating and arrangement of teaching resources have made great strides towards dynamic. Especially in the update and utilization of the Internet and multimedia teaching resources, the media resource base of physical education courses has been enriched to a great extent. The construction of cloud computing-assisted physical education and the application of resources will change the mode and method of traditional physical education teaching and realize the deep integration of information and physical education [20], so as to achieve

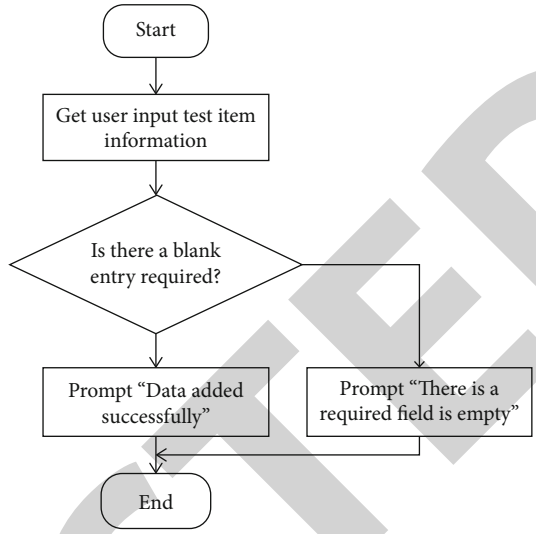


Figure 1: Program flow of new test items.

the purpose of improving teaching quality and efficiency and realize the "revolutionary influence" of informatization on school physical education. In the construction of cloud space, physical education teachers all over the world can upload some excellent multimedia resources to the education cloud, and teachers everywhere can download these multimedia resources from the education cloud at will. The communication between teachers and students is an important way to ensure that teachers master students' learning situation in time. Through the construction of cloud space, the communication of teaching becomes closer, and at the same time, the communication between teachers and students breaks the boundary, so as to realize the synchronous development of the communication between teachers and students inside and outside the school.

Sports test project management refers to the management of information related to sports test projects, which includes the following functional modules: the addition of sports test projects, the modification and deletion of sports test projects, and weight setting. The database tables involved in the implementation process mainly include test project information tables. Only detailed implementation instructions will be provided here for the addition of test items. The implementation process is shown in Figure 1.

The basic objectives of the course are determined according to the basic requirements of most students and refer to the objectives determined and worked hard for some students with special skills. The development goal is to continue to develop on the basis of realizing the basic goal. Based on the results of the first stage of work, the number of students who passed or failed was counted, as shown in Table 2.

In different stages of the proposed different guiding ideologies of physical education, these guiding ideologies have promoted the university sports theory to play a significant change. Table 3 shows the survey and statistics of the degree of the realization of physical education objectives in physical education teaching in colleges and universities.

Table 2: Statistics.

	Long-distance race	Dash	Long jump	High jump
Passing number	128	141	118	136
Number of failures	72	59	82	64

Table 3: Survey and statistics on the achievement of educational goals of physical education in ordinary colleges and universities.

Degree of realization	Fully realized	Partial realization	Not implemented
Number	72	89	39
Proportion (%)	36	44.5	19.5

There is a multi-index evaluation system composed of n evaluated objects $u_1, u_2, \cdots, u_n$ and m indicators $x_1, x_2, \cdots, x_m$, and $x_{ij} = x_j(x_i)\ (i = 1, 2, \cdots, m)$ is the observed value of the evaluated object u_i on the index x_i. The evaluation data matrix can be expressed as

$$A = \left[x_{ij}\right]_{n\times m} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{bmatrix}. \quad (1)$$

Among them, $m, n \geq 3$, the data in A is the preprocessed standardized data, and the physical education evaluation process is described as a general transformation:

$$y_i = f(x_{i1}, x_{i2}, \cdots, x_{\text{in}}), \quad i \in N. \quad (2)$$

The script operation module analyzes and classifies the content of the event and determines what type of event it is. Modifying the code is inevitable, because we always have to correct errors, and requirements may change. A vertex without a predecessor vertex in the correlation graph of an interface component is called a starting vertex, and a vertex without a successor vertex in the correlation graph is called a final vertex. Figure 2 is an association diagram of interface components, where "$v1$" is the starting vertex and "$v8$" is the final vertex.

Automation users can create automation objects, access objects provided by the automation server, get or set object properties, or call object methods. The interaction between automation objects and automation users is shown in Figure 3.

Through the use of modern cloud computing technology, teaching management can be effectively improved, and the informatization development of teaching course management, education management, student management, and teacher management is realized, thus further improving the operational difficulty and convenience of teaching management. Before learning enterprise courses, the design concept

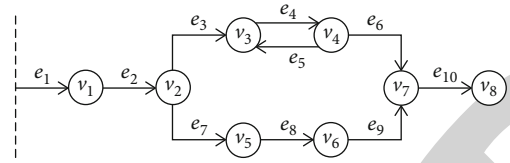


Figure 2: Interface component association.

of physical education courses can be first told to the learners, so that the students can effectively participate in their own course construction and have a very strong interest in learning physical education courses [21]. For example, Table 4 is a survey and statistics on the degree to which the physical education teaching in colleges and universities has achieved the educational objectives of physical education. Table 5 is a survey of students' satisfaction with school physical education.

Let a_{ij} and β_{ij} be the column dominance and row dominance, respectively, of the evaluated object $u_i\ (i \in N)$ on the index $x_j\ (j \in M)$ and satisfy

$$\begin{cases} a_{ij} = \dfrac{1}{n-1}\displaystyle\sum_{k\neq i}\left(x_{ij} - x_{kj}\right), & i \in N, j \in M, k \in N, \\ \beta_{ij} = \dfrac{1}{m-1}\displaystyle\sum_{p\neq i}\left(x_{ij} - x_{ip}\right), & i \in N, j \in M, p \in N. \end{cases} \quad (3)$$

If $\lambda_{ij} = \mu\alpha_{ij} + \eta\beta_{ij}\ (i \in N, j \in M)$, call λ_{ij} the autonomous superiority of the evaluated object $u_i\ (i \in N)$ with respect to the index $x_j\ (j \in M)$, where μ is the competitive target coefficient and η is the developmental target coefficient, $\mu, \eta \in [0, 1]$, $\mu + \eta = 1$.

The column strength reflects the difference in strength between the jth index of the evaluated object and the other $n-1$ evaluated objects as a whole. The row advantage quantity reflects the overall advantage difference between the jth index of the evaluated object and the other $m-1$ indexes.

The column advantage $\alpha_{ij}\ (i \in N, j \in M)$ reflects the strength difference between the jth index of the evaluated object u_i and the other $n-1$ evaluated objects as a whole. The line dominance reverse β_{ij} reflects the overall dominance difference between the jth index of the evaluated object u_i and the other $m-1$ indicators.

Traditional online courses have the disadvantages of higher development costs and being too static, and once the content of traditional online courses is developed, it is difficult to change them. Now that the speed of people's knowledge update and the pace of life are getting faster and faster, there is an urgent need to support learners' personalized sports learning in online courses. In specific physical education teaching practice, if physical education teachers want to implement personalized teaching strategies, they must completely change the traditional teaching mode in the past [22]. The cloud computing system interface module mainly presents relevant content pages to users for different operations of different users and uses natural language processing, semantic query, cloud computing technology, etc.

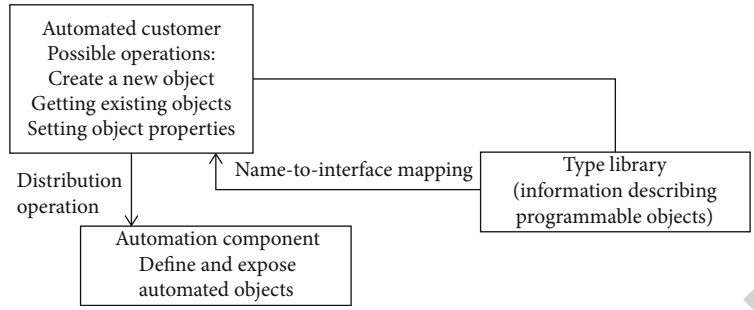


Figure 3: Interaction between automation users and automation components.

Table 4: Survey and statistics on the achievement of physical education goals in college physical education.

Degree of realization	Fully realized	Partial realization	Not implemented
Selected number	9	26	25
Proportion (%)	15	43.3	41.7

to provide an operation interface to realize human-computer interaction and realize the interaction between the user and the system.

3. Construction of the Personalized Physical Education Teaching System

Cloud computing provides a development environment as a service. Users can use middlemen's equipment to develop their own resources. Physical education teachers in colleges and universities can flexibly use the existing network platform for network teaching. Teaching students in accordance with their aptitude is an important part of modern educational philosophy. In China's traditional physical education, all students are educated by the same educational program, which will make the students' talents and potentials be ignored. University leaders and teachers must attach great importance to the current development of teaching resources and the construction of guarantee mechanisms and actively set up a special department for the construction of educational information cloud hosting to realize the good construction and development of educational resources in cloud hosting. Personalized physical education is based on the educational concept of teaching students in accordance with their aptitude. Different methods are selected for education and training of different students according to their characteristics. In doing so, it is more beneficial to the personal development of students [23]. The teacher platform mainly provides the management interface of the teaching resource database and teaching rule database, presenting the analysis information obtained by the personalized data analysis module. The student platform mainly presents teaching materials recommended by the information scheduling module for different students. The administrator platform mainly presents an interface for managing user information, user rights, and various resources.

Mastering sports skills is of course a necessary condition for students majoring in physical education, but under such conditions, only students majoring in physical education can be trained but not talents. This is one of the main problems in the training of physical education professionals in colleges and universities in China. Most PE teachers and students support the introduction of outward bound training in PE teaching, as shown in Table 6.

For any two evaluated objects $u_i', u_i''(i', i \in N,\ i' \neq i'')$, let $\omega_j^*(i', i'')$ be a random variable that obeys a certain distribution on the interval $[\min(\omega_{i'j}, \omega_{i''j}), \max(\omega_{i'j}, \omega_{i''j})]$ and call $s(u_i' > u_i'')$ as the superiority of u_i' to u_i'', as follows:

$$s\left(u_i' > u_i''\right) = p\left(f\left(u_i'\right) > f\left(u_i''\right)\right) + 0.5p\left(f\left(u_i'\right) = f\left(i_i''\right)\right). \quad (4)$$

In the formula, the aggregate function represents the event probability:

$$f\left(u_i'\right) = \sum_{j=1}^{m} \lambda_i' \omega_j^*\left(i', i''\right), \quad (5)$$

$$f\left(u_i''\right) = \sum_{j=1}^{m} \lambda_i'' \omega_j^*\left(i', i''\right). \quad (6)$$

The merit matrix S of several evaluated objects can be obtained, and $S_{ij} = s(u_i, u_j)$ is as follows:

$$S = \left[s_{ij}\right]_{n \times n} = \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1n} \\ s_{21} & s_{22} & \cdots & s_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ s_{n1} & s_{n2} & \cdots & s_{nn} \end{bmatrix}. \quad (7)$$

The backwardness of practical teaching content is undoubtedly the main reason for the decline of practical teaching quality. Even students majoring in physical education, as future physical education teachers, are also responsible for the healthy growth of future children. Personalized services can be divided into service organization personalization and service content personalization according to different levels. Personalization of service organization means that the platform has a set of services to realize various

Table 5: Survey of students' satisfaction with school physical education.

Degree of satisfaction	Very satisfied	Satisfied	Commonly	Not very satisfied	Dissatisfied
Number	15	52	216	117	50
Proportion (%)	3.3	11.6	48	26	11.1

Table 6: Survey statistics on whether it is necessary to carry out outreach training in physical education.

Choice item	Number of students	Proportion (%)	Number of teachers	Proportion (%)
Very necessary	22	8.8	11	22
Be necessary	147	58.8	30	60
Commonly	65	26	6	12
Unnecessary	16	6.4	3	6

functions. Service composition can be realized on demand according to different roles, activities, customized contents, and permissions of users. Basic courses are mainly divided into theory and practice. The teaching of basic courses should be innovated in traditional physical education. The user data collection module is the basis for realizing personalization of the whole online teaching assistance system and is mainly responsible for collecting relevant information and data of users. The personalized resource service module of the body is based on the labeled basic education resource database, processes the service requests of users, and uses the technologies of data mining and intelligent push to deeply understand user needs and actively discover potential needs so as to better provide personalized services [24]. The personalized cloud computing analysis module uses different cloud computing algorithms to mine and analyze the information of the students' original database, and the obtained results are stored in the teaching rule base after being normalized, and the personalized information is transmitted to the information scheduling module.

In the actual curriculum reform practice process, we must always take students' physical fitness and physical quality as the main line of development and actively carry out physical education courses suitable for the needs of society and students. After the intensive training period, blood ALP data before and after experimental recovery are shown in Figure 4.

In the process of human body modeling and simulation, these structural characteristics of the human body, such as freedom of movement, should be fully considered. For the range of motion, the simulation is designed by imitating the structure of the human body, and the simulation can have the same degree of freedom as the human body and complete the same motion as the human body. Based on the analysis of human body structure and actual motion, it is necessary to create constraint pairs between rigid bodies in the human skeleton model to ensure that rigid bodies with relative motion in the model can perform motion simulation according to the real motion mode of the human body. The tension-velocity relationship of the model is shown in Figure 5.

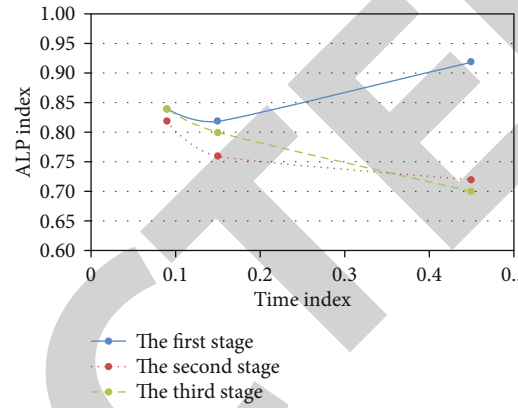


Figure 4: Blood ALP data.

The significance of developmental teaching evaluation lies in changing the old concepts, methods, and means of physical education teaching evaluation. The purpose of evaluation is to examine the students' situation comprehensively, stimulate their enthusiasm for learning, and promote their all-round development. It is also a powerful means for teachers to reflect on and improve their teaching. The students in the experimental group intervened from the beginning of training. Hemoglobin in the experimental group increased significantly. The maximum exercise capacity and anaerobic power of the experimental group increased significantly. During the intensive training period, the athletes in the experimental group are full of energy. The data index of anaerobic work before and after experimental recovery is shown in Figure 6.

From the perspective of competitive sports, the latest scientific and technological achievements will soon be transformed into concrete application on the sports field, thus bringing positive factors to create new sports results. From the point of view of school physical education, the curriculum, teaching objectives, teaching content system, teaching methods, assessment methods, and standards are also constantly influenced by contemporary pedagogy theories, showing a flexible development. Under the new situation of the PE teaching reform, the reform of PE teachers' teaching evaluation is urgent. Once there is no compulsory physical education, a considerable number of students will be isolated from physical education and their physical fitness will naturally decrease rapidly. The systematicness and necessity of teachers' selection of teaching contents within the prescribed class hours are inevitably guaranteed, which increases the randomness of teachers' teaching. In addition, schools at all levels seldom consider the main needs of students in the connection of physical education teaching.

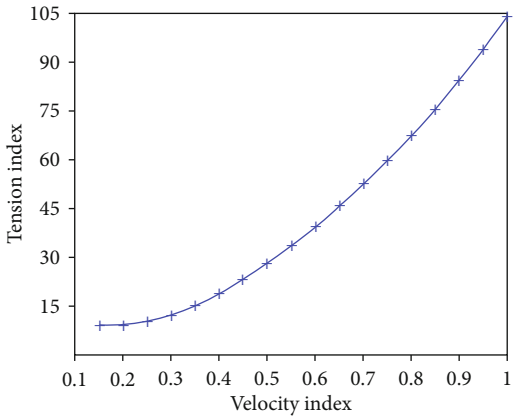


Figure 5: Tension and speed data.

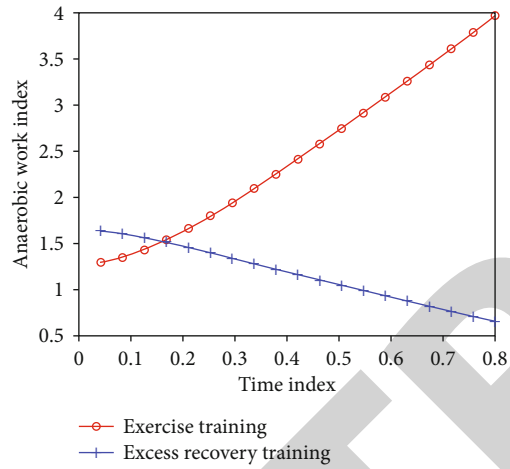


Figure 6: Anaerobic work data before and after experimental recovery.

The demand and problems of sports skills are the center, highlighting the cultivation of ability and quality. According to the imbalance of the overall level of students, different teaching plans of physical education courses are designed to maximize the teaching results. Physical education teachers' teaching viewpoints on students, curriculum, teaching materials, teaching methods, and other aspects will not only affect teachers' judgment and decisions in teaching activities but also affect teachers' teaching behaviors in teaching plans, teaching interactions, and teaching reflections [25]. The function of the information scheduling module is to transfer the corresponding learning contents from the knowledge base and the teaching resource base to the system interface module according to the rules obtained from the teaching rule base and the personalized data analysis module, so as to present personalized learning resources for different students and recommend corresponding exercises for students according to their weak links to strengthen and consolidate [26]. At present, there are quite a lot of learning resources in online physical education courses, including some very high-quality resources. However, these resources are effectively integrated in the learning process of physical education courses, thus serving the learning of this course well. In physical education classroom teaching, physical education teachers should respect and understand each student's needs and personality differences and encourage and praise students' unique opinions and different requirements.

4. Conclusions

Nowadays, with the popularization of information technology, in the daily teaching activities of colleges and universities, the traditional classroom teaching mode alone can no longer meet the requirements of both teachers' teaching and students' learning, especially there are great differences in the learning ability, learning foundation, and efforts of different students in reality. Cloud computing, as a large-scale resource integration and storage technology, provides convenience for the development and utilization of digital teaching resources. The traditional classroom teaching mode ignores the personalized features of the learning process, and the reform of the teaching mode is imperative. The development of cloud hosting has a very important impact on the development of physical education in our country and has made outstanding contributions to the construction of a benign interactive environment and the innovative development of resources. At present, the traditional class teaching system is still widely used in college physical education, but with the development of the times, this class teaching system cannot meet the internal needs of the development of education. In the process of cloud space development, teaching methods and contents are constantly changing, and the communication between teachers and students and between colleges and universities is becoming increasingly close. Physical education teachers in colleges and universities should actively explore the influence of cloud computing on network teaching methods, increase the construction of digital teaching resources of physical education courses, strive to improve the construction quality and use effect of digital teaching resources of physical education courses, and speed up the process of informatization of physical education in colleges and universities.

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 regarding the publication of this paper.

References

[1] W. Li, Z. Gao, Z. Yin, P. Xiang, B. Shen, and Q. Kong, "Impact of national physical activity and health guidelines and documents on research on teaching K-12 physical education in U.S.A," *Journal of Teaching in Physical Education*, vol. 35, no. 2, pp. 85–96, 2016.

Hindawi
Wireless Communications and Mobile Computing
Volume 2020, Article ID 8856831, 11 pages
https://doi.org/10.1155/2020/8856831

Research Article

The Construction of Builder Safety Supervision System Based on CPS

Wei Jin,[1] Yu Liu,[1] Yanfeng Jin,[1] Mengxiao Jia,[2] and Lifeng Xue[1]

[1]*Shijiazhuang Posts and Telecommunications Technical College, Shijiazhuang Hebei 050021, China*
[2]*The 54th Research Institute of China Electronics Technology Group Corporation, Shijiazhuang Hebei 050081, China*

Correspondence should be addressed to Yu Liu; liuyu@cptc.cn

Received 15 June 2020; Revised 7 July 2020; Accepted 25 August 2020; Published 18 September 2020

Academic Editor: Wei Wang

In recent years, with the development of the economy, the building industry has developed rapidly. However, due to the frequent construction safety accidents caused by natural factors and human factors, construction safety issues have also attracted the attention of the government and all walks of life. The widespread use of the Internet has made CPS an advanced technology that adapts to new social development methods and strengthens national construction. In this paper, in order to reduce the losses caused by construction accidents, firstly, the types of safety accidents are analyzed. Secondly, on the basis of CPS technology, a safety system for building workers is constructed, which is mainly composed of four parts, which are personnel management system, risk crosswarning system, monitoring, early warning system, and remote monitoring system, combined with related supporting systems. Finally, the fuzzy evaluation method is used to evaluate the safety system. It is concluded that in the process of building safety construction, prejob training should be conducted to help workers establish safety awareness, and workers' contact with not safety areas should be minimized to reduce accidents in advance; the safety supervision system constructed in this paper will help improve the management level of building safety management and reduce the incidence and losses of accidents. The evaluation results prove the superiority and effectiveness of the system.

1. Introduction

For a long time, the construction industry in the world belongs to the accident-prone industry, and the security situation is very serious. The frequent occurrence of construction safety accidents not only causes heavy casualties, construction period delay, and economic loss but also brings great pain to the casualties and families, causes harm to the whole society, threatens the development of the industry and the reputation of the government, and its direct and indirect social and economic losses are very huge. The problem of construction safety has attracted wide attention from all walks of life. How to improve the safety performance of construction industry is the common focus of academia and industry.

In order to reduce the losses caused by accidents in the construction industry, both experts in relevant research fields and managers in the construction front line attach great importance to the safety supervision of builders. At present, with the birth of AR/VR technology, the construction industry will also apply AR technology to prevent accidents. Using AR technology to simulate the virtual construction site, builders only need to wear VR glasses, and they can experience the construction site firsthand and feel the possible changes of body feeling and surrounding environment before possible accidents. For example, in practical work, similar "experience" can be used to predict the danger in advance. Compared with the traditional safety training for workers, this method reduces the occurrence of accidents to a certain extent. However, it is not scientific to predict the occurrence of accidents only based on the experience of workers.

With the introduction of CPS technology, it has been widely used in many fields. For example, government work, smart campus, transportation, smart home, and even our daily communications are the embodiment of the CPS application. This paper will use the CPS technology to achieve

safety supervision of builders, in order to reduce the occurrence of accidents and reduce the loss of personal and property caused by accidents.

CPS is a comprehensive, physical and controllable multidimensional intelligent system, which can realize real-time perception, dynamic control, and information service of complex system, which has the characteristics of safety, reliability, implementation, and efficiency [1–3]. Compared with the traditional Internet of Things system, the intelligent fire control system based on CPS can realize the intelligent processing of fire control system and meet the complex work requirements of fire monitoring and intelligent watch through the crossintegration of fire control business logic, big data, and artificial intelligence.

In the construction safety management system and major laws and regulations [4, 5], European countries and some developed countries started earlier, developed more balance, and paid more attention to the staff. To investigate the causes of construction site safety accidents, there are the following main problems: unreasonable layout of construction site, which makes different working areas have a cross section [6]; insufficient safety inspection, which makes the potential safety hazards in construction site not found and eliminated in time [7]; poor management of machinery and equipment, which cannot be repaired or replaced in time [8]; the aging of power lines has not been updated in time, which leads to the occurrence of workers' electric shock accidents [9], and the level and quality of safety protection in construction need to be improved. At the same time, due to the lack of safety awareness and self-protection awareness of builders [10], measures that cannot be strictly implemented as required lead to frequent accidents. In order to reduce the occurrence of workers' safety accidents, experts and scholars from various countries have proposed solutions, for example, the "truss" structure proposed by British scholar Mayton, which is a system that renders and infers building safety through the real-time streaming of the construction site environment. Literature [11] proposes a proximity sensing and alarming technique for heavy construction equipment operation. Liu Hongling and others focused on building safety detection and designed a large-scale building safety remote detection system based on the CPS technology. The networking of database transmission and its security information processing system are given. Considering building safety, Teizer and Cheng [12] and other people proposed an automatic approach hazard indicator to collect perceptual information about the intersection of workers and mobile devices, which is helpful to identify dangerous areas in the construction site. In urban underground business district, which is an important form of comprehensive utilization of urban underground space, Tan Zhanglu established a three-in-one safety management system of urban underground business district based on CPS to ensure its safety. In the literature, wireless sensor networks are deployed on the construction site to monitor potential hazards that workers may face.

The CPS has a wide range of applications and has applications in the field of construction worker safety supervision. With the continuous development of the economy, the development of the construction industry has been accelerated. As far as cities are concerned, the construction of residential buildings, large shopping malls, and entertainment and leisure places has brought convenience to people. However, compared with manufacturing industry, construction production has the characteristics of large amount of information, many uncertainties, and complexity. The CPS technology has a wide application prospect in the field of construction.

This paper is mainly divided into five parts. In the first part, in order to solve the loss caused by frequent accidents in the construction industry, the CPS technology is applied to the construction site, and some work done by relevant scholars and the basic concepts of the CPS are introduced. The second part introduces the technical structure of the CPS and applies it to the field monitoring management, safety management, early warning of dangerous areas, and building materials in the construction process. The third part, according to the characteristics of the construction industry, regards each construction site as a whole and constructs a construction site safety supervision system which is composed of personnel management system, risk cross-early warning system, panic and early warning system, and remote monitoring system based on the CPS technology. These four subsystems have a clear division of labor, constitute a whole, and are carried out on a construction site. In the fourth part, the fuzzy evaluation method is applied to evaluate the safety supervision system and verify the rationality of the system. The last part is a summary of this paper.

2. Application of the CPS Technology in Building Construction

For occupational safety and health costs, there are prevention costs and accident costs. The relationship between them is shown in Figure 1. Among them, the amount and difference of prevention cost and accident cost are affected by the change of safety level [13].

The three curves in Figure 1 show that T is the total cost of occupational safety and health, C is the accident cost, P is the prevention cost, and the three meet $T = C + P$. It can be seen from the figure that when the accident cost and prevention cost are the value at the M point, the total cost is the lowest and is the profit maximization point from the enterprise's point of view. On the left side of point M, the accident cost investment is relatively high and the prevention cost is relatively low. When the cost input is allocated according to it, many benefits can be brought; for example, it has a great impact on the stability of the human heart, and at the same time, it will bring corresponding benefits to the enterprise. On the right of point M, the increase in preventive cost input can improve the worker's work efficiency and speed up the completion of the work, so the benefit can offset the cost of the total input; that is, the increase in the total cost does not necessarily reduce the benefit. Therefore, there is a unified relationship rather than a confrontation between production efficiency and safety level [14]. Enterprises' investment in construction can bring certain economic benefits, and it is particularly important to reduce accidents.

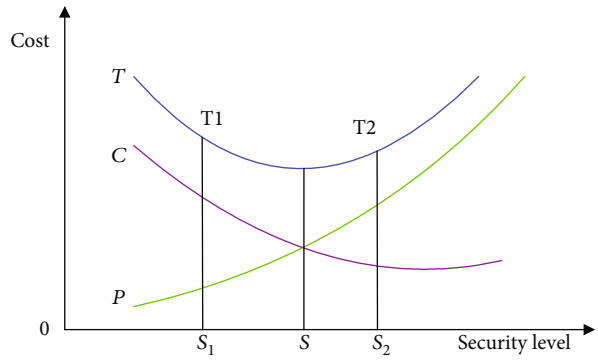


Figure 1: Occupational safety and health expense model.

The application of the CPS technology in a construction site provides better solutions to many problems. The CPS technology detects unsafe factors at the construction site through information sensing devices such as radio frequency identification (RFID), global positioning systems, gas sensors, and human behavior sensors [15]. When a dangerous situation is detected, it sends a signal to the field workers in time to let the staff make safety protection, so as to reduce the adverse effects of accidents on the construction site. The CPS technology is mainly divided into three layers [16]: the physical layer, the network layer, and the information layer, as shown in Figure 2.

The physical layer contains the hardware of CPS, such as sensors and actuators. The actuators directly or indirectly change the physical world, and the sensor samples sense the physical world.

The network layer is only responsible for the transmission of information and plays an important role in connecting the physical layer and the information layer, so it is regarded as an independent layer. The network layer includes various communication protocols and devices of the network, such as hubs, switches, and routers.

The information layer is responsible for the processing of information, including the calculation, identification, pattern matching and other processing of the information collected by the physical layer, the preservation of all kinds of information in the system, the management and configuration of the system, the making of decisions and orders based on perception, and the scheduling of tasks and resources.

2.1. On the Supervision and Management of Construction Site. In the safety management process of the construction site, the application of the CPS technology can not only effectively supervise and manage the things and construction operations on the construction site but also actively respond to and warn the unexpected safety accidents on the construction site. The CPS technology has the advantages of high sensitivity, small space occupied for laying, and resistance to electromagnetic interference. It is suitable for application in the field of construction safety monitoring. In addition, the CPS technology can also effectively monitor the operation of high-rise buildings, bridges, roads, tunnels, water conservancy, and dam projects and can make timely response when danger occurs. In order to detect and monitor the impact force, shape control, and vibration damping in the construction project, the perceptual devices belonging to the perceptual layer in the CPS technology are pasted in the wall of the monitored area and some movable devices and laid in the middle of the engineering structure. Through the analysis of the data returned by the sensors, the defects and deficiencies existing in the construction process are corrected in time.

2.2. Safety Management in the Construction Site. In the construction industry, safety accidents may occur at any time. Therefore, the safety hazards at the construction site have become the most important aspect of the daily monitoring and supervision work of the construction party and contractor. The application of the CPS technology can give early warning to the occurrence of safety accidents in the construction site. The technology of CPS is used to monitor the risk factors at the construction site, and the RFID marking is put on the protective shed near the entrance or exit or on the protective equipment at both ends of the elevator wellhead. In addition, corresponding commands such as tag number, hazard protection level, hazard parameter judgment threshold, and alarm information are input into the radio frequency identification chip, which should be consistent with the location and requirements of the building. Accordingly, the management of safety factors in the construction site is realized.

In the construction industry, it is difficult to monitor the safety of workers working at high altitude. At this time, the same RFID sign can be installed on the safety configuration of the safety helmet and seat belt of the high-altitude workers. Then, according to the building information model, the position of high-altitude workers can be accurately positioned. In the event of a dangerous accident, it is convenient for the operator to be found by the site safety management personnel in the shortest time. The management personnel take safety measures quickly to avoid the accident as far as possible.

2.3. Early Warning of Dangerous Areas. As an industry with high incidence of dangerous accidents, the environment of the construction site is usually relatively complex and scattered. In addition, most of the construction site workers are less educated. Workers have insufficient knowledge of the hazards and are prone to enter relatively dangerous work areas, posing a great threat to the safety of construction workers. When construction workers overlap with dangerous areas, it means that there are constructors entering dangerous areas. At this time, the CPS technology will immediately issue a danger warning to warm constructors to leave dangerous areas immediately.

2.4. Safety Management of Building Materials. High-quality building materials are an important guarantee for the quality of construction projects. Only when the building materials are guaranteed, the quality of building construction can tend to a good direction. Therefore, the safety management of building materials is also very important in the entire construction project. In terms of quality control of building materials, the CPS technology can identify building materials

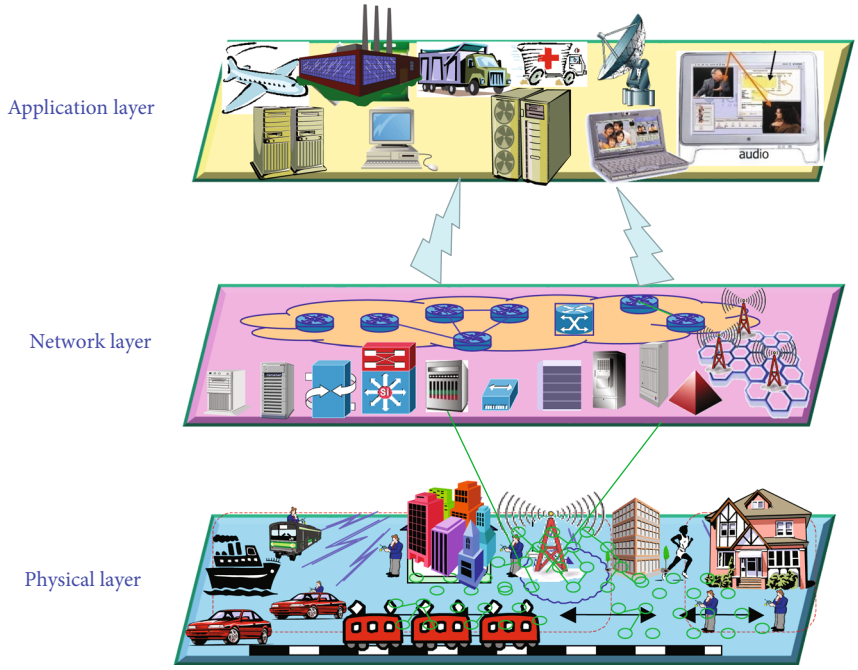


Figure 2: The structure of CPS.

by barcode sweeping, so that it is not necessary to open the outer packaging or protective layer of building materials. The CPS technology can easily identify the encoding of building material microelectronic chips by scanned code. Workers can use the CPS technology to control the quality of building materials. At the same time, Internet technology also helps to achieve transparent quality monitoring and management of building materials, which facilitates staff to grasp building materials and construction information at any time.

In short, as an emerging industry in the era, the CPS technology has been increasingly applied in the development of all walks of life. The CPS technology can play a great role in the construction safety management of the building industry. The application of the CPS technology in the construction site can help to realize the safety management of the construction site, protect the property and life safety of personnel on the construction site, and provide guarantee for the safe production of the construction industry.

3. Construction Safety System Based on CPS

This paper establishes a safety supervision system based on CPS, which is composed of personnel management system, risk crossearly warning system, alarm and early warning system, and remote monitoring system. Personnel management system is the foundation, including all the management objects. Risk crossprediction system is the center and the most important functional module. The alarm and early warning system is an information output module, which undertakes the function of information release. The remote monitoring system is the eye of the system, which pays attention to the changes of the system in real time. The four subsystems cooperate with each other and work together.

The construction safety system should follow certain principles. In the process of construction safety management, not a technology is suitable for a certain operation, but the integration of many technologies to realize the comprehensive management of people, objects, and computers in construction, so as to realize the purpose of early warning of danger and reduce accident risk. In the construction process, the selectivity faced by the staff should be minimized, that is, the uniqueness of the system implementation. The construction site is a dynamic process of continuous change, so the parameters of the equipment collected by sensors in the management system are constantly updated, which requires the continuous operation of the safety system. According to the principles of building construction safety system and the three characteristics of the CPS, perception of objects, information transmission, intelligent processing, and the CPS technology can be applied to it. The perceptual layer of the CPS perceives some parameters of construction facilities set in advance and transmits these information to the network layer. The network layer analyzes the returned information and makes corresponding judgments according to different dangerous situations by setting threshold. The application layer makes certain decisions based on the judgments made by the network layer, dangerous early warning, accident handling, and so on.

Safety management in the construction industry has the characteristics of high mobility of construction team, complex composition of staff, and staff intensive. The results of the survey of construction workers are shown in Figure 3, which are the age structure of construction workers, the professional level of construction workers, and their working years.

Through the analysis of the survey results, the construction workers have a weak understanding of building

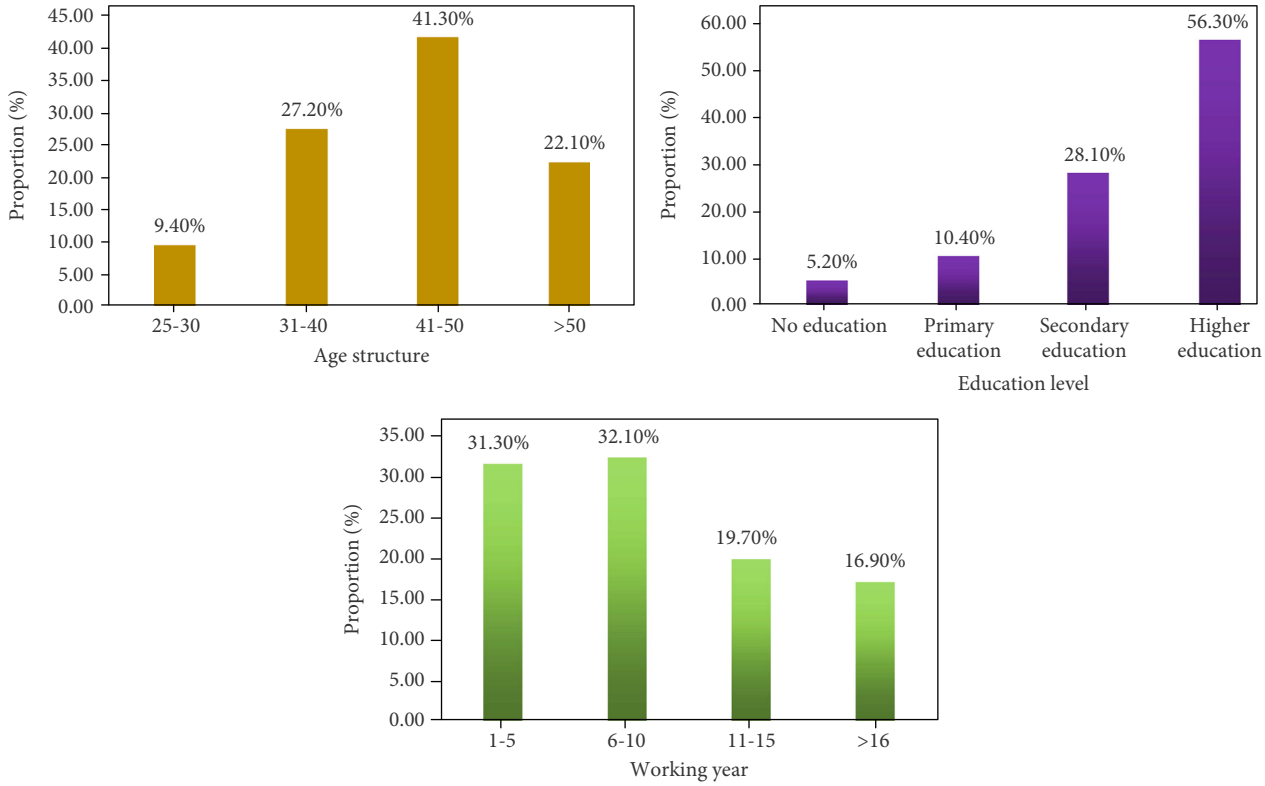


Figure 3: Survey results of construction workers.

expertise. Therefore, it is necessary to continuously train the construction workers' expertise to improve builders' awareness of dangers and to make timely and accurate respond when the danger occurs. However, the supervision of construction safety is not enough to improve the quality of workers. Therefore, it is necessary to build a safety supervision system on the basis of the CPS, consisting of personnel management system, risk crosswarning system, alarm and early warning system, and remote monitoring system.

3.1. Personal Management System. In the construction process, people are an important part, and the disorderly activities of workers have become one of the important factors inducing safety accidents. Workers have their own specific working areas, and the operation happens occasionally when someone enters a nonown area. Therefore, real-time supervision of workers is needed to avoid some accidents. Here, the personnel management system is set up to input the fingerprints of each staff member. Workers brush their fingerprints into the work and life fields. For those who do not belong to a certain range, because their fingerprints are not in the system of the work area, they cannot enter the work area. In this way, only the trained workers can enter the corresponding working area, which greatly reduces the accidents caused by the inexperience of the workers in a certain aspect. Based on this, a personal activity control system is constructed, and its structure is shown in Figure 4.

3.2. Risk Crosswarning System. The system uses the trajectory intersection theory, which considers that accidents are prone to occur when there is an overlap of time and space between an unsafe behavior of staff members and the unsafe state of a certain equipment in the construction site. For example, there will be large trucks loaded with raw materials in and out of the building construction environment. Because of the complex environment and dense personnel, there will be blind areas in the operation of trucks for drivers. At this time, if the staff, who are working in the blind area, do not pay attention to the approach of trucks, there is no time to avoid the accident.

Risk crosswarning system is used to make staff aware of the danger and take timely measures to avoid accidents. The system uses the function of information transmission and reading between the RFID reader and the electronic tag [17]. It attaches the electronic tag to the staff at the construction site and installs the dangerous reader in each device. The dangerous areas set by different devices are different. When a worker enters a dangerous area, the reader receives a signal from the electronic reactor worn by the worker and sends out a dangerous warning message to remind the worker to leave the area as soon as possible. This situation is reflected in the management center of the security department, showing the situation of crossrisk.

3.3. Monitoring and Early Warning System. The monitoring and early warning system mainly arranges wireless sensors for real-time monitoring of key positions in the construction site. The monitored data are transmitted to the integrated information processing center through the wireless sensor network to realize remote monitoring. Once the monitored

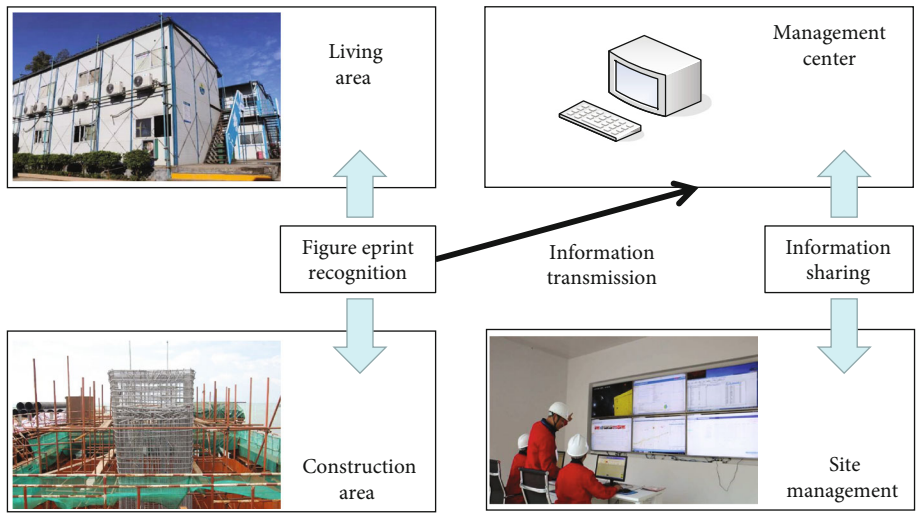


Figure 4: Personal activity control system.

data exceeds the threshold set by the system, the early warning system sends out dangerous warning signals to remind the construction site staff to take immediate measures and avoid accidents. Figure 5 is a schematic diagram of the monitoring and early warning system.

Every year, in the accidents caused by construction, the falling objects and collapses at high places account for a large proportion. Therefore, these two items become the focus of monitoring. The main monitoring objects are tower crane, scaffolding, formwork engineering, structure, earth and rock, collapse of piles, and so on. Each monitoring object has its corresponding monitoring content, such as the tower crane, which mainly observes the displacement of the tower seat, the degree of inclination, and the stress of the tower body. The data obtained from these monitoring objects are displayed in real time on the terminal in the form of text and graphics and are usually reflected to users in the form of curves, pies, histograms, and other intuitive forms. In this paper, the data of settlement observation of the main building during the construction of an office building are analyzed to determine whether the building is safe or not. According to the engineering practice, the datum points and settlement points are laid out for the office building. Figure 6 is the plan layout for setting settlement points. The black dots represent the location of the settlement induction device, with a total of 19 monitoring points.

Table 1 is the data attributes to be recorded at each monitoring point. This paper only shows the data records of some monitoring points on a certain day.

During the monitoring period, P5 monitoring point is selected and its broken line diagram deformation of deformation amount and the deformation after denoising treatment, as well as the residual diagram before and after processing, are displayed on the screen of the monitoring center, so that the staff of the monitoring center can more intuitively judge whether the building is dangerous enough, as shown in Figures 7 and 8.

Through the observation of Figures 7 and 8 , the staff can judge whether the settlement of the building will affect the safety of the building, and if it exceeds the safety limit, the staff can immediately start the danger warning.

3.4. Remote Monitoring System. Remote monitoring system is based on camera technology and image recognition technology, through the installation of camera equipment in the construction area to achieve the supervision of the operation process, construction area monitoring, security hazard detection, and other functions and to improve the efficiency of safety supervision. Among them, the supervision of the operation process mainly includes whether the workers wear safety protection devices according to the requirements, whether there is an operation that violates the construction process, and whether the worker is not concentrated in the operation process. The supervision of the hidden safety hazards removes the faults and hazards from the real-time inspection of the equipment. The safety status monitoring is used to check the situation in time when there is a danger in the construction site, and the accident scene can be controlled remotely to reduce the safety risk. The operation process of the system focuses on monitoring the work process and eliminating the hidden safety hazards in the construction site.

The four systems in the safety management system of the construction site based on the CPS technology are mutually reinforcing, and their functions are overlapping. The personnel management system guarantees the orderliness of the construction area and avoids the illegal operation of nonoperators. The risk crosswarning system plays the role of dangerous warning and avoids the injury accidents caused by unconsciousness. The monitoring and early warning system is mainly aimed at the subprojects which are more dangerous in the construction process, The monitoring and early warning of the whole time can help the management personnel to find hidden dangers of safety accidents in time and take measures in advance to avoid casualties. The remote monitoring system can not only change the current situation of safety supervision and inspection in the operation process but also help

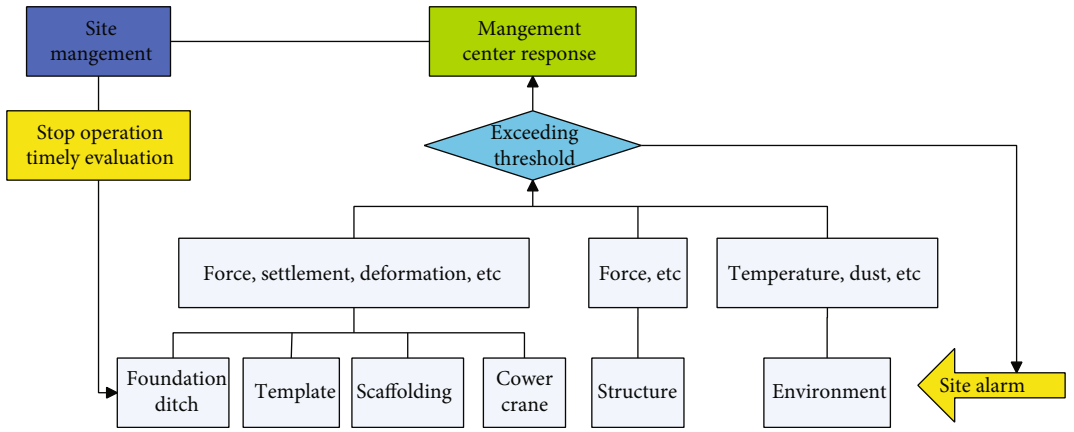


Figure 5: Schematic diagram of monitoring and early warning system.

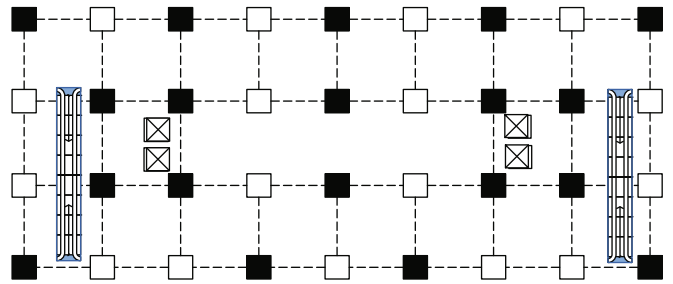


Figure 6: Plane layout of settlement observation points.

Table 1: Building settlement monitoring data.

An office building-surface settlement monitoring daily report					
Survey point number	Initial measurement (m)	Last measurement (m)	This measurement (m)	Deformation (mm)	Deformation rate (mm/d)
P1	67.14231	67.14286	67.14247	-0.39	-0.39
P2	67.13701	67.1377	67.1373	-0.43	-0.43
P3	67.04162	67.04234	67.04153	-0.81	-0.81
P4	67.00955	67.01019	67.00876	-1.43	-1.43
P5	67.00615	67.00964	67.00538	-1.56	-1.56
P6	67.05787	67.05851	67.05703	-1.48	-1.48

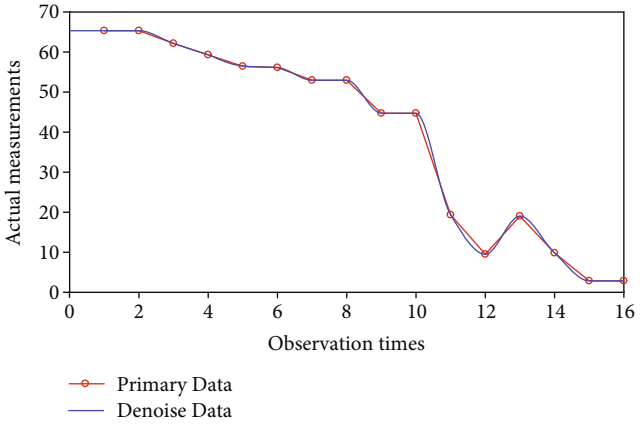


Figure 7: P5 monitoring raw data and denoising data.

managers to control the construction area as a whole. The four systems perform their respective functions and complement each other to jointly achieve the safety management objectives in the construction process.

In addition to technically avoiding accidents, workers are required to be trained before the job. The management personnel conduct regular inspections and real-time monitoring of the equipment in strict accordance with the requirements, conduct standardized management according to the system requirements, and strictly restrict unsafe behavior. Safety protection hardware on site is also one of the important factors to reduce accident injury. When the danger occurs, the emergency system will be launched to evacuate the staff on site in time and carry out emergency rescue quickly so as to reduce the damage caused by the accident.

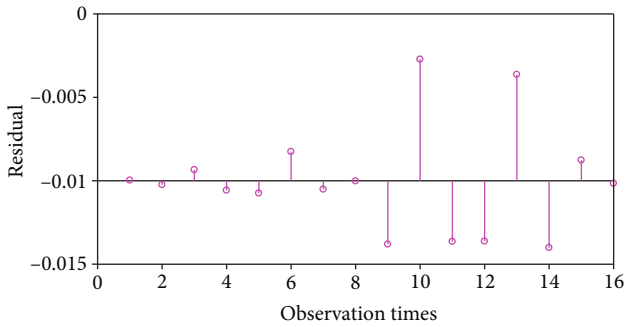


Figure 8: Residual between original data and denoised data of the P5 monitoring point.

4. System Evaluation

The building safety management system based on the CPS system is aimed at predicting engineering risks through effective management measures, preventing accidents, and reducing the loss of life and property caused by accidents. Therefore, the performance of the system is closely related to the loss caused by accidents on the construction site. Therefore, this paper examines the performance of the building safety supervision system based on the CPS.

In the process of construction, danger and safety are relative, and there is no clear boundary. Therefore, this paper uses the fuzzy evaluation method [18] to evaluate the system. The basic steps are as follows [19–24]:

Firstly, the evaluation grade of system safety is established, which is safety, safer, general, severe, and dangerous. The evaluation result is determined in a certain interval. The setting of the score is shown in Table 2.

For convenience of calculation, grade coefficients are set to integers.

Secondly, the weights are calculated. The comparison criteria are based on the Saaty scale method [25], which is shown in Table 3.

According to Table 3, the two pairs are compared between two layers, and the scale values are calculated, and the corresponding judgment matrix is calculated, as shown in

$$A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix}. \quad (1)$$

Calculate the weight according to formula (2):

$$AW = \lambda_{max} W \quad (2)$$

Among them λ_{max} is the maximum eigenvalue of matrix A, and W is the eigenvector corresponding to the maximum eigenvalue of matrix A. The weight of the index is checked by normalizing W. If the consistency is not satisfied, the original judgment matrix needs to be modified until the consistency is satisfied. Formula (3) is a consistency test formula [26].

$$CI = \frac{\lambda_{max} - n}{n - 1}. \quad (3)$$

Among them, CI is the consistency index, λ_{max} is the largest eigenvalue of matrix A, and n is the order of judgment matrix. The formula for calculating the average random consistency index CR is

$$CR = \frac{CI}{RI}. \quad (4)$$

When CR < 0.1, the judgment matrix satisfies consistency. RI is an average random consistency index, and its specific value varies with the order of the judgment matrix.

4.1. Accident Prevention Evaluation. According to the construction safety system based on the CPS technology [27], the corresponding indicators and evaluation system are established, as shown in Figure 9.

From Figure 9, it can be seen that the comprehensive evaluation index system consists of five parts, which are operating personal management system, monitoring early system, risk crossearly warning system, visual monitoring system, and relevant supporting system.

Table 4 is the parameter symbols set according to the main control content of each subsystem.

Taking reduction of the possibility of an accident as a standard, according to the judgment criteria of Saaty scale method, two or two indexes are compared to get their respective importance. Accordingly, the judgment matrix A of the personnel management system, the judgment matrix B of the monitoring and warning system, the judgment matrix C of the risk crossover and warning system, the judgment matrix D of the visual monitoring system, and the corresponding system judgment matrix E can be obtained [28], as follows:

$$A = \begin{bmatrix} 1 & 3 \\ 1/3 & 1 \end{bmatrix},$$

$$B = \begin{bmatrix} 1 & 1 & 3 & 2 & 7 & 8 \\ 1 & 1 & 4 & 5 & 3 & 6 \\ 1/3 & 1/4 & 1 & 1/2 & 1/3 & 3 \\ 1/2 & 1/5 & 2 & 1 & 3 & 5 \\ 1/7 & 1/3 & 3 & 1/3 & 1 & 6 \\ 1/8 & 1/6 & 1/3 & 1/5 & 1/6 & 1 \end{bmatrix},$$

$$C = \begin{bmatrix} 1 & 3 & 7 & 4 \\ 1/3 & 1 & 6 & 5 \\ 1/7 & 1/6 & 1 & 1/2 \\ 1/4 & 1/5 & 2 & 1 \end{bmatrix}, \quad (5)$$

$$D = \begin{matrix} 1 & 5 & 8 \\ 1/5 & 1 & 3, \\ 1/8 & 1/3 & 1 \end{matrix}$$

$$E = \begin{bmatrix} 1 & 4 & 9 \\ 1/4 & 1 & 5 \\ 1/9 & 1/5 & 1 \end{bmatrix}.$$

Table 2: Standard table of evaluation grade.

Level	1	2	3	4	5
Evaluation value	$X < 15$	$15 \leqq X < 40$	$40 \leqq X < 60$	$60 \leqq X < 80$	$80 \leqq X$
Evaluation language	Dangerous	Severe	General	Safer	Safety

Table 3: Saaty scale method.

Meaning	Scale value
X_m and X_n are equally important	1
X_m is slightly more important than X_n	3
X_m is obviously more important than X_n	5
X_m is more important than X_n	7
X_m is extremely important than X_n	9
2, 4, 6, and 8 are the two adjacent median values	

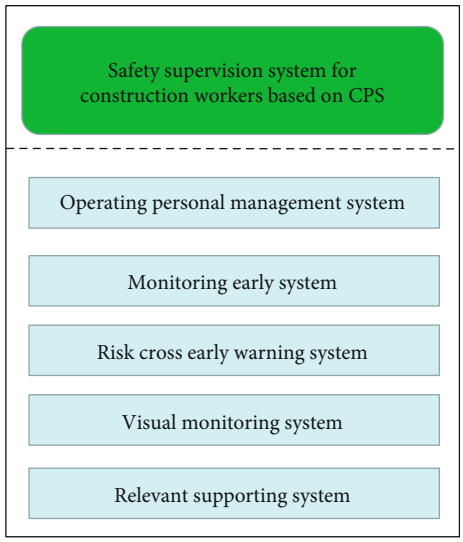


Figure 9: Comprehensive evaluation index system.

After calculation, the maximum eigenvalues of matrices A, B, C, D, and E are 2.0, 6.6, 4.2, 3.04, and 3.07, respectively, and the corresponding eigenvectors of the maximum eigenvalues are $a = (0.75,0.25)$, $b = (0.31,0.32,0.07,0.15,0.12,0.03)$, $c = (0.52,0.32,0.06,0.10)$, $d = (0.74,0.18,0.08)$, and $e = (0.70,0.24,0.06)$, respectively. The eigenvectors are obtained by normalization, so each element in the vector is the weight value of the corresponding factor. According to formula (4), the average randomness index CR of each judgment matrix is less than 0.1; that is, each judgment matrix satisfies the consistency.

The weights calculated for each index are sorted, and Table 5 is obtained.

The weights of each index are shown in a histogram. Figure 10 shows that the weights are divided into three intervals according to the size of the weights [29]. The weight in the third interval is more than 0.5, the weight in the second interval is between 0.25 and 0.5, and the weight in the first interval is between 0 and 0.25.

Looking at Figure 10, there are four indicators with weights greater than 0.5, namely, special equipment/regional access control (A_1), human and dangerous area (C_1), operation process supervision (D_1), and safety training and education (E_1). This shows that in the case of accident prevention, effective control of the four indicators in the third section is the key to prevent accidents.

Indicators with weight values in the second interval are tower crane (B_1), foundation pit engineering (B_2), and human and equipment (C_2), and the weight values of these three indicators are similar, and there are intersections among the indicators. During construction, tower crane accidents accounted for the majority of all kinds of accidents. As one of the important equipment in the construction site, safety management of tower crane has become an important measure to prevent accidents.

Most of the indicators are in the first interval, but there are still some indicators with larger weights, such as active area control (A_2), environment (B_4), template engineering (B_5), safety hidden danger monitoring (D_2), and safety protection (E_2). As a passive safety measure, safety protection can effectively protect workers in a potentially dangerous working environment and avoid accidents. The control of active area avoids construction workers entering unfamiliar environment and accidents caused by workers' wrong operation. Safety hidden danger monitoring and template engineering belong to "machine" and "environment" of human-machine-environment, respectively. As the main object of accident injury, the people's supervision and monitoring of "machine" and "environment" can avoid accidental energy release and cause human injury.

Other weight indicators are not very different, but they also play an important role in the prevention of safety accidents. In the process of safety supervision of construction site workers, any safety measure in the construction safety system based on the CPS technology is indispensable.

5. Conclusion

Firstly, in view of the losses caused by frequent accidents on the construction site, the characteristics of the construction personnel and the construction site are analyzed. This paper applies the CPS technology to construction operation and constructs a safety supervision system consisting of personnel management system, risk crosswarning system, monitoring and early warning system, and remote monitoring system. According to the characteristics of the CPS technology, the perception layer perceives the scene and transmits it to the network layer in the form of data. The network layer analyzes the data through a program set in advance and feeds back the analysis to the application layer. The application

Table 4: Indicators for each system.

System and symbolic representation		Indicators and symbolic representation
A	A_1	Special equipment/regional access control
	A_2	Active area control
B	B_1	Tower crane
	B_2	Foundation pit engineering
	B_3	Scaffolding
	B_4	Environment
	B_5	Template engineering
	B_6	Structure
C	C_1	Human and dangerous areas
	C_2	People and equipment
	C_3	Equipment and hazardous areas
	C_4	Equipment and equipment
D	D_1	Operational process supervision
	D_2	Safety hidden danger monitoring
	D_3	Safety status monitoring
E	E_1	Safety training and education
	E_2	Safety protection
	E_3	Emergency system

Table 5: Weight values of indicators and their ranking.

Target layer		Weight value	Total sort
A_1	Special equipment/regional access control	0.75	0.0317
A_2	Active area control	0.25	0.0578
B_1	Tower crane	0.311	0.0623
B_2	Foundation pit engineering	0.3215	0.0748
B_3	Scaffolding	0.0748	0.0768
B_4	Environment	0.1448	0.1027
B_5	Template engineering	0.1161	0.1161
B_6	Structure	0.0317	0.1448
C_1	Human and dangerous areas	0.5212	0.1863
C_2	People and equipment	0.3183	0.2364
C_3	Equipment and hazardous areas	0.0578	0.25
C_4	Equipment and equipment	0.1027	0.311
D_1	Operational process supervision	0.7370	0.3183
D_2	Safety hidden danger monitoring	0.1863	0.3215
D_3	Safety status monitoring	0.0768	0.5212
E_1	Safety training and education	0.7013	0.7013
E_2	Safety protection	0.2364	0.737
E_3	Emergency system	0.0623	0.75

Figure 10: Histogram of weight values for each watch.

layer makes specific responses, such as risk warning and accident command, to reduce the losses caused by accidents.

Secondly, in this paper, the fuzzy comprehensive evaluation method is used to evaluate the performance of the system; it verifies the feasibility of the safety supervision system of construction workers based on the CPS technology and further concludes that avoiding workers' contact with dangerous areas is an important factor to reduce accidents. According to the statistics of annual construction accidents, it is known that high-altitude falling and collapse account for the majority of accidents, and by controlling human contact with dangerous areas, accidents can be effectively prevented. From the evaluation results, this method has also achieved good results and advantages. Therefore, the safety supervision system constructed in this paper will improve the level of construction safety management and reduce the incidence of accidents, which is worthy of adoption by construction enterprises.

Finally, due to the limitations of research conditions, resources, and time, some aspects need further study. In the follow-up study, more scientific research methods such as observation and experiment can be used to collect data on safety behaviors and safety results. Meanwhile, as leaders and policy makers and executors, managers of construction enterprises and staff of safety supervision departments of the government are also included in the study. The theoretical contribution of this article is to provide a new method for the field of safety management and expand the research horizon of related issues.

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 regarding the publication of this paper.

Acknowledgments

This research is supported by the project of Hebei Development and Reform Commission "Research on the operation policy of Hebei Import cross-border e-commerce" of which the project no. is Z20200012. At the same time, this study is

also supported by the research project of Shijiazhuang University of Post and Telecommunications, "the decision of Hebei province to import cross-border electricity supplier operation mode," number YB2019013.

References

[1] L. Wang and X. V. Wang, *Cloud-Based "Cyber-Physical Systems in Manufacturing,"*, vol. 2 of Latest advancement in CPS and IoT applications, , Cham: Springer International-Publishing, 2018.

[2] A. Sajid, H. Abbas, and K. Saleem, "Cloud-assisted IOT-based SCADA systems security: a review of the state of the art and future challenges," *IEEE Access*, vol. 4, pp. 1375–1384, 2017.

[3] Y. Zhang, M. Qiu, C.-W. Tsai, M. M. Hassan, and A. Alamri, "Health-CPS: healthcare cyber-physical system assisted by cloud and big data," *IEEE Systems Journal*, vol. 11, no. 1, pp. 88–95, 2017.

[4] Z. B. Yang, *Research on the problems and counter measures of construction safety management*, vol. 5, Housing and Real Estate, 2017.

[5] X. R. Zhao, "Discussion on the role of construction engineering safety standardization construction in safety production," *Metallurgical Collections*, vol. 1, pp. 163-164, 2018.

[6] X. L. Song, Z. Zhang, J. P. Xu, Z. Zeng, C. Shen, and F. Peña-Mora, "Bi-stakeholder conflict resolution-based layout of construction temporary facilities in large-scale construction projects," *International Journal of Civil Engineering*, vol. 16, no. 8, pp. 941–964, 2018.

[7] Y. Zhang, "Problems and countermeasures of building safety management in China," *Theoretical Research in Urban Construction*, vol. 5, no. 13, pp. 13–17, 2015.

[8] J. Liu, "Current situation and direction of construction safety management," *Theoretical Research in Urban Construction*, vol. 22, pp. 236-237, 2015.

[9] J. Qiu, "Talking about safety management and protection countermeasure of building construction," *Real Estate Biweekly*, vol. 19, pp. 319–319, 2014.

[10] I. W. H. Fung, V. W. Y. Tam, C. P. Sing, K. K. Tang, and S. O. Ogunlana, "Psychological climate in occupational safety and health: the safety awareness of construction workers in South China," *International Journal of Construction Management*, vol. 16, no. 4, pp. 315–325, 2016.

[11] J. Li, L. Zhang, X. Feng, K. Jia, and F. Kong, "Feature extraction and area identification of wireless channel in mobile communication," *Journal of Internet Technology*, vol. 20, no. 2, pp. 545–553, 2019.

[12] J. Teizer and T. Cheng, "Proximity hazard indicator for workers-on-foot near miss interactions with construction equipment and geo-referenced hazard areas," *Automation in Construction*, vol. 60, no. 2015, pp. 58–73, 2015.

[13] A. Yang, Y. Li, F. Kong, G. Wang, and E. Chen, "Security control redundancy allocation technology and security keys based on Internet of Things," *Ieee Access*, vol. 6, no. 1, pp. 50187–50196, 2018.

[14] M. Loosemore and J. Bridgeman, "Corporate volunteering in the construction industry: motivations, costs and benefits," *Construction Management and Economics*, vol. 2017, pp. 1–13, 2017.

[15] H. Fu, G. Manogaran, K. Wu, M. Cao, S. Jiang, and A. Yang, "Intelligent decision-making of online shopping behavior based on Internet of things," *International Journal of Information Management*, vol. 50, pp. 515–525, 2020.

[16] M. L. Whiteman, P. L. Fernández-Cabán, B. M. Phillips, F. J. Masters, J. A. Bridge, and J. R. Davis, "Multi-objective optimal design of a building envelope and structural system using cyber-physical modeling in a wind tunnel," *Frontiers in Built Environment*, vol. 4, pp. 13–25, 2018.

[17] O. Chieochan, A. Saokaew, and E. Boonchieng, "An integrated system of applying the use of Internet of Things, RFID and cloud computing: a case study of logistic management of Electricity Generation Authority of Thailand (EGAT) Mae Mao Lignite Coal Mining, Lampang, Thailand," in *2017 9th International Conference on Knowledge and Smart Technology (KST),*, Chonburi, Thailand, February 2017.

[18] Q. Wu, X. Fan, W. Wei, and M. Wozniak, "Dynamic scheduling algorithm for delay-sensitive vehicular safety applications in cellular network," *Information Technology and Control*, vol. 49, no. 1, pp. 161–178, 2020.

[19] A. Yang, X. Yang, W. Wu, H. Liu, and Y. Zhuansun, "Research on feature extraction of tumor image based on convolutional neural network," *IEEE Access*, vol. 7, no. 1, pp. 24204–24213, 2019.

[20] M. K. Ghorabaee, M. Amiri, E. K. Zavadskas, and J. Antucheviciene, "A new hybrid fuzzy MCDM approach for evaluation of construction equipment with sustainability considerations," *Archives of Civil & Mechanical Engineering*, vol. 18, no. 1, pp. 32–49, 2018.

[21] M. Gajzler and K. Zima, "Evaluation of planned construction projects using fuzzy logic," *International Journal of Civil Engineering*, vol. 15, no. 4, pp. 641–652, 2017.

[22] L. Wang, H. Y. Zhang, J. Q. Wang, and L. Li, "Picture fuzzy normalized projection-based VIKOR method for the risk evaluation of construction project," *Applied Soft Computing*, vol. 64, pp. 216–226, 2018.

[23] W. Wei, M. Guizani, S. H. Ahmed, and C. Zhu, "Guest editorial: special section on integration of big data and artificial intelligence for Internet of Things," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 4, pp. 2562–2565, 2020.

[24] X. Hu, W. Pedrycz, and X. Wang, "Granular fuzzy rule-based models: a study in a comprehensive evaluation and construction of fuzzy models," *IEEE Transactions on Fuzzy Systems*, vol. 25, no. 5, pp. 1342–1355, 2017.

[25] J. Zhao, L. Yu, and L. Li, "Construction of an evaluation and selection system of emergency treatment technology based on dynamic fuzzy GRA method for phenol spill," *AIP Conference Proceedings*, vol. 1839, no. 1, 2017.

[26] L. Wang and T. Haupt, "The fuzzy in evassluation of concrete structure engineering," *Journal of Engineering Design and Technology*, vol. 16, no. 2, pp. 327–338, 2018.

[27] A. Ardeshir and M. Mohajeri, "Assessment of safety culture among job positions in high-rise construction: a hybrid fuzzy multi criteria decision-making (FMCDM) approach," *International Journal of Injury Control and Safety Promotion*, vol. 25, no. 4, pp. 1–12, 2018.

[28] Y. Jin, G. Li, and H. Zhang, "Evaluation of regional rural information environment based on fuzzy method in the era of the Internet of Things," *IEEE Access*, vol. 6, pp. 78530–78541, 2018.

[29] W. Wei, X. Xia, M. Wozniak, X. Fan, R. Damaševičius, and Y. Li, "Multi-sink distributed power control algorithm for cyber-physical-systems in coal mine tunnels," *Computer Networks*, vol. 161, pp. 210–219, 2019.

Hindawi
Wireless Communications and Mobile Computing
Volume 2020, Article ID 8861207, 14 pages
https://doi.org/10.1155/2020/8861207

Research Article

Research on the Evaluation Model of Rural Information Demand Based on Big Data

Yanfeng Jin[1,2] **Gang Li**[1] **and Jianmin Wu**[2]

[1]*School of Economics and Management, Beijing University of Posts and Telecommunications, Beijing 100876, China*
[2]*Shijiazhuang Posts and Telecommunications Technical College, Shijiazhuang 050021, China*

Correspondence should be addressed to Gang Li; jinyf@bupt.edu.cn

Received 17 June 2020; Revised 7 July 2020; Accepted 20 August 2020; Published 8 September 2020

Academic Editor: Bingxian Lu

In recent years, the imbalance of rural information supply and demand has seriously hindered the process of rural informatization. Rural information demand is a decisive factor in the relationship between rural information supply and demand. Therefore, research on the influencing factors of rural information demand has attracted much attention. The traditional rural information demand factor analysis does not consider the correlation between factors. The factors themselves carry a lot of repeated information, which seriously interferes with the objectivity of the analysis results. Proceeding from the complexity and diversity of influencing factors of rural information demand, based on the selected subjective and objective factors, based on the forward partial correlation analysis and post-ROC test, a probit discriminant model of influencing factors of rural information demand was constructed, and the relationship with Lingshou was determined. There are 24 factors that are significantly related to county rural information needs. The research results show that this method not only eliminates the factors that carry highly repetitive information and the correlation is not significant but also makes the results more reliable. At the same time, it also found that rural information supply is related to farmers' information cognition ability, acceptance awareness, and acceptance ability. This study provides new methods and new ideas for solving related problems.

1. Introduction

With the rapid development of science and technology and the advent of the big data era, the construction of smart cities at home and abroad has made remarkable achievements. At the same time, rural informatization also ushered in new opportunities for development. Digital rural areas and intelligent rural areas have become the hot spots of scholars. Of course, there are great challenges as well as opportunities. Due to the unbalanced development of regional economy, there are many problems in rural informatization. First, the collection, processing, integration, and sharing of rural information are difficult. Second, data mining cannot be carried out effectively and does not provide the information farmers need. Third, there is a contradiction between the diversity of farmers' information needs and the unity of platform information. Fourth, lack of dynamic maintenance and update mechanism, data outdated, cannot play its due role. This requires big data technology and big data thinking to improve and solve the current difficulties. The application of big data in various industries has achieved good results. The idea of big data has gradually penetrated into the process of rural informatization. With the help of big data technology, we can build a comprehensive rural information platform based on farmers' information needs.

Whether in developed countries such as Europe and the United States, or in developing countries such as Asia and Africa, there are many studies on the information needs of rural residents. Scholars' research shows that farmers' demand for information is more and more extensive, and the types of demand and access channels show a variety of characteristics [1]. Kaniki's survey of two rural communities in South Africa found that the main information needs of farmers are information needed to seek jobs or increase income, vocational or skills training opportunities, information about grants, medical and health information, legal counseling services, and so on [2]. In Asia, Raju's study found that the most common information needs of Indian farmers

were medical and health information, infrastructure information, crop improvement and yield information, product sales and market information, policy, and service information [3]. Vevrek thinks the daily information needs of the rural population in the United States are information about local government decisions, information about health services, and local news [4–6]. Domestic researchers have found that farmers pay more attention to specialized information related to agricultural production and operation. Zhang Ying (2017) based on the rural information service platform, from the perspective of farmers, found that farmers' demand for labor market information, agricultural market information, agricultural policy information, and agricultural production information decreased in turn. Li Lu (2016) surveyed the demand for agricultural technology social services and found that farmers' age, education level, and whether they went out to work would affect farmers' demand for agricultural technology social services. Young and experienced peasants paid more attention to information services related to the circulation of agricultural products. Zhou Fengtao (2016) studied the farmers' demand for information services and found that educational level and whether to participate in rural cooperatives had a significant impact on their demand for technical services, agricultural services, and information services. Lu Xinru and Li Zhigang (2017) explored the unique information needs and behaviors of farmers through questionnaires. Farmers' information demand had three characteristics: the tendency of market purchase and sale information, the necessity of policies and regulations, and the particularity of meteorological forecast. Information behavior was restricted by educational level and the overall channel was narrow. Ma Chunyan (2016) carried out an investigation and research on poverty-stricken areas. From the questionnaire, through the analysis of demand types, information access channels, personal literacy, and other aspects, it provided suggestions and countermeasures for speeding up the development of local agriculture and reversing the backward development situation in remote areas. Pan Yuchen and Huo Yucan (2018) analyzed the concept of rural information consumption, the level of demand, and the motivation of consumption, especially in the field of emotional demand, which was also a further reflection of the demand level theory. Provided guidance for the development of the whole society and related enterprises helped enterprises to improve the pertinence of information services and achieve steady growth. Wang Xiaoning and Wang Ming (2018) empirically analyzed the main channels for farmers to obtain information under the background of mobile Internet by issuing questionnaires. Through the analysis, it was concluded that mobile micromessaging, mobile QQ, and mobile microblogging are the absolute dominant advantages in information dissemination, while agricultural information website platform was not generally known to farmers. Guan Lili (2017) analyzed the information needs and constraints of farmers through questionnaires, especially the five characteristics of local farmers: the increasing variety of demand categories, the diversification of access methods, the depth of demand levels, the strong internal motivation of demand, and the strong ability of information research and judgment. It provided experience in understanding the level of rural informatization and promoting the construction of information frameworks. Cui Kai and Feng Xian (2017) combed and analyzed the relevant literature at home and abroad, and studied the significance of information dissemination, the information needs of rural residents, and the information supply in rural areas. From the perspective of information poverty alleviation, Li Gang and Qiao Haicheng (2017) proposed that the government should pay attention to information poverty alleviation through the construction of rural poverty-stricken area model and analysis of relevant data.

In summary, it has been found from the existing research that the information needs of farmers in China are increasingly strong and the demand structure is increasingly diversified, but the specialized information related to agricultural production development is still the most important component of farmers' information consumption. Affected by income levels and cultural quality, mass media such as television and broadcasting are still the main channels of information dissemination, but the proportion of computers and mobile phones is increasing, especially in economically developed areas [10–13]. Researchers summarized and analyzed the influencing factors of farmers' information demand from various angles, but the correlation analysis between the influencing factors is relatively small, and the statistics and descriptions of the factors are not comprehensive enough [14–16]. At the same time, the significant impact of various factors on rural information demand is insufficient. Aiming at the above problems, the concept model of farmer information demand of "source-flow-use" was put forward. Based on the discrete selection model of econometrics, the probit model of rural information demand was constructed. Firstly, the partial correlation analysis of the influencing factors of rural information demand was carried out, and the high coincidence factor was removed. The probit model was used for the second test. Finally, the ROC curve was used for discrimination. Eight factors with no significant influence, such as the proportion of fixed-line administrative villages, were removed. At the same time, 24 significant influencing factors were ranked according to the degree of influence. The results prove the feasibility of the method.

2. Model Building

2.1. Evaluation of Influencing Factors Based on Partial Correlation Analysis. In a system consisting of multiple elements, when studying the influence or correlation of one element on another, the influence of other elements is regarded as a constant, i.e., the close relationship between the two elements is studied separately without considering the influence of other elements, which is called partial correlation analysis [17, 18]. That is the partial correlation coefficient. In the study of rural information demand, there are many factors involved. There may be some correlations between the factors, which leads to the duplication of information reflected by two or more factors, which leads to the system being too complicated because there are unrelated factors [19]. Through partial correlation analysis, factors with repeated

information that affect rural information needs can be removed.

(1) Calculation of partial correlation coefficient.

Suppose t_{ij} is the data value of the i index of the selected village j in the region, t_{kj} is the data value of the k index of the selected village j in the region, and r_{ik} is the partial correlation coefficient between the k index and the first index. The formula is as follows:

$$r_{ik} = \frac{\sum_1^n \left(x_{ij} - \overline{x_i}\right)\left(x_{kj} - \overline{x_k}\right)}{\sqrt{\sum_1^n \left(x_{ij} - \overline{x_i}\right)^2}\sqrt{\sum_1^n \left(x_{kj} - \overline{x_k}\right)^2}}. \tag{1}$$

Among them, n denotes the number of villages in the study area, $\overline{x_i}$ denotes the average value of the i factor, and $\overline{x_k}$ denotes the average value of the k factor.

Suppose R is a correlation coefficient matrix composed of partial correlation coefficient r_{ik}, where m is the number of influencing factors, then.

$$R = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1m} \\ r_{21} & r_{22} & \cdots & r_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ r_{m1} & r_{m2} & \cdots & r_{mm} \end{bmatrix}. \tag{2}$$

Let S be the inverse matrix of the correlation coefficient matrix R.

$$S = R^{-1} = \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1m} \\ s_{21} & s_{22} & \cdots & s_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ s_{m1} & s_{m2} & \cdots & s_{mm} \end{bmatrix}. \tag{3}$$

According to the formula of partial correlation coefficient, the partial correlation coefficient r'_{ik} between the i factor and the k factor can be obtained.

$$r'_{ik} = \frac{-s_{ik}}{\sqrt{s_{ii}s_{kk}}}. \tag{4}$$

The greater the partial correlation coefficient r'_{ik} is, the greater the correlation between the and the k influencing factors is. And the smaller the r'_{ik} is, the smaller the correlation between the i and the k influencing factors is.

(2) Calculation of F value

When the correlation between the two factors is high, in order to avoid the subjective deletion of the significant factors, we can solve this problem by calculating the F value of the two factors. Assuming that F_i is the F value of the i factor, Equation (5) can be used for calculation.

$$F_i = \frac{\left(\bar{x}_j^{(0)} - \bar{x}_j\right)^2 + \left(\bar{x}_j^{(1)} - \bar{x}_j\right)^2}{\left(1/\left(n^{(0)} - 1\right)\right)\sum_{y_j=0}\left(x_{ij} - \bar{x}_j^{(0)}\right)^2 + \left(1/\left(n^{(1)} - 1\right)\right)\sum_{y_j=1}\left(x_{ij} - x_j^{(1)}\right)^2}, \tag{5}$$

$$\overline{x_i}^{(0)} = \frac{1}{n^{(0)}}\sum_{y_j=0} x_{ij}, \tag{6}$$

$$\overline{x_i}^{(1)} = \frac{1}{n^{(1)}}\sum_{y_j=1} x_{ij}, \tag{7}$$

$$\overline{x_i} = \frac{1}{n}\sum_{j=1}^{n} x_{ij}. \tag{8}$$

F_i reflects the magnitude of the influence of the i factor on rural information demand; the greater the F_i is, the greater the impact is; on the contrary, the smaller the impact on rural information demand is.

In the multivariate analysis of rural information demand factors, pure correlation analysis cannot fully reflect the correlation between the factors, because other factors interfere with these factors, so partial correlation analysis is an effective way to solve this problem [20].

(3) Set the deletion criterion based on partial correlation analysis

If the absolute value of the partial correlation coefficient of two related factors $|r_{ik}| > 0.7$, it is considered that the two factors are highly correlated, and the information of the two factors response is highly repeatable, so one of them should be deleted. If the partial correlation coefficient is greater than 0.7, the factor whose F value is less than 0.7 should be deleted.

2.2. Analysis of Influencing Factors Based on Probit Regression

2.2.1. Discrete Probit Regression Model. The probit model is a generalized linear model that follows a normal distribution [20]. The simplest probit model is that the explanatory variable Y is a 0, 1 variable, and the probability of an event occurring depends on the explanatory variable $(Y = 1) = f(x)$, that is, the probability of $Y = 1$ is a function of X, where $f()$ obeys the standard normal distribution. This paper will use the probit model to screen out the factors affecting the information demand in rural areas. When the value of dependent variable y_j is 1, it shows that independent variable has an impact on rural information demand, and when the value of dependent variable y_j is 0, it shows that independent variable has no effect on rural information demand.

(1) Introducing intermediate variables y_j^*

Because the probit model is a linear model, and the dependent variable y_j is 0 and 1, it is a discrete variable, so it cannot be directly calculated by linear regression equation. Therefore, it can be solved by introducing intermediate

variable y_j^* and fitting linear regression equation with influencing factors. y_j^* can represent a state of rural information demand; when $y_j^* > 0$ and the value of y_j is 1, think that this factor has an impact on rural information demand; when $y_j^* < 0$, think that the value of y_j is 0, and this factor has no impact on rural information demand. The linear regression equation is given below.

$$y_j^* = \sum_{i=1}^{m} \beta_i x_{ij} + \alpha + \varepsilon_j = X_j\beta + \alpha + \varepsilon_j. \quad (9)$$

y_j^* is an intermediate variable, representing the rural information demand state of the j village; β_i represents the regression coefficient of the i influencing factor; x_{ij} represents the observed value of the i influencing factor of the j village; α is a constant term; ε_j is a random variable and obeys normal distribution $\varepsilon_j \sim N(0, \sigma^2)$; $\beta = (\beta_0, \beta_1, \cdots, \beta_m)$ is a regression coefficient vector, and $X_j = (x_{1j}, x_{2j}, \cdots, x_{mj})$ is a vector composed of the influencing factors of the j village.

(2) Calculate the probability of rural information demand in each village

The intermediate variable y_j^* of Equation (10) is used to calculate the probability of rural information demand in each village. Because of $\varepsilon_j \sim N(0, \sigma^2)$, it is concluded that

$$P\left(y_j = 1 \mid X_j\right) = P\left(y_j^* > 0 \mid X_j\right) = \Phi\left(\alpha + X_j\beta\right). \quad (10)$$

Similarly, it is possible to calculate the probability of unaffected information demand in rural areas:

$$P\left(y_j = 0 \mid X_j\right) = P\left(y_j^* < 0 \mid X_j\right) = 1 - \Phi\left(\alpha + X_j\beta\right). \quad (11)$$

Where f is a normal distribution function, it can be solved by Equation (12) through maximum likelihood estimation.

$$\text{MAX} \ln L = \sum_{j=1}^{n} \left[y_j \ln\left(\Phi\left(\alpha + X_j\beta\right)\right) + \left(1 - y_j\right) \ln\left(1 - \Phi\left(\alpha + X_j\beta\right)\right)\right]. \quad (12)$$

2.2.2. Testing Based on the Probit Model. Construct a probit model, establish the Wald statistic of the influencing factors, and use the chi-square test [21, 22]. When the corresponding significance probability is greater than 0.01, the factors with the greatest significance probability are deleted. The specific steps are as follows:

(1) Calculate the regression coefficient of the probit model. The probit regression model was constructed according to Equations (9) and (12) of m factors affecting rural information demand and the corresponding observed values of rural information demand state y_i. The corresponding coefficients α, β and corresponding standard errors SE_{β_k} are solved, where $\beta = (\beta_1, \beta_2, \cdots, \beta_m)$

(2) Calculate the significance probability of each factor s, construct the Wald statistics of each factor, and test the hypothesis of the significance of each factor

Suppose H_0: $\beta_k = 0$. If H_0, the k factor has no significant impact on rural information demand.

Suppose H_1: $\beta_k \neq 0$. If H_1, then the k factor has a significant impact on the rural information demand.

Let W_k be the Wald statistical variable corresponding to the k influencing factor of rural information demand, β_k be the parameter estimation value of the k influencing factor, and SE_{β_k} be the standard error of β_k, then.

$$W_k = \left(\frac{\beta_k}{\text{SE}_{\beta_k}}\right)^2. \quad (13)$$

By constructing the Wald statistic W_k, it is possible to test whether the parameter estimation β_k of the influence factors is significantly 0. If $\beta_k = 0$, H_0 is true. W_k obeys the chi-square distribution with degree of freedom 1, that is $W_k \sim \chi^2(1)$; the corresponding significance probability value s is obtained according to the chi-square distribution table.

(i) If $s < 0.01$, the original hypothesis H_0 is rejected, which shows that this factor has a significant impact on the rural information demand

(ii) If $s > 0.01$, then accept the original hypothesis H_0, indicating that although $\beta_k = 0$, but this factor has no significant impact on rural information needs

(3) For all the influencing factors of significant probability $s > 0.01$, the maximum s value is removed. $s > 0.01$ shows that accepting the hypothesis H_0, this factor has no significant impact on rural information demand. Among all the factors that have no significant impact, the factors corresponding to the maximum s value can be removed. It should be noted that all factors affecting $s > 0.01$ cannot be deleted at one time, because each factor may be affected by multiple variables, deleting a variable; the original nonsignificant factors may become significant factors

(4) Repeat Steps (1)–(3) until the coefficients of all variables in the model meet $s < 0.01$

By solving the state variable y of rural information demand and the coefficient of probit regression equation β between influencing factors and its standard error SE_β, construct Wald statistics of influencing factors to test the significance probability of regression equation coefficient β and eliminate the factors that have little impact on rural information demand, and the regression coefficient β is not significant.

Table 1: Classification confusion matrix.

Real situation	Prediction results	
	Positive example	Counter example
Positive example	True example (TP)	False counter example (FN)
Counter example	False positive cases (FP)	True counter example (TN)

Table 2: Classification results of the probit regression model for influencing factors of rural information demand.

Actual impact	Model classification results		
	1 (influential)	0 (no impact)	Total
1 (influential)	The actual influence is determined by the number of factors that are affected by the model TP	The number of factors that actually affect but is misjudged by the model is not affected by FN	TP + FN
0 (no impact)	The number of factors that are misjudged by the model is FP	The actual number of factors that were correctly judged by the model was not affected by TN	FP + TN
Total	TP + FP	FN + TN	

2.3. Validation of Influencing Factors Based on ROC Curve

2.3.1. ROC Curve. The ROC curve refers to the receiver operating characteristic. Each point on the ROC curve reflects the sensitivity to the same signal stimulus [23, 24]. In view of the relationship between the predicted value and the true value, we can divide the sample into four parts: true positive (TP): the predicted value and the true value are all 1; false positive (FP): the predicted value is 1, and the true value is 0; true negative (TN): the predicted value and the true value are both 0; and false negative (FN): the predicted value is 0, and the true value is 1. The classification confusion matrix is shown in Table 1.

The vertical axis of the ROC curve represents true positive rate (TPR), and the horizontal axis represents false positive rate (FPR).

$$\mathrm{TPR} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}, \tag{14}$$

$$\mathrm{FPR} = \frac{\mathrm{FP}}{\mathrm{TN} + \mathrm{FP}}. \tag{15}$$

ROC curve is actually a dot plot of TPR and FPR under different thresholds. Given a threshold, we can get the corresponding TPR and FPR values. By detecting a large number of thresholds, a TPR-FPR correlation map can be obtained. In AUC (area under the curve), that is, the larger the area under the ROC curve is, the better the classifier is, the maximum value is 1.

2.3.2. Inspection of Influencing Factors of Rural Information Demand Based on ROC Curve. The ACU value of ROC curve is used to determine whether the factors affecting rural information demand selected by the probit regression model are correct [25]. According to the confusion classification matrix, the number of influential factors is recorded as TP, the number of factors misjudged as influential factors is recorded as FN, the number of factors judged as unaffected factors is recorded as FP, and the number of factors misjudged as unaffected factors is recorded as TN. The specific analysis results are shown in Table 2.

According to Equation (14), the correct discriminant rate is calculated, and the number TP which is discriminated as the influential factor is divided by the number TP + FN which is the actual number of all the influential factors. It indicates that the factors that actually affect the rural information demand are discriminated as the probability of influencing factors by the abovementioned probit model [26].

According to Equation (15), the misjudgment rate is calculated, and the number of factors which are misjudged as influential factors is divided by the number of factors that are not actually affected by the number of FP + TN. It is indicated that the factors that have no influence on rural information demand are identified as influential factors by the abovementioned probit model.

The ROC curve is plotted on the longitudinal axis and the horizontal axis, respectively, by the correct discriminant rate and false discrimination rate [27]. When the abscissa is constant, the larger the ordinate is, the greater the impact of this factor on rural information demand is, and the corresponding AUC value is also larger. Therefore, the larger the AUC value is, the better the classifier is, which means that the greater the impact of this factor on rural information needs is, the maximum value is 1. When AUC = 1, it is a ideal classifier, and with this prediction model, ideal prediction can be achieved no matter what threshold is set. When $0.9 < \mathrm{AUC} < 1$, the influence factor is better. If the threshold is set properly, the model has better predictions. When $0.7 < \mathrm{AUC} < 0.9$, the influence factors are moderate, and the model has a certain predictive value. When $0.5 < \mathrm{AUC} < 0.7$, the discriminant effect is poor, and there is basically no predictive value. Where $\mathrm{AUC} < 0.5$, the discriminant effect of the model is very poor, but it is better than random guess as long as it always goes against prediction.

Therefore, according to all the factors identified by the above probit regression model, if the AUC value is greater than 0.9, it is concluded that this factor has a significant impact on rural information demand. The research shows that the area under the ROC curve constructed by all the factors in this paper is higher than 0.9, which ensures the ability

Table 3: Influencing factors of rural information demand.

Research object	First level influencing factors	Two level influence factors
Influencing factors of rural information demand	Environmental factors	Fixed coverage of administrative villages X1
		Number of cable TV per 1000 people X2
		Optical fiber length per 100 square kilometers X3
		TV coverage rate X4
		Number of information talents per 10000 people X5
		Number of students per 1000 students X6
		The number of computers per 100 households in rural areas X7
		The number of TV sets per 100 households in rural areas X8
		The number of mobile phones per 100 households in rural areas X9
		Number of Internet users per 10000 X10
		Rural per capita postal volume X11
		Fixed investment in telecom industry accounts for the proportion of total social investment X22
		Fixed investment in the information industry accounts for the proportion of fixed asset investment in the whole society X13
	Main factors	Sex X14
		Age X15
		Marital status X16
		Health X17
		Cultural level X18
		Occupation X19
		Personal income X20
		Experience of going out for work X21
	Family factors	Number of family members X22
		Number of family labor force X23
		Number of male family members X24
		Number of family members X25
		Source of family income X26
		Family happiness index X27
	Economic factors	Per capita income of farmers X28
		Source of farmers' income X29
		Per capita disposable income of farmers X30
	Geographical factors	Distance from county highway X31
		Distance from provincial highway X32
		Distance from town center X33
		Distance from county center X34
	Cognitive factors	Knowledge of information X35
		Awareness of information acceptance X36
		The ability to receive information X37
	Policy factors	National informatization policy X38

to distinguish the influence of various factors on rural information demand.

3. Empirical Analysis of Rural Information Demand

3.1. Analysis of Influencing Factors of Rural Information Demand. Through the combing and research of domestic and foreign literatures, the factors affecting rural information demand are summarized into seven aspects: environmental factors, subject factors, family factors, economic factors, geographical factors, cognitive factors, and political factors [28, 29].

3.1.1. Environmental Factors. At the micro level, the popularity of the Internet, the number of computers, and the number of mobile phones, television, and radio coverage have become important factors affecting rural information needs.

Table 4: Distribution of selected respondents.

Village name	Sample size	Village name	Sample size	Village name	Sample size
New village	78	Lijiazhuang	55	Xichatou	85
Xituo	66	Wanli	61	Lijiagou	54
Ximufu	48	Nanyanchuan	44	Majiazhuang	53
Beijicheng	65	Sijiazhuang	46	Liatong	44
Zhushi	39	Nanbaishi	54	Zhangjiatai	35
Xiaohanlou	67	Xiqingtong	39	Xiwan	29
Niucheng	69	Ciyu	70	Niuzhuang	34
Dongchengnan	43	Dongjiazhuang	46	Zhaitou	48
Sunzhuang	57	Beitanzhuang	80	Nanying	32
Nangoutai	34	Shanmenkou	42	Manshan	26

First, rural information infrastructure and technology are the basic resources of rural information environment and an important premise of rural information environment optimization. Its construction level is an important part of rural information environment. The second is rural information talents. The optimization of rural information environment needs high-quality and professional talent team to achieve, in order to continuously promote the improvement of rural informatization level. Rural scientists and technicians are an important force in the construction of rural information environment and an important guarantee for the continuous advancement of rural informatization. Rural college students have higher professional quality and professional ability, which is an important force in the future construction and optimization of rural information environment. The third is the rural information network coverage. It reflects the application of rural information infrastructure. The four is the input and output of rural informatization.

3.1.2. Subject Factors. Individual characteristics mainly include gender, age, marital status, health status, educational level, occupation, personal income, and migrant work experience. Gender is an important factor affecting rural information need. Generally speaking, men's demand for information is more intense than that of woman. From the perspective of information economics, the subjective desire of different age structures for rural information needs is quite different. Young people are more likely than the elderly to accept new information technology and information products. The impact of marital status on rural information needs research results that are rare, and it is unclear whether there is a correlation. This paper will explore this issue through follow-up models. Health status is also a major impact on rural information needs. The cultural level affects the information quality of rural subjects to a great extent. The traditional theory of rural informatization holds that the farmers' information quality has a positive correlation with the demand and acceptance of informatization. People are engaged in agricultural and nonagricultural occupations in rural areas; the dual nature of occupation may also have an impact on rural information needs. In general, the higher the personal income is, the stronger the demand for information is. Farmers with migrant experience have a wider horizon and a stronger sense of information needs.

3.1.3. Family Factors. Family factors mainly include the number of family population, the number of family labor force, the number of male family, the number of female family, family happiness index, and family income sources. The theory of network externalities believes that as the number of users increases, utility gained by each user from the network increases. Therefore, the number of family members may also be an important factor affecting the information needs in rural areas. Statistical studies have shown that gender is an important factor affecting Internet demand. For rural households, the more males there are, the stronger the rural information demand there is. Similarly, the number of women in the family may also affect the family's demand for rural information. The quantity of household labor force is proportional to household income to a certain extent. The more the labor is, the higher the household income is, the stronger the demand for information is. On the contrary, the less the labor is, the lower the household income is, the lower the desire for information demand is. Family happiness index in a sense reflects the level of family income and indirectly affects the farmers' demand for information. At present, the relationship between happiness index and information access demand has not been found in academic and theoretical circles. However, we can see that the higher the family happiness index is, the higher the income is, so it will indirectly affect the farmers' demand for information.

3.1.4. Economic Factors. Economic factors mainly include the per capita income of farmers, the source of farmers' income, and the level of regional economy.

3.1.5. Geographical Factors. The geographical characteristics of rural information demand have great influence. The geographical features are mainly reflected in the geographical location of rural areas, including county-level roads, provincial highways, distance from township centers, and distance from county centers.

Table 5: Part village related raw data.

First level influencing factors	Two level influence factors	Related raw data					
		C1	C2	C3	C28	C29	C30
Environmental factors	Fixed coverage of administrative villages X1	80	91	68	86	69	74
	Number of cable TV per 1000 people X2	306	203	156	489	543	345
	Optical fiber length per 100 square kilometers X3	5	23	45	21	32	37
	TV coverage rate X4	90	97	88	96	76	85
	Number of information talents per 10000 people X5	123	178	45	36	234	132
	Number of students per 1000 students X6	45	67	39	44	29	56
	The number of computers per 100 households in rural areas X7	39	43	28	50	45	77
	The number of TV sets per 100 households in rural areas X8	90	93	96	97	99	99
	The number of mobile phones per 100 households in rural areas X9	156	137	269	211	169	304
	Number of Internet users per 10000 X10	1267	3304	890	1278	908	4512
	Rural per capita postal volume X11	4	7	11	6	4	2
	Fixed investment in telecom industry accounts for the proportion of total social investment X22	10	13	21	8	25	14
	Fixed investment in the information industry accounts for the proportion of fixed asset investment in the whole society X13	12	13	16	21	15	17
Main factors	Sex X14	1	1	2	2	1	1
	Age X15	3	3	4	3	4	5
	Marital status X16	1	1	1	1	1	1
	Health X17	3	4	3	4	5	4
	Cultural level X18	2	1	2	1	2	3
	Occupation X19	1	1	1	1	1	1
	Personal income X20	1	2	2	2	2	3
	Experience of going out for work X21	2	2	2	2	2	2
Family factors	Number of family members X22	2	3	3	2	3	3
	Number of family labor force X23	2	3	2	3	3	2
	Number of male family members X24	1	2	1	2	2	2
	Number of family members X25	2	1	2	1	1	1
	Source of family income X26	1	1	1	1	1	1
	Family happiness index X27	3	2	3	3	3	4
Economic factors	Per capita income of farmers X28	3247	4563	3349	5640	5530	6742
	Source of farmers' income X29	1	1	1	1	1	1
	Per capita disposable income of farmers X30	3111	4290	3150	4890	4369	5548
Geographical factors	Distance from county highway X31	12	34	18	22	9	15
	Distance from provincial highway X32	2	5	7	4	6	13
	Distance from town center X33	6	8	9	4	7	15
	Distance from county center X34	15	18	14	26	33	18
Cognitive factors	Knowledge of information X35	2	3	2	2	3	3
	Awareness of information acceptance X36	1	2	2	1	2	2
	The ability to receive information X37	2	3	3	3	2	3
Policy factors	National informatization policy X38	1	1	1	1	1	1

3.1.6. Cognitive Factors. Cognitive factors have an important impact on rural information needs. Cognitive factors mainly include the cognitive level of rural subjects to information, the awareness of information acceptance, and the ability to receive information.

3.1.7. Policy Factors. It mainly refers to the national policy information on rural informatization. Government informatization policies, such as rural revitalization strategies, rural e-commerce, digital rural areas, and smart rural areas, affect farmers' perceptions of rural information needs.

In summary, rural information needs are affected by 38 factors in 7 aspects of the environment. This paper uses partial correlation coefficient, probability model, and ROC curve to screen and identify the factors affecting rural information demand, and finally find out the real key factors affecting rural information demand. Specific factors are shown in Table 3.

Table 6: Quantitative design of qualitative variables and its implications.

Variable	Variable name	Variable value	The meaning of variable value
X14	Age	{1, 2, 3, 4, 5}	1 = 18 years old and below, 2 = 19 to 28 years old, 3 = 29-38 years old, 4 = 39 to 48 years old, and 5 = 49 years old and above
X15	Sex	{1, 2}	1 = male and 2 = female
X18	Educational level	{1, 2, 3, 4, 5}	1 = primary school and below, 2 = junior high school, 3 = high school or technical secondary school, 4 = specialist, and 5 = undergraduate and above
X20	Personal income	{1, 2, 3, 4, 5}	1 = 1000 and below, 2 = 1000-3000, 3 = 3000-5000, 4 = 5000-7000, and 5 = 7000 above
X22	Number of family members	{1, 2, 3, 4, 5}	1 = 1, 2 = 2-3, 3 = 4-5, 4 = 6-7, and 5 = 8 and above
X16	Marital status	{1, 2, 3, 4}	1 = unmarried, 2 = married, 3 = divorced, and 4 = widow
X35	Knowledge of information	{1, 2, 3, 4, 5}	1 = conflict, 2 = is unwilling, 3 = is general, 4 = is willing, and 5 = is very willing
X36	Awareness of information acceptance	{1, 2, 3, 4, 5}	1 = is very confused, 2 = does not understand, 3 = is general, 4 = understands, and 5 = knows very well
X37	The ability to receive information	{1, 2, 3, 4, 5}	1 = is very bad, 2 = is bad, 3 = is general, 4 = is strong, and 5 = is very strong

The original data are standardized according to different data types.

Table 7: Partial factors and partial correlation coefficient calculation results.

Factor name	Fixed coverage of administrative villages X1	Number of cable TV per 1000 people X2	The ability to receive information X37	National informatization policy X38
Fixed coverage of administrative villages X1	-1.00	***	***	***
Number of cable TV per 1000 people X2	-0.07	-1.00	***	***
The ability to receive information X37	0.13	0.09	-1.00	***
National informatization policy X38	0.22	0.16	0.31	-1.00

3.2. Sample Selection and Data Sources

3.2.1. Sample Selection. Because this paper studies the rural information needs, so from the regional survey object selected as the villagers of natural villages. Considering the convenience of data acquisition and the homogeneity of sample division, and covering the plain, hilly, and mountainous terrain in the regional space, this study selected 30 natural villages of 15 townships in Lingshou County, Hebei Province, as the sample. The specific distribution is shown in Table 4.

3.2.2. Data Source. The empirical data mainly come from two aspects: the first is the statistical yearbook data of Lingshou County. Second is the survey data; this part of the data mainly includes interviews with relevant personnel data and sample survey data. The specific data is shown in Table 5.

3.3. Data Standardization

3.3.1. Standardization of Data Indicators. For the data indicators including positive, negative, and interval three categories, respectively, the above formulas are used to calculate the standardized 0-1 interval data [30–32].

3.3.2. Quantitative Processing of Qualitative Data. The qualitative data are quantified by using the Likert scale principle. The specific variable design and its meaning are shown in Table 6.

3.4. Evaluation of Influencing Factors of Rural Information Demand Based on Partial Correlation Analysis. Partial correlation analysis of standardized data is carried out to avoid the correlation of indicators only existing in data and the lack of correlation of economic significance [33]. Using the data in Table 7 and according to Equations (4)–(7), the partial correlation coefficients of each factor can be calculated by SPSS software. The results are shown in Table 8. According to the calculation results, the partial correlation coefficients of six pairs of factors are greater than 0.7, so the six pairs of factors are highly correlated and there is information redundancy. Therefore, it is necessary to further calculate the F value of six pairs of related factors. The six related factors are the number of cable TV per 1000 people and the number of TV sets per 100 households in rural areas, the number of information talents per 10000 people and the number of college students per 1000 people, the number of computers per 100 households in rural areas and the number of Internet

Table 8: Factors to be deleted based on F value.

Factors with partial correlation coefficient greater than 0.8				Partial correlation coefficient	Deleting factors
Influencing factors 1	F value	Influencing factors 2	F value		
Number of cable TV per 1000 people X2	0.014	The number of TV sets per 100 households in rural areas X8	0.023	0.84	X2
Number of information talents per 10000 people X5	0.132	Number of students per 1000 students X6	0.095	0.91	X6
The number of computers per 100 households in rural areas X7	0.059	Number of Internet users per 10000 X10	0.018	0.93	X10
Personal income X20	0.004	Per capita income of farmers X28	0.012	0.88	X20
Source of family income X26	0.236	Source of farmers' income X29	0.170	0.85	X29
Distance from county highway X31	0.301	Distance from provincial highway X32	0.332	0.87	X31

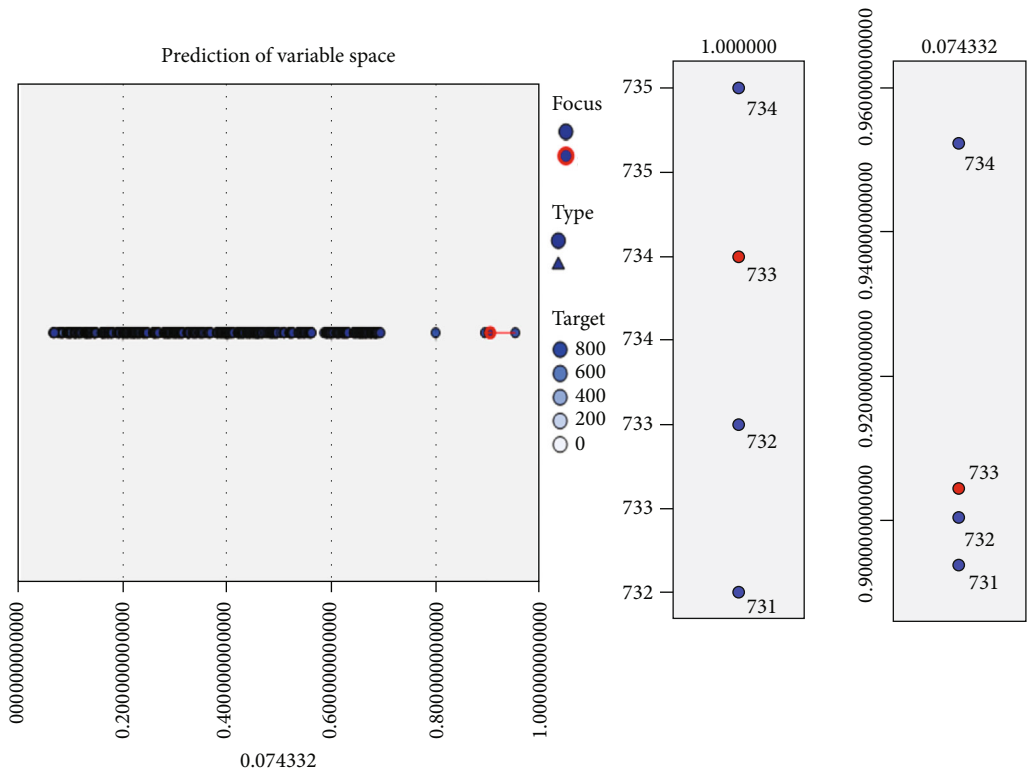


Figure 1: Clustering results of partial correlation coefficients of various factors.

users per 10000 people, personal income and per capita income of farmers, family income sources and farmers' income source, distance from county highway, and distance from provincial highway. The specific results are shown in Table 7.

The F values of six pairs of related factors are calculated, and the results are shown in Figure 1. At the same time, six pairs of factors with F values were compared, and 6 factors with smaller F value were deleted. From the data in Table 9, we can see that the F value of the number of cable TV per 1000 people is less than the F value of the number of television per 100 households in rural areas, the F value of the number of college students per 10000 people is less than the F value of the number of information personnel per 10000 people, the F value of the number of Internet users per 10000 people is less than the F value of the number of computers per 100 households in rural areas. The F value is smaller than the per capita income of farmers, and the F value of the distance between county highway and provincial highway is smaller than that of provincial highway. Therefore, six factors such as X2, X6, X10, X20, X29, and X31 with smaller F value are deleted. The specific results are shown in Table 8.

3.5. Analysis of Influencing Factors of Rural Information Demand Based on Probit Regression. On the basis of partial correlation analysis, the remaining factors are screened by using the probit regression model to find out the factors that have a greater impact on rural information demand [34]. After regression analysis of the remaining 32 factors, the relevant regression parameters were calculated. The specific results are shown in Table 9.

Table 9: Probit regression results.

Factor name	Regression coefficient	Standard error	Wald test value	Saliency probability
Fixed coverage of administrative villages X1	0.176	0.287	0.361	0.073
Optical fiber length per 100 square kilometers X3	0.384	1.236	0.445	0.129
TV coverage rate X4	0.279	0.883	0.750	0.069
Number of information talents per 10000 people X5	0.783	0.176	0.115	0.230
The number of computers per 100 households in rural areas X7	0.212	0.478	1.004	0.176
The number of TV sets per 100 households in rural areas X8	0.334	0.579	0.668	0.097
The number of mobile phones per 100 households in rural areas X9	0.913	0.336	0.654	0.209
Rural per capita postal volume X11	0.592	0.448	0.790	0.075
Fixed investment in the information industry accounts for the proportion of fixed asset investment in the whole society X13	0.398	0.667	0.085	0.148
Sex X14	-1.023	0.369	0.981	0.668
Age X15	0.359	0.783	0.245	0.189
Marital status X16	-1.382	0.450	0.695	0.033
Health X17	-0.659	0.560	0.235	0.439
Cultural level X18	0.785	0.207	0.638	0.091
Occupation X19	-3.772	0.697	0.346	0.087
Experience of going out for work X21	-2.910	0.458	0.442	0.037
Number of family members X22	0.709	0.430	0.675	0.127
Number of family labor force X23	0.559	0.650	0.283	0.076
Number of male family members X24	0.389	0.452	0.109	0.087
Number of family members X25	0.669	0.127	1.245	0.343
Source of family income X26	0.945	0.457	2.331	0.061
Family happiness index X27	0.775	0.451	0.609	0.108
Per capita income of farmers X28	0.707	0.532	0.246	0.079
Per capita disposable income of farmers X30	0.409	0.610	3.026	0.417
Distance from provincial highway X32	-1.731	0.796	0.458	0.065
Distance from town center X33	-1.088	0.569	0.337	0.098
Distance from county center X34	-3.952	0.480	0.639	0.112
Knowledge of information X35	0.515	0.707	3.041	0.046
Awareness of information acceptance X36	0.738	0.449	2.064	0.032
The ability to receive information X37	0.893	0.649	1.372	0.018
National informatization policy X38	0.776	0.985	0.689	0.050

The standard error of each factor reflects to a certain extent of the variation degree of sample average to total average [35]. The difference of standard errors of factors shows that there are certain differences in the selection of samples for each factor. However, the significance of this effect on each factor is acceptable.

In the significant probability $s > 0.01$ factor, delete the biggest factor of s value. According to this principle, we compare the s of all $s > 0.01$ factors in Table 10 to delete the largest one. Probit regression is performed on the remaining 31 factors, and the corresponding regression parameters are calculated until the s value of all the factors is less than 0.01. For example, according to the results of the first regression, all s values are less than 0.1, but the gender factor has the largest s value, so the gender factor is deleted, and then probit regression is performed again until the s value of all factors is less than 0.1. Finally, through the probit regression analysis, 8 factors such as the proportion of administrative village, gender, marital status, health status, number of family members, number of male family members, number of female family members, and family happiness index were deleted, which did not significantly affect rural information demand.

3.6. Validation and Analysis of Factors Affecting Rural Information Demand Based on ROC Curve. The data of 24 selected factors were brought into Equations (9)–(12). The probability $P(y_j = 1|X_j)$ of each village affected by relevant factors was calculated by using the probit model. When $P(y_j = 1|X_j) > 0.5$, the effect was obvious, and when $P(y_j = 1|X_j) < 0.5$, it was not.

First, the AUC value is a probability value. When you randomly select a positive sample and a negative sample, the probability that the current classification algorithm ranks the positive sample before the negative sample according to the calculated score value is the AUC value. The larger the

Table 10: AUC value of all influencing factors.

Serial number	Factor name	AUC value
1	The number of mobile phones per 100 households in rural areas X9	0.975
2	The number of computers per 100 households in rural areas X7	0.972
3	Number of family members X22	0.967
4	The ability to receive information X37	0.966
5	Optical fiber length per 100 square kilometers X3	0.96
6	Distance from county center X34	0.955
7	Number of family labor force X23	0.953
8	Awareness of information acceptance X36	0.953
9	The number of TV sets per 100 households in rural areas X8	0.947
10	Distance from town center X33	0.946
11	Experience of going out for work X21	0.942
12	National informatization policy X38	0.941
13	Knowledge of information X35	0.937
14	Source of family income X26	0.936
15	Per capita income of farmers X28	0.935
16	Fixed investment in the information industry accounts for the proportion of fixed asset investment in the whole society X13	0.933
17	Age X15	0.931
18	Distance from provincial highway X32	0.928
19	Cultural level X18	0.925
20	Occupation X19	0.921
21	Rural per capita postal volume X11	0.919
22	Number of information talents per 10000 people X5	0.917
23	Per capita disposable income of farmers X30	0.917
24	TV coverage rate X4	0.913

AUC value is, the more likely the current classification algorithm will rank the positive sample before the negative sample, so that they can be better classified.

Specifically, it is to count all $M \times N$ (M is the number of positive samples; N is the number of negative samples) positive and negative sample pairs; how many groups of positive samples have a score greater than the negative sample score. When the scores of the positive and negative samples in the binary group are equal, the calculation is performed according to 0.5. Then divide by MN. The formula for calculating the AUC value is as follows:

$$\mathrm{AUC} = \frac{\sum_{i \in \text{positive class}} \mathrm{rank}_i - (M(1+M)/2)}{M \times N}. \tag{16}$$

The ROC curve corresponding to 24 factors and the area under the curve (AUC) value were obtained by calculation. The results show that all AUC values are greater than 0.9, indicating that all factors are significantly related to rural information demand. At the same time, according to the rule that the greater the AUC value is, the more significant the demand relationship is, the order of 24 factors is ranked. The impact of every 100 households in rural areas that have mobile phones is most significant. The AUC values for the specific 24 factors are shown in Table 10.

The ROC curve is composed of dot plots of TPR and FPR corresponding to multiple critical values. Therefore, different threshold values can be used to obtain points above the multiple ROC curves, and the TPR and FPR values are used as the horizontal and vertical axes, respectively. The SPSS software draws the most significant factor. The ROC curve of the number of mobile phones per 100 households in rural areas is shown in Figure 2.

The area below the ROC curve indicates that the AUC value reflects the significant impact of the number of mobile phones per 100 households in rural areas on rural information demand. In Figure 2, AUC = 0.975 is greater than 0.9, so there are 100 rural households screened by the probit model. The number of mobile phones has a significant impact on rural information needs.

4. Conclusion

This chapter mainly analyzes and studies the information demand problem caused by the lack of rural information supply as a whole, and obtains the following conclusions:

(1) The traditional factor analysis of rural information demand does not consider the correlation between factors, so the factors themselves carry a lot of

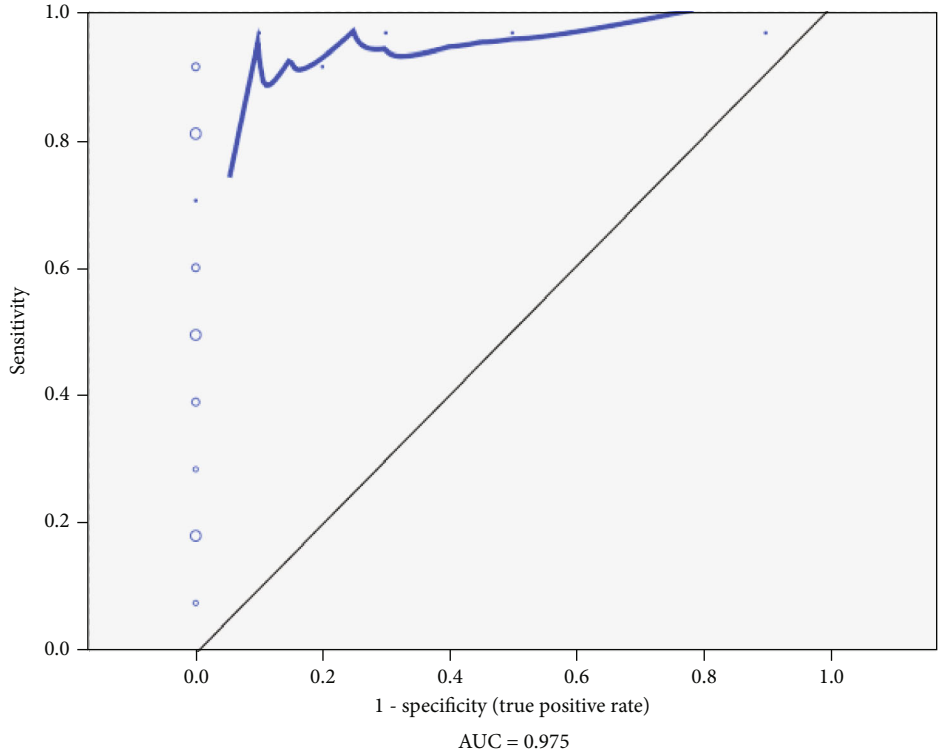


Figure 2: ROC curve of the number of mobile phones per 100 households.

redundant information, which is a certain interference to the judgment of the impact degree. Taking Lingshou County as an example, using the method of partial correlation analysis, by calculating F value, the influencing factors with highly repetitive information are eliminated, and the complexity of calculation is reduced. The probit regression model is constructed to test the influencing factors of rural information demand. Through the comparison of regression coefficient and test probability, the nonsignificant correlation of rural information demand is deleted, and ROC curve is introduced to test the above results twice, which improves the reliability of factor correlation

(2) The 24 influencing factors of rural information demand directly or indirectly affect the supply of rural information services. They provide the basis for the supply of rural information services from the seven aspects of objective environment, subject characteristics, family, economy, geography, cognition, and policy, such as improving infrastructure construction, training information service talents, and providing differentiation, and at the same time, the research results also show that the supply of rural information is related to farmers' information cognitive ability, acceptance awareness, and acceptance ability

The innovation of this paper lies in the partial correlation analysis of influencing factors of rural information demand and the ROC secondary test. It provides a new idea and method to solve the related problems.

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 regarding the publication of this paper.

Acknowledgments

This research is supported by the "Three Three Three Talent Project" funded by Hebei Province (Project No.: A202001064).

References

[1] A. Antunes, D. Bonfim, N. Monteiro, and P. M. M. Rodrigues, "Forecasting banking crises with dynamic panel probit models," *International Journal of Forecasting*, vol. 34, no. 2, pp. 249–275, 2018.

[2] A. E. Elshaikh, X. Jiao, and S.-h. Yang, "Performance evaluation of irrigation projects: theories, methods, and techniques," *Agricultural Water Management*, vol. 203, pp. 87–96, 2018.

[3] A. M. Valente, H. Binantel, D. Villanua, and P. Acevedo, "Evaluation of methods to monitor wild mammals on

Mediterranean farmland," *Mammalian Biology*, vol. 91, pp. 23–29, 2018.

[4] A. I. Bandos, B. Guo, and D. Gur, "Estimating the area under ROC curve when the fitted binormal curves demonstrate improper shape," *Academic Radiology*, vol. 24, no. 2, pp. 209–219, 2017.

[5] U. Benjamin and U. CLN, "Libraries and information in Nigerian rural development," *International Journal of Information Management*, vol. 34, no. 1, pp. 14–16, 2014.

[6] C. A. Damalas and M. Khan, "RETRACTED: Pesticide use in vegetable crops in Pakistan: insights through an ordered probit model," *Crop Protection*, vol. 99, pp. 59–64, 2017.

[7] G. Msoffe and P. Ngulube, "Farmers access to poultry management information in selected rural areas of Tanzania," *Library & Information Science Research*, vol. 38, no. 3, pp. 265–271, 2016.

[8] H. Hu, B. Tang, X. Gong, W. Wei, and H. Wang, "Intelligent fault diagnosis of the high-speed train with big data based on deep neural networks," *IEEE Transactions on Industrial Informatics*, vol. 13, no. 4, pp. 2106–2116, 2017.

[9] G. Fountas and P. C. Anastasopoulos, "A random thresholds random parameters hierarchical ordered probit analysis of highway accident injury-severities," *Analytic Methods in Accident Research*, vol. 15, pp. 1–16, 2017.

[10] G. Zhang, C. Zhang, and H. Zhang, "Improved K-means algorithm based on density canopy," *Knowledge-Based Systems*, vol. 145, pp. 289–297, 2018.

[11] H. S. Loh, Q. Zhou, V. V. Thai, Y. D. Wong, and K. F. Yuen, "Fuzzy comprehensive evaluation of port-centric supply chain disruption threats," *Ocean & Coastal Management*, vol. 148, pp. 53–62, 2017.

[12] J. A. Cook, "ROC curves and nonrandom data," *Pattern Recognition Letters*, vol. 85, pp. 35–41, 2017.

[13] J.-K. Park, S.-K. Lee, and J.-H. Kim, "Development of an evaluation method for nuclear fuel debris–filtering performance," *Nuclear Engineering and Technology*, vol. 50, no. 5, pp. 738–744, 2018.

[14] J.-F. Chen, H.-N. Hsieh, and Q. H. Do, "Evaluating teaching performance based on fuzzy AHP and comprehensive evaluation approach," *Applied Soft Computing*, vol. 28, pp. 100–108, 2015.

[15] J. J. C. Tambotoh, A. D. Manuputty, and F. E. Banunaek, "Socio-economics factors and information technology adoption in rural area," *Procedia Computer Science*, vol. 72, pp. 178–185, 2015.

[16] Y. Jin, G. Li, and H. Zhang, "Evaluation of regional rural information environment based on fuzzy method in the era of the Internet of things," *IEEE Access*, vol. 6, pp. 78530–78541, 2018.

[17] K. Kwon, J. W. Shin, and N. S. Kim, "Incremental basis estimation adopting global k-means algorithm for NMF-based noise reduction," *Applied Acoustics*, vol. 129, pp. 277–283, 2018.

[18] L. Zhang, Y. Feng, P. Shen et al., "Efficient finer-grained incremental processing with MapReduce for big data," *Future Generation Computer Systems*, vol. 80, pp. 102–111, 2018.

[19] K. Papangelis, N. R. Velaga, F. Ashmore, S. Sripada, J. D. Nelson, and M. Beecroft, "Exploring the rural passenger experience, information needs and decision making during public transport disruption," *Research in Transportation Business & Management*, vol. 18, pp. 57–69, 2016.

[20] M. de Figueiredo, C. B. Y. Cordella, D. J.-R. Bouveresse, X. Archer, J.-M. Bégué, and D. N. Rutledge, "A variable selection method for multiclass classification problems using two-class ROC analysis," *Chemometrics and Intelligent Laboratory Systems*, vol. 177, pp. 35–46, 2018.

[21] M. F. M. Firdhous and P. M. Karuratane, "A model for enhancing the role of information and communication technologies for improving the resilience of rural communities to disasters," *Procedia Engineering*, vol. 212, pp. 707–714, 2018.

[22] M. Filippini, W. H. Greene, N. Kumar, and A. L. Martinez-Cruz, "A note on the different interpretation of the correlation parameters in the bivariate probit and the recursive bivariate probit," *Economics Letters*, vol. 167, pp. 104–107, 2018.

[23] P. Mozharovskyi and J. Vogler, "Composite marginal likelihood estimation of spatial autoregressive probit models feasible in very large samples," *Economics Letters*, vol. 148, pp. 87–90, 2016.

[24] P. Matous, "Complementarity and substitution between physical and virtual travel for instrumental information sharing in remote rural regions: a social network approach," *Transportation Research Part A: Policy and Practice*, vol. 99, pp. 61–79, 2017.

[25] R. Khajouei, S. H. Gohari, and M. Mirzaee, "Comparison of two heuristic evaluation methods for evaluating the usability of health information systems," *Journal of Biomedical Informatics*, vol. 80, pp. 37–42, 2018.

[26] R. Fattahi and M. Khalilzadeh, "Risk evaluation using a novel hybrid method based on FMEA, extended MULTIMOORA, and AHP methods under fuzzy environment," *Safety Science*, vol. 102, pp. 290–300, 2018.

[27] R. H. Lange, "The predictive content of the term premium for GDP growth in Canada: evidence from linear, Markov-switching and probit estimations," *The North American Journal of Economics and Finance*, vol. 44, pp. 80–91, 2018.

[28] S. T. Yen and E. M. Zampelli, "Religiosity, political conservatism, and support for legalized abortion: a bivariate ordered probit model with endogenous regressors," *The Social Science Journal*, vol. 54, no. 1, pp. 39–50, 2017.

[29] S. Han and E. J. Vytlacil, "Identification in a generalization of bivariate probit models with dummy endogenous regressors," *Journal of Econometrics*, vol. 199, no. 1, pp. 63–73, 2017.

[30] T.-t. Gao and S.-m. Wang, "Fuzzy integrated evaluation based on HAZOP," *Procedia Engineering*, vol. 211, pp. 176–182, 2018.

[31] W. Yang, K. Xu, J. Lian, L. Bin, and C. Ma, "Multiple flood vulnerability assessment approach based on fuzzy comprehensive evaluation method and coordinated development degree model," *Journal of Environmental Management*, vol. 213, no. 1, pp. 440–450, 2018.

[32] W. Li, W. Liang, L. Zhang, and Q. Tang, "Performance assessment system of health, safety and environment based on experts' weights and fuzzy comprehensive evaluation," *Journal of Loss Prevention in the Process Industries*, vol. 35, pp. 95–103, 2015.

[33] W. Cai, L. Dou, M. Zhang, W. Cao, J.-Q. Shi, and L. Feng, "A fuzzy comprehensive evaluation methodology for rock burst forecasting using microseismic monitoring," *Tunnelling and Underground Space Technology*, vol. 80, pp. 232–245, 2018.

[34] X. Yu, W. Meng, and L. Xiang, "Comprehensive evaluation chronic pelvic pain based on fuzzy matrix calculation," *Neurocomputing*, vol. 173, Part 3, pp. 2097–2101, 2016.

[35] Y. Jin and G. Li, "Application of improved K-means algorithm in evaluation of network resource allocation," *Boletín Técnico*, vol. 55, no. 5, pp. 284–292, 2017.

Hindawi
Wireless Communications and Mobile Computing
Volume 2020, Article ID 8842694, 11 pages
https://doi.org/10.1155/2020/8842694

Research Article

Deep Reinforcement Learning-Based Content Placement and Trajectory Design in Urban Cache-Enabled UAV Networks

Chenyu Wu,[1] **Shuo Shi**,[1,2] **Shushi Gu**,[2,3] **Lingyan Zhang**,[3] **and Xuemai Gu**[1,2]

[1]*School of Electronics and Information Engineering, Harbin Institute of Technology, Harbin 150001, China*
[2]*Peng Cheng Laboratory, Shenzhen, China 518052*
[3]*Harbin Institute of Technology (Shenzhen), Shenzhen, China 518055.*

Correspondence should be addressed to Shuo Shi; crcss@hit.edu.cn

Received 27 May 2020; Revised 19 June 2020; Accepted 13 July 2020; Published 14 August 2020

Academic Editor: Bingxian Lu

Cache-enabled unmanned aerial vehicles (UAVs) have been envisioned as a promising technology for many applications in future urban wireless communication. However, to utilize UAVs properly is challenging due to limited endurance and storage capacity as well as the continuous roam of the mobile users. To meet the diversity of urban communication services, it is essential to exploit UAVs' potential of mobility and storage resource. Toward this end, we consider an urban cache-enabled communication network where the UAVs serve mobile users with energy and cache capacity constraints. We formulate an optimization problem to maximize the sum achievable throughput in this system. To solve this problem, we propose a deep reinforcement learning-based joint content placement and trajectory design algorithm (DRL-JCT), whose progress can be divided into two stages: offline content placement stage and online user tracking stage. First, we present a link-based scheme to maximize the cache hit rate of all users' file requirements under cache capacity constraint. The NP-hard problem is solved by approximation and convex optimization. Then, we leverage the Double Deep Q-Network (DDQN) to track mobile users online with their instantaneous two-dimensional coordinate under energy constraint. Numerical results show that our algorithm converges well after a small number of iterations. Compared with several benchmark schemes, our algorithm adapts to the dynamic conditions and provides significant performance in terms of sum achievable throughput.

1. Introduction

With the development of wireless communication technology, the future networks will require high-quality multimedia streaming applications and highly diversified traffic demand. Unmanned aerial vehicles (UAVs) have brought promising opportunities to assist conventional cellular communication [1]. In urban environment, wireless communication suffers from severe shadowing due to Non-Line-of-Sight (NLoS) propagation [2, 3]. Compared with conventional terrestrial infrastructures, UAVs can be deployed at flexible altitudes, which leads to high probability of Line-of-Sight (LoS) dominant link. Due to the agility and low cost, UAVs can be easily and quickly deployed in a large number of scenarios including disaster relief. In addition, the maneuverability and mobility offer new opportunities for communication enhancement. The performance of communication is significantly improved by dynamically adjusting the UAV states, including flying direction, speed, transmission scheme, and storage resources allocation. The continuous and proper control of UAVs better suits the varying communication conditions.

Thanks to the advantages of mobility, agility, and high probability of LoS propagation, UAVs can be deployed to serve many practical applications, such as base station (BS) offloading [4], Internet-of-Things (IoT) data collection [5], mobile relays [6], and massive machine type communications [7]. Extensive research efforts focus on the two- or three-dimensional deployment of UAVs which are treated as stationary aerial BSs [8–10]. Another important branch focuses on the trajectory design which fully unleashes the potential of mobility to enhance performance [11]. To take fully advantage of limited resources, there are a lot of research aiming at optimizing the resource allocation such as power control and computing offloading. Moreover, the

deployment of multiple UAVs makes it ideal for meeting Adhoc demands with more flexible network architectures.

Despite all the promising benefits mentioned above, there are still many challenges to be overcome. First of all, though UAV helps to assist wireless communication, reacting as relay nodes limits the effectiveness of UAV networks, which still brings backhaul burden. Secondly, UAVs have limited endurance due to the constraint of load bearing. Thirdly, it is hard to track users with continuous mobility to maintain the high performance of optimization algorithms.

Content caching can be an effective approach to reduce the burden of backhaul link by storing popular files during off-peak periods [12]. To overcome the aforementioned challenges, we consider cache-enabled UAV networks, where multiple UAVs serve mobile users with preloaded content. By leveraging optimization and reinforcement learning method, we give the content placement and trajectory design to maximize the achievable sum throughput.

1.1. Related Work

(1) Cache-enabled UAV networks: in cache-enabled UAV networks, content requested from the users can be obtained via the local cache of UAVs, which bypasses the backhaul bottleneck and enhances the network capacity. The existing literature has studied many problems relating to time delay, throughput, and hit rate. Content placement [13, 14], transmission scheme, and trajectory design are studied in recent work to make performance gain. In [12], the authors focus on content-centric communication system and present a caching scheme which is divided into two stages. By jointly designing the trajectory, communication scheduling, and file caching policy, the authors make trade-off between the file caching and file retrieval cost. In [15], the authors optimize the deployment of UAVs and the content to cache to maximize the quality of experience of wireless devices. The authors in [16] study the content placement and virtual reality network to meet the delay requirement. The problem of trajectory design and content placement and delivery for UAV to vehicle link is studied in [17]. However, the movement of users and the energy-efficient control of UAV are mostly neglected. Moreover, most related work in urban scenario treats UAVs as stationary BS, which ignores the efficient control of UAVs

(2) Machine learning in UAV networks: machine learning, especially reinforcement learning (RL), has become a powerful tool to tackle wireless communication problems [18]. As hardware computing efficiency increases, machine learning can be used to deal with more complex problems including nonconvex optimization and multiagent regression. Interference management, trajectory design, power control, and energy harvesting problems are tackled with machine learning. The authors in [19] proposed a deep reinforcement learning-based algorithm to ensure fair coverage of ground terminals. The proposed deep deterministic policy gradient-based algorithm is extended to a crowd sensing scenario, where UAVs aid to collect data with the limited energy reserved [20]. In [21], Q-learning-based algorithm is used to solve the three-dimensional placement of UAVs to coverage ground users. In [22], the authors take the advantage of content-centric caching. UAV trajectory and content delivery are jointly optimized through actor-critic reinforcement learning-based algorithm to reduce the average request queue. A power transfer and data collection scheme is proposed in [23]. The problem of minimizing the overall packet loss is solved using deep reinforcement learning. However, there is little work on cache-enabled UAV networks by leveraging machine-learning approach. Machine learning can play a role in the wider field of UAV-assisted communication to meet the diversity of urban communication services

1.2. Our Contributions. In summary, the main contributions of the paper are as follows:

(1) We propose a scheme for cache-enabled UAV networks with mobile users in urban scenario. We formulate an optimization problem to maximize the sum achievable throughput under the storage capacity and energy constraints. The target function is multiobjective and nonconvex which is hard to solve using traditional optimization methods. Thus, we propose a deep reinforcement learning-based joint content placement and trajectory design algorithm (DRL-JCT) to solve this problem

(2) At the offline content placement stage, we formulate a link-based caching strategy to maximize the sum hit rate of users' file requirements. The objective function is nonlinear with binary variables, which is NP-hard. Caching scheme is obtained through approximation and convex optimization. The proper caching strategy obtains great gain in file hit rate by making trade-off between file popularity and diversity

(3) Considering the mobility of users and the constraint of UAV endurance, we propose a Double Deep Q-Network- (DDQN-) based online trajectory design algorithm to track real-time users. To the best of our knowledge, the mobility of users and energy-efficient UAV control have not been considered in most current research on cache-enabled UAV networks

(4) We compared our algorithm with several benchmark schemes. We demonstrate that the proposed algorithm shows a fast convergence and great gain. Numerical results shows that our algorithm achieves significant performance gain in terms of achievable throughput.

1.3. Organization. The rest of papers is organized as follows. In Section 2, we describe the system model and formulate the

Table 1: List of notations.

Notations	Descriptions
K	Number of UAVs
T	Total time slots
$Vmax$	Maximum speed of UAVs
$x_i^k(t)y_i^k(t)$	Coordinate of users
F	Total number of files
q_f^k	Caching probability
$P_{\text{LoS}}, P_{\text{NLoS}}$	LoS and NLoS probability
$r_i^k(t)$	Instantaneous rate of user
B	Total bandwidth of each UAV
$\mu_{\text{LoS}}, \mu_{\text{NLoS}}$	Additional path loss for LoS, NLoS
$P_0, P_1, P_2, v_0, U_{\text{Tip}}$	Propulsion power relevant parameters
$\mathcal{K}_u(i)$	UAV set with link to user i
U	Number of users
δ	Time slot length
H	Altitude of UAVs
γ	Zipf exponent
P_f	Content popularity
C^k	Storage capacity
θ_i^k	Elevation angle
N_0	Noise power spectral density
a, b	Environmental parameters (urban)
α	Path loss exponent
$E_{\max}$	Battery capacity
s_t, a_t, r_t	State, action, and reward in RL

problem of content placement and trajectory design. In Section 3, the efficient content placement policy is presented. In Section 4, we propose our RL-based trajectory design with roaming users. The simulation results are illustrated and analyzed in Section 5. The conclusions are presented in Section 6. In addition, the list of notations is shown in Table 1.

2. System Model and Problem Formulation

We consider the downlinks of cache-enabled UAV networks in urban environment as shown in Figure 1. Multiple UAVs are deployed as aerial BSs to cache files and serve the users of our target area. The UAVs connect with the core network by terrestrial BS. After the UAVs complete the content placement, the drones use online tracking algorithm to obtain achievable maximum service rate. There are a set $\mathcal{K}$ of K UAVs serving a set $\mathcal{U}$ of U mobile users. We take the fixed link scheme that the links are predetermined by random or signal-to-noise ratio- (SNR-) based user allocation algorithms. The network topology and the maximum users that a UAV can serve are predetermined. Each user is served by at least one UAV, and we denote $\mathcal{K}_u(i)$ as the set of UAVs with link to user i. We denote U_k as the total amount of users

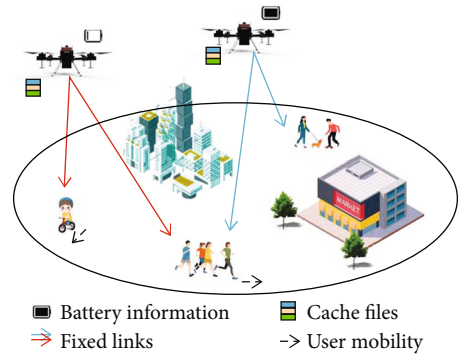


Figure 1: Illustration of cache-enabled UAV networks in urban environment. Multiple UAVs serve mobile users with fixed links. We maximize the achievable sum rate under storage capacity and energy constraints.

served by UAV k and u_i^k as the i-th user. Users affiliated to the same UAV are served simultaneously by frequency division multiple access (FDMA).

2.1. Mobility and Caching Model. We assume that the users are distributed as the Homogeneous Poisson Point Process (HPPP) [24]. The users may move continuously during the task period of UAVs. Mobility models can be random walk models or deterministic models. Considering the high mobility of UAVs, we divide the flying time horizon into T equal time slots with sufficiently small slot length δ, indexed by t. The UAVs are assumed to fly at a fixed altitude H with a speed not exceeding maximum speed $V_{\max}$. The two-dimensional coordinate of user at time slot t can be expressed as $[x_i^k(t), y_i^k(t)]^T$, while the horizontal coordinate of UAV k is $[x^k(t), y^k(t)]^T$. The position set of UAVs during the flying period is denoted as $\boldsymbol{x}$ and $\boldsymbol{y}$. Using time and trajectory discretization, the distance between UAV k to user u_i^k can be assumed to be a constant during one time slot t which is

$$d_i^k(t) = \sqrt{H^2 + \left[x^k(t) - x_i^k(t)\right]^2 + \left[y^k(t) - y_i^k(t)^2\right]}. \quad (1)$$

For the file database system, we assume that there are F contents with the same normalized size. The popularity which indicates the users' interest in a certain file is modeled using the Zipf law [25, 26]. The popularity p_f of content f, $f = \{1, 2, \cdots, F\}$ is modeled as

$$p_f = \frac{1/f^\gamma}{\sum_{j=1}^{F} 1/j^\gamma}, \quad (2)$$

where γ is the Zipf exponent describing the probability of content reuse. The larger the γ is, the more concentrated the files are.

Due to the fact that UAVs have limited storage capacity, we further assume that the capacity of UAV k is C^k and thus that each UAV can proactively cache no more than C^k popular contents. The caching probability of UAV k for

content f is denoted as q_f^k, and then we have the following constraint:

$$\sum_{f=1}^{F} q_f^k \le C^k \quad \forall k \epsilon \mathscr{K}. \tag{3}$$

The caching strategy set with $K * F$ binary variables is denoted as $\boldsymbol{q}$. Considering that the UAVs are not able to cache all required files, an appropriate content placement strategy is of vital importance. Proactive caching and proper content placement save transmission resource and improve the achievable sum coverage throughput.

2.2. Transmission Model. The downlinks between UAVs and users can be regarded as a LoS dominant air-to-ground channel. In urban environment, the LoS link may be occasionally blocked by obstacles like high buildings and towers. We use Probabilistic LoS Channel model by taking account of the shelter [27–29]. The LoS probability can be expressed as

$$P_{\mathrm{LoS}}\left(\theta_i^k\right) = \frac{1}{1 + a\ \exp\left(-b\left(\theta_i^k - a\right)\right)}, \tag{4}$$

where $\theta_i^k = \sin^{-1}(H/d_i^k)$ is the elevation angle UAV k and user u_i^k and a and b are parameters related to the environment. The probability of not having direct link to user u_i^k is thus given by $P_{\mathrm{NLoS}}(\theta_i^k) = 1 - P_{\mathrm{LoS}}(\theta_i^k)$. Intuitively, P_{LoS} increases as the elevation angle increases and approaches 1 as θ_i^k gets sufficiently large.

Then, the channel's power gain between UAV k and user u_i^k can be expressed as

$$g_i^k(t) = \left(\frac{4\pi f_c}{c}\right)^{-2} \left[d_i^k(t)\right]^{-\alpha} \left[P_{\mathrm{LoS}}\left(\theta_i^k\right)\mu_{\mathrm{LoS}} + P_{\mathrm{NLoS}}\left(\theta_i^k\right)\mu_{\mathrm{NLoS}}\right]^{-1}, \tag{5}$$

where f_c is the carrier frequency, c is the speed of light, α is the path loss exponent, and μ_{LoS} and μ_{NLoS} are the attenuation factors of LoS and NLoS links, respectively.

For simplicity, we assume that the bandwidth B for each UAV is equally allocated to the associated users; thus, the licensed spectrum for all the U^k user is $B_i^k = B/U^k$. Also, the maximum transmit power $P_{\max}$ for each UAV k is also uniformly allocated; thus, the transmit power allocated to user u_i^k is $P_i^k = P/U^k$. By calculating the received SINR, the achievable rate of user u_i^k at time slot t can be expressed in bit/s as

$$r_i^k(t) = B_i^k \log_2\left(1 + \frac{p_i^k g_i^k(t)}{\sigma^2}\right), \tag{6}$$

where $\sigma^2 = B_i^k N_0$ is the variance of Addictive White Gaussian Noise (AWGN) and N_0 is the noise power spectral density.

2.3. Energy Consumption Model. The energy consumption of a UAV-aided communication system generally consists of two parts: transmission energy, which is related to communication, and propulsion energy, which aims at supporting the movement of UAVs. Compared to the flight energy consumption, the communication-related power is small enough to be negligible. In this paper, we consider the rotary-wing UAV propulsion energy consumption model which depends on the instantaneous velocity in [30]. The propulsion power of UAV k with scalar velocity v is

$$e^k(v) = P_0\left(1 + \frac{3v^2}{U_{\mathrm{tip}}^2}\right) + P_1\left(\sqrt{1 + \frac{v^4}{4v_0^4}} - \frac{v^2}{2v_0^2}\right)^{1/2} + P_2 v^3, \tag{7}$$

where P_0 and P_1 are two constants representing the blade profile power and induced power when hovering, respectively. U_{tip} denotes the tip speed of the rotor blade, and P_2 and v_0 are parameters related to fuselage. We denote $E_{\max}$ as the onboard energy, and the total energy consumption should not exceed it for the sake of safe return or landing.

2.4. Problem Formulation. To investigate the benefits brought by cache-enabled UAVs, we optimize the following formulated problem to maximize the achievable sum service rate R_{sum}.

$$\max_{\boldsymbol{x},\boldsymbol{y},\boldsymbol{q}} R_{\mathrm{sum}} = \sum_{f=1}^{F}\sum_{i=1}^{U}\sum_{t=1}^{T} p_f r_{i,f}^k(t) \mathbb{1}\left\{\sum_{k \epsilon K_u(i)} q_f^k \ge 1\right\}, \tag{8a}$$

$$s.t \sum_{f=1}^{F} q_f^k \le \mathrm{C}^k, \tag{8b}$$

$$q_f^k \epsilon \{0, 1\} \quad \forall k, f, \tag{8c}$$

$$\sum_{t=1}^{T} e^k(t) \le E_{\max} \forall k, \tag{8d}$$

$$0 \le v^k(t) \le V_{\max} \forall k, t, \tag{8e}$$

where $\mathbb{1}\{\bullet\}$ is the indicator function and guarantees the file request from user i which can be responded by any of the neighboring UAVs. Equation (8b) represents the probability of content placement is binary variables. $r_{i,f}(t)$ is the service rate of UAVs associated with user i which is positively relevant to $r_i^k, k \in \mathscr{K}_u(i)$. Equation (8d) ensures that the sum propulsion consumption will not exceed the battery capacity. Equation (8e) is the constraint of maximum flying speed for UAV control.

From (8a) we can see that the achievable service rate is related to content placement strategy and instantaneous position of UAVs. A proper caching scheme makes tradeoff between cooperation gain and content diversity gain, thus increasing the hit rate of users' requirements. In real scenario, due to the mobility of users and energy constraint of UAVs, appropriate control policy of UAVs can enhance the endurance and throughput, which both improve the overall system service rate.

To solve the optimizing problem is challenging due to the nonconvex target function and restricted conditions. Any search-based algorithms will be of high computational complexity. To solve this problem, we proposed our algorithm DRL-JCT to jointly optimize content placement and online trajectory.

3. Offline Content Placement

In this section, we present our offline content placement strategy based on file popularity and existed links.

Considering that the users move continuously and randomly, it is hard to predict the real-time positions of users. Hence, it is difficult to predefine which content to cache according to user locations. A link-based algorithm is a good substitute to deal with offline content placement. We assume the channel between UAVs and users are fixed according to frequency spectrum and user-side information. Let l_{ik} denotes whether there is a link between user i and UAV k. Then, the responsible UAV set for serving user i is $\mathcal{K}_u(i) = \{l_{ik} = 1, i \in U\}$. When the initial locations of users are known, the links can be allocated according to SNR or throughput. When there is no user-side information, the links can be allocated randomly according to the numbers of UAV and the maximum users that a UAV can serve.

We use deterministic caching model which is commonly used in cache-enabled networks [31, 32]. To take the full advantages of caching capacity, we transform the constraint F of binary variables q_f^k to equality, which is $\sum_{f=1}^{F} q_f^k = C^k$. We aim to maximize the hit rate of users' file requirements, which is formed as

$$\max H(\boldsymbol{q}) = \sum_{i=1}^{U}\sum_{f=1}^{F} p_f \mathbb{1}\left\{\sum_{k\epsilon\mathcal{K}_u(i)} q_f^k \geq 1\right\}, \tag{9}$$

However, the hit rate is also related to the arrival rates of files which varies with the activity of users. We assume the arrival rates are normalized for simplicity. The indicated function which presents the ability of the adjacent UAVs to provide content f for user i can also be written as $\mathbb{1}\{\sum_{k\epsilon\mathcal{K}_u(i)} q_f^k \geq 1\} = 1 - \prod_{k\epsilon K_u(i)}(1 - q_f^k)$. The problem is a binary optimization problem with nonlinear objective funtion which is proved to be NP-hard. To tackle this problem, we introduce nonnegative slack variables μ_m and reformed the problem as

$$\min H'(\boldsymbol{q}) = -\left\{\sum_{i=1}^{U}\sum_{f=1}^{F} p_f\left[1 - \prod_{k\epsilon\mathcal{K}_u(i)}\left(1 - q_f^k\right)\right] + \mu_m\left(1 - q_f^k\right)q_f^k\right\}. \tag{10}$$

Note that the maximization problem (9) is identical to (10), and the introduced slack variables do not affect the optimal value. We then relax the variables q_f^k to closed interval [0,1] to form $H'(\widehat{\boldsymbol{q}})$ with continuous variables.

Lemma 1. *$H'(\widehat{\boldsymbol{q}})$ is convex [33] when the slack variables satisfies*

$$\mu_m > \frac{1}{2}\sum_{j\neq m}\left|\mathscr{H}'(\widehat{\boldsymbol{q}})_{mj}\right| \; m = 1, 2, \cdots, KF, \tag{11}$$

where $\mathscr{H}'(\widehat{\boldsymbol{q}})_{mj}$ is the mj-th term of the Hessian matrix of $H'(\widehat{\boldsymbol{q}})$.

Proof. We can calculate that

$$\frac{\partial^2 H'(\widehat{q}_m)}{\partial \widehat{q}_m^2} = \frac{\partial^2[-\mu_m(1 - q_m)q_m]}{\partial \widehat{q}_m^2} = 2\mu_m. \tag{12}$$

Then, the diagonal elements of the Hessian matrix is $2\mu_m$. According to the Gershgorin Theorem, the range of eigenvalues of a square matrix **A** satisfies

$$|\lambda - a_{ii}| \leq \sum_{j\neq i} a_{ji}. \tag{13}$$

Then, the lower bound of each eigenvalue of the Hessian is $2\mu_m - \sum_{j\neq m}|\mathscr{H}^i(\widehat{q})_{mj}|$. Given the constraint in (11), all the eigenvalues are nonnegative, which means the Hessian matrix is positive semidefinite. According to the properties of convex function, $H'(\widehat{\boldsymbol{q}})$ is convex.

By properly choosing slack variables, we can efficiently solve the convex problem by tools like CVX. Note that the solution contains the original problem (9) since it searches on a larger scale. Since the optimal solution may be fractional and we need a binary solution, we can approximate the optimal solution by the greedy method. Specifically, we choose one for solution bigger than 1/2 and inversely, zero.

It can be seen that the larger the overlapping part of the users, the more decentralized the content should be to satisfy multiple content requirements. When the number of the users that are served by a single UAV is larger, the optimal solution tends to be the most popular strategy. Ultimately, when all the users are covered by only one UAV, the product terms in (10) vanished, which makes the problem linear.

The link-based caching strategy is independent from user locations when the popularity of each content is identical so that the caching stage can be completed before taking off. However, the optimal achievable throughput relies on the relative position between UAVs and users. In the next section, we propose our online trajectory design to meet the need of mobile users.

4. Online Trajectory Design

In the actual scene, the users tend to move continuously which may lead to throughput reduction. Actually, there are no traditional solutions to track the users efficiently by calculating. Exhaustive search may cause big calculation time and high latency. Thus, the reinforcement learning is invoked to track real-time users. In this section, we first introduce

some preliminaries of Deep Reinforcement learning, and then we present our Double Deep Q-learning-based algorithm to maximize the sum throughput.

4.1. Deep Reinforcement Learning Background. Reinforcement learning contains basic elements including environment, agent, state, action, and reward. In reinforcement learning, an agent interacts with the environment with discrete decision epochs. Our training agent is selected as each UAV. The state can be set as all relevant parameters such as speed, location, and energy, and actions are chosen according to the current state. The process can be modeled as a Markov Decision Process (MDP) with a set, $\mathscr{M} = <\mathscr{S}, \mathscr{A}, \mathscr{R}, \Pr(s_{t+1} \mid s_t, a)$ where $\mathscr{S}, \mathscr{A}$, and $\mathscr{R}$ are the set of state, action, and reward, respectively. $\Pr(s_{t+1} \mid s_t, a_t)$ denotes the transmit probability set from s_t to s_{t+1} when action a_t is taken.

Q-learning is one of the simple algorithms of reinforcement learning used in UAV control. The basic idea for Q-learning is to maintain a table to record and maximize the long-term discounted cumulative reward.

$$\max C = \mathbb{E}^{\pi}\left(\sum_{t=1}^{\infty} \gamma_d^{t-1} r(s_{t+1} \mid s_t, a_t)\right), \tag{14}$$

where $\pi = \arg\max_{a_t \epsilon \mathscr{A}} Q(s_t, a_t)$ is the policy to choose action.

A good tip to choose the best action is to adopt the ϵ -greedy policy in order to explore the environment. γ_d is the discount factor for future state. Following the Bellman equation to find the best decision of MDP process, the Q table which is also known as value function is updated by

$$Q_{t+1}(s_t, a_t) = (1-\alpha) Q_t(s_t, a_t) + \alpha\left(r_t + \gamma_d \max_{a'} Q_t\left(s_{t+1}, a'\right)\right), \tag{15}$$

where α is a small positive fraction indicating learning rate. Q-learning algorithm is proved to be convergent. However, since this algorithm requires a value table for each action and state pair, when the state space get larger, the table gets extremely huge, which causes a curse of dimensionality [34] problem. Also, Q-learning is unable to deal with continuous space problem.

Combined Q-learning with neural network, Deep Q-Network (DQN) [35] can be seen as a "deep" version of Q-learning. DQN uses a neural network to estimate the huge Q table. The DQN is trained by minimizing the loss function:

$$L\left(\theta^Q\right) = \mathbb{E}\left[r_t + \gamma_d Q'\left(s_{t+1}, \pi(s_{t+1}) \mid \theta^{Q'}\right) - Q\left(s_t, a_t \mid \theta^Q\right)\right]^2, \tag{16}$$

where the first part $y_t = r_t + \gamma_d Q'(s_{t+1}, \pi(s_{t+1}) \mid \theta^{Q'})$ is the target value to reach and θ^Q is the weight vector of the DQN. The network updates θ^Q from the derivative $\nabla L(\theta^Q)$ by common methods like gradient decent and back propagation.

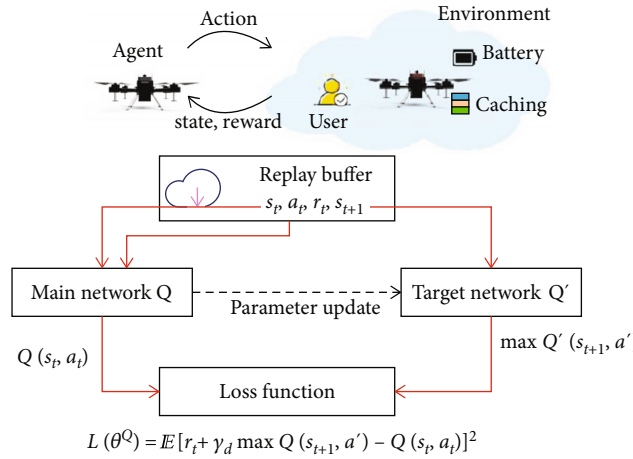


Figure 2: The architecture and basic idea of DQN.

Additionally, DQN adopts two techniques, experience replay and target network, to mitigate the impact of correlations between data. The experience replay buffer stores past training data and packs them in a batch. Experience replay is conducted to randomly choose a minibatch with the size of B_s from the experience replay buffer. Moreover, DQN uses target network with the same structure as the original neural network. The parameters of target network are updated using weights of original network with delay. The architecture and basic idea of DQN are shown in Figure 2.

4.2. DDQN for Trajectory Design. Based on the idea of DQN, our agent, state, action, and reward are defined as follows:

(1) Agent: our training agents are UAV $k, k = 1, 2, \cdots, K$

(2) State: for each training epoch t (the training epoch can be seen as time slot or training step), we define $s_t = [x^k(t), y^k(t), x_1^k(t), x_2^k(t), \cdots, x_{U_k}^k(t), E^k(t)]$. The state for each agent k consists of the dynamic position of UAV and users and also the current energy available. Thus, the agent can take actions according to its battery capacity and the current user-side information

(3) Action: the actions represent the flying velocity and direction of each UAV. The instantaneous speed can be discretized to several options and the upper bound $V_{\max}$. Also, the agent can choose to hover at one point. The UAV can fly to eight directions: forward, backward, left, right, northeast, northwest, southeast, and southwest

(4) Reward: the reward of epoch t is defined as:

$$r(t) = \frac{\sum_{i=1}^{U_k} r_i^k(t)}{e_i^k(t)}, \tag{17}$$

which is the current energy efficiency. Using this, the agent can make a trade-off between sum throughput and energy consumption, which further improves the endurance.

1: Initialize content placement $\boldsymbol{q}$.
2: Randomly initialize value function Q with weight $\boldsymbol{\theta}$.
3: Initialize target value function $Q^{'}$ with weight $\boldsymbol{\theta}^{-}$.
4: Initialize replay memory $\mathcal{D}$ to size N, replay buffer size to B_s.
5: **for** episode $m = 1, 2, \cdots, M$ **do**
6: Initialize environment and state to s_1.
7: **while** available energy > 0 **do**
8: **if**random $\leq \epsilon$ **then**
9: choose action $a_t = \arg\max_a Q(s_t, a; \boldsymbol{\theta})$.
10: **else**
11: randomly choose an action.
12: **end if**
13: Execute a_t and observe s_{t+1}, r_t.
14: store transition (s_j, a_j, r_j, s_{j+1}) in $\mathcal{D}$.
15: sample random minibatch (s_j, a_j, r_j, s_{j+1}) with size B_s from $\mathcal{D}$.
16: Calculate target value: $y_j = r_j + \gamma_d Q^{'}(s_{j+1}, \arg\max_a Q(s_{j+1}, a \mid \boldsymbol{\theta}) \mid \boldsymbol{\theta}^{-})$.
17: Loss function $L(\boldsymbol{\theta}^{Q}) = \sum_{j=1}^{B_s} [y_i - Q(s_j, a_j \mid \boldsymbol{\theta})]^2$
18: update $\boldsymbol{\theta}$ using $\nabla L(\boldsymbol{\theta}^{Q})$ by gradient decent.
19: Every B_{up} steps reset $\boldsymbol{\theta}^{-} = \boldsymbol{\theta}$.
20: **end while**
21: **end for**

Algorithm 1: Deep reinforcement learning-joint caching and trajectory design (DRL-JCT).

Table 2: Simulation parameters.

Notations	Descriptions	Value
δ	Time slot length	1 s
$V_{\max}$	Maximum speed of UAVs	40 m/s
H	Altitude of UAVs	100 m
N_0	Noise power spectral density	-174 dBm
B	Total bandwidth of each UAV	1 MHz
a, b	Environmental parameters(urban)	10, 0.15
μ_{LoS}	Additional path loss for LoS	2
μ_{NLoS}	Additional path loss for NLoS	100
α	Path loss exponent	2
P_0, P_1	Parameters of blade profile	580, 790
U_{Tip}	Tip speed of the rotor blade	200
P_2, v_0	Parameters related to fuselage	0.79, 7.2

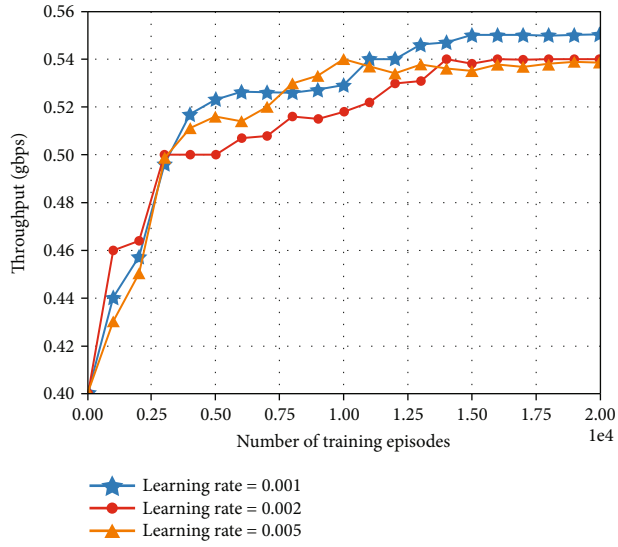


Figure 3: Convergence of DDQN algorithm with different learning rates versus the number of training episodes.

We also use the Double Deep Q-Network (DDQN) [36] which is an improved version of DQN. This architecture makes small changes at the action chosen but brings significant improvement. Traditional Deep Q-learning has the drawback of over estimating the value function. The Double Deep Q-network uses separate architecture for choosing action and estimating the value brought by actions, which eliminates the correlation between the two networks. Thus, the estimated value function is closer to the true value. Instead of choosing action according to the target net, DDQN finds the action corresponding to the maximum Q value according to the current network.

5. Simulation Results

We conducted extensive simulations to evaluate our proposed solution: DRL-JCT. In this section, we introduce our simulation settings at first and then present results and analysis.

5.1. Simulation Settings. Our experiments are performed with TensorFlow 1.0 and Python 3.7. In our simulation, we set the target area to be square with the size of 800 × 800 m. The simulation parameters are summarized in Table 2. We compared DRL-JCT with some commonly used baselines. For caching strategy, we choose the Uniform Distributed Caching (UDC)

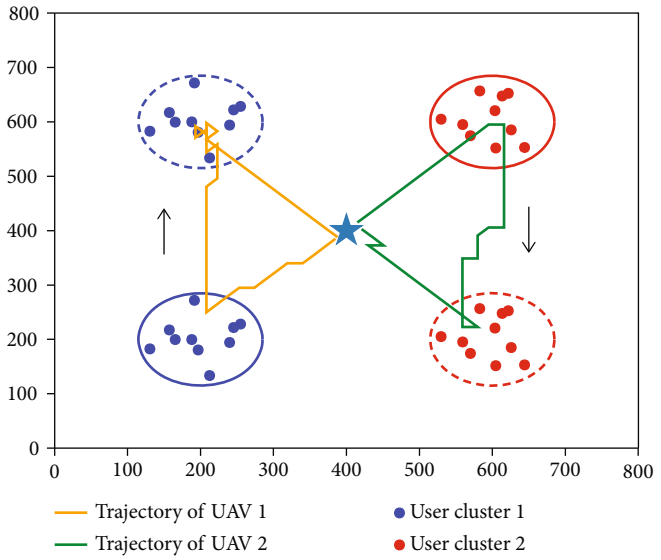


Figure 4: Trajectory of UAVs when users are roaming. The blue star stands for the starting point. The coordinate stands for the testing place.

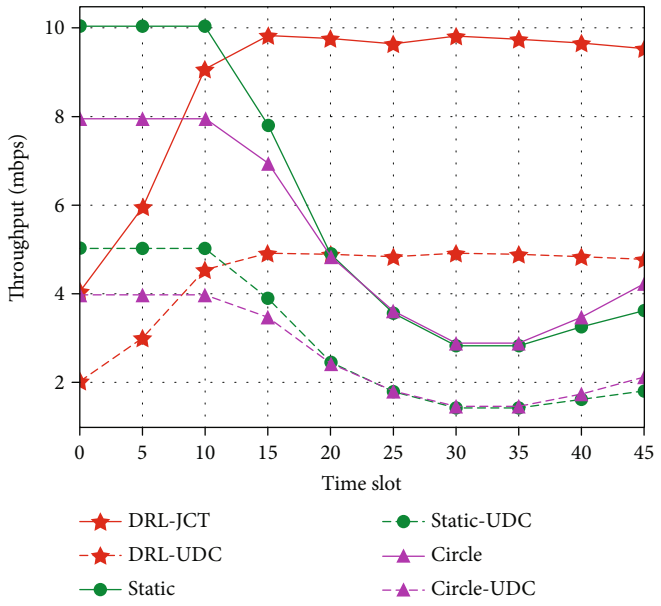


Figure 5: Comparing between DRL-JCT and other baselines over instantaneous throughput in movement scenario.

[37] and Most Popular Contents (MPC) as baselines. Using UDC, each UAV randomly selects C^k different files to store according to the uniform probability $q_f^k = C^k/F$. Using MPC, each UAV dependently caches C^k most popular contents.

For trajectory design, we choose 2 common baselines:

(1) Hovering: to avoid collision, the UAVs hover at the cluster centers of the users

(2) Circular flight trajectory: the UAVs periodically move around the cluster centers with radius of 100 m and constant velocity.

We find the appropriate hyperparameters in neural networks by a great number of experiments. We set the learning rate as 0.001, batch size $B_s = 32$, update iteration $B_{\text{up}} = 200$, memory size $N = 2000$, and discount factor $\gamma_d = 0.9$. We use a two-layer fully connected neural network to serve as the target and evaluate networks, and the number of neural units of the hidden layer is 120.

5.2. Results and Analysis. Figure 3 shows that the sum throughput with the number of training episodes. We set the total energy available as 46 kJ. RL contains many hyperparameters such as learning rate, discount factor, and memory size as mentioned above. Choosing appropriate hyperparameters can improve the performance of DDQN. Among them, we demonstrate the influence of learning rate. Learning rate can be neither too large nor small since large

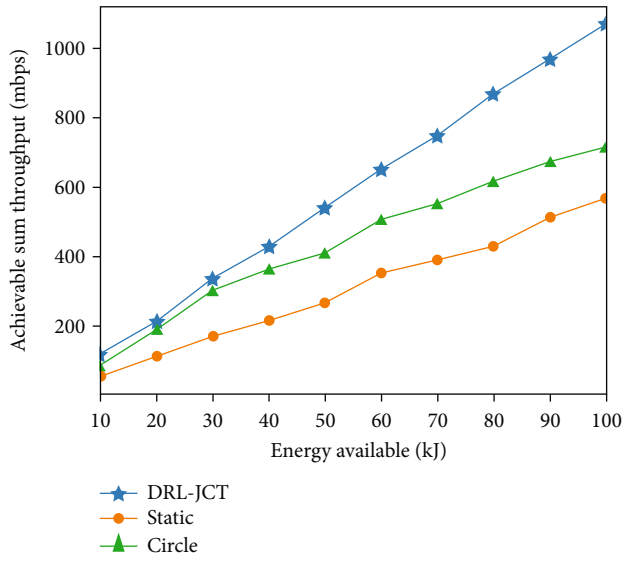


Figure 6: Achievable throughput versus the energy available.

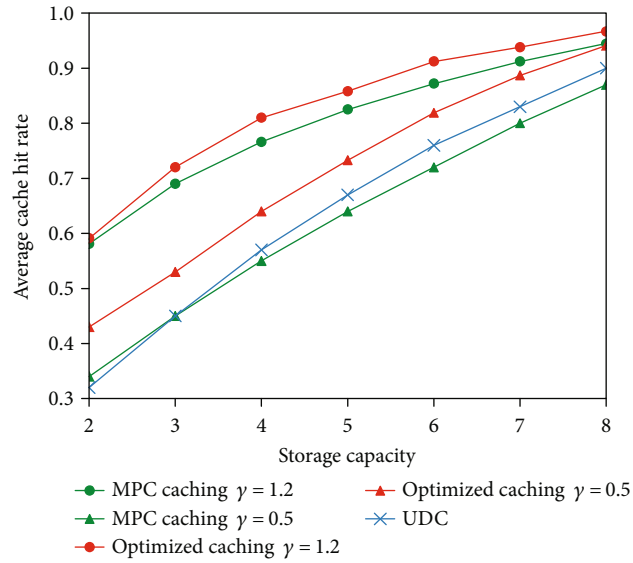


Figure 7: Average cache hit rate versus the storage capacity and the Zipf parameter γ.

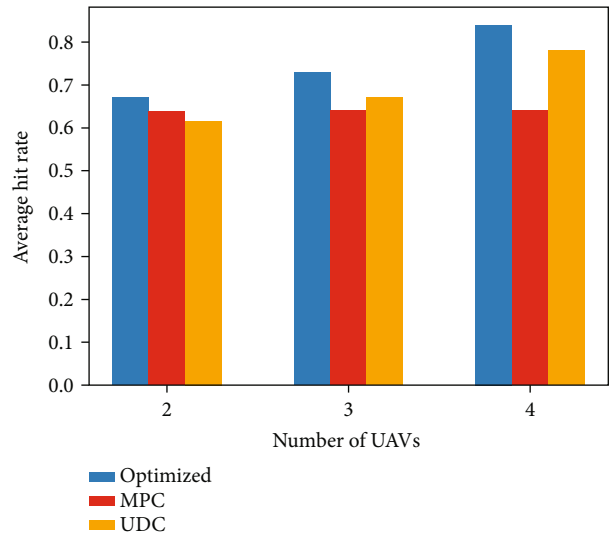


Figure 8: Average cache hit rate versus the number of UAVs.

learning rate causes fluctuation while small one tends to cause slow training speed. We choose different learning rates 0.001, 0.002, and 0.005 to train the neural network. We can see from the figure that 0.001 is the best choice in this scene since the throughput tends to increase smoothly to convergence with the training episodes. The other two learning rates can also reach a relatively small convergence throughput. Figure 4 plots the trajectories of UAVs derived from the proposed approach under the circumstance of moving users. There are totally $F = 10$ files, and the storage capacity $C^k = 5$. For simplification, we assume that there are 20 users which are clustered into 2 groups, demonstrated as red and blue solid circle. Two UAVs take off from the airport displayed as the blue star. The users follow deterministic moving model. The users of the first cluster move from (200, 200) at the 20th time slot to (400, 400) while the second cluster moves from (600, 600) to (600, 200). Since the users tend to at relatively slow speed, the UAVs may hover around to wait the users thus forming polyline.

Figure 5 characterizes the real-time throughput of one UAV derived from different algorithms in the movement scenario shown in Figure 4. We compared DRL-JCT with other baseline trajectory with optimized content caching and UDC separately.

We can observe that the throughput of DRL-JCT keeps increasing and maintains relatively constant since the algorithm can track the mobile users. The throughput of the static UAV is high at the beginning because it is deployed at the cluster center initially, and the throughput approximates the optimal coverage. But, as the users start to roam, neither the static deployment nor circle trajectory can meet the need of tracking mobile users. We can also analysis that DRL-JCT improves the performance of the system in terms of optimal content placement compared with the random caching strategy.

Figure 6 demonstrates the achievable sum throughput versus the available energy over different trajectories. We consider a more general situation where the users move following the random walk model and the UAVs carry optimized cache content. The moving direction and speed of each user are uniformly distributed. It can be observed that the sum throughput increases as the rise of endurance and DRL-JCT achieves the best performance over the benchmarks. Note that hovering at a still position is not a good choice in UAV control since it consumed more energy in unit time slot, and this is confirmed by the result of the figure.

Figure 7 shows the average cache hit rate versus the storage capacity and the Zipf parameter γ. In this scenario, 15 users are served by 3 UAVs. The total files $F = 10$. The links are established according to SNR threshold $\tau = 500$ kbps. Intuitionally, the hit rate increases with larger storage space. For different parameters, the optimized caching scheme outperforms the random caching and most populer content

caching. When γ gets larger, the files are more concentrated, which means the popular files are more often required, so that the MPC approximates the optimized solution.

Figure 8 compares the average cache hit rate with different numbers of serving UAVs. The UAVs are deployed separately using circle packing algorithm [38] to maximize the coverage range. The links are chosen based on throughput threshold $\tau = 500$ kbps. The number of users, the maximum users a UAV can serve, the total files, and the caching capacity are 15, 10, 10, and 5, respectively. It is clear that the hit rate increases with more serving UAVs. The behavior of MPC remains unchanged since it does not exploit the diversity of content. As the number of UAV gets larger, there are more overlapping areas and more stable links. The UDC scheme may perform well but waste the resource of air vehicles.

6. Conclusion

Cache-enabled UAV networks have been an appealing technology in wireless communication. However, the efficient use of UAVs meets greet challenge due to power and storage constraints. In this paper, we propose the DRL-JCT algorithm to jointly optimize the caching strategy and trajectory in cache-enabled UAV networks. We give the optimal caching scheme and trajectory design, respectively, using convex approximation and deep reinforcement learning approaches. Numerical results show that our algorithm has much better performance than the baselines in terms of achievable sum throughput. In our future work, we will consider a more complicated scenario of real-time transmission and content requirements in urban cache-enabled UAV networks.

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 is no conflict of interest regarding the publication of this paper.

Acknowledgments

This work is supported by the National Natural Sciences Foundation of China, under grant 61701136, and the project "The Verification Platform of Multi-tier Coverage Communication Network for Oceans" (LZC0020).

References

[1] Y. Zeng, J. Lyu, and R. Zhang, "Cellular-connected UAV: potential, challenges, and promising technologies," *IEEE Wireless Communications*, vol. 26, no. 1, pp. 120–127, 2019.

[2] X. Liu and X. Zhang, "NOMA-based resource allocation for cluster-based cognitive industrial internet of things," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 8, pp. 5379–5388, 2020.

[3] X. Liu, C. Sun, M. Zhou, C. Wu, B. Peng, and P. Li, "Reinforcement learning-based multislot double-threshold spectrum sensing with Bayesian fusion for industrial big spectrum data," *IEEE Transactions on Industrial Informatics*, pp. 1–10, 2020.

[4] J. Lyu, Y. Zeng, and R. Zhang, "UAV-aided offloading for cellular hotspot," *IEEE Transactions on Wireless Communications*, vol. 17, no. 6, pp. 3988–4001, 2018.

[5] M. Samir, S. Sharafeddine, C. M. Assi, T. M. Nguyen, and A. Ghrayeb, "UAV trajectory planning for data collection from time-constrained IoT devices," *IEEE Transactions on Wireless Communications*, vol. 19, no. 1, pp. 34–46, 2020.

[6] Q. Wu, Y. Zeng, and R. Zhang, "Joint trajectory and communication design for multi-UAV enabled wireless networks," *IEEE Transactions on Wireless Communications*, vol. 17, no. 3, pp. 2109–2121, 2018.

[7] M. N. Soorki, M. Mozaffari, W. Saad, M. H. Manshaei, and H. Saidi, "Resource allocation for machine-to-machine communications with unmanned aerial vehicles," in *2016 IEEE Globecom Workshops (GC Wkshps)*, pp. 1–6, Washington, DC, USA, December 2016.

[8] 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.

[9] J. Lyu, Y. Zeng, R. Zhang, and T. J. Lim, "Placement optimization of UAV-mounted mobile base stations," *IEEE Communications Letters*, vol. 21, no. 3, pp. 604–607, 2017.

[10] M. M. Azari, F. Rosas, K. Chen, and S. Pollin, "Optimal UAV positioning for terrestrial-aerial communication in presence of fading," in *2016 IEEE Global Communications Conference (GLOBECOM)*, pp. 1–7, Washington, DC, USA, December 2016.

[11] X. Xu, Y. Zeng, Y. L. Guan, and R. Zhang, "Overcoming endurance issue: UAV-enabled communications with proactive caching," *IEEE Journal on Selected Areas in Communications*, vol. 36, no. 6, pp. 1231–1244, 2018.

[12] C. Zhan, Y. Zeng, and R. Zhang, "Energy-efficient data collection in UAV enabled wireless sensor network," *IEEE Wireless Communications Letters*, vol. 7, no. 3, pp. 328–331, 2018.

[13] Y. Zhu, G. Zheng, L. Wang, K.-K. Wong, and L. Zhao, "Content placement in cache-enabled sub-6 gHz and millimeter-wave multi-antenna dense small cell networks," *IEEE Transactions on Wireless Communications*, vol. 17, no. 5, pp. 2843–2856, 2018.

[14] G. Qiao, S. Leng, S. Maharjan, Y. Zhang, and N. Ansari, "Deep reinforcement learning for cooperative content caching in vehicular edge computing and networks," *IEEE Internet of Things Journal*, vol. 7, no. 1, pp. 247–257, 2020.

[15] M. Chen, M. Mozaffari, W. Saad, C. Yin, M. Debbah, and C. S. Hong, "Caching in the sky: proactive deployment of cache-enabled unmanned aerial vehicles for optimized quality-of-experience," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 5, pp. 1046–1061, 2017.

[16] M. Chen, W. Saad, and C. Yin, "Echo-liquid state deep learning for 360° content transmission and caching in wireless vr networks with cellular connected UAVs," *IEEE Transactions on Communications*, vol. 67, no. 9, pp. 6386–6400, 2019.

[17] H. Wu, J. Chen, F. Lyu, L. Wang, and X. Shen, "Joint caching and trajectory design for cache-enabled UAV in vehicular networks," in *2019 11th International Conference on Wireless*

Communications and Signal Processing (WCSP), pp. 1–6, Xi'an, China, China, October 2019.

[18] N. C. Luong, D. T. Hoang, S. Gong et al., "Applications of deep reinforcement learning in communications and networking: a survey," *IEEE Communications Surveys Tutorials*, vol. 21, no. 4, pp. 3133–3174, 2019.

[19] C. H. Liu, Z. Chen, J. Tang, J. Xu, and C. Piao, "Energy-efficient UAV control for effective and fair communication coverage: a deep reinforcement learning approach," *IEEE Journal on Selected Areas in Communications*, vol. 36, no. 9, pp. 2059–2070, 2018.

[20] C. H. Liu, Z. Chen, and Y. Zhan, "Energy-efficient distributed mobile crowd sensing: a deep learning approach," *IEEE Journal on Selected Areas in Communications*, vol. 37, no. 6, pp. 1262–1276, 2019.

[21] X. Liu, Y. Liu, and Y. Chen, "Reinforcement learning in multiple-UAV networks: deployment and movement design," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 8, pp. 8036–8049, 2019.

[22] S. Chai and V. K. N. Lau, "Online trajectory and radio resource optimization of cache-enabled UAV wireless networks with content and energy recharging," *IEEE Transactions on Signal Processing*, vol. 68, pp. 1286–1299, 2020.

[23] K. Li, W. Ni, E. Tovar, and A. Jamalipour, "On-board deep Q-network for UAV-assisted online power transfer and data collection," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 12, pp. 12215–12226, 2019.

[24] E. Bastug, M. Bennis, and M. Debbah, "Cache-enabled small cell networks: modeling and tradeoffs," in *2014 11th International Symposium on Wireless Communications Systems (ISWCS)*, pp. 649–653, Barcelona, Spain, August 2014.

[25] J. Song, H. Song, and W. Choi, "Optimal content placement for wireless femto-caching network," *IEEE Transactions on Wireless Communications*, vol. 16, no. 7, pp. 4433–4444, 2017.

[26] L. Breslau, P. Cao, L. Fan, G. Phillips, and S. Shenker, "Web caching and Zipf-like distributions: evidence and implications," in *IEEE INFOCOM '99. Conference on Computer Communications. Proceedings. Eighteenth Annual Joint Conference of the IEEE Computer and Communications Societies. The Future is Now (Cat. No.99CH36320)*, pp. 126–134, New York, NY, USA, USA, March1999.

[27] Q. Feng, E. K. Tameh, A. R. Nix, and J. McGeehan, "Wlcp2-06: modelling the likelihood of line-of-sight for air-to-ground radio propagation in urban environments," in *IEEE Globecom 2006*, pp. 1–5, San Francisco, CA, USA, December 2006.

[28] A. Al-Hourani, S. Kandeepan, and A. Jamalipour, "Modeling air-to-ground path loss for low altitude platforms in urban environments," in *2014 IEEE Global Communications Conference*, pp. 2898–2904, Austin, TX, USA, December 2014.

[29] A. Al-Hourani, S. Kandeepan, and S. Lardner, "Optimal lap altitude for maximum coverage," *IEEE Wireless Communications Letters*, vol. 3, no. 6, pp. 569–572, 2014.

[30] Y. Zeng, J. Xu, and R. Zhang, "Energy minimization for wireless communication with rotary-wing UAV," *IEEE Transactions on Wireless Communications*, vol. 18, no. 4, pp. 2329–2345, 2019.

[31] S. Krishnendu, B. N. Bharath, and V. Bhatia, "Cache enabled cellular network: algorithm for cache placement and guarantees," *IEEE Wireless Communications Letters*, vol. 8, no. 6, pp. 1550–1554, 2019.

[32] J. Yao and N. Ansari, "Joint content placement and storage allocation in c-rans for IoT sensing service," *IEEE Internet of Things Journal*, vol. 6, no. 1, pp. 1060–1067, 2019.

[33] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, Cambridge, MA, USA, 2013.

[34] Y. Duan, X. Chen, R. Houthooft, J. Schulman, and P. Abbeel, "Benchmarking deep reinforcement learning for continuous control," in *International Conference on Machine Learning*, pp. 1329–1338, New York, NY, USA, 2016.

[35] V. Mnih, K. Kavukcuoglu, D. Silver et al., "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, pp. 529–533, 2015.

[36] H. Van Hasselt, A. Guez, and D. Silver, "Deep reinforcement learning with double q-learning," in *Thirtieth AAAI conference on artificial intelligence*, pp. 2094–2100, Phoenix, AZ, USA, 2016.

[37] S. Tamoorulhassan, M. Bennis, P. H. J. Nardelli, and M. Latvaaho, "Modeling and analysis of content caching in wireless small cell networks," in *2015 International Symposium on Wireless Communication Systems (ISWCS)*, pp. 765–769, Brussels, Belgium, August 2015.

[38] Z. Gáspár and T. Tarnai, "Upper bound of density for packing of equal circles in special domains in the plane," *Periodica Polytechnica Civil Engineering*, vol. 44, no. 1, 2000.

Hindawi
Wireless Communications and Mobile Computing
Volume 2020, Article ID 8892838, 9 pages
https://doi.org/10.1155/2020/8892838

Research Article

Efficient and Privacy-Preserving Outsourcing of 2D-DCT and 2D-IDCT

Dezhi An, Shengcai Zhang, Jun Lu, and Yan Li

School of Cyber Security, Gansu University of Political Science and Law, Lanzhou 730070, China

Correspondence should be addressed to Dezhi An; adz6199@gsli.edu.cn

Received 12 May 2020; Revised 2 June 2020; Accepted 16 June 2020; Published 27 July 2020

Academic Editor: Wei Wang

As a subset of discrete Fourier transform (DFT), discrete cosine transform (DCT), especially two-dimensional discrete cosine transform (2D-DCT), is an important mathematical tool for digital signal processing. However, the computational complexity of 2D-DCT is quite high, which makes it impossible to meet the requirements in some signal processing fields with large signal sizes. In addition, to optimize the 2D-DCT algorithm itself, seeking help from a cloud platform is considered to be an excellent alternative to dramatically speeding up 2D-DCT operations. Still, there are three key challenges in cloud computing outsourcing that need to be addressed, including protecting the privacy of input and output data, ensuring the correctness of the returned results, and ensuring adequate local cost savings. In this paper, we explore the design of a practical outsourcing protocol for 2D-DCT and 2D-IDCT, which well solves the above three challenges. Both theoretical analysis and simulation experiment results not only confirm the feasibility of the proposed protocol but also show its outstanding performance in efficiency.

1. Introduction

DFT is a common tool for frequency domain analysis of discrete signals and systems, but it is inconvenient to process image and voice data due to the need for complex domain operations. To solve this problem, a real domain transform, called DCT, is constructed based on DFT by preserving only the cosine term in the Fourier series. In addition to the general orthogonal transform properties of DCT, the basis vector of its transform matrix can well describe the relevant features of image signals and human voice signals. Therefore, DCT is considered to be a quasioptimal tool for transforming image and voice signals and is widely used in various fields such as media compression [1]–[3], digital watermarking [4]–[6], and wireless communication [7, 8]. 2D-DCT can directly transform two-dimensional data, so it is quite suitable for the analysis and processing of two-dimensional signals, such as static images. For example, 8×8 2D-DFT is adopted as a standard by the JPEG still image compression algorithm [9].

The computational complexity of 2D-DCT and 2D-IDCT is dominated by computing the product of three matrices. Such high-computational overhead makes it impossible to perform high-efficiency processing on signals when the local computing equipment is insufficient or the signals are large in size. Wireless camera sensor networks, for example, are widely used in habitat monitoring, target detection, and espionage [10, 11], where nodes typically use processors with low cost, micro power consumption, and poor performance. In contrast, these nodes are often burdened with complex image processing tasks, a considerable portion of which require the participation of 2D-DCT or 2D-IDCT.

In order to improve the efficiency of the transformation, the usual method is to divide the signal into blocks, then perform 2D-DCT or 2D-IDCT operations in each block, and finally merge the blocks. Nevertheless, the method of using blocks is not a once and for all solution, because smaller blocks are necessary for greater efficiency, but too small blocks will lead to serious block effects. In addition, various fast 2D-DCT and 2D-IDCT algorithms have been proposed [12]–[14], which can reduce the computational complexity by more than half on the basis of using a block method.

We emphasize that in addition to efficiency optimization from an algorithmic perspective, requesting computing assistance from the cloud platform is also an excellent alternative.

For example, the image processing tasks of nodes in a wireless sensor network can be delegated to a public cloud server. The cloud platform has a large amount of hardware and software resources, and the temporary use rights of partial computing resources it owns are transferred to clients by way of fee lease. The process by which clients rent cloud resources to help them complete their own computing is called cloud outsourcing. On the one hand, the client can achieve significant computing overhead savings and efficiency improvement through computing outsourcing, so as to quickly complete high-complexity computing tasks. On the other hand, the client can significantly reduce costs by eliminating the expense of purchasing and maintaining large amounts of computing equipment. At the same time, cloud platforms can also reap considerable economic benefits from resource rentals. Therefore, computing outsourcing is considered a win-win move for both the client and the cloud.

However, computing outsourcing between the client and the cloud faces three key challenges. First, we must protect the client's privacy from being stolen. The client's input and output data contains private information, which may be personally identifiable information, trade secrets, or core technical parameters. Since the cloud is untrusted, we must protect the plaintext of the input and output data from being obtained by the cloud. Meanwhile, the high concentration of information makes the cloud platform vulnerable, which is also a possible way to leak private information. Second, we must verify the results returned by the cloud. The cloud platform is run for profit, and a malicious cloud will deliberately return random error results to the client to extract cost savings. Even if the cloud is honest, calculation errors may occur due to software bugs or hardware errors. Therefore, it is necessary to design an efficient verification algorithm to strictly control the correctness of the returned results. Third, we must ensure that the client can realize significant savings from computing outsourcing. In other words, the total complexity of decryption, encryption, and result verification algorithms must be far less than that of solving the original problem directly; otherwise, there is no need for the client to outsource the computing task. In general, a qualified computing outsourcing protocol must be secure, verifiable, and efficient.

In this paper, a protocol capable of solving the above three challenges is designed for the outsourcing of 2D-DCT and 2D-IDCT. Before giving the protocol, we discuss two possible designs in an exploratory way. Multiround communication and block encryption severely damage the efficiency of both designs. In response to the problems in these two designs, we propose the formal outsourcing protocol, which greatly reduces the client's communication overhead and key management overhead. In the proposed protocol, the original two-dimensional signal matrix is integrally encrypted without affecting the block flexibility of the cloud, i.e., the cloud can still perform 2D-DCT or 2D-IDCT of any block size according to the client's will. This is an important point that affects the versatility of the protocol, because different application scenarios may require the operation of 2D-DCT or 2D-IDCT with different block sizes. Subsequently, we carried out theoretical analysis to confirm that the proposed protocol meets the requirements of security, verifiability, and efficiency. Finally, the simulation results show that the proposed protocol is not only better than block 2D-DCT and 2D-IDCT in efficiency but also faster than the corresponding fast 2D-DCT and 2D-IDCT algorithms.

To summarize, our main contributions include the following:

(i) To the best of our knowledge, we are the first to propose such an outsourcing protocol for 2D-DCT and 2D-IDCT, and we are also the first to accelerate the operational efficiency of 2D-DCT and 2D-IDCT from the perspective of cloud computing outsourcing

(ii) Through carrying out theoretical analysis and simulation experiments, it is shown that the proposed protocol handles the three challenges faced in computing outsourcing well

(iii) In the proposed protocol, only a single round of communication is required and the key management work is simple, which is suitable for the outsourcing of 2D-DCT and 2D-IDCT with arbitrary block size

The rest of this paper is organized as follows. Section 2 describes the related work, and Section 3 gives the problem statement. 2D-DCT and 2D-IDCT are briefly introduced in Section 4. Section 5 presents the possible designs and the formal outsourcing protocol. Section 6 provides theoretical analysis, followed by experiments in Section 7. Finally, Section 8 concludes this paper.

2. Related Work

There are two ways to implement outsourcing for complexity calculations. On the one hand, the theoretical cryptography community considers designing a universal design that covers all problems, i.e., any outsourcing of computing can be realized through this design. The basic approach to achieve this goal is to use some sophisticated basic cryptographic tools, such as Yao's garbled circuits [15] and Gentry's fully homomorphic encryption (FHE) schemes [16]. In these designs [17]–[20], the original problem is converted into a Boolean circuit over {0, 1}, and then, the client encrypts the converted problem using a FHE algorithm and sends the encrypted problem to the cloud, who solves the problem homomorphically and returns the result. Finally, the client decrypts the result and checks its correctness. However, such designs are far from practical applications because of the extremely high complexity of FHE operations and the pessimistic circuit sizes.

On the other hand, the security engineering community focuses on designing different outsourcing protocols to deal with different practical problems. Under this idea, the encryption of the original problem is generally achieved through some ingenious data conversion. Meanwhile, in order to face practical applications, the efficiency of the protocol is given special consideration to ensure that the client

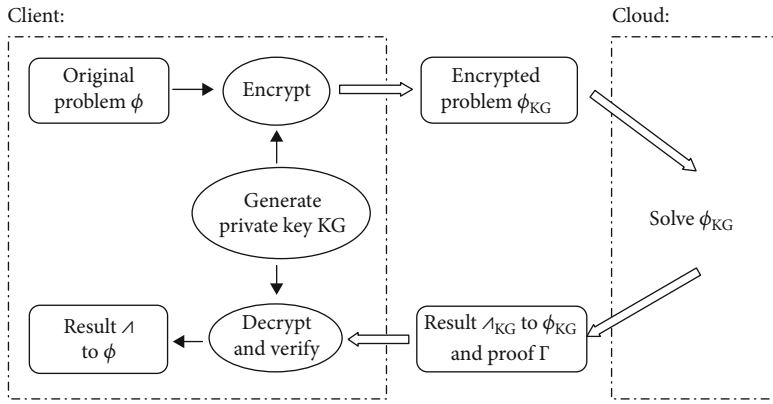


Figure 1: System model.

can acquire considerable cost savings from outsourcing. According to this trend, a large number of practical outsourcing protocols have been proposed. Among these protocols, there are outsourcing of basic mathematical calculations. Lei et al. proposed an outsourcing protocol for matrix inversion [21], Chen et al. proposed an outsourcing protocol for linear regression [22], and Li et al. proposed an outsourcing protocol for ID3 decision tree [23]. Several different outsourcing protocols for linear equations were presented in [24–26]. In addition, some outsourcing protocols are directly oriented to practical engineering application scenarios. An outsourcing protocol for Markowitz portfolio was proposed by Zhang et al. in [27], an outsourcing protocol for data classification was proposed by Li et al. in [28], an outsourcing protocol for the storage and statistics of smart meter data was proposed by Zhang et al. in [29], and an outsourcing protocol for biometric identification was proposed by Zhu et al. in [30]. Moreover, Li et al. proposed an optimal GPU-accelerated multimedia processing service pricing strategy in [31].

By using a homomorphic cryptography algorithm, some digital signal processing operations are transferred to the ciphertext domain to protect the privacy of the signal. Bianchi et al. implemented ciphertext domain operations on DFT and DCT in [32, 33], respectively. Zheng and Huang proposed a method for implementing discrete wavelet transform (DWT) and multiresolution analysis (MRA) in the homomorphic ciphertext domain [34]. Pedrouzo-Ulloa et al. realized secure number theoretic transform (NTT) in a distrustful environment [35]. Recently, Han et al. improved the homomorphic DFT with batch homomorphic encryption [36]. However, because of the high complexity of homomorphic operations, outsourcing with these solutions cannot meet the requirements of efficiency, i.e., the client fails to acquire the required savings. By contrast, data processing tools can also be efficiently outsourced, mainly to speed up signal processing. For example, Xiao et al. realized efficient and secure outsourcing of DFT, IDFT, and circular convolution [37], and Zhang et al. designed two outsourcing protocols for compressed sensing and sparse robustness decoding service in [38, 39], respectively. Nevertheless, the efficient outsourcing of 2D-DCT and 2D-IDCT has not been designed by predecessors.

3. Problem Statement

3.1. System Model. As illustrated in Figure 1, the system consists of two entities: the client and the cloud. Their roles are elaborated as follows.

(1) *Client*, The client has a complex problem Φ that needs to be solved. In this paper, the problem Φ is considered to be the 2D-DCT or 2D-IDCT operation. In order to acquire computing overhead savings, the client plans to outsource the solution of Φ to the cloud platform. To achieve the purpose of privacy protection, the client firstly uses the locally generated private key KG to encrypt the original problem Φ and generate the encrypted problem Φ_{KG}. Subsequently, the user sends Φ_{KG} to the cloud, who solves Φ_{KG} to obtain the result Λ_{KG} and a proof Γ used for verification and returns Λ_{KG} and Γ to the client. After the client receives Λ_{KG}, the private key KG is used to decrypt Λ_{KG} to obtain the result Λ of problem Φ. Finally, the client uses Γ to verify the correctness of Λ. If the verification passes, the client accepts Λ; otherwise, the client rejects Λ.

(2) *Cloud*. The cloud has a large number of hardware and software resources used for computing. It charges the rental fee of the client and provides computing support services for the client with part of its own resources. After receiving the encrypted problem Φ_{KG}, the cloud solves it and returns Λ_{KG} and Γ.

3.2. Threat Model and Design Goals. The threat mainly comes from the cloud's noncredibility, and here, we assume that the cloud is malicious. On the one hand, the cloud tries to obtain plaintext about problem Φ and result Λ. On the other hand, the cloud tries to save resources by returning a random erroneous result in the expectation that it will not be discovered by the client. In this case, we summarize the design goals of an efficient and privacy-preserving outsourcing protocol as follows.

(1) *Correctness*. The client must be able to acquire the correct answer if both the client and the cloud follow the protocol carefully.

(2) *Privacy*. The cloud cannot steal any private information of the client from the input and output data.

(3) *Soundness*. The client must be able to verify the correctness of the returned results.

(4) *Efficiency*. The client must be able to realize significant savings from computing outsourcing.

3.3. Framework. Syntactically, an outsourcing protocol includes the following five algorithms.

(1) *KeyGen* (1^λ). On input of a security parameter λ, the client uses this algorithm to generate a private key KG.

(2) *ProbEnc* (Φ;*KG*). On input of the original problem Φ and the key KG, the client uses this algorithm to encrypt Φ to generate the encrypted problem Φ_{KG} and then sends Φ_{KG} to the cloud.

(3) *ProbSolve* (Φ_{KG}). On input of the encrypted problem Φ_{KG}, the cloud uses this algorithm to solve Φ_{KG}. Then, the cloud returns the result Λ_{KG} back, together with a proof Γ.

(4) *ResultDec* (Λ_{KG};*KG*). On input of the result Λ_{KG} of Φ_{KG} and the key KG, the client uses the algorithm to decrypt Λ_{KG} to obtain the result Λ of the original problem Φ.

(5) *ResultVerify* (Λ;Γ). On input of the result Λ of Φ and the proof Γ, the client uses the algorithm to verify the correctness of Λ.

4. 2D-DCT and 2D-IDCT

4.1. 2D-DCT. Suppose that the original two-dimensional signal can be represented by a matrix $f \in R^{N\times N}$. In practice, if the original signal is not square, the transformation is usually done after the complement, and the original signal is subsequently obtained by removing the complement after the reconstruction. We consider performing block 2D-DCT of $K \times K$ on the matrix f

, where N is divisible by K.

First, the matrix f needs to be partitioned as

$$f = \begin{pmatrix} f(1,1) & f(1,2) & \cdots & f(1,N) \\ f(2,1) & f(2,2) & \cdots & f(2,N) \\ \cdots & \cdots & \cdots & \cdots \\ f(N,1) & f(N,2) & \cdots & f(N,N) \end{pmatrix} = \begin{pmatrix} f_{1,1} & f_{1,2} & \cdots & f_{1,N/K} \\ f_{2,1} & f_{2,2} & \cdots & f_{2,N/K} \\ \cdots & \cdots & \cdots & \cdots \\ f_{N/K,1} & f_{N/K,2} & \cdots & f_{N/K,N/K} \end{pmatrix}, \tag{1}$$

where $f(i,j), i,j \in [1,N]$ and $f_{i,j} \in R^{K\times K}, i,j \in [1,N/K]$ are the elements and the blocks of f, respectively, and the relation between them is

$$f_{i,j} = \begin{pmatrix} f(Ki-K+1, Kj-K+1) & \cdots & f(Ki-K+1, Kj) \\ \cdots & \cdots & \cdots \\ f(Ki, Kj-K+1) & \cdots & f(Ki, Kj) \end{pmatrix}. \tag{2}$$

Subsequently, an orthogonal matrix $A \in R^{K\times K}$ is calculated by

$$A(i,j) = c(i)\cos\left[\frac{(j+0.5)\pi}{N} i\right], \tag{3}$$

where $c(i)$ is defined by

$$c(i) = \begin{cases} \sqrt{\dfrac{1}{N}}, & i = 0, \\ \sqrt{\dfrac{2}{N}}, & i \neq 0. \end{cases} \tag{4}$$

Finally, the operation result $F \in R^{N\times N}$ of $K \times K$ 2D-DCT on f can be expressed as

$$F = \begin{pmatrix} A\cdot f_{1,1}\cdot A^T & A\cdot f_{1,2}\cdot A^T & \cdots & A\cdot f_{1,N/K}\cdot A^T \\ A\cdot f_{2,1}\cdot A^T & A\cdot f_{2,2}\cdot A^T & \cdots & A\cdot f_{2,N/K}\cdot A^T \\ \cdots & \cdots & \cdots & \cdots \\ A\cdot f_{N/K,1}\cdot A^T & A\cdot f_{N/K,2}\cdot A^T & \cdots & A\cdot f_{N/K,N/K}\cdot A^T \end{pmatrix}. \tag{5}$$

4.2. 2D-IDCT. 2D-IDCT is the inverse process of 2D-DCT. Thus, similarly, the 2D-IDCT operation is represented as

$$f = \begin{pmatrix} A^T\cdot F_{1,1}\cdot A & A^T\cdot F_{1,2}\cdot A & \cdots & A^T\cdot F_{1,N/K}\cdot A \\ A^T\cdot F_{2,1}\cdot A & A^T\cdot F_{2,2}\cdot A & \cdots & A^T\cdot F_{2,N/K}\cdot A \\ \cdots & \cdots & \cdots & \cdots \\ A^T\cdot F_{N/K,1}\cdot A & A^T\cdot F_{N/K,2}\cdot A & \cdots & A^T\cdot F_{N/K,N/K}\cdot A \end{pmatrix}. \tag{6}$$

5. Protocol Construction

5.1. Attempt One. Since the efficiency of 2D-DCT is mainly limited by the calculation of three matrix multiplication, we naturally think that the key of design is to realize the outsourcing of $A\cdot f_{i,j}\cdot A^T$, where $i,j \in [1,N/K]$. Several different outsourcing protocols for the multiplication of two matrices have been proposed [40]–[42]. Therefore, a direct assumption is to first use these protocols to outsource $A\cdot f_{i,j}$ to obtain the product S and then outsource $S\cdot A^T$, thereby acquiring calculation result of $A\cdot f_{i,j}\cdot A^T$.

Apart from other flaws of the protocol, the most obvious drawback of the protocol is the need for two rounds of communication between the cloud and the client. This will bring not only multiplied communication overhead to the client but also multiplied computing overhead for the client, because the client needs to perform independent encryption, decryption, and result verification operations for each round of outsourcing. As a result, the existence of two-round communication will make it difficult for the outsourcing protocol to meet the requirements of high efficiency.

5.2. Attempt Two. We attempt to implement an efficient outsourcing protocol that required only one round of communication by using the privacy-preserving matrix multiplication as adopted in [40, 43, 44]. First, to protect the privacy of input and output data, the private key matrices $P_1, P_2, \cdots, P_{4N^2/K^2} \in R^{K\times K}$ are locally generated by the client as follows:

$$P_n(i,j) = p_i \delta_{\pi_n(i),j}, \quad n \in \left[1, \frac{4N^2}{K^2}\right], \tag{7}$$

where p_i is a random number and $\pi_n(\cdot)$ is a permutation function that maps an original index i to its permuted index. Besides, $\delta_{i,j}$ is the Kronecker delta function given by

$$\delta_{i,j} = \begin{cases} 1, & i = j, \\ 0, & i \neq j. \end{cases} \tag{8}$$

We emphasize that the inverse of P_n can be accessed simply by

$$P_n^{-1}(i,j) = \frac{1}{P_i} \delta_{\pi_n^{-1}(i),j}. \tag{9}$$

Afterwards, for any block $A \cdot f_{i,j} \cdot A^T$ of the matrix F in (5), the client encrypts the matrices A, $f_{i,j}$, and A^T as

$$\begin{cases} E(A) = P_{(4N/K)(i-1)+4j-3} \cdot A \cdot P^{-1}_{(4N/K)(i-1)+4j-2}, \\ E\left(f_{i,j}\right) = P_{(4N/K)(i-1)+4j-2} \cdot f_{i,j} \cdot P_{(4N/K)(i-1)+4j-1}, \\ E\left(A^T\right) = P^{-1}_{(4N/K)(i-1)+4j-1} \cdot A^T \cdot P_{(4N/K)(i-1)+4j}. \end{cases} \tag{10}$$

The encrypted matrices $E(A)$, $E(f_{i,j})$, and $E(A^T)$ are then sent to the cloud, who perform

$$\begin{aligned} E\left(F_{i,j}\right) &= E(A) \cdot E\left(f_{i,j}\right) \cdot E\left(A^T\right) \\ &= P_{(4N/K)(i-1)+4j-3} \cdot F_{i,j} \cdot P_{(4N/K)(i-1)+4j}. \end{aligned} \tag{11}$$

Finally, to decrypt the returned $E(F_{i,j})$, the client only needs to calculate

$$F_{i,j} = P^{-1}_{(4N/K)(i-1)+4j-3} \cdot E\left(F_{i,j}\right) \cdot P^{-1}_{(4N/K)(i-1)+4j}. \tag{12}$$

This design avoids the trouble of multiround communication existing in Attempt One and seems to be feasible. However, this is not the case, because the complex key management severely slows down the efficiency of the protocol. In Attempt Two, for an original data matrix with dimension $N \times N$, the client needs to generate $4N^2/K^2$ mutually independent key matrices with size $K \times K$ for encryption. Meanwhile, $2N^2/K^2$ key matrices in them need to be saved in order to be invoked during decryption. Besides, we emphasize that in 2D-DCT and 2D-IDCT operations, only f and F are private in the input and output data, while A and A^T are public and can be generated independently by anyone. Therefore, an efficient protocol should manage to encrypt only f or F on input.

5.3. Formal Protocol

5.3.1. Outsourcing Protocol for 2D-DCT. We now present the formal protocol, which solves the flaws of Attempt One and Attempt Two well, i.e., not only requires only one round of communication but also greatly simplifies the key management workload.

First, in terms of privacy protection, we adopted the privacy-preserving matrix addition as discussed in [26, 41, 45]. The client generates two random vectors u and v with size $N \times 1$ as the private key. To protect the privacy information in the original two-dimensional signal, the client encrypts the matrix f as follows:

$$E(f) = f + u \cdot v^T. \tag{13}$$

After the encryption of (13), any block $f_{i,j}$ in the matrix f satisfies

$$E\left(f_{i,j}\right) = f_{i,j} + u_i \cdot v_j^T, \tag{14}$$

where vector $u_i \in R^{K\times 1}$ is a vector consisting of the $Ki - K + 1$th to Kith elements of vector u and vector $v_j^T \in R^{1\times K}$ is a vector consisting of the $Kj - K + 1$th to Kjth elements of vector v^T.

Afterwards, the encrypted matrix $E(f)$ is sent to the cloud who perform $K \times K$ 2D-DCT as

$$\begin{aligned} E\left(F_{i,j}\right) &= A \cdot E\left(f_{i,j}\right) \cdot A^T = A \cdot \left(f_{i,j} + u_i \cdot v_j^T\right) \cdot A^T \\ &= A \cdot f_{i,j} \cdot A^T + A \cdot u_i \cdot v_j^T \cdot A^T = F_{i,j} + A \cdot u_i \cdot v_j^T \cdot A^T. \end{aligned} \tag{15}$$

We emphasize that matrices A and A^T are not private and can be generated independently by the cloud itself by (3). The calculated results $E(F_{i,j}) \forall i, j \in [1, N/K]$ are then combined into matrix $E(F)$ and returned to the client, who uses the private key vectors u and v to decrypt $E(F)$ efficiently through

$$F_{i,j} = E\left(F_{i,j}\right) - (A \cdot u_i) \cdot \left(v_j^T \cdot A^T\right). \tag{16}$$

Finally, the client quickly verifies the correctness of $F_{i,j}$ by the inference of Parseval's theorem, i.e.,

$$\sum_{m=1}^{K}\sum_{n=1}^{K} f_{i,j}^2(m,n) = \sum_{m=1}^{K}\sum_{n=1}^{K} F_{i,j}^2(m,n). \tag{17}$$

If (17) holds within the allowable error range, block $F_{i,j}$ passes the verification. As long as all the blocks in the matrix F pass the verification, the client accepts the calculation result F; otherwise, the client rejects F.

We summarize the outsourcing protocol for 2D-DCT as follows.

(1) *KeyGen* (1^λ). On input of a security parameter λ, the client generates the private key $\mathrm{KG} = (u, v)$.

(2) *ProbEnc* (Φ;*KG*). On input of the original problem $\Phi = f$ and the key KG, the client performs (13) to encrypt Φ to generate the encrypted problem $\Phi_{\mathrm{KG}} = E(f)$ and then sends Φ_{KG} to the cloud.

(3) *ProbSolve* (Φ_{KG}). On input of the encrypted problem Φ_{KG}, the cloud solves Φ_{KG} by (15). Then, the cloud returns the result $\Lambda_{\mathrm{KG}} = E(F)$ back, together with an empty proof Γ.

(4) *ResultDec* (Λ_{KG};*KG*). On input of the result Λ_{KG} of Φ_{KG} and the key KG, the client performs (16) to decrypt Λ_{KG} to obtain the result $\Lambda = F$ of the original problem Φ.

(5) *ResultVerify* (Λ;Γ). On input of the result Λ of Φ and the proof Γ, the client verifies the correctness of Λ by (17).

Compared with Attempt One and Attempt Two, on the one hand, the formal protocol requires only one round of communication. On the other hand, the difficulty of key management is greatly reduced because the key is just two vectors of length N. Meanwhile, only the private matrix f is encrypted on input, eliminating the encryption and transmission of A and A^T, which further improves the efficiency of outsourcing. Besides, the original signal matrix f is encrypted in its entirety, making the encryption process more efficient and concise compared with block encryption. However, the integral encryption operation does not affect the block flexibility of the cloud, i.e., the cloud can still perform 2D-DCT of any block size based on negotiation with the client.

5.3.2. Outsourcing Protocol for 2D-IDCT. The outsourcing protocol of 2D-IDCT is similar to that of 2D-DCT. First, the client also generates vectors u and v locally as the private key. To protect the privacy of the input privacy matrix F, the client performs

$$E(F) = F + u \cdot v^T. \tag{18}$$

Afterwards, the encrypted matrix $E(F)$ is sent to the cloud who perform $K \times K$ 2D-IDCT as

$$E\left(f_{i,j}\right) = A^T \cdot E\left(F_{i,j}\right) \cdot A. \tag{19}$$

The calculated result $E(f)$ is then returned to the client who uses the private key to decrypt $E(f)$ efficiently through

$$f_{i,j} = E\left(f_{i,j}\right) - \left(A^T \cdot u_i\right) \cdot \left(v_j^T \cdot A\right). \tag{20}$$

Finally, the correctness of resulting f is also verified by equation (17). We summarize the outsourcing protocol for 2D-IDCT as follows.

(1) *KeyGen* (1^λ). On input of a security parameter λ, the client generates the private key $\mathrm{KG} = (u, v)$.

(2) *ProbEnc* (Φ;*KG*). On input of the original problem $\Phi = F$ and the key KG, the client performs (18) to encrypt Φ to generate the encrypted problem $\Phi_{\mathrm{KG}} = E(F)$ and then sends Φ_{KG} to the cloud.

(3) *ProbSolve* (Φ_{KG}). On input of the encrypted problem Φ_{KG}, the cloud solves Φ_{KG} by (19). Then, the cloud returns the result $\Lambda_{\mathrm{KG}} = E(f)$ back, together with an empty proof Γ.

(4) *ResultDec* (Λ_{KG};*KG*). On input of the result Λ_{KG} of Φ_{KG} and the key KG, the client performs (20) to decrypt Λ_{KG} to obtain the result $\Lambda = f$ of the original problem Φ.

(5) *ResultVerify* (Λ;Γ). On input of the result Λ of Φ and the proof Γ, the client verifies the correctness of Λ by (17).

6. Protocol Construction

6.1. Privacy Analysis. For the outsourcing of 2D-DCT, the input privacy matrix is f and the output privacy matrix is F. In the proposed protocol, f is encrypted with two key vectors u and v by the privacy-preserving matrix addition. According to the theoretical proof in [26], the resulting matrix $E(f)$ is computationally indistinguishable in value from a random matrix R. That is, any probabilistic polynomial time distinguisher cannot distinguish the element $E(f)(i,j)$ of $E(f)$ from the element $R(i,j)$ of R for any $\forall\, i, j \in [1, N]$, except with negligible success probability. The success probability of the distinguisher decreases linearly with the expansion of the value range of the key vector. Therefore, the cloud cannot access any private plaintext contained in f from $E(f)$, thereby successfully protecting input privacy. Subsequently, the cloud performs block 2D-DCT of $K \times K$ on the encrypted matrix $E(f)$, resulting in $E(F)$. F is the result of performing $K \times K$ 2D-DCT operation on f. Since the polynomial adversary cannot find any valuable association between $E(f)$ and f in terms of value, the association between $E(F)$ and F is believed to be untraceable, i.e., the output privacy is also successfully protected.

For the outsourcing of 2D-IDCT, the input privacy matrix is F and the output privacy matrix is f. Based on the

consistency of the implementation approach between 2D-IDCT outsourcing and 2D-DCT outsourcing, we conclude that the privacy of input and output is also well protected in the outsourcing of 2D-IDCT.

6.2. Verification Analysis. In terms of result verification, two effects should be achieved. On the one hand, the correct results returned from the cloud should be accepted by the client with a probability close to 1. On the other hand, the erroneous results returned from the cloud should be rejected by the client with a probability close to 1.

In the proposed outsourcing protocol for 2D-DCT and 2D-IDCT, the returned results are all verified by the inference of Parseval's theorem as shown in (17). On the one hand, any correct 2D-DCT or 2D-IDCT transformation result obviously satisfies the inference of Parseval's theorem, thus reaching the first required effect. On the other hand, we emphasize that the cloud is unable to intentionally generate an erroneous result satisfying (17), because the cloud cannot even obtain the plaintext of the input privacy matrix and therefore cannot know the correct sum of squares of elements. In addition, we discuss the situation where the cloud returns an erroneous result at random, but it just meets (17). In this case, the sum of squares of the elements in the erroneous result matrix happens to be correct. For a grayscale block image with $K \times K$ pixels, since the grayscale value of each pixel is taken from the interval [0, 255], the possible values of the sum of squares of elements are in the interval $[0, K^2 \cdot 255^2]$. As long as K is not too small, the probability that the sum of squares of the elements in a randomly selected result matrix happens to be equal to the correct value is minimal. Moreover, the cloud will not try to avoid verification at any cost, because the motivation for the cloud to return erroneous results is considered to be cost-saving. Therefore, we believe that the proposed protocol also conforms to the second required effect.

6.3. Efficiency Analysis. In this section, we conduct a theoretical evaluation for the efficiency performance of the protocol. The evaluation is carried out from three parts: client-side overhead, cloud-side overhead, and communication overhead. Since the efficiency of 2D-DCT outsourcing is consistent with that of 2D-IDCT outsourcing, we take the outsourcing of 2D-DCT as an example for evaluation.

6.3.1. Client-Side Overhead. First, the client takes time $O(2N)$ on generating the private key vectors u and v and $O(N)$ on encrypting the two-dimensional signal matrix f. Then, the client performs (16) to decrypt the result $E(F)$ returned by the cloud, which takes time $O((2 + (1/K))N^2)$. Finally, the client verifies the decrypted result by (17), which takes time $O(2N^2)$. In general, the client needs to spend a total of $O((4 + (1/K))N^2 + 3N)$ time on the outsourcing of 2D-DCT.

6.3.2. Cloud-Side Overhead. The only operation required in the cloud is to perform block 2D-DCT operation of $K \times K$ on the encrypted matrix $E(f)$, which takes time $O(2KN^2)$.

6.3.3. Communication Overhead. During the outsourcing process, only the matrix f and matrix F need to be transmitted between the client and the cloud; thus, the communication overhead is quite small.

Note that the computation overhead in the cloud is the same as it would be for the client to perform the 2D-DCT locally. Based on the above analysis, we conclude that as long as $K \geq 3$, the client can get computing savings from outsourcing. Moreover, the larger the value of K, the more cost savings the client can achieve. For the most common case of $K = 8$, the client can reduce the computing overhead to about one quarter by outsourcing.

7. Simulation Experiment

In this section, we carry out a simulation experiment to further confirm the efficiency of the proposed protocol. The experiment was performed using MATLAB 2016b on a laptop with an Intel Core i5 processor and 8 GB RAM simulating a client. We simulate the situations of users performing 8×8 2D-DCT, fast 8×8 2D-DCT, and computing outsourcing, respectively. The fast 2D-DCT algorithm used for comparison here is proposed by Cho and Lee in [13]. We, respectively, calculated the time required for the client to perform 2D-DCT operation on a two-dimensional signal of different sizes in the three cases. We also assume that the communication between the client and the cloud can be achieved in a negligible time, so the experiment does not count the time spent on communication. The experimental results are shown in Table 1, which is the average value of 10 repeated experiments. The meaning of the parameters in Table 1 is as follows.

(i) $t_{\text{2D–DCT}}$ represents the time taken by the client to perform block 2D-DCT of 8×8 using the definition method

(ii) $t_{\text{fast 2D–DCT}}$ represents the time taken by the client to perform block 2D-DCT of 8×8 using the fast algorithm method

(iii) $t_{\text{outsourcing}}$ represents the time taken by the client to perform block 2D-DCT of 8×8 using the outsourcing method

(iv) $t_{\text{2D–DCT}}/t_{\text{fast 2D–DCT}}$ represents the performance gain to the client using the fast algorithm method compared to the definition method

(v) $t_{\text{2D–DCT}}/t_{\text{outsourcing}}$ represents the performance gain to the client using the outsourcing method compared to the definition method

(vi) $t_{\text{fast 2D–DCT}}/t_{\text{outsourcing}}$ represents the performance gain to the client using the outsourcing method compared to the fast algorithm method

As can be seen from Table 1, on the one hand, compared with the definition method, regardless of the value of size N, the outsourcing protocol can bring the client a performance gain of more than 4.43 times, which is consistent with the result of previous theoretical analysis. On the other hand, even compared with the fast 2D-DCT algorithm, outsourcing

Table 1: Experimental results.

Size N	t_{2D-DCT} (sec)	$t_{\text{fast }2D-DCT}$ (sec)	$t_{outsourcing}$ (sec)	$\frac{t_{2D-DCT}}{t_{\text{fast }2D-DCT}}$	$\frac{t_{2D-DCT}}{t_{outsourcing}}$	$\frac{t_{\text{fast }2D-DCT}}{t_{outsourcing}}$
256	0.5530	0.2070	0.1110	2.6715	4.9820	1.8649
512	1.9076	0.7979	0.4301	2.3908	4.4352	1.8551
1024	8.7805	2.8735	1.7091	3.0557	5.1375	1.6813
2048	33.0577	11.9314	6.9559	2.7706	4.7525	1.7153
4096	127.8909	47.5288	28.0907	2.6908	4.5528	1.6920

can still bring more than 1.68 times of performance gain to the client, which further demonstrates the outstanding efficiency of the outsourcing method.

8. Conclusions

In this paper, we speed up 2D-DCT and 2D-IDCT from the point of view of computing outsourcing rather than algorithm optimization. Based on the privacy-preserving matrix addition, we realize the privacy protection of input and output data. Meanwhile, the inference of Parseval's theorem is used to verify the returned results. Furthermore, compared with the two possible solutions Attempt One and Attempt Two, the proposed formal protocol's features of one-round communication and simple key management guarantee the protocol's efficiency performance. The proposed outsourcing protocol is confirmed by theoretical analysis to solve three key challenges in computing outsourcing. Experimental results show that outsourcing is even more efficient than the fast 2D-DCT and 2D-IDCT algorithms. Thus, the adoption of the proposed outsourcing protocol could indeed be an excellent alternative for speeding up 2D-DCT and 2D-IDCT operations. In the future, we will strive to achieve efficient and secure outsourcing of other signal processing tools, such as DWT and NTT.

Data Availability

No data were used to support the findings of the study.

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 Team Project of Collaborative Innovation in Universities of Gansu Province (No. 2017C-16) and the Major Project of Gansu University of Political Science and Law (No. 2016XZD12).

References

[1] F. Ernawan, M. N. Kabir, Z. Mustaffa, K. Moorthy, and M. Ramalinga, "An Improved Image Compression Technique using Large Adaptive DCT Psychovisual Thresholds," in *2019 IEEE 2nd International Conference on Knowledge Innovation and Invention (ICKII)*, pp. 561–564, Seoul, Korea (South), 2019.

[2] H. Kaur and R. Kaur, "Speech compression and decompression using DWT and DCT," *International Journal of Computer Technology and Applications*, vol. 3, no. 4, pp. 1501–1503, 2012.

[3] S. Heng, C. So-In, and T. G. Nguyen, "Distributed Image Compression Architecture over Wireless Multimedia Sensor Networks," *Wireless Communications and Mobile Computing*, vol. 2017, 21 pages, 2017.

[4] H.-J. Ko, C.-T. Huang, G. Horng, and S.-J. Wang, "Robust and blind image watermarking in DCT domain using inter-block coefficient correlation," *Information Sciences*, vol. 517, pp. 128–147, 2020.

[5] B. Y. Lei, I. Y. Soon, and Z. Li, "Blind and robust audio watermarking scheme based on SVD-DCT," *Signal Processing*, vol. 91, no. 8, pp. 1973–1984, 2011.

[6] L.-S. Liu, R.-H. Li, and Q. Gao, "A robust video watermarking scheme based on DCT," in *2005 International Conference on Machine Learning and Cybernetics*, vol. 8, pp. 5176–5180, Guangzhou, China, 2005.

[7] F. Cruz-Roldán, J. Piñeiro-Ave, J. L. Rojo-Álvarez, and M. Blanco-Velasco, "Simple Algorithms for Estimating the Symbol Timing Offset in DCT-Based Multicarrier Systems," *Wireless Communications and Mobile Computing*, vol. 2018, 8 pages, 2018.

[8] N. Al-Dhahir, H. Minn, and S. Satish, "Optimum DCT-based multicarrier transceivers for frequency-selective channels," *IEEE Transactions on Communications*, vol. 54, no. 5, pp. 911–921, 2006.

[9] G. K. Wallace, "The JPEG still picture compression standard," *IEEE Transactions on Consumer Electronics*, vol. 38, no. 1, pp. xviii–xxxiv, 1992.

[10] D. Estrin, D. Culler, K. Pister, and G. Sukhatme, "Connecting the physical world with pervasive networks," *IEEE Pervasive Computing*, vol. 1, no. 1, pp. 59–69, 2002.

[11] G. J. Pottie and W. J. Kaiser, "Wireless integrated network sensors," *Communications of the ACM*, vol. 43, no. 5, pp. 51–58, 2000.

[12] C. Ma, "A Fast Recursive Two Dimensional Cosine Transform," in *Intelligent Robots and Computer Vision VII*, Boston, MA, USA, 1989.

[13] N. I. Cho and S. U. Lee, "Fast algorithm and implementation of 2-D discrete cosine transform," *IEEE Transactions on Circuits and Systems*, vol. 38, no. 3, pp. 297–305, 1991.

[14] A. C. Hung and T. H.-Y. Meng, "A comparison of fast inverse discrete cosine transform algorithms," *Multimedia Systems*, vol. 2, no. 5, pp. 204–217, 1994.

[15] A. C. Yao, "Protocols for secure computations," in *23rd Annual Symposium on Foundations of Computer Science (sfcs 1982)*, pp. 160–164, Chicago, IL, USA, 1982.

[16] C. Gentry, "Fully homomorphic encryption using ideal lattices," in *Proceedings of the 41st annual ACM symposium on Symposium on theory of computing - STOC '09*, pp. 169–178, Bethesda, MD, USA, 2009.

[17] R. Gennaro, C. Gentry, and B. Parno, "Non-interactive Verifiable Computing: Outsourcing Computation to Untrusted Workers," in *Advances in Cryptology – CRYPTO 2010*, pp. 465–482, 2010.

[18] M. Barbosa and P. Farshim, "Delegatable Homomorphic Encryption with Applications to Secure Outsourcing of

Computation," in *Lecture Notes in Computer Science*, pp. 296–312, 2012.

[19] K.-M. Chung, Y. Kalai, and S. Vadhan, "Improved Delegation of Computation Using Fully Homomorphic Encryption," *Advances in Cryptology – CRYPTO 2010*, pp. 483–501, 2010.

[20] B. Parno, J. Howell, C. Gentry, and M. Raykova, "Pinocchio: Nearly Practical Verifiable Computation," in *2013 IEEE Symposium on Security and Privacy*, pp. 238–252, Berkeley, CA, USA, 2013.

[21] X. Lei, X. Liao, T. Huang, H. Li, and C. Hu, "Outsourcing Large Matrix Inversion Computation to A Public Cloud," *IEEE Transactions on Cloud Computing*, vol. 1, no. 1, pp. 1–1, 2013.

[22] F. Chen, T. Xiang, X. Lei, and J. Chen, "Highly Efficient Linear Regression Outsourcing to a Cloud," *IEEE Transactions on Cloud Computing*, vol. 2, no. 4, pp. 499–508, 2014.

[23] Y. Li, Z. L. Jiang, X. Wang, J. Fang, E. Zhang, and X. Wang, "Securely Outsourcing ID3 Decision Tree in Cloud Computing," *Wireless Communications and Mobile Computing*, vol. 2018, 10 pages, 2018.

[24] C. Wang, K. Ren, J. Wang, and Q. Wang, "Harnessing the cloud for securely outsourcing large-scale systems of linear equations," *IEEE Transactions on Parallel and Distributed Systems*, vol. 24, no. 6, pp. 1172–1181, 2013.

[25] X. Chen, X. Huang, J. Li, J. Ma, W. Lou, and D. S. Wong, "New algorithms for secure outsourcing of large-scale systems of linear equations," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 1, pp. 69–78, 2015.

[26] S. Salinas, C. Luo, X. Chen, W. Liao, and P. Li, "2015 Efficient secure outsourcing of large-scale sparse linear systems of equations," in *EEE Conference on Computer Communications (INFOCOM)*, vol. 4no. 1, pp. 26–39, Kowloon, Hong Kong, 2015.

[27] Y. Zhang, J. Jiang, Y. Xiang, Y. Zhu, L. Wan, and X. Xie, "Cloud-assisted privacy-conscious large-scale Markowitz portfolio," *Information Sciences*, vol. 527, pp. 548–559, 2020.

[28] X. Li, Y. Zhu, J. Wang, Z. Liu, Y. Liu, and M. Zhang, "On the soundness and security of privacy-preserving SVM for outsourcing data classification," *IEEE Transactions on Dependable and Secure Computing*, vol. 15, no. 5, pp. 906–912, 2018.

[29] Z. Zhang, M. Dong, L. Zhu et al., "Achieving Privacy-Friendly Storage and Secure Statistics for Smart Meter Data on Outsourced Clouds," *EEE Transactions on Cloud Computing*, vol. 7, pp. 638–649, 2017.

[30] Y. Zhu, X. Li, J. Wang, and J. Li, "Cloud-assisted secure biometric identification with sub-linear search efficiency," *Soft Computing*, vol. 24, no. 8, pp. 5885–5896, 2020.

[31] H. Li, K. Ota, M. Dong, A. Vasilakos, and K. Nagano, "Multimedia processing pricing strategy in GPU-accelerated cloud computing," *IEEE Transactions on Cloud Computing*, 2017.

[32] T. Bianchi, A. Piva, and M. Barni, "On the implementation of the discrete fourier transform in the encrypted domain," *IEEE Transactions on Information Forensics and Security*, vol. 4, no. 1, pp. 86–97, 2009.

[33] T. Bianchi, A. Piva, and M. Barni, "Encrypted Domain DCT Based on Homomorphic Cryptosystems," *EURASIP Journal on Information Security*, vol. 2009, 12 pages, 2009.

[34] P. Zheng and J. Huang, "Discrete Wavelet Transform and Data Expansion Reduction in Homomorphic Encrypted Domain," *IEEE Transactions on Image Processing*, vol. 22, no. 6, pp. 2455–2468, 2013.

[35] A. Pedrouzo-Ulloa, J. R. Troncoso-Pastoriza, and F. Perez-Gonzalez, "Number Theoretic Transforms for Secure Signal Processing," *IEEE Transactions on Information Forensics and Security*, vol. 12, no. 5, pp. 1125–1140, 2017.

[36] K. Han, M. Hhan, and J. H. Cheon, "Improved Homomorphic Discrete Fourier Transforms and FHE Bootstrapping," *IEEE Access*, vol. 7, pp. 57361–57370, 2019.

[37] X. Xiao, J. Huang, Y. Zhang, and X. He, "Efficient and Secure Outsourcing of DFT, IDFT, and Circular Convolution," *IEEE Access*, vol. 7, pp. 60126–60133, 2019.

[38] Y. Zhang, Y. Xiang, L. Y. Zhang, L.-X. Yang, and J. Zhou, "Efficiently and securely outsourcing compressed sensing reconstruction to a cloud," *Information Sciences*, vol. 496, pp. 150–160, 2019.

[39] Y. Zhang, J. Zhou, Y. Xiang et al., "Computation Outsourcing Meets Lossy Channel: Secure Sparse Robustness Decoding Service in Multi-Clouds," *IEEE Transactions on Big Data*, 2017.

[40] X. Lei, X. Liao, T. Huang, and F. Heriniaina, "Achieving security, robust cheating resistance, and high-efficiency for outsourcing large matrix multiplication computation to a malicious cloud," *Information Sciences*, vol. 280, pp. 205–217, 2014.

[41] X. Zhang, S. Liu, H. Cui, and T. Chen, "Reading the Underlying Information From Massive Metagenomic Sequencing Data," *Proceedings of the IEEE*, vol. 105, no. 3, pp. 459–473, 2017.

[42] S. Fu, Y. Yu, and M. Xu, "A Secure Algorithm for Outsourcing Matrix Multiplication Computation in the Cloud," in *Proceedings of the Fifth ACM International Workshop on Security in Cloud Computing - SCC '17*, pp. 27–33, Abu Dhabi United Arab Emirates, 2017.

[43] X. Lei, X. Liao, T. Huang, and H. Li, "Cloud Computing Service: The Case of Large Matrix Determinant Computation," *IEEE Transactions on Services Computing*, vol. 8, no. 5, pp. 688–700, 2015.

[44] W. Liao, C. Luo, S. Salinas, and P. Li, "Efficient Secure Outsourcing of Large-Scale Convex Separable Programming for Big Data," *IEEE Transactions on Big Data*, vol. 5, no. 3, pp. 368–378, 2017.

[45] Y. Zhang, X. Xiao, L.-X. Yang, Y. Xiang, and S. Zhong, "Secure and Efficient Outsourcing of PCA-Based Face Recognition," *IEEE Transactions on Information Forensics and Security*, vol. 15, pp. 1683–1695, 2020.

Hindawi
Wireless Communications and Mobile Computing
Volume 2020, Article ID 8887933, 10 pages
https://doi.org/10.1155/2020/8887933

Research Article

An Intelligent Planning-Based Modeling Method for Diagnosis and Repair

Chuang Li,[1,2,3] Dantong Ouyang,[1,3] Xiaoyu Wang,[2] and Wei Wei[4]

[1]*College of Computer Science and Technology, Jilin University, Changchun 130012, China*
[2]*College of Computer, Jilin Normal University, Siping 136000, China*
[3]*Key Laboratory of Symbolic Computation and Knowledge Engineer of Ministry of Education, Changchun 130012, China*
[4]*Public Computer Education and Research Center, Jilin University, Changchun 130012, China*

Correspondence should be addressed to Wei Wei; wei_wei@jlu.edu.cn

Received 24 April 2020; Revised 26 May 2020; Accepted 22 June 2020; Published 24 July 2020

Academic Editor: Wei Wang

Planning and model-based diagnosis are both branches of artificial intelligence. In model-based diagnosis, multiresults may be gotten which lead to an uncertain diagnosis. We use the landmark method from planning to designing an event sequence to get a reaction. A method that uses planning to repair in local results of incremental diagnosis is proposed. Firstly, a model is established on model-based diagnosis and planning. Incremental diagnosis results are used as the initial state of planning, and the heuristic search method is used to find the solution to an unfaulty state. Two algorithms with different strategies are designed for diagnosis and repair: one is to repair all possible faults and use controllable events to repair them, and the other is to test through the feedback of controllable events and observable events to get the only solution and repair them. At the same time, the efficiency of the incremental diagnosis method is improved based on heuristics.

1. Introduction

AI planning and model-based diagnosis are developing at the forefront of AI research. Diagnostic problems originate from medicine and then extend to the detection of equipment failure or behavior deviation. As machinery equipment gradually replaces manual work, the significance of safety is gradually highlighted, so the broad diagnosis problems are also paid attention to in the field of control and artificial intelligence [1]. Model-based diagnosis is a diagnostic method of common reasoning based on the equipment model and behavior. Because of its low equipment dependence and high efficiency, it has developed rapidly in theory and application since it was put forward.

The model-based diagnosis method was originally used in static systems [2]; then, the conflict-based method was used in other problems such as knowledge representing [3] and planning [4]. After the system shutdown, the fault of the current system was inferred based on the online observations. However, as the scale of equipment increases, the behavior becomes more and more complex and the loss caused by shutdown increases. The dynamic diagnosis that can be monitored online is put forward [5, 6] and has been extended to various industries.

The goal of dynamic diagnosis is to discover the possible faults online and present the faults and their causes as the results; the results are useful to the industrial system [7, 8]. Furthermore, when the system failure is identified, we hope the system can be effectively intervened and repaired online [9]. Intelligent repair of the system is the follow-up work of fault diagnosis, which further improves the security of the system.

Dynamic model-based diagnosis methods are usually based on discrete event system models, which decompose the behavior of the system into persistent states and abrupt events [10]. Faults are defined as special events. When running online, only part of the system's behavior can be detected, which is called an observable event. The diagnostic system infers whether there is a fault in the current state of the system according to the discrete event system model

and the sequence of observable events. If there is a fault in the system, the process of the system's behavior given by the diagnostic results is a path in which states are triggered by events. Diagnosis usually assumes that the fault is a persistent fault, that is, when the fault occurs, without interference, the fault will not disappear by itself. Then, the fault is repaired from the fault state to the nonfault state through controllable events in the device.

When the system failed, the diagnostic system tries to use controllable events to repair it [11], and we hope these faults can be attached. Path planning uses state and action to describe the current environment and system and accomplish the expected tasks. The path planning method is combined with the diagnosis model. The planning method is used to design a path from a faulty state to unfaulty, and by the landmark, incremental diagnosis is provided a heuristic value. On the result of the dynamic diagnosis, the fault is repaired by using path planning through controllable events and known models of the system. The merging strategy in the field of gesture recognition is used for reference; the gestures are collected by different equipment but used by the same framework [12–16].

The article is organized as follows: Section 2 defines the model framework; Section 3 gives the process of diagnosis and planning; Section 4 gives the corresponding algorithm; Section 5 is the experimental results.

2. Model

A discrete event system is used to model dynamic diagnosis and map from real world to logical system. Firstly, the basic model is established.

Definition 1 (model). Model is defined as a five-tuple $G = (Q, E, T, I, F)$, in which Q is a finite state set, E is a finite event set, T is a transition set, $T \subseteq Q \times E \times Q$, I is an initial state set, $I \subseteq Q$, and F is a final state set, $F \subseteq Q$.

State is usually in the process of system stability. Event represents the behavior of the system, which is instantaneous relative to the state. Transfer represents the process in which the system is triggered by the behavior from one continuous state to another state. Event triggering may be passive or active.

In classical diagnostic methods, event sets are usually divided into three subsets according to their properties: observable event set E_o, normal unobservable event set E_n, and faulty event set E_f, and they satisfy $E = E_o \cup E_n \cup E_f$, $E_o \cup E_n = \varnothing, E_o \cup E_f = \varnothing, E_n \cup E_f = \varnothing$. If the system can directly change the state of the system by controlling some events, the event set also includes such special events.

Definition 2 (controllable events). Controllable events are events that can be controlled and executed by the diagnostic system, marked as E_c, $E_c \subseteq E$.

Through controllable events and models, the faults are repaired. A path is formed in the model.

Definition 3 (trajectory). Trajectory s is the transfer sequence of state triggered by events $s = <q_0, e_0, q_1 \cdots q_n>$, and $\forall t$, $t = q_i \times e_{i+1} \times q_{i+1}, t \in T$, $S(G)$ is defined as the set of all possible trajectories on model G.

In actual operation, the system is carried out along a certain trajectory. Some of the events in the trajectory are visible, collected by the diagnostic system, and searched in the predefined model to obtain the determined trajectory. If there is a fault event in the trajectory, the system is considered to be in a fault state. The incremental search path planning is used to repair and restore the system to a normal state.

Definition 4 (repair path). Repair path is an extension of the model, defined as a three tuple $R = (G_p, Q_{\text{start}}, Q_{\text{goal}})$, among which, G_p is a local system model, $G_p = (Q_p, E_p, T_p)$, and $Q_p \subseteq Q, E_p \subseteq E, T_p \subseteq T$; Q_{start} is the finite state set with fault label, $Q_{\text{start}} \subseteq Q, \forall q \in Q_{\text{start}}, q = q \times \{e_{f0}, e_{f1} \cdots e_{fi}\}$; Q_{goal} is the finite state set with fault label, $Q_{\text{goal}} \subseteq Q, \forall q \in Q_{\text{goal}}$, $q = q \times \varnothing$.

The repair path is generated according to the diagnoses, and the initial state is obtained from Q_{start}. According to the model using the incremental method in reverse, the heuristic search is conducted. The path composed of controllable events is obtained, and the local system model is generated. Finally, the set Q_{goal} is obtained, the system is restarted, and the diagnosis is made.

The system is event-driven, and the possible sequence of events is defined as the language of the system.

Definition 5 (language). Language on system G is a sequence of possible events $L(G) = <e_i^*>, e_i^* \in 2^E$.

In practice, only part of the behavior in the system can be observed, which is a sequence of observable events corresponding to the model.

Definition 6 (observable mapping). Observable mapping is defined between event sets and observable event sets.

$$P_o(e) = \begin{cases} e & e \in E_o, \\ \varepsilon & e \in E \setminus E_o. \end{cases} \tag{1}$$

Observable mappings are defined linguistically.

$$P_o(l) = P_o\left(l'e\right) = P_o\left(l'\right)P_o(e). \tag{2}$$

And further, observable mappings are defined on trajectories.

$$P_o(s) = P_o\left(s'e_iq_i\right) = P_o\left(s'\right)P_o(e_i). \tag{3}$$

The sequence of observation events actually obtained during the system operation is recorded as Obs. The same mapping can also be defined between the event set and the

fault event set. Conversely, the inverse mapping of an observable mapping is defined as a set of languages.

Definition 7 (observable inverse mapping). $P_o^{-1}(l') = \{l \mid P_o(l) = l'\}$.

The relationship between observable mapping and inverse mapping is not a one-to-one correspondence, and there may be more than one path to repair faults through controllable events, but a fault-free state can be reached, which is acceptable to the diagnosis system.

Definition 8 (fault label). Fault label is a set of fault events, $l = \{e_i \mid e_i \in E_f\}$.

The state carries one or more fault labels, which are normal if and only if all fault labels are empty sets. Otherwise, it will be a failure state. When the diagnosis system gets the conclusion of the fault, the fault path will be given, which is actually the state-event sequence with the fault label.

Controllable events confirm repair by modifying the fault label in the state, so adding and deleting effects are set for controllable events. In the state q, if triggered by controllable events, the adding effect of controllable events is set to reduce or increase the fault label of the state and transfer the state. The deletion effect of controllable events is set to withdraw from the target state to the current state and add or subtract the fault label of the current state.

Definition 9 (diagnosis). Diagnosis is defined as a compatible path between model and observation $\Delta(G, \text{Obs}) = \{s \mid s \in S(G), P_o(s) = \text{Obs}\}$.

The diagnosis is definitive, but in the actual system, sensors are limited, leading to more than one diagnosis. If only one result is needed, the most probable diagnoses will be chosen from the candidate set. But when the choice is wrong, the diagnosis will backtrack and choose another. The backtrack will depress the efficiency, so we use a landmark as a heuristic function to improve the chosen procedure.

Definition 10 (landmark). A landmark q_m in diagnosis is a state with a value, $q_m = q \times n, q \in Q, n \in \mathbb{N}^+$.

The landmark is used to mark the state which leads to more than one trajectories; the value in the state expresses how many times the state is passed.

A path is obtained by the diagnosis. When the path s contains a fault event, the subsequent fault status label must contain the event. Searching forward from path s, when there is a path that can get a nonfault state through controllable events, the fault is considered to be repaired. The repair behavior is defined as in the model. An example is given to explain the definitions.

The digraph in Figure 1 is a model with $Q = \{q_0 \cdots q_7\}$, $E = \{e_0 \cdots e_5\}$, $T = \{q_0 \times e_0 \times q_2, \cdots q_5 \times e_5 \times q_6\}$, $I = \{q_0\}$, and $F = \{q_6, q_7\}$. The controllable events can be some special events that can change some states. In the example, it can be e_3. $s = q_0 e_0 q_2 e_2 q_7$ is a trajectory. If q_5 is a faulty state, q_1 is a normal state; a trajectory which contains $t = q_5 e_4 q_1$ may be a repair path. Events in a trajectory can be a language, $e_1 e_3 e_5$ is a language, but a prefix $e_1 e_3$ is a language. In event set E, if e_3 is observable, the observable map of language $e_1 e_3 e_5$ is e_3, and the observable inverse mapping of e_3 is $e_1 e_3 e_5$ and $e_1 e_2 e_4 e_3$. If the event e_5 is faulty, the state q_6 has a label of fault e_5, $q_6\{e_5\}$. The diagnoses of Obs = $<e_3>$ will be $q_0 q_1 q_4 q_6$ and $q_0 q_1 q_3 q_5 q_1$. Some of the states at the key position is defined as a benchmark, such as q_1.

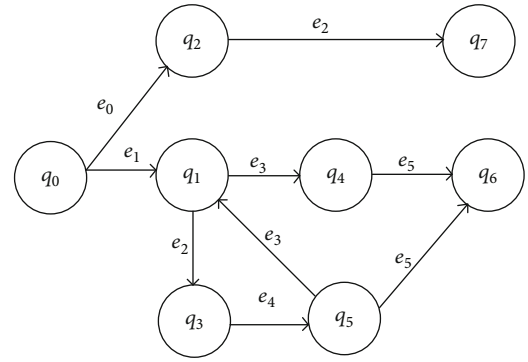


Figure 1: An example of a model.

3. Method

Discuss the diagnosis and repair of the system using the planning method according to the situation. The incremental method is used to diagnose the system. With the arrival of the observation, one or more diagnostic results may be produced at a time, and the diagnostic results may or may not contain faults. There are four situations as follows:

(1) There is only one diagnosis result, which does not include fault. Keep the current path and wait for the next observation

(2) There is only one diagnosis result, which includes the fault. Path planning based on the heuristic search is carried out from the final state of diagnosis results until a repair path is found

(3) There are many diagnostic results, which do not include faults. The heuristic method is used to select a path as the diagnosis result and continue to wait for the next observation

(4) There are many diagnostic results, including faults. Combined with the multiobjective programming method, the paths that may cover the most diagnostic results are tested. After determining the only fault path, it is repaired

The first case and the third case have been solved in incremental diagnosis, so they are not considered in this paper. Next, we will discuss the situation (2) and (4). They both occur one or some faults that should be repaired, so a planning path should be deduced. The controllable events

will carry other behaviors to change the path. So when the planning path is found, the system can be repaired or done some different behaviors.

3.1. Single Candidate Trajectory. When the diagnosis only obtains one trajectory and there is a fault event in the trajectory, go backward from the current model and search for possible repair trajectories. The heuristic search method for multiple target paths is used.

The initial state Q_{start} of the repair path is determined. Only one path in the single candidate path is compatible with the observation. Then, the final state q of the current diagnosis trajectory is selected into the set Q_{start}. The goal is to obtain the local model $G_p = (Q_p, E_p, T_p)$, to make the repair path $R = (G_p, Q_{\text{start}}, Q_{\text{goal}})$, and $q \in Q_{\text{start}}$.

With elements in the set Q_{start} as the initial state, the controllable events are searched for an extension. The controllable event defines its heuristic value according to the difficulty of its control, adds or deletes the fault label of the state according to the adding or deleting effect of controllable events, and generates the trajectory. In the process of searching, the model G is regarded as the global environment, and the local model G_p that has been obtained is regarded as the local environment. Control events may change the local environment and transfer the state to a specific follow-up state. Incremental methods are used to deal with the search problem in a variable environment.

The search begins in the open table. Controllable events are added according to heuristic values, and the impact on current status fault labels is judged. When a state without a faulty label is obtained, it is added to Q_{goal}. As a final state, all controllable event sequences from Q_{start} to Q_{goal} constitute a path, which is a repair sequence.

3.2. Multiple Candidate Trajectories. Multiple candidate trajectories have the same observation sequence, but they contain different faults. Therefore, before repairing, a test is carried out as far as possible through controllable events, and the set of candidate trajectories is reduced. A unified or compatible repair path can be obtained by multiobjective programming on the reduced set of paths.

If the current set of candidate diagnostic trajectories is S_i and the observation sequence is Obs_i, then $\forall s_i \in S_i, P_o(s_i) = \text{Obs}_i$. The initial states of the trajectories are the same and the fault labels of the states are different. The repair path $R_i = (G_p, Q_{\text{start}}, Q_{\text{goal}})$. The final states of all diagnostic trajectories are added to the initial state set Q_{start} of the repair path. Controllable events are used to test the trajectories, and the candidate diagnostic trajectories are further distinguished according to the observations returned again. There may be more than one state in Q_{start} and more controllable events that can be selected. The controllable events where adding effects can reduce the maximum number of failure events are selected.

That is, in Q_{start}, there are multiple states and their fault labels $q_0 \times l_{00,}l_{01}, \cdots q_n \times l_{n0}, l_{n1}$. Among all controllable event sets that can trigger states, the controllable events that can delete the most faults should be selected. At this time, two kinds of methods are considered for the properties of the system: (1) All repairs are made directly through controllable events until they are fully restored to a normal state. (2) Candidate diagnostic sets are tested by controllable events, and the candidate diagnostic sets are reduced according to the observations obtained from the interaction. The problem is reduced to the repair problem of a single candidate trajectory. Next, we will discuss the situation further.

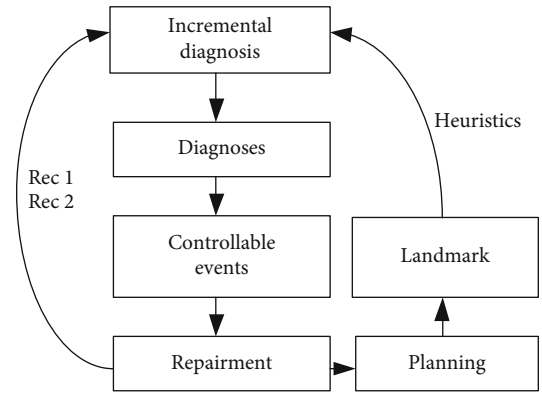


Figure 2: The relationship of the methods.

3.2.1. Multiple Candidate Repair. If the current set of candidate diagnostic paths is S_i and the observation sequence is Obs_i, the initial state set Q_{start} of repair is the final state set of all paths, and it has a fault label. For $\forall q \in Q_{\text{start}}$, all controllable events and adding effects are generated, and events that can produce maximum additive effects are selected. The corresponding state in Q_{start} is replaced, and the replacement process is added to model G_p. The above process is repeated until all the states in Q_{start} are normal. The relationship between each candidate diagnostic trajectories is "and." The final state set generates Q_{goal}.

3.2.2. Test and Repair. In the current final state of candidate trajectories, all controllable events and adding effects are generated. If $\exists e_c \in E_c$ enables at least two paths to produce different observable events, then the event is controlled. In the effective time, if the determined observation events are obtained, the corresponding paths are retained and other paths are deleted. The process is repeated until the only diagnosis result can be selected, and then, the repair of a single candidate path is dealt with.

3.3. Landmark and Heuristic Incremental Diagnosis. When the model becomes more complex, diagnosis based on an overall model will be more difficult. Then, the incremental method was proposed by [17]; the diagnosis is gotten incrementally. Incremental diagnosis is gotten from a part of the model and corresponding observations and expresses as the newest result, when another new observations are gotten, the diagnosis will be updated upon the last diagnosis.

But when the multiple trajectories appear, some periodical diagnoses are wrongly chosen, some incremental method backtrack to initial state, and some method backtrack to a

Input: System model $G = (Q, E, T, I, F)$, Observation sequence Obs = <$\text{obs}_0, \text{obs}_1 \cdots$.>.
Output: Diagnostic results Δ_0, Δ_1
1. $G_f = \text{Comp}(G)$ //Add fault labels to the model
2. while(Obs $\neq \varnothing$)
3. $\Delta_i = \text{Dia}(G_f, \text{obs}_i)$ //Diagnose the model
4. if$(\exists e_f \in \Delta_i)$
5. $\text{Rec}(G_f, \Delta_i)$ //If there is a fault, repair it
6. Return $\Delta = \Delta_0 \Delta_1 \cdots$ //Return all diagnostic results

Algorithm 1: Dia and Rec.

former state. If we want to improve the efficiency, we need to reduce the number of backtracks.

To any path s, if there is a set $S(G)$, such that $\forall s \in S(G)$, $S(G) = \{s \mid P_o(s) = P_o(s_i)\}$, it gives the basic heuristic value a to the end states in paths, and the heuristic value of the set is a^2. When the diagnostic system runs, the heuristic value a is changed by the chosen times and backtrack times.

We use the heuristic value to get a more possible trajectory: at the beginning of diagnosing, the value depends on the number of the trajectories with the same observations, and when the diagnostic system begins to run, the value depends on the number of trajectories was passed. Once backtrack happened in one chosen trajectory, the value will be reduced, and when the recovery planning passes the trajectory, the value will be increased.

The heuristic value is used in diagnosis and planning, which accelerates the diagnostic process. The experiments will be given in part 5.

4. Algorithm and Explanation

Online diagnosis and repair mainly include three parts: compilation of offline models, diagnosis based on online observation, and repair based on the diagnosis. Offline model compilation generates fault labels based on models. Online diagnosis uses incremental methods. Diagnosis and repair include two strategies: repair all possible paths and verify the repair of a single path by testing. The relationship of the following methods is described in Figure 2.

Firstly, the whole algorithm is given.

Next, the key algorithms are described in detail. Firstly, offline compilation is carried out according to the model to generate fault labels. Based on the computational diagnosability method, the fault label is generated. On the basis of the computation diagnosability method, fault labels are generated.

The algorithm Comp generates a model with fault label in the offline state. In the online process, if the state with fault label is obtained from the observation, the current system is in a fault condition. The algorithm of online incremental diagnosis is as follows:

In fact, multiple candidate diagnostic trajectories may be obtained for each diagnosis. According to heuristics or other principles, the only path is selected as the result of this diagnosis. The initial state of the next diagnosis result is set to the final state of this diagnosis result. When there is a fault state

Input: System model $G = (Q, E, T, I, F)$
Output: Model with fault label $G_f = (Q \times L, E, T, I, F \times L)$
1. Initialization: $BI = I$, $Q_f = \{q \mid q = q \times l \times v\}$, $l = \varnothing$, $v = 0$
2. While$(BI \neq \varnothing \| BI \notin F)$
3. $\forall q \in BI, \exists t = q \times e \times q'$
4. if$(e \in E_f)$
5. if$(v' == 0)$
6. $l' = l' \cup \{e\}$
7. else
8. $l' = l' \times \{e\}$
9. $BI = BI - \{q\} \cup \{q'\}$
10. Return G_f

Algorithm 2: Comp.

in the diagnosis result, it is repaired. The repair algorithm Rec is given. According to the strategy of repairing all possible faults or the unique path obtained by a test, two algorithms are given: Rec1 (repair all possible faults) and Rec2 (repair the unique path obtained by a test).

If it is a single candidate fault trajectory, a trajectory composed of control events is returned. If the control system triggers controllable events in turn, the fault can be repaired. If the control system triggers controllable events in turn, the fault can be repaired. If there are multiple candidate fault paths, multiple paths may cross or be independent are returned. If the control system triggers controllable events in turn, all possible faults can be repaired. If no control event can be found to repair all candidate diagnostic trajectories, the trajectory with the highest heuristic value is selected and the single candidate fault path is processed instead.

Another method uses heuristic values to select candidate diagnostic trajectories and reduces the set of candidate diagnosis according to the feedback observations until the only diagnostic result is obtained. On this result, the fault is repaired.

According to the continuous feedback observation, the candidate diagnosis set is reduced, and the only diagnosis trajectory is finally obtained, and then, the repair is carried out. Compared with the algorithm Rec1, Rec2 takes a long time to interact continuously, and controlling controllable events for testing may consume more time or system workload. However, when Rec2 has more candidate diagnostic paths, the

Input: Model with fault label $G_f = (Q \times L, E, T, I, F \times L)$ Real-time observation sequence Obs = <obs$_0$, obs$_1$ ⋯ obs$_i$ ⋯ >
Output: Candidate diagnostic results Δ_i
1. Initialization: $QI = I$, $QF = \varnothing$, $S = \varnothing$
2. $\forall q \in Q, s = <q, e_0, \cdots q'>$
3. if$(P_o(s) == \text{obs}_i)$
4. $S = S \cup \{s\}$
5. $s' = \text{Choose}(S)$ //choose the most possible path
6. $QF = q_n{}'$
7. $QI_{i+1} = q_0{}'$
8. $\Delta_i = S$

Algorithm 3: Dia.

Input: Model with fault label $G_f = (Q \times L, E, T, I, F \times L)$
Real-time observation sequence Obs = <obs$_0$, obs$_1$ ⋯ obs$_i$ ⋯ >
Current diagnostic results Δ_i
Output: Repair path S_c
1. Initialization: $S_c = \varnothing, N_q = \varnothing$
2. if$(|\Delta_i| > 1)$
3. $\forall s \in \Delta_i, N_q = \{q \mid s = <q_0 e_0 q_1 \cdots q>\}$ //Get the final state of the trajectory
4. while$(\forall q \in N_q, l_q \neq \varnothing)$
5. $e_c = \text{Choose}(q)$ //Select the control event with maximum adding effect
6. $G_c = \text{Change}(N_q, e_c)$ //Update status and delete faults
7. $S_c = S_c e_c$ //Record a sequence of controllable events
8. if$(\text{Dead}(N_q))$ //If it cannot be expanded
9. Back(N_q, e_c) //Backtrack and recover failures
10. $S_c = S_c - e_c$
11. if$(\nexists S_c)$ //If there is no path to full recovery
12. $S_c = \max(\Delta_i)$, break //Choose the diagnosis with the greatest heuristic value and quit multi-fault repair
13. else
14. $s = <q_0 \cdots q>, N_q = \{q\}$ //Get the final state of the path
15. while$(l_q \neq \varnothing)$
16. $S_c = \text{Find}(N_q, e_c)$ //Get the repair status of the path
17. if$(\nexists S_c)$
18. Return Fail //If the repair path cannot be found, fail to exit

Algorithm 4: Rec1.

size of the path set is reduced by testing, and the amount of computation is less. Rec1 is prone to generate more branches, which eventually leads to calculation failure.

Rec1 and Rec2 select in the current set of candidate diagnostic trajectories and select controllable events for the next step. The heuristic method is chosen to evaluate the current trajectory. In Rec1, the principle of choosing controllable events is to cover as many paths as possible and repair more faults. In Rec2, the principle of choosing controllable events is to generate as many different observations as possible, so as to distinguish candidate diagnosis trajectories and reduce the set of candidate diagnosis trajectories.

Next, the above methods are tested to get the time and space complexity of the algorithm. The efficiency of repair results of the Rec1 method and the Rec2 method under different conditions (such as the controllable number of events and controllable event efficiency) is compared.

5. Experimental Results and Analysis

In this paper, the experimental machine configuration is as follows: processor: Inter i7-6700, 3.4 GHz, memory: 8 GB RAM, Window7 Home 64-bit operating system.

5.1. Test Problem Description. The frameworks of three tanks were used in the data set, but the behavior set was expanded, and the controllable event set was added. Using the simulation model to test, the current system has M states and N events, among which the number of normal events, observable events, and fault events is N_1, N_2, and N_3, respectively, the number of controllable events is N_c, the execution time is t_c, and the time of generating and acquiring observable events is t_o. Generating model G is complete and diagnosable, real-time observation timing is normal, and controllable event control is normal.

Input: Model with fault label $G_f = (Q \times L, E, T, I, F \times L)$
Real-time observation sequence Obs = <$\text{obs}_0, \text{obs}_1 \cdots \text{obs}_i \cdots$ >
Current diagnostic results Δ_i
Output: Repair results S_c
1. Initialization: S_c = null, $N_q = \varnothing$, $w = \varnothing$,$\Delta' = \varnothing$
2. while($|\Delta_i| > 1$)
3. $\forall s \in \Delta_i, N_q = \{q \mid s = <q_0 e_0 q_1 \cdots q>\}$
4. e_c = Choose(q) //Choose the path with the most different observations
5. Do(e_c)
6. w = Next(obs_i)
7. Δ' = Dia(Δ_i, w) //New diagnosis is obtained according to new observations
8. Δ_i = Delete(Δ_i) //Update candidate diagnostic results
9. while($l_q \neq \varnothing$)
10. S_c = Find(N_q, e_c) //Get the repair status of the path
11. if($\nexists S_c$)
12. Return Fail

Algorithm 5: Rec2.

Table 1: Time efficiency of diagnosis and planning methods.

Time (s)	Rec1		Rec2
	Calculation	Control	
$M = 10$	<0.001	1.06	1.24
$M = 50$	0.02	3.31	3.68
$M = 100$	0.49	6.11	7.02
$M = 1000$	1.98	13.76	15.99
$M = 5000$	4.39	39.20	44.11
$M = 10000$	9.83	72.15	108.65

In the process of running, the model randomly chooses the trajectory to advance and generates the corresponding observation as output. After receiving the observation, the diagnosis system diagnoses the model. After obtaining the fault, it is repaired according to algorithm Rec1 and Rec2, respectively. The efficiency of the Rec1 algorithm is divided into two parts: calculating the time and space efficiency of repairing path, the time efficiency of control event execution, and heuristic selection of controllable events with the most repair states. The efficiency of the Rec2 algorithm includes the overall test path, the time and space efficiency of the calculation and repair, and heuristic selection of controllable events that cause the most different observations.

5.2. Analysis of Experimental Results. The test is divided into two parts. Table 1 shows the efficiency of using algorithms Rec1 and Rec2, respectively, when $M : N_1 : N_2 : N_3 : N_c = 10 : 5 : 3 : 2 : 3$, $t_c = 1s$, $t_o = 0.2s$. The time efficiency of Rec1 is divided into two parts. The generated models may be different, so the time is the average value of multiple models.

The time of Rec1 is mainly spent on controllable events. If the time of controllable events is small, the choosing Rec1 is more reasonable. The time of Rec2 is mainly spent on controllable events and in waiting for further observation events. If the critical value n of system diagnosability is small, the choice of Rec2 is more reasonable.

Table 2 shows the ratio of state to event when $M = 1000$ and the time efficiency of diagnosis and repair under the same scale model when controllable events and observation time change. The former value is the time efficiency of using Rec1, and the latter value is the time efficiency of using Rec2.

Similarly, under the same proportion of event states, the time complexity of using Rec1 and Rec2 remains unchanged. When the ratio of $M : (N_1 + N_2 + N_3)$ is larger, the path branches are usually less and the diagnosis efficiency is higher. When the proportion of observable events is higher, the diagnostic efficiency is relatively high because there is relatively little backtracking. When the proportion of controllable events is higher, there are more candidate events. And there is less backtracking when using the Rec1 method, and the efficiency is higher. When the ratio of $M : (N_1 + N_2 + N_3)$ is small, there are more branches and the same diagnosis efficiency is relatively low.

5.3. Contrast of Heuristic Diagnosis and Incremental Diagnosis. We contrast the heuristic diagnosis with the landmark and the normal incremental diagnosis; three kinds of model are used: the normal model, the long chain with short local model, and the short chain with long local model. The results are shown in Figures 3–5. The blue line is the normal incremental method, and the red dotted line is the landmark method.

It can be seen from the comparison of Figures 3–5, the heuristic algorithm of using landmark improves the efficiency of diagnosis in different types of models. In incremental diagnosis, the new diagnoses begin from the former result. But when the former result was wrong, the real faults will not be obtained and backtrack will occur. The landmark is used to define the heuristic value to the diagnostic trajectories and reduce the time complexity.

Since the landmark is used to choose a trajectory, the backtrack caused by misdiagnosis is reduced. In the model according to Figure 4, more backtrack is avoided, so the method of the landmark is more efficient.

Table 2: Time efficiency with different percentages of different kinds of events.

$M : N_1 : N_2 : N_3 : N_c$	$t_c = 1$ $t_o = 0.2$	$t_c = 1$ $t_o = 0.5$	$t_c = 1$ $t_o = 1$	$t_c = 0.2$ $t_o = 0.2$	$t_c = 0.5$ $t_o = 0.2$
10 : 5 : 3 : 2 : 3	15.74/15.99	15.74/26.45	15.74/28.23	6.56/9.62	8.28/10.13
10 : 10 : 3 : 1 : 3	16.01/16.63	16.01/20.06	16.01/21.54	7.01/8.96	8.83/11.04
10 : 10 : 5 : 2 : 1	12.06/12.34	12.06/22.35	12.06/24.91	6.89/8.77	7.96/9.29

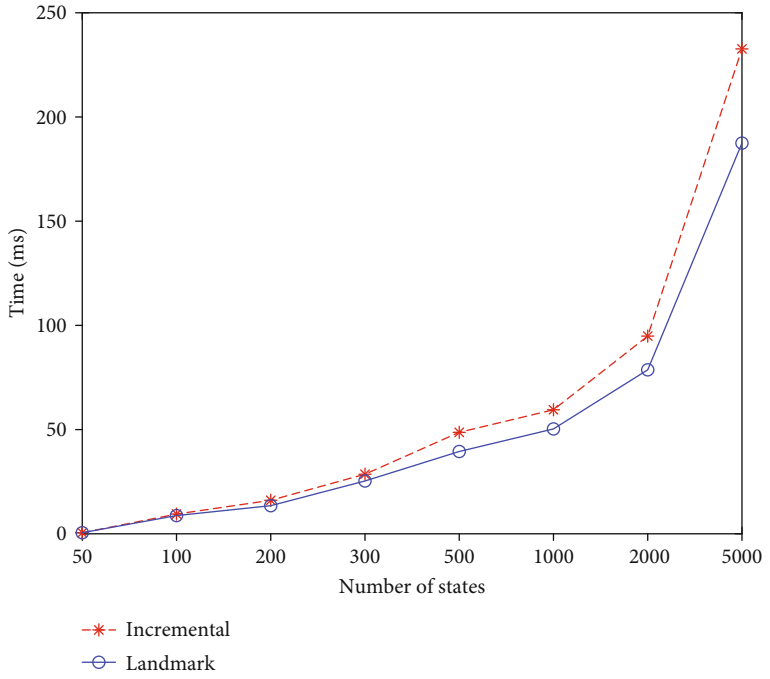


Figure 3: Contrast of basic model.

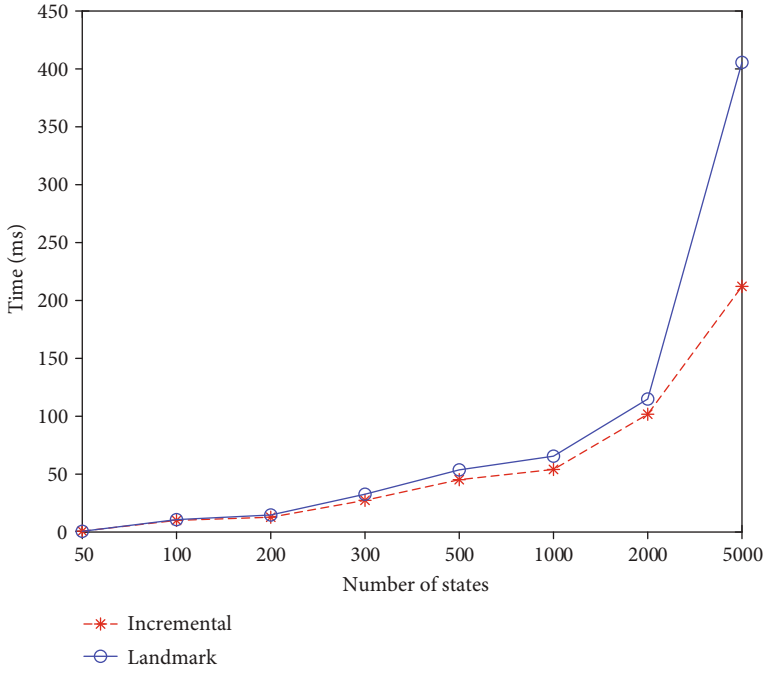


Figure 4: Contrast of long chain with short model.

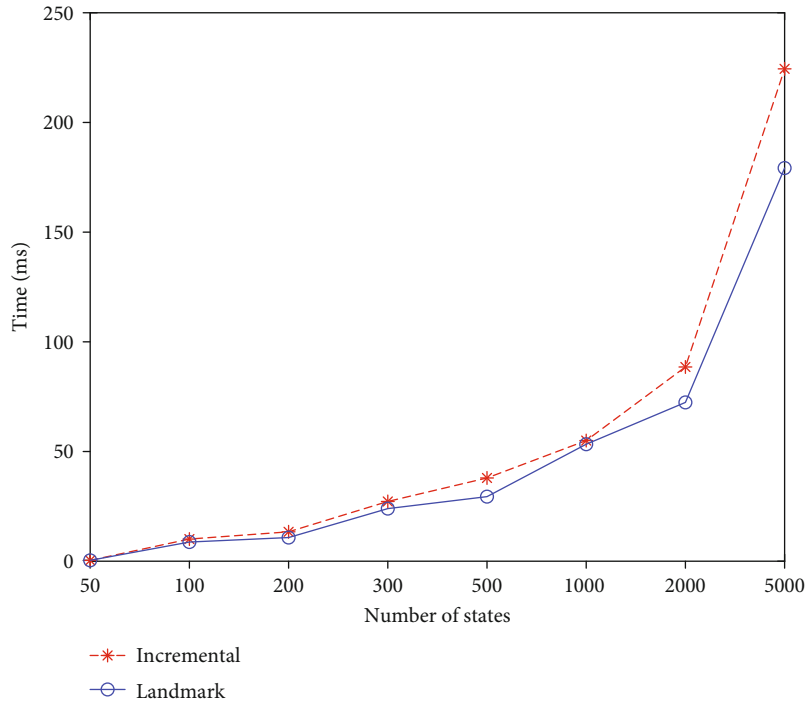


Figure 5: Contrast of short chain with long model.

6. Conclusion

Planning and model diagnosis are both symbolic branches of artificial intelligence. A method is proposed to merge the planning model and diagnosis model which is used to solve the fault testing and diagnosis acceleration. Based on the merged model, when an uncertain fault occurs, the landmark is used to generate a heuristic which leads to a better choice, and the controllable events are used to judge whether an event occurs or not. Two kinds of testing methods using controllable events are proposed to cater different cases. If the system has more expensive observation and less expensive controllable events, that means the system is easy to control and hard to observe; then, we use a controllable event to get a reaction. And another method is used to opposite case. A concept of a landmark in planning is used to generate a heuristic value, which accelerates diagnosis. When multiple diagnoses occur, the value will guide a better choice. If the testing method cannot get more observations or controllable events, the landmark can find a more precise result.

In the process of online diagnosis, the most possible path is marked with landmarks to improve diagnosis efficiency. After obtaining the diagnosis results, the landmarks are used to control the planning process and select the possible repair path or test points. A diagnosis-repair framework is proposed. However, if there are few controllable events, even there are no controllable events, the planning will not be able to backtrack, repair, and detect. This method cannot be implemented, which is a disadvantage.

Data Availability

We will not publish our data for open resource.

Conflicts of Interest

The authors declare that they have no conflict of interest.

Acknowledgments

This work was supported by a project grant from the National Natural Science Foundation of China (Grant Nos. 61502197, 61502199, and 61672261), the Science and Technology Project of "13th Five-Year" supported by the Jilin Provincial Education Department (Grant Nos. JJKH20191000KJ and JJKH20180763KJ), Science and Technology Development Project supported by Siping City of Jilin Province (Grant No. 2017093), and the Graduate Innovative Research Project of Jilin Normal University (Grant No. 201949).

References

[1] L. Travé-Massuyès, "Bridging control and artificial intelligence theories for diagnosis: a survey," *Engineering Applications of Artificial Intelligence*, vol. 27, pp. 1–16, 2014.

[2] R. Reiter, "A theory of diagnosis from first principles," *Artificial Intelligence*, vol. 32, no. 1, pp. 57–95, 1987.

[3] D. Benavides, S. Segura, and A. Ruiz-Cortés, "Automated analysis of feature models 20 years later: a literature review," *Information Systems*, vol. 35, no. 6, pp. 615–636, 2010.

[4] J. Y. Halpern, *Reasoning about uncertainty*, MIT Press, Cambridge, Massachusetts, USA, 2017.

[5] M. Sampath, R. Sengupta, S. C. Lafortune, K. Sinnamohideen, and D. Teneketzis, "Diagnosability of discrete-event systems," *IEEE Transactions on Automatic Control*, vol. 40, no. 9, pp. 1555–1575, 1995.

[6] M. Sampath, R. Sengupta, S. Lafortune, K. Sinnamohideen, and D. C. Teneketzis, "Failure diagnosis using discrete-event models," *IEEE Transactions on Control Systems Technology*, vol. 4, no. 2, pp. 105–124, 1996.

[7] L. K. Carvalho, Y.-C. Wu, R. Kwong, and S. Lafortune, "Detection and mitigation of classes of attacks in supervisory control systems," *Automatica*, vol. 97, pp. 121–133, 2018.

[8] A. G. C. Gonzalez, M. V. S. Alves, G. S. Viana, L. K. Carvalho, and J. C. Basilio, "Supervisory control-based navigation architecture: a new framework for autonomous robots in industry 4.0 environments," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 4, pp. 1732–1743, 2018.

[9] C. Edwards, S. K. Spurgeon, and R. J. Patton, "Sliding mode observers for fault detection and isolation," *Automatica*, vol. 36, no. 4, pp. 541–553, 2000.

[10] C. Cassandras and S. Lafortune, *Introduction to Discrete Event Systems*, vol. 11 of Kluwer International Series in Discrete Event Dynamic Systems, Kluwer Academic Publisher, Boston, MA, USA, 1999.

[11] H. Ibrahim, P. Dague, A. Grastien, L. Ye, and L. Simon, "Diagnosability planning for controllable discrete event systems," in *31st National Conference on Artificial Intelligence AAAI-17*, San Francisco, USA, 2017.

[12] W. Cheng, Y. Sun, G. Li, G. Jiang, and H. Liu, "Jointly network: a network based on CNN and RBM for gesture recognition," *Neural Computing and Applications*, vol. 31, no. S1, Supplement 1, pp. 309–323, 2019.

[13] B. Li, Y. Sun, G. Li et al., "Gesture recognition based on modified adaptive orthogonal matching pursuit algorithm," *Cluster Computing*, vol. 22, no. S1, Supplement 1, pp. 503–512, 2019.

[14] G. Li, D. Jiang, Y. Zhou, G. Jiang, J. Kong, and G. Manogaran, "Human lesion detection method based on image information and brain signal," *IEEE Access*, vol. 7, pp. 11533–11542, 2019.

[15] Y. He, G. Li, Y. Liao et al., "Gesture recognition based on an improved local sparse representation classification algorithm," *Cluster Computing*, vol. 22, no. S5, Supplement 5, pp. 10935–10946, 2019.

[16] D. Jiang, G. Li, Y. Sun, J. Kong, and B. Tao, "Gesture recognition based on skeletonization algorithm and CNN with ASL database," *Multimedia Tools and Applications*, vol. 78, no. 21, pp. 29953–29970, 2019.

[17] A. Grastien, M. O. Cordier, and C. Largouet, "Incremental diagnosis of discrete-event systems[C]," in *Proceedings of the 19th International Joint Conference on Artificial Intelligence*, pp. 1564–1665, Edinburgh, Scotland, 2005.

Hindawi
Wireless Communications and Mobile Computing
Volume 2020, Article ID 8877731, 10 pages
https://doi.org/10.1155/2020/8877731

Research Article

Multistep-Ahead Prediction of Urban Traffic Flow Using *Ga*TS Model

Benchao Wang,[1,2] Pan Qin,[1] and Hong Gu[1]

[1]*Faculty of Electronic Information and Electrical Engineering, Dalian University of Technology, Dalian 116024, China*
[2]*Department of Administration and Supervision, Liaoning Police College, Dalian 116036, China*

Correspondence should be addressed to Pan Qin; qp112cn@dlut.edu.cn

Received 27 April 2020; Revised 3 June 2020; Accepted 24 June 2020; Published 17 July 2020

Academic Editor: Wei Wang

The mathematical models for traffic flow have been widely investigated for a lot of application, like planning transportation and easing traffic pressure by using statistics and machine learning methods. However, there remains a lot of challenging problems for various reasons. In this research, we mainly focused on three issues: (a) the data of traffic flow are nonnegative, and hereby, finding a proper probability distribution is essential; (b) the complex stochastic property of the traffic flow leads to the nonstationary variance, i.e., heteroscedasticity; and (c) the multistep-ahead prediction of the traffic flow is often of poor performance. To this end, we developed a Gamma distribution-based time series (*Ga*TS) model. First, we transformed the original traffic flow observations into nonnegative real-valued data by using the Box-Cox transformation. Then, by specifying the generalized linear model with the Gamma distribution, the mean and variance of the distribution are regressed by the past data and homochronous terms, respectively. A Bayesian information criterion is used to select the proper Box-Cox transformation coefficients and the optimal model structures. Finally, the proposed model is applied to the urban traffic flow data achieved from Dalian city in China. The results show that the proposed *Ga*TS has an excellent prediction performance and can represent the nonstationary stochastic property well.

1. Introduction

As the main driving force of development, traffic has significant effects on the flow of production factors and the daily life of the urban system. The intelligent transportation system (ITS) can effectively provide innovative services relating to different modes of transport and traffic management [1, 2], such as transportation planning [3], traffic pressure easing [4], and traffic accident evaluation [5]. It enables transport networks to be more informed, more coordinated, and more efficient for various users. ITS requires a reliable prediction of traffic information in real time. Thus, how to accurately and timely predict traffic is a challenging task, which has gained more and more attention.

The traffic flow is full of complex dynamics and is stochastic [6], which make analysis and prediction mainly depend on timely or historical traffic data. As we illustrate in the latter, the urban traffic flow data is a nonstationary stochastic process with heterogeneous variance. Thus, the time series models are preferred for the prediction of traffic flow. On the other hand, the control operations, like variable speed limits (VSLs), are always embedded into ITS [7]. This fact suggests that the prediction models in ITS should be of concise structures natural to conduct the control and operation design. Thus, our studies focus on developing the data-driven time series model, which can predict the nonstationary distribution of the traffic flow and is of concise structures for the stochastic control design for ITS.

To construct the statistical model for the nonnegative traffic flow data, we first investigate which probability distribution is feasible for describing the uncertainty of the traffic flow. By detecting the change point of the traffic flow in 24 hours, we divide the traffic flow into four groups, whose distributions are separable from each other. According to the characters of the four groups, we proposed

Gamma distribution-based time series (*Ga*TS) models motivated by the generalized linear model [8].

We take the original observations and their Box-Cox transformation [9, 10] as the response variable, which can be considered as random variables generated by a Gamma distribution-based stochastic process. Moreover, we extracted the homochronous term from the historical observations used as explanatory variables.

We use the Bayesian information criterion (BIC) to select the optimal model structure. By these means, the proposed model can predict not only the mathematical expectation but also the nonstationary variance from the past observations. Furthermore, using the homochronous term makes our model of outstanding accuracy in multistep-ahead prediction. Finally, the real data collected from Dalian city in China are used to validate the performance of the proposed *Ga*TS. The computational results from the real-world data indicate that the homochronous term is helpful to enhance multistep-ahead prediction performance. Meanwhile, *Ga*TS is of the linear structure. Thus, *Ga*TS is efficient and convenient for further control design in ITS.

The rest of this paper is organized as follows: In Section 2, we review the studies on short-term traffic flow prediction. In Section 3, we present the *Ga*TS methodology for generating traffic flow data as building blocks for prediction. In Section 4, we discuss the experimental results. At last, concluding remarks are described in Section 5.

2. Literature Review

Over the past few decades, a lot of mathematical models have been developed by using statistics and machine learning methods for traffic flow prediction. The regressive type model, including the autoregressive models and the support vector regression (SVR), has been used as the parametric methods. The nonlinear model, like the artificial neural network (ANN) model, has also been applied to the prediction of the traffic flow. Besides the parametric methods, the nonparametric models, including the k-nearest neighbour (KNN) model, were also constructed.

In the family of the regression-type parametric models, the autoregressive integrated moving average (ARIMA) models were widely used for predicting the traffic flow [11–17]. Besides, the extensions of ARIMA have been studied for the prediction of traffic flow. The space-time autoregressive integrated moving-average model was proposed to fulfil the internal relationships of the links [18]. Stathopoulos and Karlaftis [19] designed a model for predicting the traffic congestion on the basis of a multivariate time-series state-space model. Meanwhile, SVR is also used for traffic prediction [20–23]. These regressive models mainly focused on the prediction of the tendency (mathematical expectation) of the traffic flow and ignored the statistics for the dispersion (variance).

The ANN model is one of the most commonly used nonlinear models in artificial intelligence methods. Most of the researchers proposed to apply ANN for the traffic prediction problem using a different architecture of the ANN models or to treat the ANN model as a baseline for comparing a wide variety of classification methods [7, 24–28]. It has been observed recently that many researchers in the field have proposed to integrate ANN with different preprocessing methods like fuzzy methods to improve its performances [29–32]. Wang et al. [33] proposed to design a prediction model for traffic flow by integrating a fuzzy ANN using the Taguchi method. This work employed the Taguchi method to fix up a count for sensors along the roadside. They proved the benefits of the information collected through the detectors. On similar lines, Quek et al. [34] proposed to utilize a fuzzy-based ANN to the problem of estimation of short-term traffic flow. The reported results indicated that the performance of the proposed model was promising in comparison to the backpropagation-based trained feed-forward (FF) ANN. The ANN and deep learning networks have achieved excellent performance on the traffic flow prediction problem [35–37]. However, they are relatively difficult to develop further traffic control designs for their complex structures.

Effectively modelling traffic flow variance can produce more accurate confidence intervals for short-term traffic flow forecasts and thus improves prediction reliability. Because the generalized autoregressive conditional heteroscedasticity (GARCH) model can be used to describe the time-varying volatility structure of the time series data, it was used by Kamarianakis et al. [38] for prediction of the conditional variance of speed with the mean equation of the ARIMA model. Similarly, GARCH was used for dependent variance prediction to 15 min volume based on a seasonal ARIMA model [39, 40]. Furthermore, Tsekeris and Stathopoulos used a fractionally integrated asymmetric power GARCH model with the mean equation of an autoregressive fractionally integrated moving average model for traffic volatility prediction and found that the combined model outperformed the ARIMA-GARCH model [41]. Because of the stochastic characteristics in traffic flow series, another volatility model, the stochastic volatility model, was proposed by Tsekeris and Stathopoulos [42] for urban traffic variability prediction. The evaluation results showed that the stochastic volatility model could produce a more accurate forecast speed variance than GARCH.

KNN is an essential method in the family of nonparametric methods. KNN has the ability to predict the sampled data based on a number k without formulating a model [43, 44]. KNN can most benefit the situation with little prior knowledge. Keeping its simplicity and better performance into consideration, the popularity of the applied algorithm is increasing in the field of traffic prediction. Yu et al. [45] proposed a KNN model for regression of estimation of multiple-time-step prediction. The parameters are measured for each minute by a loop detector. Hou et al. [46] presented a model for determining the flow of short-term traffic based on KNN. The major limitation of the algorithm is that it requires tremendous computational resources for a massive amount of historical data. The algorithm also suffers a limitation of the sensitivity of the outliers of archival data.

Besides the works mentioned above, the Kalman filtering method [47, 48], advanced techniques for kernel regression [49, 50], and mixtures of multivariate Gaussian processes [51] were also used to the prediction of the traffic flow.

3. Methodology

3.1. Data Description and Transformation. This research focused on urban traffic flow data. The data under study were collected at Donglian road, which is an important business centre of Dalian city of China. The microwave radar vehicle detectors were used to record the traffic flow every 15 minutes from Jan. 5, 2016, at 00:15 to Jan. 15, 2016, at 00:00, including a total of 960 samples (Figure 1). As shown in Figure 1, the data are of apparent periodicity, which suggests that the information in the corresponding period of the past days can be helpful for the prediction. This fact motivates us to improve the multistep-ahead prediction by using the homochronous term.

Figure 2 summarizes the boxplots for the corresponding sampling points of 10 days. The ranges of the boxes at each corresponding sampling point show that the variance of traffic flow is small at night and is large in the daytime. Thus, the traffic data is of time-varying variance, i.e., heteroscedasticity. Furthermore, we divide the data into four segments by using the change point detection method for the periodic time series [52]. Figure 3 illustrates the histograms for the four segments. Figures 3(a)–3(c) suggest that the distribution for the positive random variable, like the log-normal distribution and the Gamma distribution, can be used. However, Figure 3(d) is the histogram of the data collected from midnight to early morning and is of a single right tail, which cannot be approximated by the density function of the log-normal distribution. Thus, we use the Gamma distribution to build the time series model.

Furthermore, we use the Box-Cox transformation to find a proper positive real-valued time series data as the following:

$$y_t = \begin{cases} \dfrac{(x_t+1)^{\lambda}-1}{\lambda}, & \text{for } \lambda \neq 0, \\ \log(x_t+1), & \text{for } \lambda = 0. \end{cases} \tag{1}$$

We use BIC to select the proper λ.

3.2. Model Structure. Because the transformed y_t is nonnegative real-valued, we assume that y_t obeys a Gamma distribution. Let $f(y_t \mid \mu_t, \sigma_t)$ denote the probability density function, with μ_t and σ_t being, respectively, the location and scale parameters. Consequently, the conditional probability density of y_{t+j} can be formulated as the following:

$$f\left(y_{t+j} \mid \mu_t, \sigma_t\right) = \frac{1}{\left(\sigma_{t+j}^2 \mu_{t+j}\right)^{1/\sigma_{t+j}^2}} \frac{y_{t+j}^{\left(1/\sigma_{t+j}^2\right)-1} e^{-\left(y_{t+j}/\sigma_{t+j}^2 \mu_{t+j}\right)}}{\Gamma\left(1/\sigma_{t+j}^2\right)}, \tag{2}$$

with $\Gamma(\cdot)$ being the Gamma function.

The time-varying σ_t implies that the stochastic process generating y_t is nonstationary. To predict such nonstationary on the basis of historical data, μ_t and σ_t are regressed as follows:

$$\begin{cases} \mu_{t+j} = \beta_{10} + \beta_1^T u_{1t}, \\ \log\left(\sigma_{t+j}\right) = \beta_{20} + \beta_2^T u_{2t}, \end{cases} \tag{3}$$

where u_{it} is the explanatory vector given by $[y_t, \cdots, y_{t-l_{iy}}, c_{t+j}, \cdots, c_{t+j-l_{\mathrm{is}}}]^{\top}$ with l_{iy} and l_{is} being the maximum time lags of each variable for $i = 1, 2$. j means that (3) is used for j-step-head prediction. c_t is called the homochronous term, which is the mean of the observations at time t in the past five days up to the day containing y_t. Then, for the data set $\{(y_t, c_t) \mid t = 1, 2, \cdots, T\}$, the likelihood can be formulated as the following:

$$\begin{aligned} L(\mathscr{B} \mid) &= f(y_0, y_1, \cdots, y_t \mid \mathscr{B}, c_0, c_1, \cdots, c_t) \\ &= f(y_0, y_1, \cdots, y_{l-1}) \\ &\quad \times \prod_{t=l}^{T-j} f\left(y_{t+j} \mid \mathscr{B}, y_0, y_1, \cdots, y_{l-1}, c_0, c_1, \cdots, c_{t+j}\right) \\ &= f(y_0, y_1, \cdots, y_{l-1}) \times \prod_{t=l}^{T-j} f\left(y_{t+j} \mid \mathscr{B}, u_{it}\right), \end{aligned} \tag{4}$$

where $f(y_0, y_1, \cdots, y_{l-1})$ is the initial joint distribution and $\mathscr{B} = \{\beta_{i0}, \beta_i \mid i = 1, 2\}$ is the set of unknown parameters. Note that the initial joint distribution is not the function about $\mathscr{B}$. The parameter set $\mathscr{B}$ can be estimated by solving the following maximum likelihood estimation problem:

$$\widehat{\mathscr{B}} = \arg\max L(\mathscr{B}). \tag{5}$$

3.3. Evaluation Criteria for Prediction Performance. To comprehensively test the prediction performance of the models, several evaluation criteria are calculated. We use the mean absolute error (MAE) [53] and root mean square error (RMSE) [54] to show the scale of the prediction error:

$$\begin{aligned} \mathrm{MAE} &= \frac{\sum_{t=1}^{n} |y_t - \widehat{y}_t|}{n}, \\ \mathrm{RMSE} &= \sqrt{\frac{\sum_{t=1}^{n} (y_t - y\wedge_t)^2}{n}}. \end{aligned} \tag{6}$$

Because the above two criteria cannot be used to evaluate the models crossing the data sets, the coefficient of determination R^2 calculated from the observation y_t and estimated value $\widehat{y}_t$ is used as the following:

$$R^2 = 1 - \frac{\sum_t (y_t - y\wedge_t)^2}{\sum_t (y_t - \bar{y})^2}. \tag{7}$$

Here, $\bar{y}$ is the sample mean of y_t. From (7), we can know that R^2 does not consider the time-varying variance. Thus, it

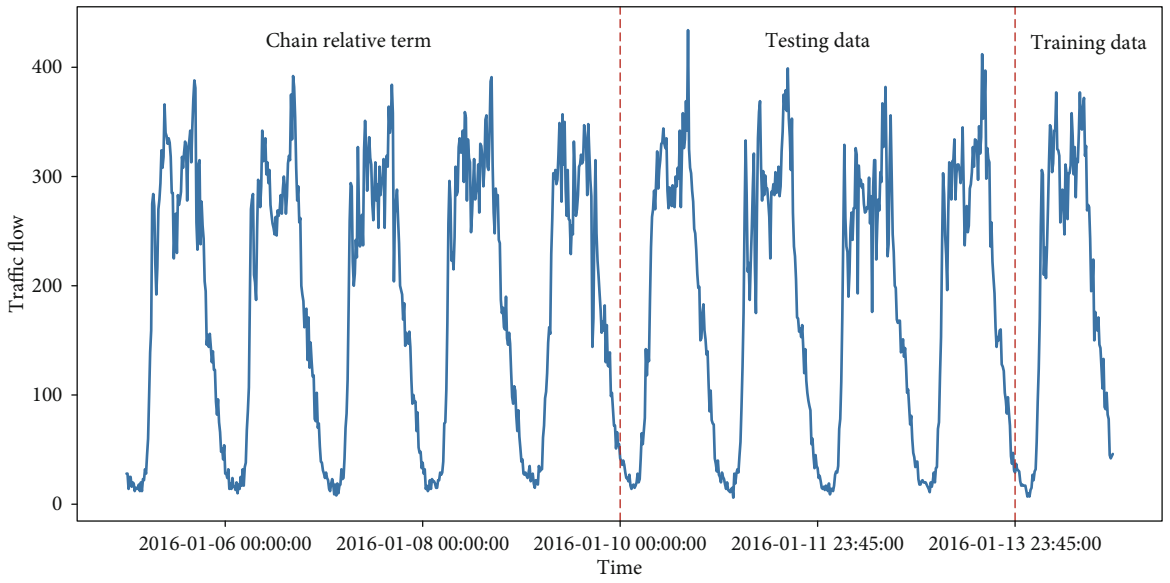


Figure 1: Time series of traffic flow in ten days. (Jan. 6, 2016, to Jan. 15, 2016).

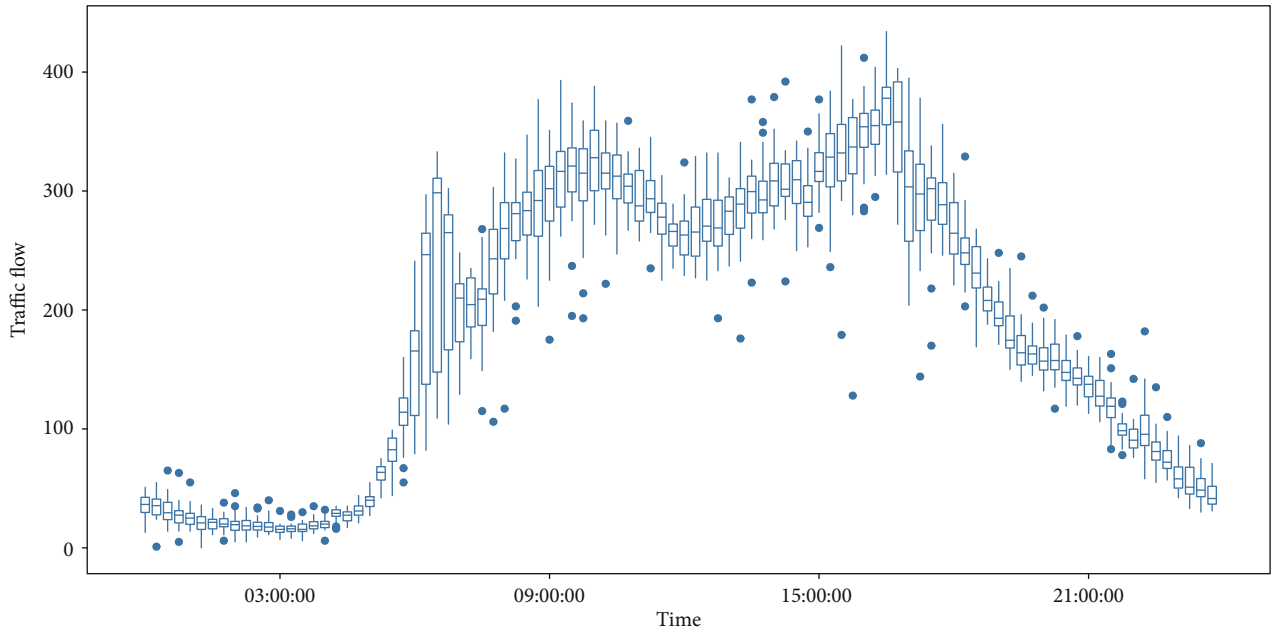


Figure 2: Boxplots for each sampling point in 24 hours summarized by 10-day time series.

is not proper for evaluating *Ga*TS models with a time-varying variation. To solve this problem, we use the adjusted coefficient of determination R_H^2 [55] as the following:

$$R_H^2 = 1 - \frac{\sum_t (y_t - y\wedge_t)^2 / \widehat{\sigma}_t^2}{\sum_t (y_t - \bar{y}_\omega)^2 / \widehat{\sigma}_t^2}, \tag{8}$$

with

$$\bar{y}_\omega = \frac{\sum_t y_t / \widehat{\sigma}_t^2}{\sum_t 1 / \widehat{\sigma}_t^2} \tag{9}$$

being weighted mean. Note that $R^2 = R_H^2$ if the scale parameter σ_t is estimated as constant.

The appropriate model structure determined by the time lags is crucial for prediction. Note that both R^2 and R_H^2 are monotonic increasing with the complexity of the model. Therefore, they cannot be used for model structure selection. Instead,

$$BIC = -2 \sum_{t=l}^{T-j} \log f\left(y_{t+j} \middle| \widehat{\mathscr{B}}, u_{it}\right) + \left|\widehat{\mathscr{B}}\right| \log\left(T - j - l + 1\right) \tag{10}$$

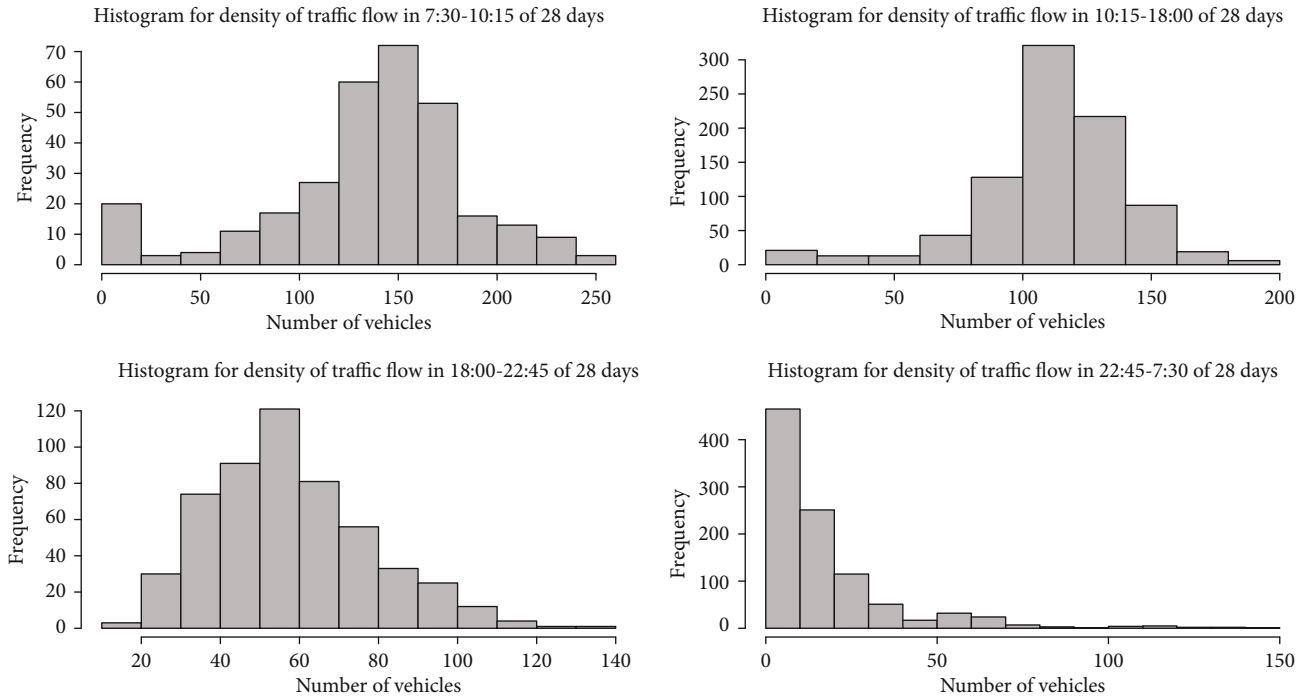


Figure 3: Histograms of four segments divided by changing points.

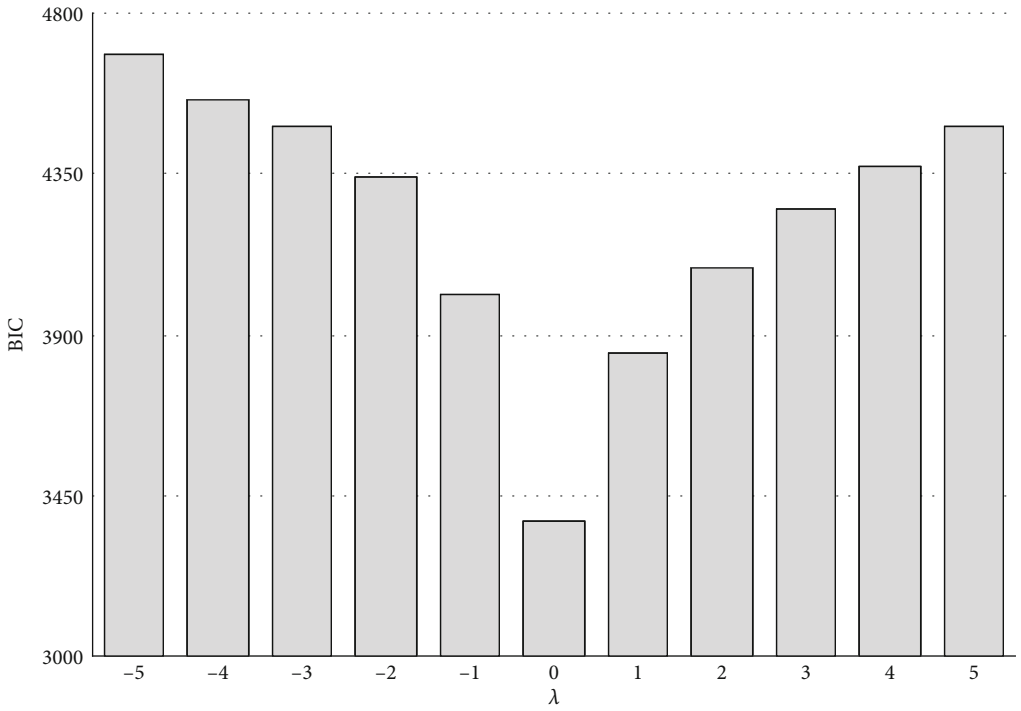


Figure 4: BIC values corresponding to several λ's for Box-Cox transformation.

is used to evaluate *Ga*TS, where $|\widehat{\mathscr{B}}|$ means the number of the estimated parameters. We computed BIC to evaluate different models. The optimal λ in the Box-Cox transformation illustrated in Figure 4 and the model structures with the minimum BIC are selected. For fairness, all the evaluations for various changes are conducted by using the original data. That is, we hold inverse transformation of the Box-Cox transformation (1) for the prediction of the transformed data, and all the evaluation criteria are computed by using the original data.

Table 1: Results of six models $\mathscr{M}_i$, $i = 1, 2, \cdots, 6$, from one- to four-step-ahead.

Prediction	Location μ_t		Scale σ_t		$R^2_{H\text{tr}}$	$R^2_{H\text{te}}$	MAE	RMSE	BIC
	y_t	c_t	y_t	c_t					
1-step ahead									
$\mathscr{M}_1$	9	×	×	×	0.913	0.9209	23.1128	35.2989	3624.776
$\mathscr{M}_2$	9	×	1	×	0.9758	0.9816	21.2866	30.8052	3553.057
$\mathscr{M}_3$	4	1	1	0	**0.9856**	**0.9906**	15.5009	**20.4802**	**3356.326**
$\mathscr{M}_4$	5	1	1	0	0.9794	0.9853	16.1826	21.9583	3466.493
$\mathscr{M}_5$	4	1	1	0	0.9804	0.9869	15.9534	21.5957	3443.754
$\mathscr{M}_6$	5	1	1	0	0.9828	0.9892	**15.5007**	20.7218	3370.927
2-step ahead									
$\mathscr{M}_1$	9	×	×	×	0.8054	0.7984	33.1687	56.364	3790.313
$\mathscr{M}_2$	9	×	1	×	0.9584	0.966	31.569	47.9216	3767.335
$\mathscr{M}_3$	4	1	1	0	**0.9835**	0.9884	**16.7219**	**22.9739**	**3416.928**
$\mathscr{M}_4$	5	1	1	0	0.9753	0.9795	18.6151	26.5485	3578.103
$\mathscr{M}_5$	4	1	1	0	0.9817	0.9845	17.1965	24.5677	3493.702
$\mathscr{M}_6$	4	1	1	0	0.9834	**0.9888**	16.8457	23.5447	3423.929
3-step ahead									
$\mathscr{M}_1$	9	×	×	×	0.6415	0.5814	41.5642	81.2282	3944.636
$\mathscr{M}_2$	9	×	1	×	0.7395	0.7138	36.8849	63.09	3901.592
$\mathscr{M}_3$	3	1	1	0	**0.9823**	**0.9869**	17.5917	**24.303**	**3422.813**
$\mathscr{M}_4$	4	1	1	0	0.9705	0.9748	19.194	27.9442	3598.335
$\mathscr{M}_5$	4	1	1	0	0.9778	0.9816	18.2077	25.8022	3505.349
$\mathscr{M}_6$	3	1	1	0	0.9822	0.9863	**17.3481**	24.7383	3431.816
4-step ahead									
$\mathscr{M}_1$	9	×	×	×	0.5572	0.5127	54.2951	115.5521	4042.217
$\mathscr{M}_2$	9	×	1	×	0.6141	0.5618	45.6247	77.9117	3993.805
$\mathscr{M}_3$	3	1	1	0	0.9856	**0.9859**	**17.9045**	**25.4079**	**3425.344**
$\mathscr{M}_4$	4	1	1	0	0.9769	0.9656	20.0474	28.9373	3603.713
$\mathscr{M}_5$	4	1	1	0	0.984	0.9779	18.8877	26.9388	3521.464
$\mathscr{M}_6$	3	1	1	0	**0.9864**	0.9858	18.0882	25.606	3434.907

4. Experiments

4.1. Model Structure Setting Up. Table 1 shows the one- to four-step-ahead (15×1 minutes to 15×4 minutes) prediction results obtained by three kinds of models. ($\{\mathscr{M}_i \mid i = 1, 2, 3\}$ defined by the autoregressive variable y_t and homochronous term c_t, where $\mathscr{M}_1$ is of time-invariant scale variable $\sigma_t = \sigma$, $\mathscr{M}_2$ and $\mathscr{M}_3$ are of time-varying scale parameter σ_t, and the homochronous term is applied to $\mathscr{M}_3$. For $\mathscr{M}_1$ to $\mathscr{M}_3$, the proper λ's for the Box-Cox transformation are selected by BIC. Figure 4 shows the BIC values for the λ candidates. $\mathscr{M}_4$ is the normal distribution-based model. $\mathscr{M}_5$ is *Ga*TS by using the original data without the Cox-Box transformation. $\mathscr{M}_6$ is the log-normal distribution-based time series model.) The 2nd to 5th columns are the maximum lags of each variable selected by BIC. Here, × means that the variables are unnecessary to be estimated; 0 means that the variables are not chosen by BIC. $R^2_{H\text{tr}}$ is calculated from the training data, and $R^2_{H\text{te}}$ is derived from the testing data.

4.2. Expectation Prediction. The difference between $R^2_{H\text{tr}}$ values and $R^2_{H\text{te}}$ values is minimal, which suggests that all the models have been well estimated without overfitting or underfitting. Bold numbers are the best values for each evaluation criterion. They indicate that $\mathscr{M}_3$'s with time-varying scale parameters and homochronous terms are the optimal models for one- to four-step-ahead predictions. Meanwhile, $R^2_{H\text{te}}$ values show that $\mathscr{M}_3$'s are of the best prediction performance. In $\mathscr{M}_3$'s, the selected regression structures for σ_t are more straightforward than those for μ_t. This is similar to the results in work [55]. The homochronous terms c_t represent the periodicity of μ_t rather than that of σ_t, which are

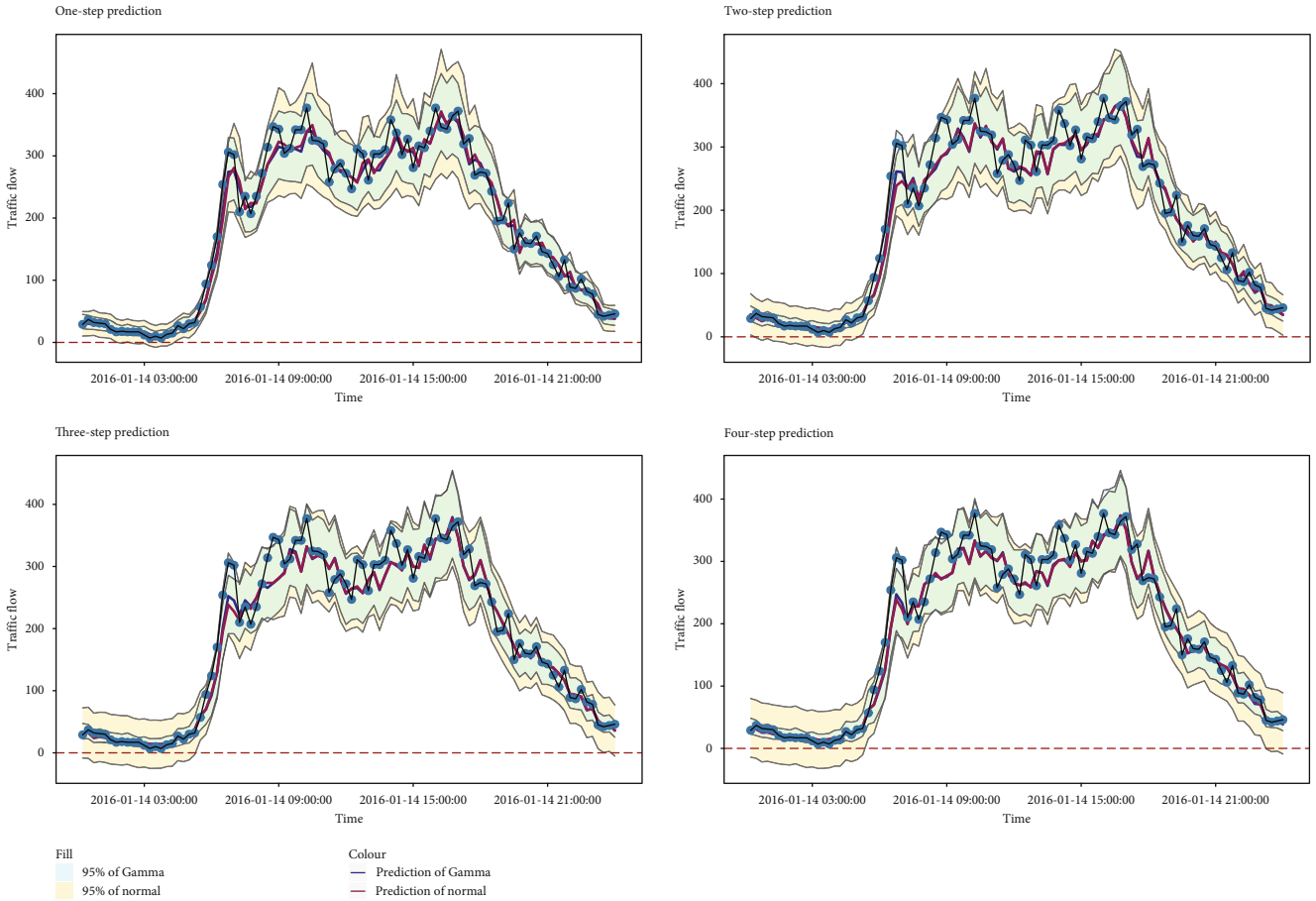


FIGURE 5: Range estimation by using Gamma and normal distribution-based models from one- to four-step-ahead.

the mean of the traffic flow at the same time of 5 successive days before the day containing prediction time t. Therefore, the homochronous terms have not been selected to regress to σ_t by BIC, as we expected.

We also construct normal distribution, *Ga*TS without the Cox-Box transformation and log-normal distribution-based models, denoted by $\mathscr{M}_4$'s, $\mathscr{M}_5$, and $\mathscr{M}_6$ in Table 1. By comparing $\mathscr{M}_i$'s for $i = 3, 4, 5, 6$, we find that $\mathscr{M}_3$'s are optimal models according to the minimum BIC values. This suggests that *Ga*TS are more optimal than the other distribution-based ones.

4.3. Range Estimates. Figures 5 and 6 illustrate the range estimation results for one day on Jan. 14, 2016, in which 95% confidence intervals (CI) are obtained by the predicted $\widehat{\mu}_t$ and $\widehat{\sigma}_t$ along with t. In Figure 5, the green fields are collected by four $\mathscr{M}_3$'s in Table 1, and the yellow areas are obtained by the corresponding normal distribution-based models. Figure 5(a) shows that *Ga*TS have a similar CI prediction performance, compared with the normal distribution-based models, for one-step-ahead prediction. However, Figures 5(b)–5(d) indicate that the *Ga*TS-based models are of a more narrow range than the normal distribution-based models for multistep-ahead prediction. Furthermore, the normal distribution-based models for multistep-ahead prediction even obtained the negative lower bound of CI, when the traffic flow values are small. This is contradicting to the fact that the traffic flow is positive-valued.

In Figure 6, the green fields are also collected by four $\mathscr{M}_3$'s in Table 1, and the yellow areas are obtained by the corresponding log-normal distribution-based models with the identical structures of four $\mathscr{M}_3$'s. From Figures 5(a)–5(d), we can see that both the two models of exponential transformation are more stabilized than using normal distribution directly from one- to four-step. Therefore, the variance of the log-normal distribution is a wider range than the Gamma distribution in the daytime, which is contrary to the real traffic flow state. That will influence the estimation and prediction of the traffic state by the ITS. Furthermore, the CI obtained by *Ga*TS can well approximate the heteroscedasticity shown in Figure 2. Thus, *Ga*TS can achieve more rational CI predictions.

5. Conclusion

This research mainly focused on the prediction issue on the urban traffic flow. By specifying GLM with the Gamma distribution, we proposed *Ga*TS to predict the nonstationary stochastic process of the traffic flow. The objective of *Ga*TS is to predict the probability distribution of the traffic flow in real time. To this end, the Gamma distribution presents the stochastic properties of nonnegative-valued traffic flow

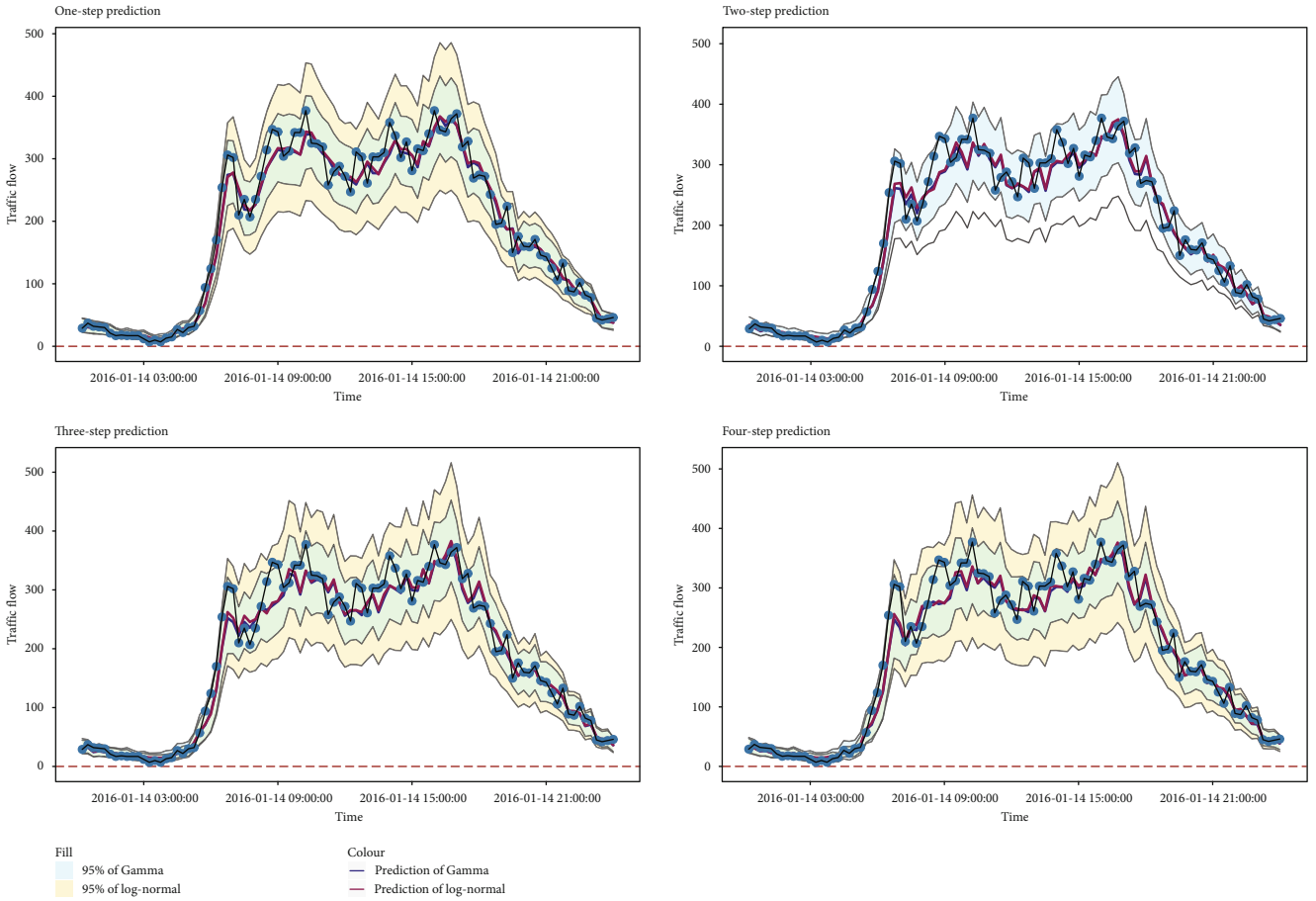


Figure 6: Range estimation by using Gamma and log-normal distribution-based models from one- to four-step-ahead.

data and one-side-tailed histogram for the midnight period. The GLM structure supports the heteroscedasticity of the traffic flow. The homochronous term improves the accuracy of the multistep-ahead prediction of the mathematical expectation.

The traffic flow data in this research were collected from Dalian, which is a large port in northern China, as well as a major destination for Chinese tourists [56]. The aggregation of a large number of different types of crowds not only brings traffic and environmental problems but also makes the state of the use of the region's urban public space become complicated and contradictory [57]. Because it is relatively difficult to improve transportation infrastructure, we are focusing on developing intelligent software control and management to relieve traffic congestion. ITS need more accurate historical information and future prediction of the road network [2]. Furthermore, to control and regularize the traffic flow, the model with precise accuracy should be of simple structures. Thus, our proposed *Ga*TS is more proper than the research focus on the models of ANN and deep learning. Furthermore, a series of *Ga*TS can be extended to model the joint distribution for the joint prediction of the multiple sensors.

Several research topics can be further considered. We successfully specified GLM by using the Gamma distribution for the urban zone. However, the probability distributions for other zones, like the highways, should be further summarized. The potential external factors, which have a relation with the traffic flow and can be governed by ITS, should be investigated to be embedded into *Ga*TS. On the basis of *Ga*TS with the external factors, the cost function for the control and regularization of the traffic flow should be constructed, and the corresponding optimization solver should be developed.

Data Availability

The data that support the findings of this study are available from the ITS database of the traffic police department in Dalian city of China. But restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data are however available from the authors upon reasonable request and with permission of the traffic police department.

Conflicts of Interest

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

Acknowledgments

This research was funded by the National Natural Science Foundation of China (Grant Nos. U1560102, 61633006, and 61502074).

References

[1] D. Pavlyuk, "Feature selection and extraction in spatiotemporal traffic forecasting: a systematic literature review," *European Transport Research Review*, vol. 11, no. 1, 2019.

[2] H. Li, Z. Li, R. White, and X. Wu, "A real-time transportation prediction system," *Applied Intelligence*, vol. 39, no. 4, pp. 793–804, 2013.

[3] W. Ma and Z. Qian, "On the variance of recurrent traffic flow for statistical traffic assignment," *Transportation Research Part C: Emerging Technologies*, vol. 81, pp. 57–82, 2017.

[4] S. Sarkar, S. Chawla, S. Ahmad et al., "Effective urban structure inference from traffic flow dynamics," *IEEE Transactions Big Data*, vol. 3, no. 2, pp. 181–193, 2017.

[5] P. Greibe, "Accident prediction models for urban roads," *Accident Analysis & Prevention*, vol. 35, no. 2, pp. 273–285, 2003.

[6] S. G. Wang, D. X. Yu, X. G. Ma, and X. Xing, "Analyzing urban traffic demand distribution and the correlation between traffic flow and the built environment based on detector data and POIs," *European Transport Research Review*, vol. 10, no. 2, 2018.

[7] Z. H. Khan and T. A. Gulliver, "A macroscopic traffic model for traffic flow harmonization," *European Transport Research Review*, vol. 10, no. 2, 2018.

[8] J. A. Nelder and R. W. Wedderburn, "Generalized linear models," *Journal of the Royal Statistical Society. Series A (General)*, vol. 135, no. 3, p. 370, 1972.

[9] G. E. P. Box and D. R. Cox, "An analysis of transformations," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 26, no. 2, pp. 211–252, 1964.

[10] R. M. Sakia, "The Box-Cox transformation technique: a review," *The Statistician*, vol. 41, no. 2, pp. 169–178, 1992.

[11] S. Ishak and H. Al-Deek, "Performance evaluation of short-term time-series traffic prediction model," *Journal of Transportation Engineering*, vol. 128, no. 6, pp. 490–498, 2002.

[12] B. M. Williams, P. K. Durvasula, and D. E. Brown, "Urban freeway traffic flow prediction: application of seasonal autoregressive integrated moving average and exponential smoothing models," *Transportation Research Record*, vol. 1644, no. 1, pp. 132–141, 1998.

[13] B. M. Williams and L. A. Hoel, "Modeling and forecasting vehicular traffic flow as a seasonal ARIMA process: theoretical basis and empirical results," *Journal of Transportation Engineering*, vol. 129, no. 6, pp. 664–672, 2003.

[14] S. V. Kumar and L. Vanajakshi, "Short-term traffic flow prediction using seasonal ARIMA model with limited input data," *European Transport Research Review*, vol. 7, no. 3, 2015.

[15] H. Zhang, X. Wang, J. Cao, M. Tang, and Y. Guo, "A multivariate short-term traffic flow forecasting method based on wavelet analysis and seasonal time series," *Applied Intelligence*, vol. 48, no. 10, pp. 3827–3838, 2018.

[16] X. Luo, L. Niu, and S. Zhang, "An algorithm for traffic flow prediction based on improved SARIMA and GA," *KSCE Journal of Civil Engineering*, vol. 22, no. 10, pp. 4107–4115, 2018.

[17] H. Zhang, X. Wang, J. Cao, M. Tang, and Y. Guo, "A hybrid short-term traffic flow forecasting model based on time series multifractal characteristics," *Applied Intelligence*, vol. 48, no. 8, pp. 2429–2440, 2018.

[18] W. Min and L. Wynter, "Real-time road traffic prediction with spatio-temporal correlations," *Transportation Research Part C: Emerging Technologies*, vol. 19, no. 4, pp. 606–616, 2011.

[19] A. Stathopoulos and M. G. Karlaftis, "A multivariate state space approach for urban traffic flow modeling and prediction," *Transportation Research Part C: Emerging Technologies*, vol. 11, no. 2, pp. 121–135, 2003.

[20] M. Castro-Neto, Y. S. Jeong, and M. K. Jeong, "Online-SVR for short-term traffic flow prediction under typical and atypical traffic conditions," *Expert Systems with Applications*, vol. 36, no. 3, pp. 6164–6173, 2009.

[21] W. C. Hong, "Traffic flow forecasting by seasonal SVR with chaotic simulated annealing algorithm," *Neurocomputing*, vol. 74, no. 12-13, pp. 2096–2107, 2011.

[22] M. W. Li, W. C. Hong, and H. G. Kang, "Urban traffic flow forecasting using Gauss-SVR with cat mapping, cloud model and PSO hybrid algorithm," *Neurocomputing*, vol. 99, pp. 230–240, 2013.

[23] Y. S. Jeong, Y. J. Byon, M. M. Castro-Neto, and S. M. Easa, "Supervised weighting-online learning algorithm for short-term traffic flow prediction," *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, no. 4, pp. 1700–1707, 2013.

[24] G. Zhang, B. E. Patuwo, and M. Y. Hu, "Forecasting with artificial neural networks:: The state of the art," *International Journal of Forecasting*, vol. 14, no. 1, pp. 35–62, 1998.

[25] J. Mackenzie, J. F. Roddick, and R. Zito, "An evaluation of htm and lstm for short-term arterial traffic flow prediction," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 5, pp. 1847–1857, 2019.

[26] G. G. Zhang, C. Zhang, and W. D. Zhang, "Evolutionary echo state network for long-term time series prediction: on the edge of chaos," *Applied Intelligence*, vol. 50, no. 3, pp. 893–904, 2020.

[27] M. S. Dougherty and M. R. Cobbett, "Short-term inter-urban traffic forecasts using neural networks," *International Journal of Forecasting*, vol. 13, no. 1, pp. 21–31, 1997.

[28] W. Wang, J. Chen, J. Wang, J. Chen, J. Liu, and Z. Gong, "Trust-enhanced collaborative filtering for personalized point of interests recommendation," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 9, pp. 6124–6132, 2020.

[29] T. Zhou, G. Han, X. Xu, C. Han, Y. Huang, and J. Qin, "A learning-based multimodel integrated framework for dynamic traffic flow forecasting," *Neural Processing Letters*, vol. 49, no. 1, pp. 407–430, 2019.

[30] G. Shen, X. Kong, and X. Chen, "Short-term traffic flow intelligent hybrid forecasting model and its application," *Journal of Control Engineering and Applied Informatics*, vol. 13, no. 3, pp. 65–73, 2011.

[31] J. Z. Zhu, J. X. Cao, and Y. Zhu, "Traffic volume forecasting based on radial basis function neural network with the consideration of traffic flows at the adjacent intersections," *Transportation Research Part C: Emerging Technologies*, vol. 47, pp. 139–154, 2014.

[32] A. Stathopoulos, L. Dimitriou, and T. Tsekeris, "Fuzzy modeling approach for combined forecasting of urban traffic flow," *Computer-Aided Civil and Infrastructure Engineering*, vol. 23, no. 7, pp. 521–535, 2008.

[33] X. X. Wang, L. H. Xu, and K. X. Chen, "Data-driven short-term forecasting for urban road network traffic based on data processing and LSTM-RNN," *Arabian Journal for Science and Engineering*, vol. 44, no. 4, pp. 3043–3060, 2019.

[34] C. Quek, M. Pasquier, and B. B. S. Lim, "POP-TRAFFIC: a novel fuzzy neural approach to road traffic analysis and prediction," *IEEE Transactions on Intelligent Transportation Systems*, vol. 7, no. 2, pp. 133–146, 2006.

[35] J. Gao, Y. L. Murphey, and H. Zhu, "Multivariate time series prediction of lane changing behavior using deep neural network," *Applied Intelligence*, vol. 48, no. 10, pp. 3523–3537, 2018.

[36] G. Yu and J. Liu, "A hybrid prediction approach for road tunnel traffic based on spatial-temporary data fusion," *Applied Intelligence*, vol. 49, no. 4, pp. 1421–1436, 2019.

[37] X. Ma, Z. Tao, Y. Wang, H. Yu, and Y. Wang, "Long short-term memory neural network for traffic speed prediction using remote microwave sensor data," *Transportation Research Part C: Emerging Technologies*, vol. 54, pp. 187–197, 2015.

[38] Y. Kamarianakis, A. Kanas, and P. Prastacos, "Modeling traffic volatility dynamics in an urban network," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1923, no. 1, pp. 18–27, 2005.

[39] J. H. Guo, W. Huang, and B. M. Williams, "Adaptive Kalman filter approach for stochastic short-term traffic flow rate prediction and uncertainty quantification," *Transportation Research Part C: Emerging Technologies*, vol. 43, no. 1, pp. 50–64, 2014.

[40] W. Huang, W. W. Jia, J. Guo et al., "Real-time prediction of seasonal heteroscedasticity in vehicular traffic flow series," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 10, pp. 3170–3180, 2018.

[41] C. Chen, J. Hu, and Q. Meng, "Short-time traffic flow prediction with ARIMA-GARCH model," in *2011 IEEE Intelligent Vehicles Symposium (IV)*, pp. 607–612, Baden-Baden, Germany, June 2011.

[42] T. Tsekeris and A. Stathopoulos, "Short-term prediction of urban traffic variability: stochastic volatility modeling approach," *Journal of Transportation Engineering*, vol. 136, no. 7, pp. 606–613, 2010.

[43] H. H. Xie, X. H. Dai, and Y. Qi, "Improved K nearest neighbor algorithm for short-term traffic flow prediction," *Journal of Traffic and Transportation Engineering (English Edition)*, vol. 14, no. 3, pp. 87–94, 2014.

[44] Z. Zheng and D. Su, "Short-term traffic volume forecasting: A _k_ -nearest neighbor approach enhanced by constrained linearly sewing principle component algorithm," *Transportation Research Part C: Emerging Technologies*, vol. 43, no. SI, pp. 143–157, 2014.

[45] B. Yu, X. Song, F. Guan, Z. Yang, and B. Yao, "K-nearest neighbor model for multiple-time-step prediction of short-term traffic condition," *Journal of Transportation Engineering*, vol. 142, no. 6, p. 04016018, 2016.

[46] X. Hou, Y. Wang, and S. Hu, "Short-term traffic flow forecasting based on two-tier K-nearest neighbor algorithm," *Procedia - Social and Behavioral Sciences*, vol. 96, pp. 2529–2536, 2013.

[47] W. He, C. Zhong, M. A. Sotelo, X. Chu, X. Liu, and Z. Li, "Short-term vessel traffic flow forecasting by using an improved Kalman model," *Cluster Computing*, vol. 22, no. S4, pp. 7907–7916, 2019.

[48] A. Emami, M. Sarvi, and S. A. Bagloee, "Using Kalman filter algorithm for short-term traffic flow prediction in a connected vehicle environment," *Journal of Modern Transportation*, vol. 27, no. 3, pp. 222–232, 2019.

[49] D. Ngoduy, "Kernel smoothing method applicable to the dynamic calibration of traffic flow models," *Computer-Aided Civil and Infrastructure Engineering*, vol. 26, no. 6, pp. 420–432, 2011.

[50] R. Huang and S. Sun, "Kernel regression with sparse metric learning," *Journal of Intelligent Fuzzy Systems*, vol. 24, no. 4, pp. 775–787, 2013.

[51] S. Sun, "Infinite mixtures of multivariate Gaussian processes," in *2013 International Conference on Machine Learning and Cybernetics*, pp. 1011–1016, Tianjin, China, July 2013.

[52] R. Lund, X. L. Wang, Q. Q. Lu, J. Reeves, C. Gallagher, and Y. Feng, "Changepoint detection in periodic and autocorrelated time series," *Journal of Climate*, vol. 20, no. 20, pp. 5178–5190, 2007.

[53] C. J. Willmott and K. Matsuura, "Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance," *Climate research*, vol. 30, no. 1, pp. 79–82, 2005.

[54] R. J. Hyndman and A. B. Koehler, "Another look at measures of forecast accuracy," *International Journal of Forecasting*, vol. 22, no. 4, pp. 679–688, 2006.

[55] P. Qin and R. Nishii, "Statistical Prediction of Dst Index by Solar Wind Data and t -Distributions," *IEEE Transactions on Plasma Science*, vol. 43, no. 11, pp. 3908–3915, 2015.

[56] L. Guo and H. Li, "Effect of foreign trade and FDI on Dalian's upgrading of industrial structure-an analysis based on the empirical data in 1990-2010," *Journal of Lanzhou Commercial College*, vol. 5, p. 20, 2012.

[57] N. He and S. C. Zhao, "Urban road traffic impedance function - Dalian city case study," *Journal of Highway and Transportation Research and Development (English Edition)*, vol. 8, no. 3, pp. 90–95, 2014.

Hindawi
Wireless Communications and Mobile Computing
Volume 2020, Article ID 8865110, 11 pages
https://doi.org/10.1155/2020/8865110

Research Article

Superresolution Reconstruction of Video Based on Efficient Subpixel Convolutional Neural Network for Urban Computing

Jie Shen,[1] **Mengxi Xu**,[2] **Xinyu Du**,[1,3] **and Yunbo Xiong**[1]

[1]*College of Computer and Information, Hohai Univerisity, Nanjing 211100, China*
[2]*College of Computer Engineering, Nanjing Institute of Technology, Nanjing 211167, China*
[3]*Bank of Shanghai, Shanghai 200120, China*

Correspondence should be addressed to Jie Shen; shenjie_2003045@hhu.edu.cn

Received 22 April 2020; Revised 24 May 2020; Accepted 9 June 2020; Published 11 July 2020

Academic Editor: Wei Wang

Video surveillance is an important data source of urban computing and intelligence. The low resolution of many existing video surveillance devices affects the efficiency of urban computing and intelligence. Therefore, improving the resolution of video surveillance is one of the important tasks of urban computing and intelligence. In this paper, the resolution of video is improved by superresolution reconstruction based on a learning method. Different from the superresolution reconstruction of static images, the superresolution reconstruction of video is characterized by the application of motion information. However, there are few studies in this area so far. Aimed at fully exploring motion information to improve the superresolution of video, this paper proposes a superresolution reconstruction method based on an efficient subpixel convolutional neural network, where the optical flow is introduced in the deep learning network. Fusing the optical flow features between successive frames can compensate for information in frames and generate high-quality superresolution results. In addition, in order to improve the superresolution, a superpixel convolution layer is added after the deep convolution network. Finally, experimental evaluations demonstrate the satisfying performance of our method compared with previous methods and other deep learning networks; our method is more efficient.

1. Introduction

Superresolution reconstruction is generating high-resolution results from the low-resolution images using construction models. In contrast to hardware, algorithm-based image construction can efficiently generate high-resolution images at a low cost. Due to a large number of images and video samples, more and more researchers focus on image reconstruction.

Recently, superresolution reconstruction methods have been proposed. For example, Farsiu et al. [1] proposed a regularized reconstruction method for solving inverse problems of ill-posed problems. A fast superresolution reconstruction algorithm is proposed for the pure translational motion and space invariant ambiguity [2]. The principle of neighborhood embedding [3] is to make the assumption that the local spatial structure between the low-resolution image block and the high-resolution image block is similar. The low-resolution image block represents the low-dimensional data, and the high-resolution image block represents the high-dimensional space. Therefore, the low-resolution image can be linearly represented and mapped to the high-dimensional image block to reconstruct the high-resolution image. Yu and Zhang proposed an improved glowworm swarm optimization algorithm for superresolution reconstruction of video images [4]. Combined with the characteristics of superresolution reconstruction, the algorithm's swarm input, firefly's luciferase, and location update equation are redefined; the optimization objective function criterion is set.

Although the above methods can solve some problems of superresolution reconstruction, the disadvantage of the neighborhood embedding method is that the number k of image blocks selected for low-resolution image blocks is artificially specified, which will affect the reconstruction effect by the supervisor and may cause the phenomenon of underfitting and overfitting. Therefore, we proposed a

new superresolution reconstruction method based on optical flow and efficient subpixel convolutional neural network (ESPCN), which can solve the traditional reconstruction methods' defects.

The rest of the paper is organized as follows. The related work is discussed in Section 2. The methodology is described in Section 3. The details of our method are proposed in Section 4. Section 5 introduced the experiments and result analysis. Finally, the conclusions related to the paper are outlined in Section 6.

2. Related Works

There are three kinds of superresolution methods: interpolation method, reconstruction method, and learning-based method.

The interpolation method is to estimate the missing position of pixels by using the prior in original low-resolution images. Herein, the key of interpolation operation is to establish the mapping relationship between low-resolution and high-resolution images. In general, the interpolation methods include three categories: nearest-neighbor interpolation [5], bilinear interpolation [6], and bicubic interpolation [7]. However, the disadvantage of interpolation-based methods lies in the low-quality reconstruction results that commonly suffer from textual noises.

The reconstruction-based method is to model the degeneration process of image samples and generate a high-resolution image by inverse transformation. The degeneration model establishment is the key, which can fuse multiple images to reconstruct high-resolution images. There are three typical methods in the reconstruction methods: iterative back projection [8], projections onto convex sets [9], and maximum posterior probability [10]. In 2011, Zhang et al. improved the method based on convex set projection for sequential frame image [11] and applied the model to the image sequence of panoramic transformation. This method realizes the geometric transformation of the sequence frame accurately and improves the speed of reconstruction. In 2012, Wallach et al. improved the map method [12] and improved the contrast and resolution of a reconstructed image by fusing global information.

Learning-based methods intend to learn the correlation between the low-resolution and high-resolution images. Three phases are included in these methods: feature extraction, feature learning, and reconstruction. Typical methods include neighbor embedding [3] and sparse coding [13–15]. The representative method based on sparse representation is the superresolution image reconstruction method [16] proposed by Yang et al. in 2008. High- and low-resolution dictionaries are trained for LR image and HR image resolution, so that all low-resolution images to be reconstructed can extract sparse representation from dictionaries. But the style and edge of the reconstructed image need to be similar to the data set used in learning, and the quality and efficiency of reconstruction are not satisfactory.

In recent years, deep learning-based methods have demonstrated excellent performance on superresolution reconstruction [17, 18]. Based on the sparse coding method, Dong et al. first proposed a superresolution method using a deep convolutional network, which is termed as the superresolution convolutional neural network (SRCNN) [17]. A 3-layer convolutional neural network (CNN) was designed to learn the mapping relation from the low-resolution to high-resolution images. Aiming to accelerate the reconstruction process, Dong et al. further proposed a fast superresolution convolutional neural network (FSRCNN) that was based on an hourglass-shaped CNN comprising more layers than previous methods but had fewer parameters. The advantage of FSRCNN lies in that it eliminates the requirement of the image magnification and increases the efficiency of the reconstruction process.

Shi et al. proposed an efficient subpixel convolutional neural network method [18] that extracted features in the low-resolution domain and replaced the classic bicubic upsampling operation with an efficient subpixel convolution. Moreover, ESPCN was characterized by the subpixel convolution that generated the feature map in r^2 channel. Through periodic activation in subpixel convolution, the feature map that was of the size $H \times W \times r^2$ was reconstructed to generate the high-resolution images with the size $rH \times rW \times 1$. These image features were extracted by the hidden layers in the deep architecture. In practice, ESPCN was more efficient in contrast to other methods, thus had more opportunities to cater to the real-time tasks. Talab et al. used ESPCN and CNN for super-low-resolution face recognition [19].

In order to expand the receptive field for image reconstruction, Kim et al. proposed a very deep superresolution reconstruction method (VDSR) that increased the kernel size from 13×13 to 41×41 [20]. The sparsity property was used to accelerate the convergence process. Besides, VDSR could generate superresolution results with multiple scales. In contrast to a previous method, VDSR could capture more details of images to improve the performance of the resolution reconstruction.

Ledig et al. introduced the generative adversarial network (GAN) to the superresolution reconstruction and proposed the superresolution generative adversarial network (SRGAN) [21]. SRGAN comprised of a generative block and an adversarial block, where the first was used for reconstruction and the second was used to classify the quality of the reconstructed images. Although SRGAN could not generate competitive results in the PSNR scores, it could obtain results that were in line with the real images, providing a visually satisfying effect.

Sun et al. introduced and updated the residual network (ResNet) into the superresolution reconstruction [22]. Herein, the number of residual layers was increased from 16 to 32, expanding the model scale to generate better results.

3. Methodology

To improve the superresolution reconstruction method in speed and quality, this paper proposed a new method based on optical flow and ESPCN. The motion information between frames is considered in the reconstruction process to improve the reconstruction quality. The optical flow can calculate the motion of an object in a very small time [23].

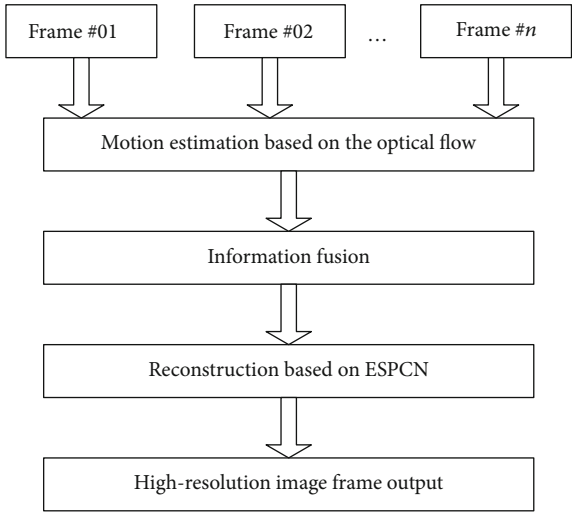


Figure 1: The framework of our superpixel reconstruction method.

It is suitable for motion estimation between video frames. Therefore, the optical flow performs well in accuracy and effectiveness compared with other motion estimation methods. Because of the subpixel convolution layer, ESPCN has very little additional computational cost compared to other deep learning methods. ESPCN can be used in real-time to improve the efficiency of urban computing and intelligence. In specific, the optical flow feature is extracted, which is jointly combined with the image to input into the network. Finally, the subpixel convolutional layer is applied to produce a high-resolution image from the feature map. The framework of our method is shown in Figure 1; there are three phases: motion information estimation, information fusion, and reconstruction.

Optical flow is defined as the instant velocity of the pixels including those in moving objects [24–26]. For the video sequence, the optical flow extraction is realized by estimating the changes between two frames. The correlation between successive frames is estimated to identify the changes in the time axis. The optical flow feature is represented as the two-dimensional vector that demonstrates the ratio of the intensity change.

Assume there are L layers in the ESPCN network, the front L-1 layers can be expressed as follows:

$$\begin{aligned} f^1\left(\mathbf{I}^{\mathrm{LR}}; W_1, b_1\right) &= \phi\left(W_1 * \mathbf{I}^{\mathrm{LR}} + b_1\right), \\ f^1\left(\mathbf{I}^{\mathrm{LR}}; W_{1:l}, b_{1:l}\right) &= \phi\left(W_l * f^{l-1}\left(\mathbf{I}^{\mathrm{LR}}\right) + b_l\right). \end{aligned} \tag{1}$$

In which $W_l b_l$ are the weight and bias of each layer, respectively; the value of l is between 1 and L-1; W_l is a two-dimensional convolution tensor with size of $n_{l-1} \times n_l \times k_l \times k_l$; n_l is the feature dimension of the lth layer; the dimension of n_0 is the channel number C, k_l is the convolution kernel size of the lth layer; ϕ is the activate function; and the last layer of the network f^L maps the low-resolution image to the high-resolution image $\mathbf{I}^{\mathrm{SR}}$.

The subpixel convolutional method is first proposed in ESPCN where the convolutional kernel is used to activate different parts of images. When shifting the kernel, the subpixels are periodically activated according to their locations which can be mathematically defined as follows:

$$\mathbf{I}^{\mathrm{SR}} = f^L\left(\mathbf{I}^{\mathrm{LR}}\right) = PS\left(W_L * f^{L-1}\left(\mathbf{I}^{\mathrm{LR}}\right) + b_L\right), \tag{2}$$

where $f^L()$ and $f^{L-1}()$ are the mapping functions in the L and $L-1$ layers, SR and LR denote the terms of the superresolution and low-resolution, W_L is the kernel in the L layer, b_L is the moderation factor, and $PS()$ is devoted to the periodical transformation which can map the feature map to the high-resolution image:

$$PS(T)_{x,y,c} = T_{[x/r],[y/r],C\cdot r\cdot \mathrm{mod}\ (y,r)+C\cdot r\cdot \mathrm{mod}\ (x,r)+c}, \tag{3}$$

where x and y locate the pixels in the high-resolution images; the size of the kernel is $n_{L-1} \times r^2 C \times k_L \times k_L$. Since the nonlinear mapping layer is eliminated in the bottom of the deep convolutional network, the subpixel convolution is operated on the low-resolution image when $k_L = (k_s/r)$ and $\mathrm{mod}\ (k_s, r) = 0$.

We use the mean squared error (MSE) as the training strategy. The computing method is as follows:

$$\ell(W_{1:L}, b_{1:L}) = \frac{1}{r^2 HW} \sum_{x=1}^{rH} \sum_{x=1}^{rW} \left(\mathbf{I}_{x,y}^{\mathrm{HR}} - f_{x,y}^{L}\left(\mathbf{I}^{\mathrm{LR}}\right)\right)^2. \tag{4}$$

Meanwhile, we used the tanh function as the activation function, by introducing nonlinearity into the network with the tanh function; the output of the upper network is mapped to the input of the lower network; and the expression ability of the neural network is enhanced. The tanh function is defined as

$$\tanh\ (x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}. \tag{5}$$

tanh function is a kind of hyperbolic function with fast convergence.

Moreover, the subpixel convolution can eliminate the initial interpolation process, which can generate a satisfying reconstruction result with low cost. This architecture nicely adapts to the superresolution reconstruction of the video with a large number of frames.

4. Algorithm

The phases of the superresolution reconstruction are shown in Figure 2. Successive 5 frames are selected as the input for our deep convolutional network. Here, we define the time scale of the third frame as n, while the time scales of the previous two frames are $n-2$ and $n-1$; the two time scales of the last two frames are $n+2$ and $n+1$. The motion feature estimation is based on the frame n; the optical flow calculation is operated on the frames $n-2$, $n-1$, $n+1$, and $n+2$ to generate the feature maps for these frames. Then, these feature maps are combined with the original frames to obtain

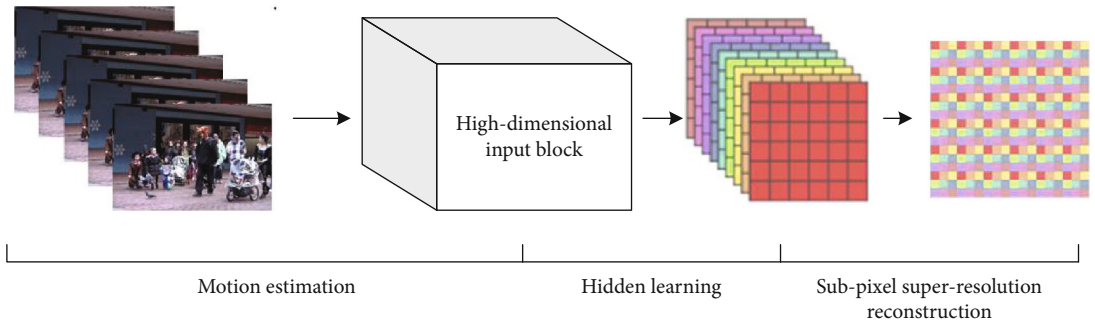


Figure 2: The phases of our algorithm.

Figure 3: The video frames in the training database: (a) news broadcast, (b) cars, (c) a patrol boat, (d) plants, and (e) a rugby game.

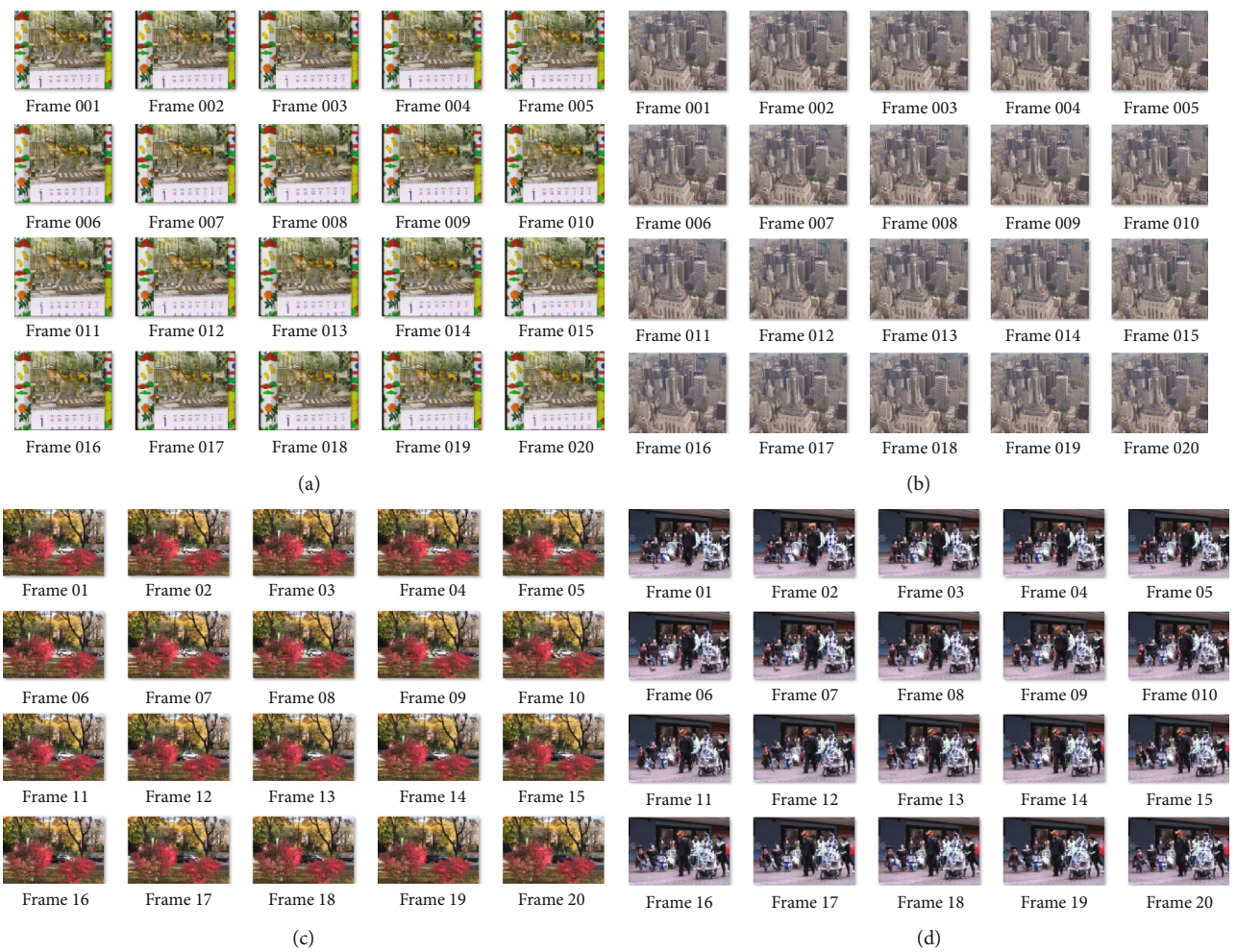


Figure 4: The testing database: (a) calendar, (b) building, (c) foliage, and (d) walk.

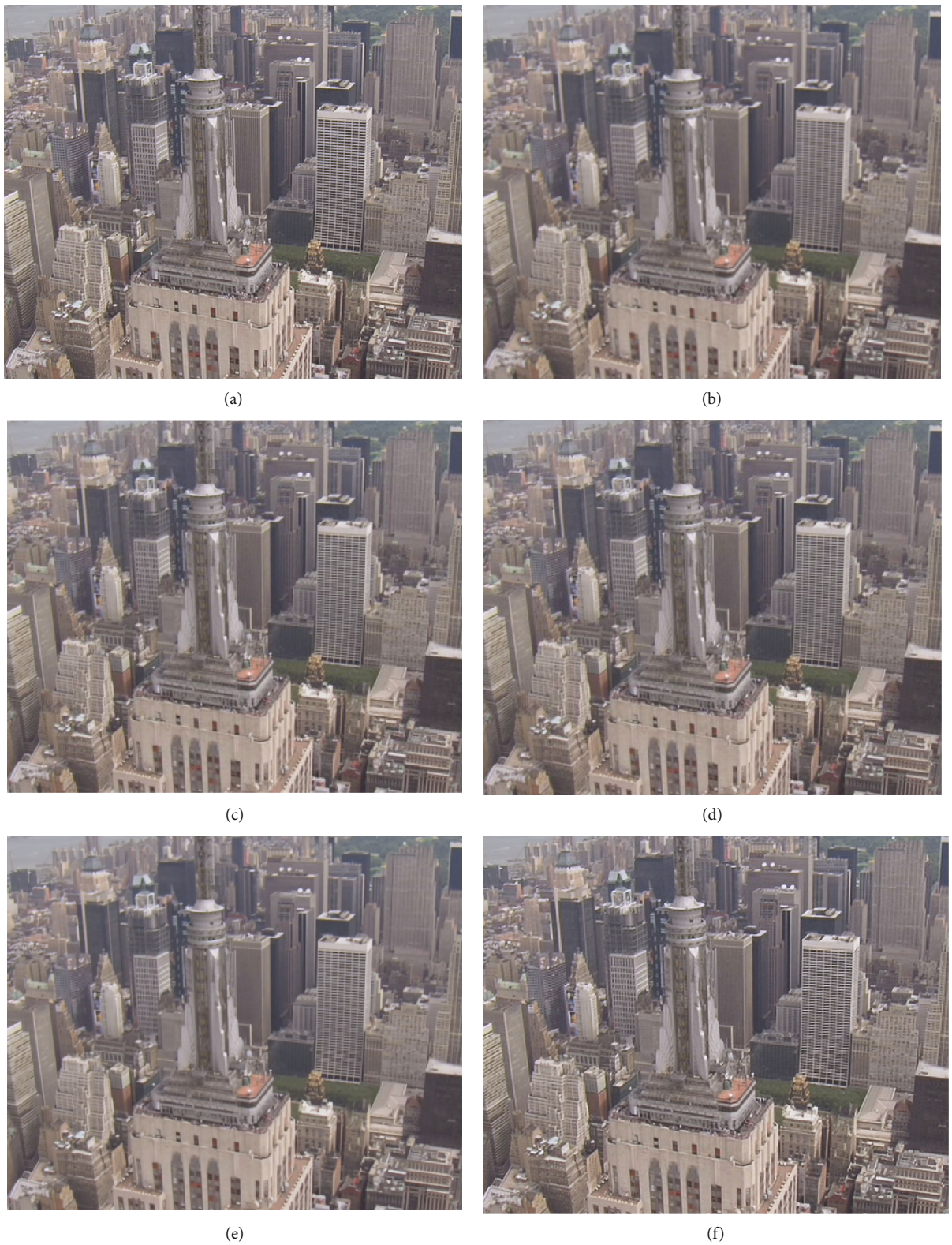
(a) (b) (c) (d) (e) (f)

Figure 5: Superresolution reconstruction for the walk video: (a) original frame, (b) bicubic interpolation, (c) POCS, (d) sparse coding, (e) SRCNN, and (f) ours.

the input block of the deep convolutional network where the subpixel convolutional layer is applied to reconstruct the superresolution images [27].

The architecture of the deep convolutional network is constructed with a 4-layer structure that comprises 3 convolutional layers and 1 subpixel convolutional layer. We transfer the color space of the original frame from the RGB space to the Ycb space. The first layer comprises 64 kernels that are $5 \times 5 \times 15$, which generate the 64-channel feature maps. The second layer comprises 32 kernels that are 3×3, which generate the 32-channel feature maps. $3 \times r^2$ kernels that are of 3×3 are included in the third layer that generates the $3 \times r^2$ feature map. The superresolution reconstruction is realized by the last subpixel convolutional layer.

In specific, we assume that the size of frames is $H \times W$, where H is the height of the frame and W is the width. Accordingly, the dimension of the input RGB frames is $H \times W \times 3$. In our model, successive 5 frames are jointly



(a) (b) (c) (d) (e) (f)

Figure 6: Superresolution reconstruction for the building video: (a) original frame, (b) bicubic interpolation, (c) POCS, (d) sparse coding, (e) SRCNN, and (f) ours.

considered to estimate the optical flow feature. As a result, the size of the input block is $H \times W \times (3 \times 5)$, and the size of the filter in the first layer is $f_1 \times f_1 \times 15$ and $f_1 = 5$. After the mapping obtained by 4 hidden layers, $H \times W \times r^2$ feature maps are generated to further reconstruct the high-resolution images for the frame n.

5. Experimental Evaluations

5.1. Database. Training data is collected from the Xiph database which includes 10 video sequences with the length of 2695 frames [28]. The resolution of the frame is 144×176. The training data shown in Figure 3 contains different contents in urban video surveillance, such as news broadcast, cars, a patrol boat, plants, and a rugby game. For training video sequences, the video segmentations are extracted and degenerated by adding the Gaussian noise and the downsampling processes, which can generate the low-quality and low-resolution samples. These samples are then inputted into the model. It is noticed that the input training data of the convolutional neural network is an image block composed of five consecutive frames after motion estimation. The output of the convolutional neural network is a high-resolution image. Meanwhile, the network generates the errors with the reference frame to update the network. Finally, the superresolution reconstruction model is going to be achieved after the network convergence.

The testing database shown in Figure 4 is given by the VideoSet4 database with 4 videos: calendar, buildings, foliage, and walk [29]. The building video includes 34 frames with the size of 704×576; the calendar video includes 41 calendar images with the size of 720×576. In this training sample, the framework contains a large number of characters and regular textures, which can be evaluated by the details of edge reconstruction. Foliage video shows traffic scenes including vehicles and streets. The resolution of the original frame is 720×480. Moreover, the contents of the foliage are quite complicated which can evaluate the robustness of our method against the dynamic noises. The walk video includes the walking pedestrians; the dataset of this video comprises 47 frames with the size of 720×480. This dataset is characterized by slowly moving objects, thus can evaluate the performance of our method for motion estimation.

5.2. Parameter Setting. The input block is the combination of the motion estimation and original images, which is 15-dimensional. With respect to the network architecture, the first layer is constructed by 64 kernels with the size of $5 \times 5 \times 15$ and generates 64-dimensional feature maps. The second layer is constructed by 32 kernels which are applied to handle the feature map extracted by the first layer and generate 32-dimensional feature maps. The third layer comprises $3 \times r^2$ kernels with the size of 3×3, which generates $3 \times r^2$-dimensional feature maps. At last, the subpixel convolutional layer is used to reconstruct the superresolution images. The learning ratio is set as 0.001, the maximum epoch number is 100, and the batch size of the network is 32.

5.3. Experimental Comparisons. The comparison is divided into two parts. The first is comparison of traditional methods and the proposed method. The second is comparison between superresolution with motion estimation and without motion estimation. The data for experimental evaluation is selected from VideoSet4. The PSNR and time cost are selected as the metrics. The PSNR is calculated as follows:

Table 1: PSNR scores given by different methods (dB).

Methods	City	Calendar	Foliage	Walk
Bicubic interpolation	25.84	20.30	24.13	26.47
POCS	28.04	21.87	26.94	29.98
Sparse coding	28.53	22.09	27.33	30.58
SRCNN	30.23	23.39	28.78	32.43
Our method	30.44	23.59	29.00	32.61

Table 2: Improvement of PSNR (dB) value of video frame by motion estimation.

Methods	City	Calendar	Foliage	Walk
Bicubic interpolation	25.84	20.30	24.13	26.47
Our method	30.44	23.59	29.00	32.61
ESPCN	30.29	23.56	28.83	32.49

$$\mathrm{PSNR} = 10 \log_{10}\left(\frac{\mathrm{MAX}_I^2}{\mathrm{MSE}}\right), \tag{6}$$

in which MAX is the image threshold and MSE is the mean square error.

5.3.1. The First Comparison. Traditional methods include the bicubic interpolation, POCS, and the sparse coding-based superresolution reconstruction methods, while the compared deep learning-based method is the SRCNN.

Figures 5 and 6 qualitatively present the results for frame #13 in the walk video and frame #8 in the building video. From the results, we can find that the reconstruction results obtained by traditional methods, i.e., sparse coding and POCS-based methods, are relatively more blurred in contrast to the results obtained by our methods. For example, our reconstructed images can clearly distinguish the details of the contour of the pedestrians and the texture of the baby carriage. Moreover, a significant advantage of our method can be demonstrated in Figure 5. In Figure 5, the windows of the building generate many complicated textures which cause periodic noise in the results obtained by compared methods. However, this effect is largely removed by our method due to the use of the motion features.

Table 1 quantitatively shows the PSNR scores obtained by different methods. From Table 1, we can find that our method generates better reconstruction results with respect to the PSNR scores. In contrast to the traditional reconstruction methods, the advantage of our method is that motion feature learning can improve the correctness of reconstruction. Although SRCNN and our methods are commonly based on convolutional operations, our method is the best. Different from the SRCNN, motion information is estimated and introduced in our method, which is the reason for our better performance.

5.3.2. The Second Comparison. We use the same video training set to train them. Table 2 shows the PSNR of four videos of VideoSet4 reconstructed by two different models. In the case of two times of magnification, it can be seen that the

(a) (b) (c) (d) (e)

Figure 7: Comparison of different reconstruction methods to reconstruct the 19th frame of foliage: (a) original frame, (b) LR images, (c) bicubic interpolation, (d) ESPCN, and (e) ours.

algorithm without motion estimation of adjacent frames is generally about 5 dB higher than bicubic interpolation. The average increase of our method is about 0.12 compared with SRCNN (without motion estimation).

Figure 7 shows the reconstruction quality of the 19th frame of the foliage sequence in VideoSet4 under different algorithms. From the results, it can be found that the biggest difference between the ESPCN model with motion estimation and the ESPCN model without motion estimation is that for moving objects, adding motion estimation can produce better reconstruction effect, which is more obvious in high-speed moving objects. In Figure 7, the black car with the midview motion has obvious distortion and halo (Figure 7(d)). In Figure 7(e), in the reconstruction image using motion estimation, the edge of the black car is relatively straight and clear. This is because the new model takes into account the motion information before and after the frame, and it can better consider the temporal coherence of the video than the single frame image with single information.

(a) (b) (c) (d) (e)

Figure 8: Comparison of different reconstruction methods to reconstruct the 31st frame of calendar: (a) original frame, (b) LR images, (c) bicubic interpolation, (d) ESPCN, and (e) ours.

The same phenomenon is shown in Figure 8. On the 31st frame of the calendar video sequence, the characters in the bicubic interpolation image calendar in Figure 7(c) are blurred. In Figure 8(d), although the superresolution method of the single frame image improves the character edge to a certain extent, there is a certain degree of dislocation between smaller characters. In Figure 7(e), in the reconstructed image of the model combined with motion estimation, the edges of the smaller characters become more separated and the deformation degree is lower. The difference between foliage and calendar is that foliage mainly reflects the main motion of the object, while calendar reflects the relative motion of the object caused by the lens motion, which shows that the motion estimation has a certain effect on the two kinds of motion.

6. Conclusions

Aimed at improving the performance of the reconstruction of videos, this paper proposes deep convolutional network-based reconstruction methods where the motion information is extracted and introduced. Moreover, the proposed method introduces the subpixel convolution, which can significantly speed up the reconstruction process. Experimental results demonstrate that our method generates better reconstruction results in contrast to previous methods, the SRCNN, and the ESPCN. The proposed method can be applied to hardware architecture composed of integrated circuit chips such as digital signal processor (DSP) and programmable system on chip (SOPC) in the form of embedded program. So, our method can be configured in the front-end device of video surveillance. In the future, it will be used as a form of edge intelligent and provide a feasible method for reducing the computing load of the centre system of urban computing, and we will evaluate the contribution of other types of motion information for superresolution reconstruction.

Data Availability

The 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 partly by the National Natural Science Foundation of China (Nos. 51979085 and 61903124), the Fundamental Research Funds for the Central Universities (No. B200202186), and the National Natural Science Foundation of China (No. 61563036).

References

[1] S. Farsiu, M. D. Robinson, M. Elad, and P. Milanfar, "Fast and robust multiframe super resolution," *IEEE Transactions on Image Processing*, vol. 13, no. 10, pp. 1327–1344, 2004.

[2] N. K. Bose, M. K. Ng, and A. C. Yau, "A fast algorithm for image super-resolution from blurred observations," *EURASIP Journal on Advances in Signal Processing*, vol. 2006, no. 1, Article ID 035726, 2006.

[3] H. Chang, D. Y. Yeung, and Y. Xiong, "Super-resolution through neighbor embedding," in , Article ID I-I*Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004*, vol. 1, Washington, DC, USA, June 2004.

[4] W. Yu and M. Zhang, "Super resolution reconstruction of video images based on improved glowworm swarm optimization algorithm," in *2018 IEEE 3rd International Conference on Image, Vision and Computing (ICIVC)*, pp. 331–335, Chongqing, China, June 2018.

[5] W. T. Freeman, T. R. Jones, and E. C. Pasztor, "Example-based super-resolution," *IEEE Computer Graphics and Applications*, vol. 22, no. 2, pp. 56–65, 2002.

[6] J. D. van Ouwerkerk, "Image super-resolution survey," *Image and Vision Computing*, vol. 24, no. 10, pp. 1039–1052, 2006.

[7] D. Glasner, S. Bagon, and M. Irani, "Super-resolution from a single image," in *2009 IEEE 12th International Conference on Computer Vision*, pp. 349–356, Kyoto, Japan, September 2009.

[8] S. Dai, M. Han, Y. Wu, and Y. Gong, "Bilateral back-projection for single image super resolution," in *Multimedia and Expo, 2007 IEEE International Conference on*, pp. 1039–1042, Beijing, China, July 2007.

[9] C. Fan, C. Wu, G. Li, and J. Ma, "Projections onto convex sets super-resolution reconstruction based on point spread function estimation of low-resolution remote sensing images," *Sensors*, vol. 17, no. 2, p. 362, 2017.

[10] G. K. Chantas, N. P. Galatsanos, and N. A. Woods, "Super-resolution based on fast registration and maximum a posteriori reconstruction," *IEEE Transactions on Image Processing*, vol. 16, no. 7, pp. 1821–1830, 2007.

[11] Z. Zhang, X. Wang, J. Ma, and G. Jia, "Super resolution reconstruction of three view remote sensing images based on global weighted POCS algorithm," in *2011 International Conference on Remote Sensing, Environment and Transportation Engineering*, Nanjing, China, June 2011.

[12] D. Wallach, F:. Lamare, G. Kontaxakis, and D. Visvikis, "Super-resolution in respiratory synchronized positron emission tomography," *IEEE Transactions on Medical Imaging*, vol. 31, no. 2, pp. 438–448, 2012.

[13] J. Yang, J. Wright, T. S. Huang, and Y. Ma, "Image super-resolution via sparse representation," *IEEE Transactions on Image Processing*, vol. 19, no. 11, pp. 2861–2873, 2010.

[14] Z. Chen, Y. Zhang, C. Wu, and B. Ran, "Understanding individualization driving states via latent Dirichlet allocation model," *IEEE Intelligent Transportation Systems Magazine*, vol. 11, no. 2, pp. 41–53, 2019.

[15] C. Xu, "A novel recommendation method based on social network using matrix factorization technique," *Information Processing & Management*, vol. 54, no. 3, pp. 463–474, 2018.

[16] J. Yang, J. Wright, T. Huang, and Y. Ma, "Image super-resolution as sparse representation of raw image patches," in *2008 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–8, Anchorage, AK, USA, June 2008.

[17] K. Umehara, J. Ota, N. Ishimaru et al., "Super-resolution convolutional neural network for the improvement of the image quality of magnified images in chest radiographs," in

Proceedings Volume 10133, Medical Imaging 2017: Image Processing, Orlando, Florida, USA, February 2017.

[18] W. Shi, J. Caballero, F. Huszár et al., "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1874–1883, Las Vegas, NV, USA, June 2016.

[19] M. A. Talab, S. Awang, and S. A. M. Najim, "Super-low resolution face recognition using integrated efficient sub-pixel convolutional neural network (ESPCN) and convolutional neural network (CNN)," in *2019 IEEE International Conference on Automatic Control and Intelligent Systems (I2CACIS)*, pp. 331–335, Selangor, Malaysia, June 2019.

[20] J. Kim, J. K. Lee, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1646–1654, Las Vegas, NV, USA, June 2016.

[21] C. Ledig, L. Theis, F. Huszár et al., "Photo-realistic single image super-resolution using a generative adversarial network," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4681–4690, Honolulu, HI, USA, July 2017.

[22] B. Sun, J. Lu, and X. Wei, "Image super-resolution reconstruction based on deep residual network," in *Proceedings of the 2018 2nd International Conference on Artificial Intelligence: Technologies and Applications (ICAITA 2018)*, Osaka, Japan, March 2018.

[23] W. L. Zhong, *Research on Application of Convolution Neural Network in Video Super Resolution*, vol. 5, University of Electronic Science and Technology, 2018.

[24] E. Ilg, N. Mayer, T. Saikia, M. Keuper, A. Dosovitskiy, and T. Brox, "Flownet 2.0: evolution of optical flow estimation with deep networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2462–2470, Honolulu, HI, USA, July 2017.

[25] X. Zhu, S. Zhang, R. Hu, Y. Zhu, and J. Song, "Local and global structure preservation for robust unsupervised spectral feature selection," *IEEE Transactions on Knowledge and Data Engineering*, vol. 30, no. 3, pp. 517–529, 2018.

[26] W. Zheng, X. Zhu, Y. Zhu, R. Hu, and C. Lei, "Dynamic graph learning for spectral feature selection," *Multimedia Tools and Applications*, vol. 77, no. 22, pp. 29739–29755, 2018.

[27] X. Y. Du, *Research on Super-Resolution Reconstruction Method Based on Deep Learning*, vol. 5, Hohai University, 2019.

[28] P. V. Vu, *'New Approaches and a Subjective Database for Video Quality Assessment', Doctoral Dissertation*, Oklahoma State University, 2015.

[29] D. Li and Z. Wang, "Video superresolution via motion compensation and deep residual learning," *IEEE Transactions on Computational Imaging*, vol. 3, no. 4, pp. 749–762, 2017.

Hindawi
Wireless Communications and Mobile Computing
Volume 2020, Article ID 8861886, 9 pages
https://doi.org/10.1155/2020/8861886

Research Article

An Encoder-Decoder Network Based FCN Architecture for Semantic Segmentation

Yongfeng Xing,[1,2] **Luo Zhong**,[1] **and Xian Zhong**[1]

[1]*School of Computer Science and Technology, Wuhan University of Technology, Wuhan 430070, China*
[2]*School of Software, Nanyang Institute of Technology, Nanyang 473000, China*

Correspondence should be addressed to Yongfeng Xing; xingyongfeng@163.com

Received 19 April 2020; Revised 29 May 2020; Accepted 9 June 2020; Published 7 July 2020

Academic Editor: Wei Wang

In recent years, the convolutional neural network (CNN) has made remarkable achievements in semantic segmentation. The method of semantic segmentation has a desirable application prospect. Nowadays, the methods mostly use an encoder-decoder architecture as a way of generating pixel by pixel segmentation prediction. The encoder is for extracting feature maps and decoder for recovering feature map resolution. An improved semantic segmentation method on the basis of the encoder-decoder architecture is proposed. We can get better segmentation accuracy on several hard classes and reduce the computational complexity significantly. This is possible by modifying the backbone and some refining techniques. Finally, after some processing, the framework has achieved good performance in many datasets. In comparison with the traditional architecture, our architecture does not need additional decoding layer and further reuses the encoder weight, thus reducing the complete quantity of parameters needed for processing. In this paper, a modified focal loss function is also put forward, as a replacement for the cross-entropy function to achieve a better treatment of the imbalance problem of the training data. In addition, more context information is added to the decode module as a way of improving the segmentation results. Experiments prove that the presented method can get better segmentation results. As an integral part of a smart city, multimedia information plays an important role. Semantic segmentation is an important basic technology for building a smart city.

1. Introduction

Convolution neural network is the part and parcel of image recognition, detection, and segmentation. The image semantic segmentation can provide a strong foundation for the construction of a smart city and has received much attention and research in recent years. Semantic segmentation is aimed at classifying all pixels in the image according to a specific category, which is commonly referred to as dense prediction. It is different from image classification because we do not classify the entire image into one class but all pixels. Thus, we boast a set of predefined categories and we need to distribute a tag to all pixels of the image according to the context of various objects in the image [1]. Deep neural network is no secret to the innovation of computer vision, particularly image classification. Since 2012, it has surpassed its predecessors by a large margin. In fact, artificial intelligence is superior to human in image classification. Inevitably, we adopted the same technology for semantic segmentation. Therefore, we put forward a network structure on the basis of encoder-decoder and atrous spatial pyramid pooling [2]. At the same time, a combination of multiple loss functions is used to be the ultimate loss function.

A relatively naive approach to construct the neural network architecture is simply stacking several convolutions, using the same padding to preserve that the dimensions remain the same and then output an ultimate segmentation map. Through a series of feature mapping transformations, the corresponding mapping of segmentation results can be learned directly from the input image. But it is quite expensive in computation to keep the whole resolution in the whole network. This architecture is illustrated in Figure 1.

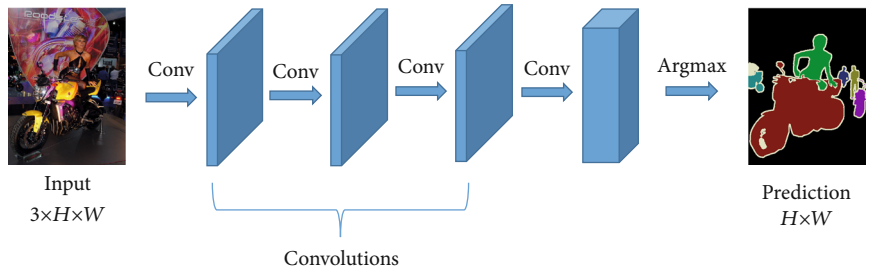


Figure 1: A simple method of constructing a neural network structure.

2. Related Works

In the deep convolution networks, the first layer studies the low-level notions, and the second layer studies the high-level feature mapping. As a way of maintaining the expression ability, the quantity of feature maps (channels) is usually increased while deepening the network. Different from the image classification which only needs the target category, image segmentation needs the location information of each pixel, so it cannot use pooling or trided convolutions to reduce the computation as safely as in the classification task. Image segmentation needs a whole-resolution semantic prediction. A popular image segmentation model is based on an encoder-decoder structure. In the encoder part, down sampling is adopted to reduce the input spatial resolution, so as to generate a lower resolution feature mappings (which is computationally efficient and can effectively distinguish different categories); in the decoder part, these feature representations are up sampled and restored to the full-resolution segmentation map.

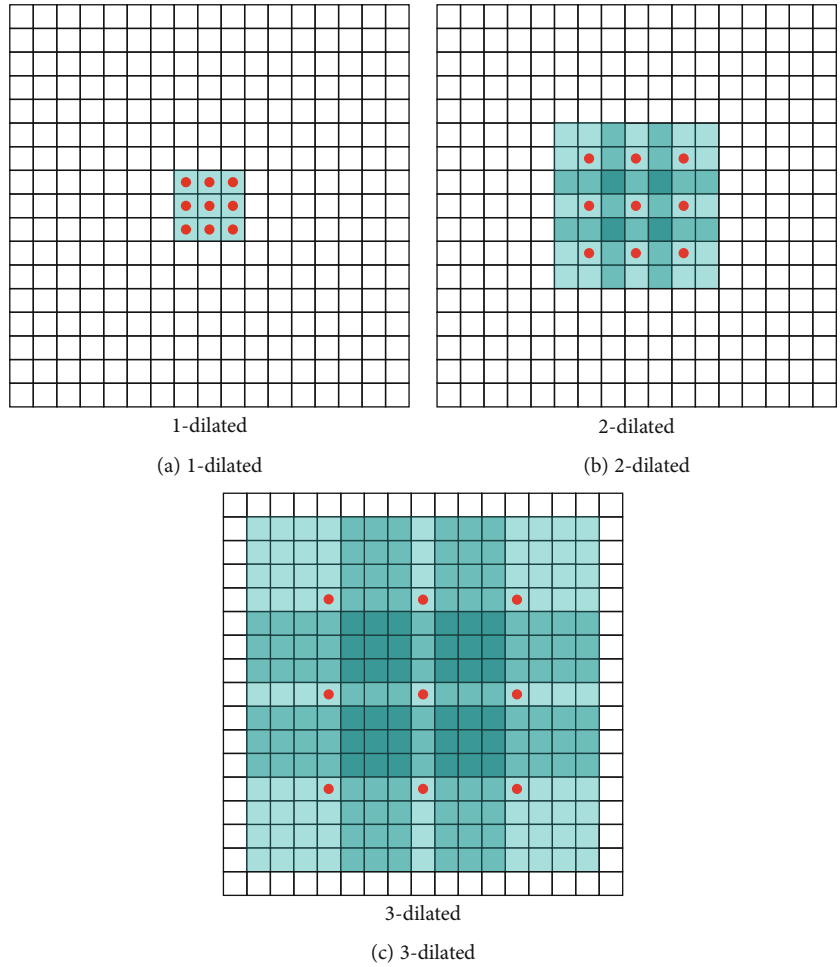
2.1. Fully Convolutional Network. Long et al. introduces the way to utilize end-to-end, pixel-to-pixel image segmentation task trained by the fully convolutional network at the end of 2014. In this paper, the author proposes to use the existing and well-researched image classification network as the encoder module of the network, adds transpose convolution layer in the decoding module, and upgrades the coarse feature mapping to the full-resolution segmentation mapping [3]. Full convolution network (FCNs) has achieved great success in the application of dense pixel prediction in semantic segmentation. The algorithm is required for predicting a variable for all pixels of the input image, a basic task in advanced computer vision understanding [1, 3]. Some of the most attractive applications include automatic driving [4], human-computer interaction [2, 5, 6], intelligent transportation system [7], auxiliary photo processing [8], and medical imaging [9]. The great achievements of FCNs come from the powerful characteristics picked up by CNNs. It is important that the convolution computer system makes the calculation efficiency of training and reasoning very high.

2.2. Encoder-Decoder. The encoder-decoder structure is a common architecture of current semantic segmentation algorithms. The structure is composed of an encoder and decoder. Classic image semantic segmentation algorithms such as FCN, U-net, and DeepLab all adopt this structure. The encoder is usually a network (VGG, Resnet, Xcepiton, etc); it consists of a deconvolution layer and upper sampling layer. Down sampling is aimed at capturing semantic or context information, while up sampling is aimed at recovering spatial information. Common decoders include bilinear interpolation, deconvolution, and dense up sampling convolution.

2.3. Dilated Convolution. In FCNs, because of continuous max pooling and down sampling operations, the feature resolution is greatly reduced. Finally, the feature mapping recovered by up sampling loses the detail sensitivity of the input image. In the full convolution network, the extended convolution is used instead of the standard convolution, so that the convolution network can accurately control the resolution of the image when calculating the feature response [10]. At the same time, the receiving the field of the filter is effectively expanded without adding the quantity of parameters and the amount of computing. Many experiments show that the algorithm uses more context information to obtain more dense features, thus improving the image semantic segmentation accuracy. It can be seen from Figure 2 that this is an expansion convolution filter with three different expansion rates: each element in the filter is a (a) 1-expansion convolution and a 3×3 receptive field, (b) 2-expansion convolution and a 7×7 receptive field, and (c) 3-expansion convolution and a 15×15 receptive field. The quantity of parameters related to each layer is the same. The receptive field increases exponentially and the number of parameters increases linearly [11].

Under the same size of convolution kernel, the receiving field of the convolution kernel can be increased by increasing the input stripe, as shown in Figure 3.

FCNs is a kind of deep convolution neural network, which has achieved good performance in pixel-level recognition tasks, but it still faces challenges in this changing and complex world. FCN is not a fully connected layer. The original method is to use the same size convolution layer stack as a way of mapping the input image to the output image. It produced strong results, but it was very expensive, because they cannot utilize any subsampling or pooling layers, because this will screw up the location of the instance. As a way of maintaining the resolution of the image, they must add many layers in a way that learns the low-level and high-level features. That means it is inefficient. For addressing this problem, they presented an encoder-decoder architecture. The encoder is a typical pretraining convolution



(a) 1-dilated (b) 2-dilated

(c) 3-dilated

Figure 2: 3×3 expansion convolution, the expansion rate is different: 1, 2, and 3.

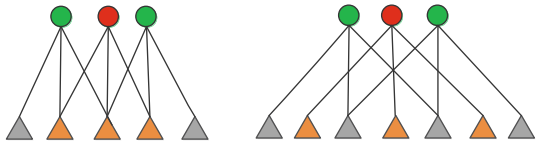


Figure 3: Illustration of the hole algorithm. 3×3 dilated convolutions with rate = 2).

network while a decoder consists of a deconvolutional layer and an upper sampling layer. Down sampling is aimed at capturing semantic or context information, while up sampling is aimed at recovering spatial information. Because the encoder lessens the image resolution, the segmentation has too few well-defined edges, meaning that the boundaries between the images are not clearly defined.

In [8], the final image prediction is usually reduced by 32 times in several stages of stride convolution and spatial pool, resulting in the loss of fine image structure information and inaccurate prediction, especially at the object boundary. DeepLab [12, 14–16] uses atrous (also names dilation) convolution to expand the receptive field while maintaining the high-resolution feature map, or use the encoder-decoder architecture to solve this problem. It regards the backbone network as an encoder and is responsible for encoding the original input image as a low-resolution feature map.

2.4. Atrous Spatial Pyramid Pooling (ASPP). The ASPP module was first proposed in [17] and further revised in [12]. In ASPP module, as shown in Figure 4, different atrous rates are used to extract multiple scale information. In conclusion, one 1×1 convolution block and three 3×3 convolution blocks have different shrinkage rates (6, 12, and 18, respectively), and one GAP block is employed in parallel. ASPP with different sampling rates and multiple views can capture objects at multiple scales.

It can be found that the receptive field has changed from 3 to 5, approximately doubled; the convolution kernel size is still 3×3, and the input stripe is 2, which is now called dilate rate [12, 14].

3. Our Approach

In this part, we introduce our presented network architecture and then explain the formation of each module in detail. We also propose a loss function as a way of further improving the performance of semantic segmentation.

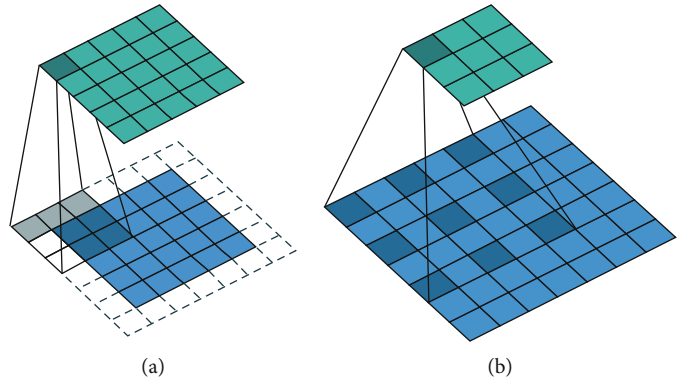


Figure 4: (a) The standard convolution of 3 × 3 kernel. (b) Expansion convolution of 3 × 3 kernel (expansion rate = 2).

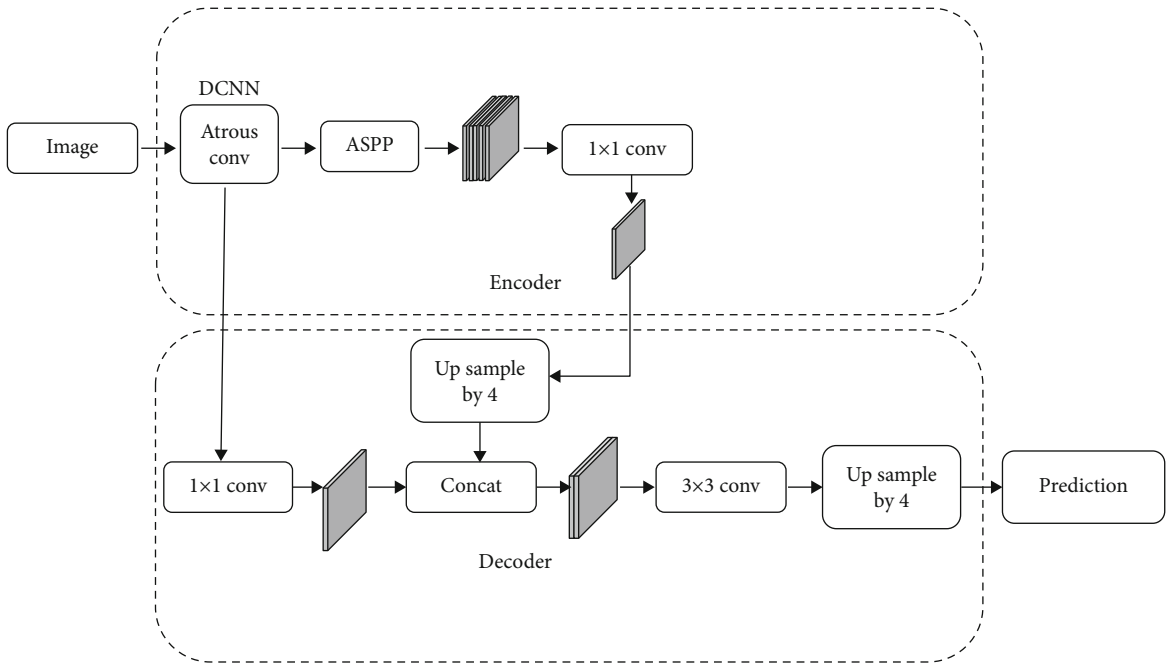


Figure 5: An image segmentation network architecture based on encoder-decoder structure.

3.1. Network Architecture. Figure 5 shows the network architecture including two parts: the encoder is used to extract the feature map and the decoder is used for recovering the resolution of the channel. The amount of parameters in the ASPP part and the decoder part are also huge. Therefore, all the ordinary convolutions are replaced by the depthwise separable convolution. At the same time, the number of channels in ASPP and decoder is also decreased. The backbone network and the ASPP module together constitute the encode module of the network. Input any size of image to obtain the corresponding high-level feature map. Then, through the bilinear up sampling and the low-level feature map of one layer of the encode module, the decode module of the network is formed. Finally, the up sampling is back to the original map size, and the corresponding segmentation map is obtained through the softmax classification layer. This is to decouple spatial information and depth information. It is found that the effect of detail set 1/2 of the size of the feature map and the decoder feature are fused, and finally good results are achieved.

3.2. Backbone Network. Over the past few years, some backbone networks of CNN have achieved great progress in visual missions, showing the most advanced level. It is stacked in the order of convolutional layer, pooling layer, activation function layer, and a fully connected layer. CNN can output the classification score corresponding to the image by inputting the image. In 2012, AlexNet [18] won the title of ILSVRC [19]. AlexNet addresses the problem of image classification and creates a new situation of computer vision. Then, top competitors put forward various CNN architectures, GoogLeNet [8], ResNet [20], DenseNet [21], etc [22]. These network structures can well extract the feature mapping of an image, which lays a solid foundation for semantic

Table 1: Common image classification network infomation summary.

Name	AlexNet	VGG	GoogLeNet	ResNet	Inception	Xception	EfficientNet
Year	2012	2014	2014	2015	2015	2016	2019
Layer	8	19	22	152	/	/	/
Conv	5	16	21	151	/	/	/
Top 5 (error)%	16.4	7.32	6.67	3.57	3.5	5.5	2.9

segmentation [23, 24]. Our network architecture uses Xception to be the feature extractor. Some common classification networks are shown in Table 1 [25]. We came to a conclusion in the experiment. With high calculation complexity, recognition accuracy is allowed to be low; with many parameters, recognition accuracy is allowed to be low. A good network structure design is very important. Different models have different parameter utilization efficiencies.

3.3. Cross-Entropy Loss and Focal Loss. The common loss function of classification problem is cross-entropy loss. It shows the distance between two probability distributions. The closer they are to the cross-entropy, the closer they are. The cross-entropy approach is a novel general method for combinatorial optimization, multipole optimization, and rare event simulation. The standard loss of binary classification is cross-entropy.

Sometimes we will meet the task of image segmentation, which is that the background accounts for a large proportion, but the object accounts for a small proportion of the seriously imbalanced dataset. At this time, we need to carefully use the loss function. The most commonly used loss functions are as follows:

$$\mathrm{CE}(p, y) = -y \log(p) - (1 - y) \log(1 - p), \tag{1}$$

where $y=y_{\text{truth}}$, $p=y_{\text{pred}}$

$$\mathrm{CE}(p, y) = \begin{cases} -\log(p) & y = 1, \\ -\log(1 - p) & \text{otherwise.} \end{cases} \tag{2}$$

From the above formula, we can draw a conclusion: when $y = 1$, the larger y' is, the closer it is to y, that is, the more accurate the prediction is, the smaller the loss is. When $y = 0$, the smaller y' is, the closer it is to y, that is, the more accurate the prediction is, the smaller the loss is. The final loss is the sum of $y = 0$ and $y = 1$. This method has one obvious drawback. While the number of positive samples is far less than the negative samples, that is to say, the number of $y = 0$ is far greater than the number of $y = 1$, and its components will dominate the loss function. The model is heavily biased towards the background.

$$\mathrm{FL}(p_t) = -(1 - p_t)^{\gamma} \log(p_t). \tag{3}$$

We define p_t:

$$p_t = \begin{cases} y' & y = 1, \\ 1 - y' & \text{otherwise,} \end{cases} \tag{4}$$

and rewrite $\mathrm{CE}(p, y) = \mathrm{CE}(p_t) = -\log(p_t)$.

First of all, the proportion of positive and negative samples should be balanced without using negative sample mining and other means. In this paper, we directly multiply a parameter α in front of the CE loss, so that we can easily control the proportion of negative and positive samples.

We get the balanced cross-entropy loss as

$$\mathrm{CE}(p_t) = -\alpha \log(p_t). \tag{5}$$

In practice, α is a decimal between [0, 1]; it is a fixed value and does not participate in training.

Although the above formula can control the weight of positive and negative samples, it cannot control the weight of easy samples and hard samples.

The γ here is called a focusing parameter, $\gamma > = 0$. A modulating factor $(1 - p_t)^{\gamma}$ is called the modulating factor. In practice, we usually add a parameter α in front of the focal loss:

$$\mathrm{FL}(p_t) = -\alpha(1 - p_t)^{\gamma} \log(p_t). \tag{6}$$

In the process of semantic segmentation, there are more categories corresponding to semantic segmentation than the two classification problem in target detection. If the selected parameters λ and γ are not suitable, the cross-entropy loss weight of these pixels will be reduced. Combined with the above analysis, we propose to increase the weight of difficult samples and keep the weight of simple samples almost unchanged. We find that the best results can be obtained by setting $\alpha = 0.5$ and $\gamma = 2$ in our experimental network.

Focal loss was first proposed in the RetinaNet model [26] to solve the imbalance and difficulty of classification in the training process. In practical application, the combination of focal loss and dice loss usually needs to scale them to the same order of magnitude. Use -log to enlarge dice loss and use alpha to reduce focal loss.

4. Experiments and Results

As a way of proving the effectiveness of our presented framework, we evaluated it on the basis of the benchmark dataset (PASCAL VOC 2012) and the latest methods. In the paper,

we report the experimental outcomes of three mainstream semantic segmentation datasets: PASCAL VOC2012, CamVid [27], and Cityscapes [28].

The mean intersection on union (MIoU) is the standard measure of semantic segmentation. The intersection and union ratio of two sets is calculated. In semantic segmentation, the two sets are base truth value and prediction segmentation. This proportion can be morphed to TP (intersection set) over TP, FP, and FN (union set). Calculate the IoU of each class and take the average.

$$\mathrm{MIoU} = \frac{1}{k+1}\sum_{i=0}^{k}\frac{p_{ii}}{\sum_j^k p_{ij} + \sum_j^k p_{ij} - p_{ji}}, \tag{7}$$

is equivalent to

$$\mathrm{MIoU} = \frac{1}{k+1}\sum_{i=0}^{k}\frac{\mathrm{TP}}{\mathrm{FN}+\mathrm{FP}+\mathrm{TP}}, \tag{8}$$

First, calculate the intersection and union ratio of each category, and then get the average. TP is the positive sample that has a correct sort, TN is the positive sample that has a wrong sort. FP is the negative sample of sort error. TP can be understood as the intersection of prediction results and labels, while TP + TN + FP is preunion of test results and labels. The closer the intersection is to the union, the more accurate the segmentation is.

We also used several widely used data augmentation strategies in our training, including 50% probability of horizontal flipping and random scaling of images, scale factor between 0.5 and 2.0 in steps of 0.25, fill and randomly crop the scaled image to 513×513. Finally, with a fine tuning learning rate of $2e$-4 is implemented in the model. When we segment some small target parts, we find that the effect of detail segmentation is very poor. To improve the details, 1/2 of the size of the feature map and the decoder feature are fused, and good results are obtained. In the training, the loss function used is an improved version, focal loss. The results show that the improved focal loss can improve semantic scores. The accuracy of the segmentation and the nonequilibrium of the sample are alleviated.

4.1. PASCAL VOC 2012. PASCAL VOC 2012 includes 20 foreground object classes and one background class, including photos from private collections. There are six indoor classes, seven cars, and seven creatures. The dataset contains 1464 columns, 1449 validation, and 1456 variable size test images. We use 512×512 crops as a way of dividing the learning rate of pretraining weight by 8. All other superparameters are the same as those in [16] experiment. Table 2 shows the performance of our algorithm on VOC 2012, and the detailed results comparison with other methods are displayed in Table 3.

According to the evaluation samples on the test set of PASCAL VOC2012 validation set dataset, we can see that the proposed method is applicable to animals, people, and objects. The edge of equal targets can be segmented carefully, which improves the classification accuracy of the stool, animal, bicycle, and so on. The evaluation of the abovementioned classification index shows that its effect is better than many segmentation methods, as shown in Figure 6. Please note that we do not use CRFs for postprocessing, which can smooth the output, but it is too slow in practice, especially for large-scale images.

Table 2: Performance on PASCAL VOC2012 test set.

Method	MIoU
FCN-8s	62.2
ResSegNet	80.4.7%
RefineNet	84.2%
PSPNet	85.4%
DeepLabv3+	87.8%
Ours	85.6%

4.2. Cityscapes. The Cityscapes dataset is a very large image dataset, which focuses on the semantic understanding of street scene. It contains the road driving images of 50 cities in spring, summer, and autumn. There are 19 classes in the dataset, including good weather and moderate weather, many dynamic objects, different scene layouts, and different backgrounds. We have carried out experiments on 5000 fine-labeled images, which are divided into 2975 training images, 500 verification images, and 1525 test images. The resolving power of all images is 1024×2048. It contains 5000 high-quality pixel level annotations of size 1024×2048 (2975, 500, and 1525 for training, verification, and test sets, respectively) and 2975, 500, and 1525 (training, verification, and test sets separately).

As shown in Figure 7, finally, the method achieves 81.79% MIoU precision on Cityscapes test set on 1024×2048 image. Table 4 shows the performance of our algorithm on Cityscapes 2012 test set.

4.3. CamVid. As a way of further proving the effectiveness and robustness of this method, we also assess its performance on the CamVid dataset. The Cambridge-driving Labeled Video Database (CamVid) is the first video collection with object l class semantic tags. The ground truth labels provided by the database associated each pixel with one of the 32 semantic classes. The CamVid dataset contains images of city road driving scenes. We use 11 classes, including 367 training, 101 verification, and 233 test images. The resolution of all images is 720×960.

We train all models from random initialization and fine tune the pretrained parameters on ImageNet. In the training process, the size of random clipping is 512×512, and the batch size is 16. All other superparameters are the same as PASCAL VOC 2012 experiment. After 30000 iterations on the training set, the model in this paper achieves 77.61% MIoU on the validation set and 69.39% MIoU on the test set.

We can see that the models in this paper can get very accurate semantic segmentation results. Whether it is a small target, or some targets with occlusion and overlap, the method in this paper can accurately segment them.

Table 3: Our highest scoring entry in each column is shown in italic. Results in a performance of 85.6% on PASCAL VOC 2012 test set.

Category	FCN-8s	ResSegNet	RefineNet	PSPNet	DeepLabv3+	Ours
Bicycle	34.2	65.2	73.2	72.7	77.1	*78.2*
Chair	21.4	37.4	43.7	43.1	56.9	*57.1*
Sheep	72.4	85.9	92.9	94.4	92.9	*94.4*
Mean	62.2	80.4	84.2	85.4	87.8	85.6

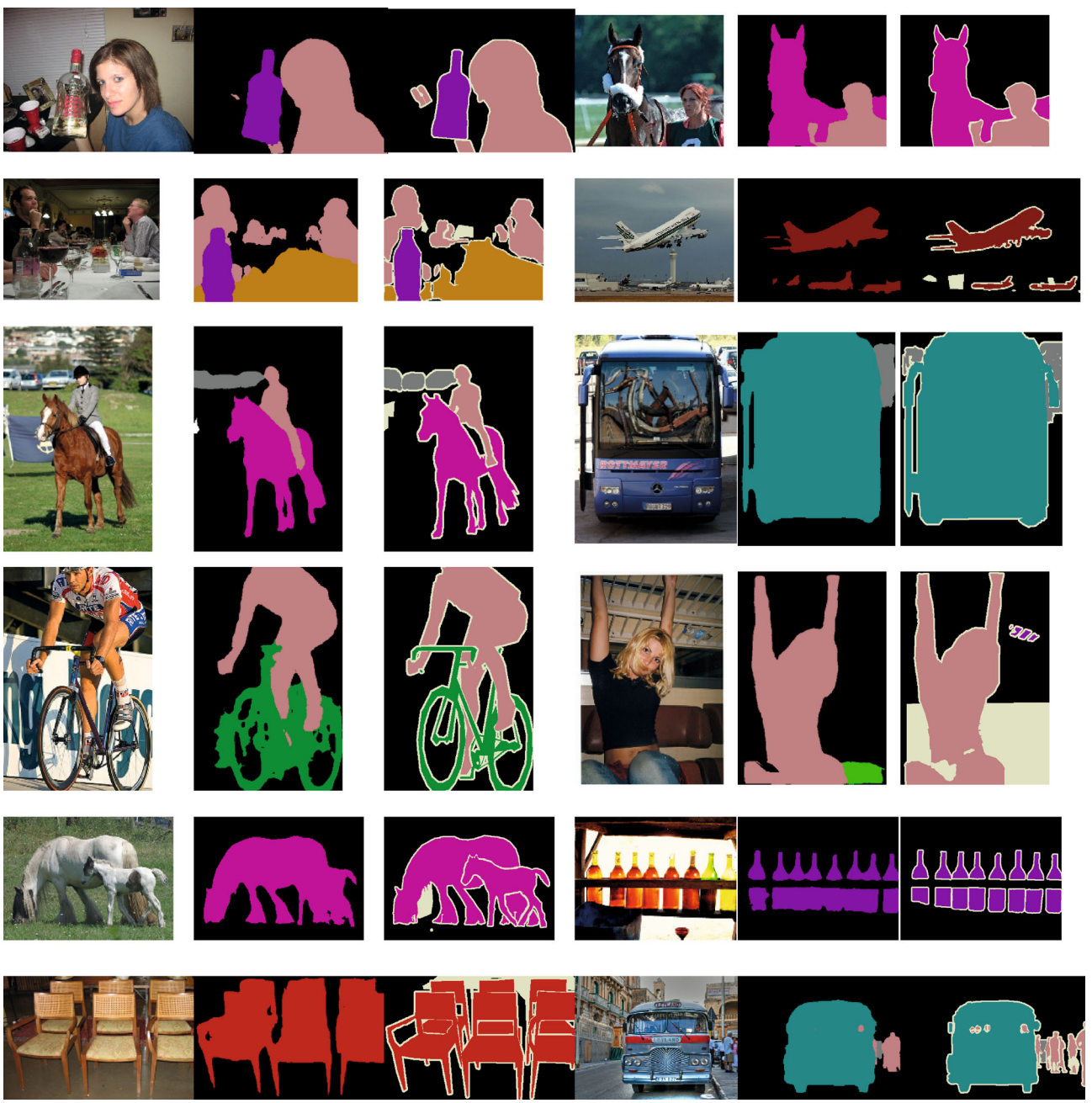


Figure 6: The visualization results on the PASCAL VOC2012 validation set using our methods.

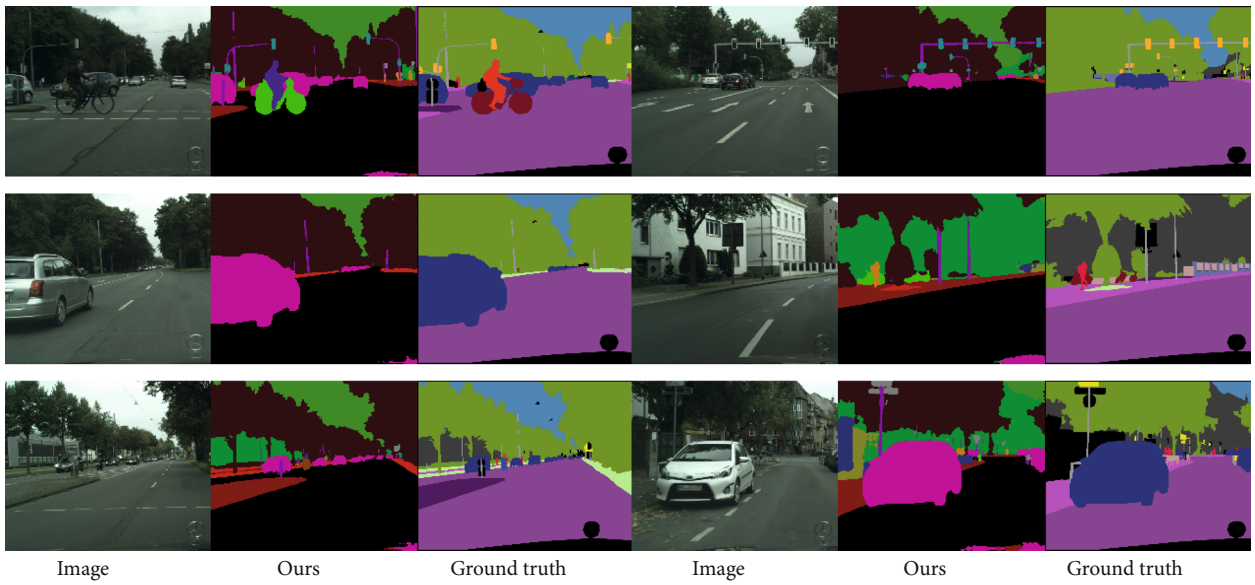


Figure 7: The visualization results on the Cityscapes data using our methods.

Table 4: Segmentation results on Cityscapes test set (https://www.cityscapes-dataset.com/benchmarks/).

Method	MIoU(%)
FCN-8s	65.3
Dilation10	67.1
ShuffleNet v2 + DPC	70.3
MobileNetV2Plus	70.7
ML-CRNN	71.2
Ladder DenseNet	74.3
TuSimple	77.6
DeepLabv3+	82.1
Ours	81.79%

5. Conclusion

We introduce a simpler yet robust network for improving semantic segmentation tasks. Combining ASPP and a classical encoder-decoder structure, an improved loss function more suitable for the application is proposed. The experimental outcomes show the superiority of this method. It not only effectively improves the segmentation performance but also significantly improves the imbalance of training data. As a way of improving the learning ability of this method, we will focus more on weak supervised learning and metalearning down the road. We believe that semantic segmentation can provide a good practice for future smart city construction.

Data Availability

The data used to find the study can be available upon request to the corresponding author.

Conflicts of Interest

The authors declared that they have no conflicts of interest to this work.

Acknowledgments

This work was supported by the Hubei Natural Science Foundation (2015CFB525), the National Natural Science Foundation (6130329), and the Hubei Natural Science Foundation Innovation Research Group (2017CFA012). We would like to thank those anonymous commentators who helped promote the quality of their papers.

References

[1] D. C. Cirean et al., "Deep neural networks segment neuronal membranes in electron microscopy images," *Advances in Neural Information Processing Systems*, vol. 25, pp. 2852–2860, 2012.

[2] M. Oberweger, P. Wohlhart, and V. Lepetit, *Hands deep in deep learning for hand pose estimation*, Computer Vision Winter Workshop (CVWW), 2015.

[3] R. Mottaghi, X. Chen, X. Liu et al., "The role of context for object detection and semantic segmentation in the wild," in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, OH, USA, June 2013.

[4] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? The KITTI vision benchmark suite," in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, Providence, RI, USA, June 2012.

[5] B. Luo, Y. Sun, G. Li, D. Chen, and Z. Ju, "Decomposition algorithm for depth image of human health posture based on brain health," *Neural Computing and Applications*, vol. 32, no. 10, pp. 6327–6342, 2020.

[6] D. Jiang, Z. Zheng, G. Li et al., "Gesture recognition based on binocular vision," *Cluster Computing*, vol. 22, Supplement 6, pp. 13261–13271, 2019.

[7] S. Segvic, K. Brkic, Z. Kalafatic, and A. Pinz, "Exploiting temporal and spatial constraints in traffic sign detection from a moving vehicle," *Machine Vision and Applications*, vol. 25, no. 3, pp. 649–665, 2014.

[8] 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.

[9] Y. Wu and K. He, "Group normalization," in *Computer Vision – ECCV 2018*, Springer, 2018.

[10] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 2, pp. 295–307, 2016.

[11] M. I. Razzak, S. Naz, and A. Zaib, *Deep learning for medical image processing: overview, challenges and future*, 2017.

[12] L. C. Chen et al., "Semantic image segmentation with deep convolutional nets and fully connected CRFs," *Computer Science*, vol. 4, pp. 357–361, 2014.

[13] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," in *International Conference on Learning Representations (ICLR)*, 2016.

[14] 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.

[15] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," 2017, https://arxiv.org/abs/1706.05587.

[16] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proceedings of the European conference on computer vision (ECCV)*, pp. 801–818, 2018.

[17] Z. Tian, T. He, C. Shen, and Y. Yan, "Decoders matter for semantic segmentation: data-dependent decoding enables flexible feature aggregation," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, CA, USA, June 2019.

[18] B. Ma and A. Entezari, "An interactive framework for visualization of weather forecast ensembles," *IEEE Transactions on Visualization and Computer Graphics*, vol. 25, no. 1, pp. 1091–1101, 2019.

[19] L. Zhou, C. Zhang, and M. Wu, "D-Linknet: linknet with pretrained encoder and dilated convolution for high resolution satellite imagery road extraction," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Salt Lake City, UT, USA, June 2018.

[20] S. Gupta et al., *Learning rich features from RGB-D images for object detection and segmentation*, 2014.

[21] 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.

[22] X. Shen, A. Hertzmann, J. Jia et al., "Automatic portrait segmentation for image stylization," *Computer Graphics Forum*, vol. 35, no. 2, pp. 93–102, 2016.

[23] S. Bianco, R. Cadene, L. Celona, and P. Napoletano, "Benchmark analysis of representative deep neural network architectures," *IEEE Access*, vol. 6, no. 1, pp. 64270–64277, 2018.

[24] B. Ma and A. Entezari, "Volumetric feature-based classification and visibility analysis for transfer function design," *IEEE Transactions on Visualization and Computer Graphics*, vol. 24, no. 12, pp. 3253–3267, 2018.

[25] B. Ma, S. K. Suter, and A. Entezari, "Quality assessment of volume compression approaches using isovalue clustering," *Computers & Graphics*, vol. 63, pp. 18–27, 2017.

[26] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, "Focal loss for dense object detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 2, pp. 318–327, 2020.

[27] V. Badrinarayanan, A. Handa, and R. Cipolla, "SegNet: a deep convolutional encoder-decoder architecture for robust semantic pixel-wise labelling," *Computerence*, vol. 39, 2015.

[28] M. Cordts, M. Omran, S. Ramos et al., "The cityscapes dataset for semantic urban scene understanding," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, June 2016.

Hindawi
Wireless Communications and Mobile Computing
Volume 2020, Article ID 8865298, 9 pages
https://doi.org/10.1155/2020/8865298

Research Article

Smart City Moving Target Tracking Algorithm Based on Quantum Genetic and Particle Filter

Zhigang Liu, Jin Shang, and Xufen Hua

Control Technology Institute, Wuxi Institute of Technology, Wuxi 214121, China

Correspondence should be addressed to Jin Shang; shangj@wxit.edu.cn

Received 28 March 2020; Revised 19 May 2020; Accepted 25 May 2020; Published 20 June 2020

Academic Editor: Bingxian Lu

In the application of moving target tracking in smart city, particle filter technology has the advantages of dealing with nonlinear and non-Gaussian problems, but when the standard particle filter uses resampling method to solve the degradation phenomenon, simply copying the particles will cause local optimization difficulties, resulting in unstable filtering accuracy. In this paper, a particle filter algorithm combined with quantum genetic algorithm (QGA) is proposed to solve the above problems. Aiming at the problem of particle exhaustion in particle filter, the algorithm adopts the method of combining evolutionary algorithm. Each particle in particle filter is regarded as a chromosome in genetic algorithm, and the fitness of each chromosome corresponds to the weight of particle. For each particle state with weight, the particle is first binary coded with qubit and quantum superposition state, and then quantum rotation gate is used for selection, crossing, mutation, and other operations, after a set number of iterations, the final particle set with accuracy and better diversity. In this paper, the filter state estimation and RMSF of $N = 50$ and $N = 100$ for nonlinear target tracking and the comparison of real state and state estimation trajectory in time-constant model under nonlinear target tracking are given. It can be seen that in nonlinear state, the quantum genetic and particle filter (QGPF) algorithm can achieve a higher accuracy of state estimation, and the filtering error of QGPF algorithm at each time is relatively uniform, which shows that the algorithm in this paper has better algorithm stability. Under the time-constant model, the algorithm fits the real state and realizes stable and accurate tracking.

1. Introduction

Target tracking is widely used in many fields of smart city. Tasks such as video surveillance, human-computer interaction, automatic vehicle control, and human behavior analysis make great use of target tracking [1]. The moving state of the target is usually a nonlinear, non-Gaussian problem with mobility. The development of target tracking algorithm can be divided into four categories: the first category is mainly based on particle filter correlation algorithm; the second category is based on sparse representation theory; and the third category is based on correlation filter tracking algorithm. With the outstanding performance of deep-learning method in feature modeling, the fourth category is mainly based on deep-learning tracking method. Particle filter is a sequential Monte Carlo method proposed by Ulam. The main idea of this method is to fit the distribution of actual samples as much as possible by using the method of weighted sampling in a sequence, so as to calculate the state distribution probability of the target at the current time through the previous sequence state distribution, and take the state corresponding to the maximum probability value as the result of the target state prediction [2, 3]. Particle filter is widely used in the actual target tracking because it does not need to assume the linear and Gaussian state transfer equation and observation equation. In the process of practical application, particle filter has the problem of particle weight degradation. In order to solve this problem, resampling technology is introduced into the particle filter. The purpose of particle resampling is to solve the degradation of particles in the classical Monte Carlo method. The main idea is to resample the probability density function of particles and their corresponding weights. It is realized by increasing the particles with larger weight and reducing the particles with smaller weight. But directly copying high weight particles and discarding low weight particles will lead to particle dilution. On the other hand, in order to

achieve high-precision state estimation, more particles are usually needed to participate, and the number of particles affects the efficiency of the algorithm. In view of the above problems, scholars at home and abroad have conducted a lot of research. The extended Kalman filter [4] and the unscented Kalman filter [5] integrate the latest observation information to provide the density function for particle filter. Although the accuracy of particle filter is improved to a certain extent, the amount of calculation is also greatly increased. The adaptive partial resampling method, which only resamples a part of the particle set, makes a moderate compromise between the weight degradation and particle dilution [6]. The sensitive resampling algorithm keeps the particles with the largest weight and propagates them by the quasi-Monte Carlo method, which restrains the lack of samples but increases the calculation time [7]. The deterministic resampling algorithm should consider both the particle weight and the particle state value when sampling, avoiding the nondetection discard of low weight particles and avoiding the sample dilution to some extent [8]. After the resampling of the PF algorithm by using the GA algorithm, Kao and Zahara updated the particles by cross-mutation to alleviate sample degradation, which improved the accuracy of PF algorithm to a certain extent, but made the calculation efficiency of PF algorithm lower [9]. In reference [10], a robust regularization particle filtering (RPF) method is proposed to reduce multimodal loss by mixing local RPF. In reference [11], an improved method of spherical simple unscented particle diversity (ssupf) is proposed and applied to high-precision astronomical navigation in deep-space environment. It is worth noting that the number of sampled particles has a great impact on PF performance. Reference [12] shows that too few sampling particles will lead to divergence, while too many particles will increase the computational burden and limit the real-time performance of navigation. In order to improve the efficiency of PF, a novel adaptive PF method based on the Kullback Leibler distance (KLD) is proposed in [13, 14] and applied to mobile robot positioning. In reference [15], many PF techniques for adaptive sampling particle number are reviewed. The above research shows that it is feasible to control the number of particles to ensure the estimation quality and reduce the calculation cost. In reference [16], a resampling step is introduced, and an algorithm of sampling importance resampling (SIR) is proposed. The particles with high weight are resampled, and the particles with low weight are discarded, and then a new set of particles is generated by state space model and measurement model. Li et al. proposed a lightweight particle filter tracking method [17], introduced the response graph of correlation filter into the calculation of particle weight, and used fewer particles to reduce the calculation burden in the sampling process. Firouznia et al. introduced the chaos theory into the particle filter framework, effectively reducing the number of particles and search space [18]. Jenkins et al. introduced the multibag subspace recovery mechanism to solve the problem of updating the appearance model in particle filter tracking and resetting in case of tracking drift, so as to improve the robustness of the model [19]. Although the tracking algorithm based on particle filter framework is constantly improving [20–22], there are still some problems in the existing algorithm, such as poor universality in complex scenes, and the tracking accuracy needs to be further improved. In [23], based on the PF algorithm, the principle of the mean-shift algorithm is integrated to realize the target tracking model of the fusion algorithm. This kind of algorithm can improve the calculation efficiency and real-time performance of the PF algorithm, but how to choose the appropriate fusion strategy of the mean-shift algorithm is also one of the difficulties; In addition, how to prevent tracking drift should be considered in this algorithm. In [24], the relevant knowledge expressed by sparse theory is applied to the PF algorithm to solve the target tracking task. The experiment shows that the improved method can improve the accuracy of the target tracking model to a certain extent, but the real-time performance of the algorithm is usually affected by too much calculation when solving the problem. With the rapid development of depth learning technology and the outstanding performance of depth model in image feature representation, the fusion target tracking algorithm based on depth learning method has been proposed one after another. In [25, 26], the combination of the depth model algorithm and PF principle is applied to the target tracking task, which has achieved quite good results. However, the depth learning model usually needs a large number of training samples with supervision information to train the model, so how to design a suitable structure to improve the calculation speed to meet the real-time requirements of the algorithm is also an urgent problem to be solved [27–29]. In this paper, a particle filter method combined with quantum genetic algorithm is proposed to solve the above problems. Aiming at the problem of particle exhaustion in particle filter, the algorithm adopts the method of combining evolutionary algorithm to carry out cross and mutation operation on particles. The particles with high posterior probability have high adaptability, so as to improve the accuracy of filter results. Compared with the classical genetic algorithm, the advantage of quantum genetic algorithm lies in the selection of population coding mode and evolution strategy. The core of the population coding method of quantum genetic algorithm is to code individuals by using qubits and quantum superposition states, so that each individual can represent information of multiple states, which greatly enriches the diversity of the population; at the same time, the population is updated by using quantum rotation gate, and the individual adaptability is improved to guide evolution. In the iterative process, the superposition state of each qubit will collapse to a certain state, which tends to be stable and converge, and finally achieve the purpose of optimization. Because of the unique coding and updating method, the quantum genetic algorithm has more diversity, faster convergence speed, and higher convergence accuracy than the classical genetic algorithm. The genetic mechanism can not only select excellent individuals by selection operator but also generate new individuals by cross operator and mutation operator. Therefore, properly adjusting the selection probability, crossover probability and mutation probability can ensure the effectiveness of particles and take into account the diversity of particles. The specific combination process of quantum genetic algorithm and particle filter is

as follows: each particle of the particle filter is regarded as the chromosome in genetic algorithm, and the fitness of each chromosome corresponds to the weight of particle; for each particle state with weight, firstly, the particle is binary coded with qubit and quantum superposition state, and then it is selected, crossed, and selected by quantum revolving gate After the set number of iterations, the final particle set is more accurate and diverse. In this paper, the QGPF algorithm is proposed by combining QGA with PF; the algorithm synthesizes the local search ability and calculation efficiency of QGA and the nonlinear filtering principle of PF algorithm. The simulation experiments of nonlinear target tracking model and time-constant model show that the algorithm has high precision and good numerical stability and can complete the accurate tracking of smart city moving targets.

2. Particle Filter Algorithm

Particle filter is a kind of filter algorithm based on the Bayesian theory and Monte Carlo simulation. Its core idea is to use a group of random samples to approximate the posterior conditional probability density of state variables. It is assumed that the nonlinear dynamic process is expressed as follows:

$$\begin{cases} x_k = f(x_{k-1}, v_{k-1}) \\ z_k = h(x_k, w_k), \end{cases} \tag{1}$$

in which, x is the state value of the system, $f(\cdot)$ is the state transfer equation, z is the state observation value, $h(\cdot)$ is the system observation equation, w is the system noise, and v is the observation noise.

The recursive Bayesian filtering includes two processes: prediction and update.

Prediction process: the posterior probability density $p(x_{k-1}|z_{1:k-1})$ of $k-1$ time is used to predict the predicted probability density $p(x|z_{1:k-1})$ of k time.

$$p(x_k|z_{1:k-1}) = \int p(x_k|x_{k-1})p(x_{k-1}|z_{1:k-1})dx_{k-1}. \tag{2}$$

Update process: update the posterior probability density $p(x_k|z_{1:k})$ of k time by the observation information z_k of k time.

$$p(x_k|z_{1:k}) = \frac{p(z_k|x_k)p(x_k|z_{1:k-1})}{\int p(z_k|x_k)p(x_k|z_{1:k-1})dx_k}. \tag{3}$$

Because of the integral operation in the Bayesian filtering algorithm, it is difficult to obtain closed analytical solutions for the nonlinear and non-Gaussian systems. By using the Monte Carlo method, the integral operation is transformed into the sample weighted sum operation. The posterior probability density can be approximated by the sample particle x_k^i and its weight w_k^i.

$$p(x_k|z_k) = \sum_{i=1}^{N} w_k^i \delta(x_k - x_k^i), \tag{4}$$

in which $\delta(\cdot)$ is the Dirac function.

However, it is very difficult to get samples from the posterior probability density under normal circumstances. In this paper, an importance distribution function $q(x_{0:k}|z_{1:k})$, which is easy to sample, is introduced. In order to simplify the calculation and make the weight calculation of particles recursive, the importance distribution function satisfies the following decomposition:

$$q(x_{0:k}|z_{1:k}) = q(x_{0:k}|x_{0:k-1}, z_{1:k})q(x_{0:k-1}|z_{1:k-1}). \tag{5}$$

Then, the weight update formula is

$$w_k^i = w_{k-1}^i \frac{p(z_k|x_k^i)p(x_k^i|x_{k-1}^i)}{q(x_k^i|x_{0:k-1}^i, z_{1:k})}. \tag{6}$$

The status output is

$$x_k = \sum_{i=1}^{N} w_k^i x_k^i. \tag{7}$$

Particle filter initializes the sample at the initial time and extracts the initialization state from the prior distribution. At the next moment, the particles transfer state according to the transfer equation and assign the weight of particles according to the actual system observation. Finally, all particles are weighted sum to get the state estimation at this time. After resampling the particles, the filtering process of the next period is continued [30, 31].

3. Quantum Genetic Algorithm

QGA is a newly developed probability evolution algorithm based on the quantum computing principle. It uses the concept and theory of quantum computing to integrate the characteristics of genetic algorithm (GA) which can maintain good population diversity. It applies the probability amplitude representation of quantum bits to chromosome coding, so that a chromosome can express the superposition of multiple states and use quantum rotation gate and quantum nongate to realize chromosome update operation, so as to achieve population optimization.

The population of QGA consists of quantum chromosomes encoded by quantum bits. Quantum bit is the smallest information unit in QGA, which is different from the classical bit in that can not only be in state 0 or 1 but also represent any superposition state of the two, so QGA has more diversity than GA. The population with N individuals and M -length quantum chromosome is expressed as

$$P(t) = \{p_1^t, p_2^t, p_3^t, \cdots p_n^t\},\ p_j^t = \begin{vmatrix} \alpha_1^t & \alpha_2^t & \cdots & \alpha_m^t \\ \beta_1^t & \beta_2^t & \cdots & \beta_m^t \end{vmatrix} (j = 1, 2, 3, \cdots, n) \tag{8}$$

where p_j^t is an individual of the t generation, α_i and β_i are complex numbers, which are called probability amplitude pairs, satisfying the normalization condition $\alpha^2 + \beta^2 = 1$, and t is genetic algebra.

Quantum gate is the executive mechanism of QGA to realize evolutionary operation. One of the key points of quantum genetic algorithm is to construct a proper quantum gate. According to the characteristics of QGA, it is more suitable to choose quantum revolving gate. The update of quantum bit is realized by quantum rotating gate as follows:

$$\begin{bmatrix} \alpha_i' \\ \beta_i' \end{bmatrix} = \begin{bmatrix} \cos\theta_i & -\sin\theta_i \\ \sin\theta_i & \cos\theta_i \end{bmatrix} \begin{bmatrix} \alpha_i \\ \beta_i \end{bmatrix}, \tag{9}$$

in which the quantum revolving gate is

$$U_i = \begin{bmatrix} \cos\theta_i & -\sin\theta_i \\ \sin\theta_i & \cos\theta_i \end{bmatrix}, \ (\alpha_i, \beta_i)^{\mathrm{T}} \ \left(\alpha_i', \beta_i'\right)^{\mathrm{T}} \tag{10}$$

are the ith quantum bit in the chromosome before and after renewal, and θ_i is the rotation angle of the quantum gate.

4. The Particle Filter Algorithm Combined with the Quantum Genetic Algorithm

The particle at a certain time is regarded as a possible state of the target's position, and the weight of the particle represents its possibility. The particle set can be regarded as the population of quantum genetic algorithm at a certain time, that is, each particle at that time is regarded as an individual in the population. A nondegenerate particle filter algorithm is implemented by combining quantum genetic algorithm and particle filter algorithm. The whole algorithm is divided into three stages: presampling, genetic operation, and particle individual selection: In the presampling stage, the presampled particle set at time K is extracted from the particle set at time $K-1$ through the importance distribution, and its weight is determined by measuring its consistency with the road sign observation. The weight of the particle represents the likelihood of the particle representing the environment, which can also be regarded as the fitness of the particle.

The fitness function is related to the function to be realized. In this paper, we study the tracking of moving objects in smart city and make the particles correspond to the candidate solutions of tracking objects. The fitness function is used to evaluate the quality of the candidate solutions (particles) given. z_k represents the observed output. The larger the probability of observing z_k is, the better the particle can represent the real position of the object, namely

$$\text{fitness}\left(x_k^i\right) = P\left(z_k \middle| x_k^i\right), \tag{11}$$

in which $\text{fitness}(x_k^i)$ represents the fitness of i particles at K time, z_k represents the observation result of the target at K time, and $P(z_k|x_k^i)$ represents the probability of observing z_k at the position of i particles at K time. Therefore, the higher the probability is, the more representative the particle is of the real position of the target, and the greater its weight.

Quantum genetics is applied to the resampling stage of particles to expand the diversity of particle sets. In the initialization stage of genetic algorithm, in order to set more genetic opportunities for particles with high fitness, the weights of all particles are sorted and normalized. In the stage of genetic generation of new particles, the quantum cross-mutation operation is used to generate new particles, and the algorithm is set to achieve high degree of freedom particles that have a higher probability to continue the offspring particles. This setting can effectively avoid particle degradation, because the cross-operation increases the diversity of particles, and the distribution of these new particles is wider than that before the cross, which is conducive to the improvement of particle positioning ability. In order to avoid too much error caused by particle dispersion, the project plans to carry out mutation operation according to the weight of particles: for particles whose weight is less than the average level, we can generate new particles with greater weight through mutation operation, while larger particles with better performance need to avoid the impact of mutation on them. After mutation, the particle set can be transferred to the real distribution of state.

The specific flowchart is shown in Figure 1. The input of the flowchart is the state of particles at K-1 time (each particle represents the position of the target), and the output is the state of particles at K time. Firstly, according to the state of particles and the visual perception information of the target at K-1 time, the state of each particle is initialized and the weight is updated, and the particles with high weight are inherited preferentially through the weight sequencing. Secondly, in the process of resampling, combined with the quantum genetic algorithm, we set the number of genetic iterations (expressed by K). In each generation, the quantum bits and superposition states are used to encode the particles, and then the quantum rotating gate is used to carry out genetic and mutation operations on the particles in the population, so as to get the new generation of particles. If the set number of times has been reached, the iteration is over. If not, a new cycle will be carried out, and finally, the diverse particles after crossing and mutation will be obtained, which can effectively solve the problem of particle depletion caused by the long-term running of the target.

5. Simulation Experiment and Result Analysis

In order to detect the performance of QGPF tracking algorithm, the standard particle filter algorithm (PF), the particle swarm optimization filter algorithm (PSOPF), and the algorithm in this paper (QGPF) are simulated and compared. The three algorithms are simulated and tested by two kinds of target models: nonlinear target tracking model and time-constant value model.

5.1. Nonlinear Target Tracking Model. Next, a nonlinear, non-Gaussian target tracking problem is simulated, and the

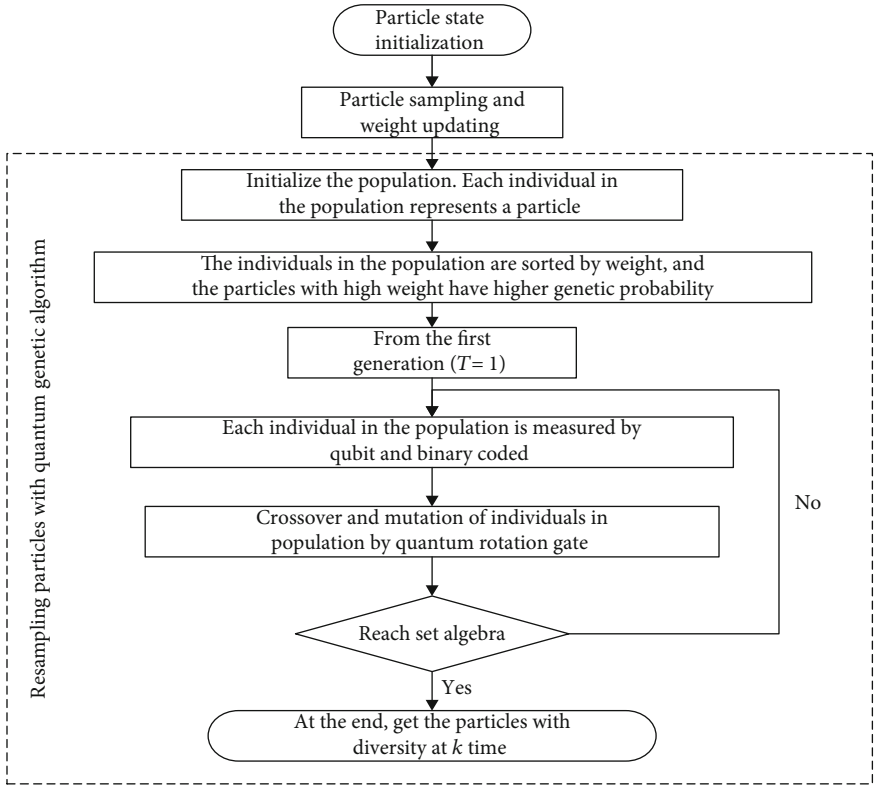


Figure 1: Flowchart of quantum genetic algorithm combined with particle filter.

motion state equation and measurement equation of the system are as follows:

$$\begin{cases} x(k) = 1 + \sin(0.03\pi k) + 0.5x(k-1) + w(k-1) \\ y(k) = \begin{cases} 0.2x^2(k) + v(k), k \le 30 \\ -2 + 0.4x(k) + v(k), k > 30. \end{cases} \end{cases} \tag{12}$$

The output results are represented by the mean value of the particle set of each particle filter algorithm:

$$\hat{x}_k = \frac{1}{S}\sum_{i=1}^{S} x_k^i. \tag{13}$$

The evaluation standard of filtering error is

$$\text{RMSE} = \sqrt{\frac{1}{T}\sum_{k=1}^{T}\left(x_k - \hat{x}_k\right)^2}, \tag{14}$$

in which $w(k-1)$ and $v(k)$, the system noise and measurement noise, are all zero mean Gaussian white noise. The choice of population size requires a trade-off between accuracy, stability, and running time, sometimes taking into account the dimensions of the problem. Generally speaking, if we focus on reducing operation time, the population size can be set at about 40; if we prefer high precision and high stability, it can be set at 50 to 80. When the population size is greater than 100, the improvement of its precision is not obvious [32, 33]. Therefore, in terms of population size, we choose 50 and 100.

Figures 2–5 shows the filtering state estimation and filtering error statistics when the number of filtering particles is 50 and 100, respectively.

It can be seen from Figures 2–5 that compared with the standard PF and PSOPF algorithm, the QGPF algorithm proposed in this paper can achieve higher precision state estimation, and the filtering error of QGPF algorithm at each time is relatively uniform, which shows that the algorithm in this paper has better algorithm stability. This is because the QGP algorithm applies quantum genetics to the resampling phase of particles, expands the diversity of particle set, uses quantum cross-mutation operation to generate new particles, and sets a higher probability to realize the high degree of freedom particles in the algorithm to continue the offspring particles, effectively avoiding particle degradation. At the same time, in order to avoid excessive error caused by particle dispersion, it advances according to the weight of particles, line mutation operation, so that the particles set to the real distribution of the state around the transfer.

QGA is integrated into PF, and the particles are binary coded with quantum bits and quantum superposition states.

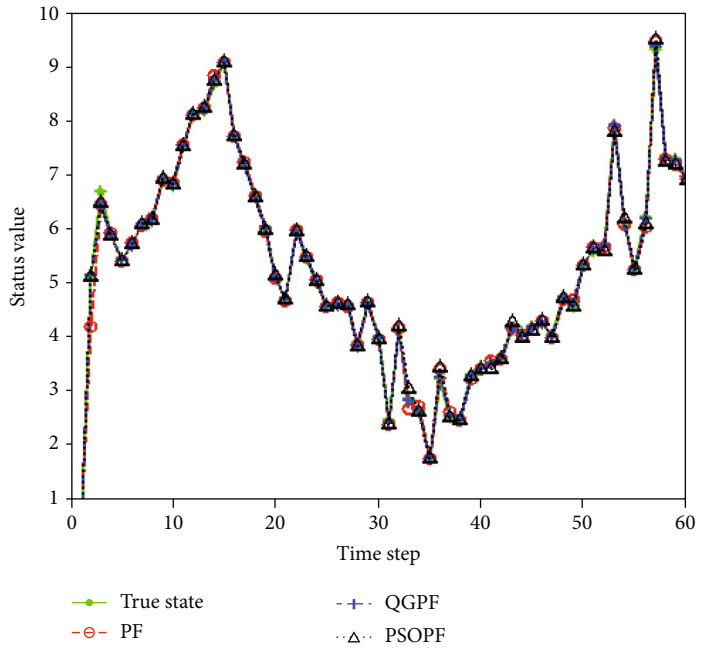


Figure 2: Filter state estimation ($N = 50$).

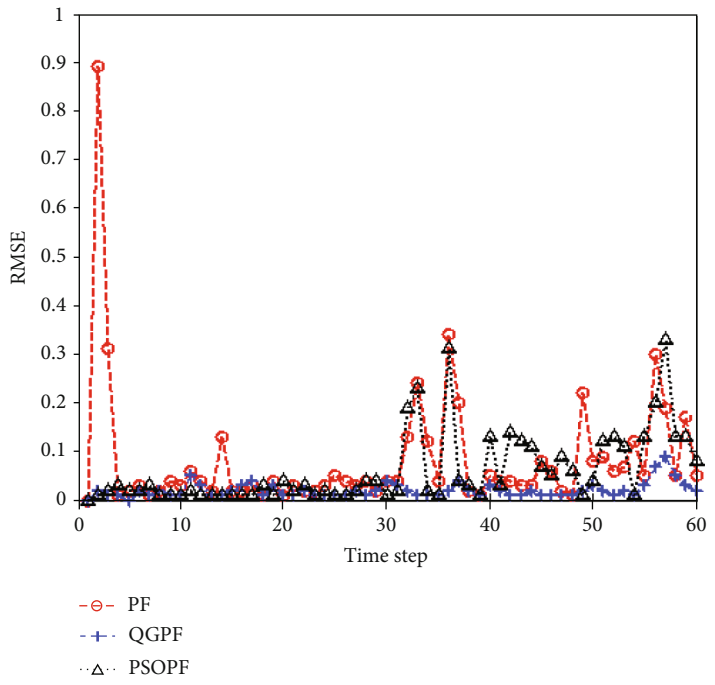


Figure 3: RMSF mean curve ($N = 50$).

The stability and accuracy of the algorithm are guaranteed by selecting the appropriate selection probability, crossover probability, and mutation probability. In the results of Figures 2–5, this conclusion has been verified.

5.2. Time-Constant Model. The traditional PF algorithm has insufficient tracking ability due to sample degradation, especially in the time-constant model. Next, the improved particle filter algorithm is simulated in the time-constant model to test its tracking ability. The system state equation and observation equation of the time-constant model are shown in formula (15).

$$\begin{cases} x(k) = \begin{cases} 5 & k \le 30 \\ 10 & 30 < k \le 60 \end{cases} \\ y(k) = x(k) + v(k). \end{cases} \tag{15}$$

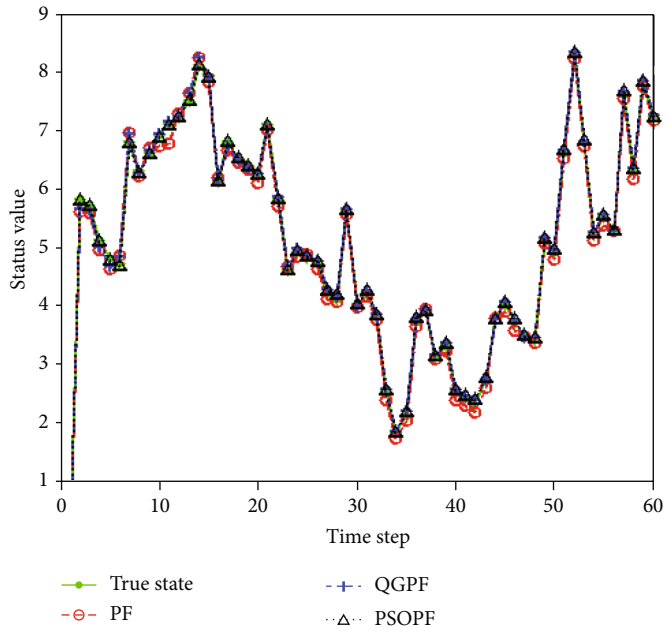


Figure 4: Filter state estimation ($N = 100$).

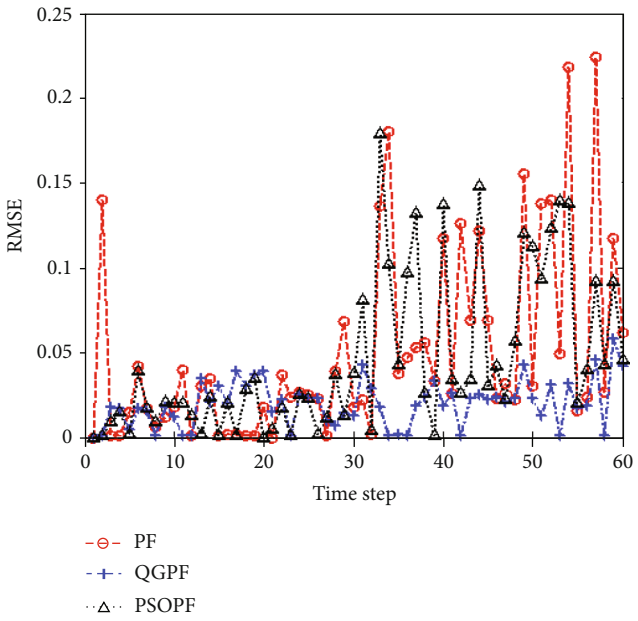


Figure 5: RMSF mean curve ($N = 50$).

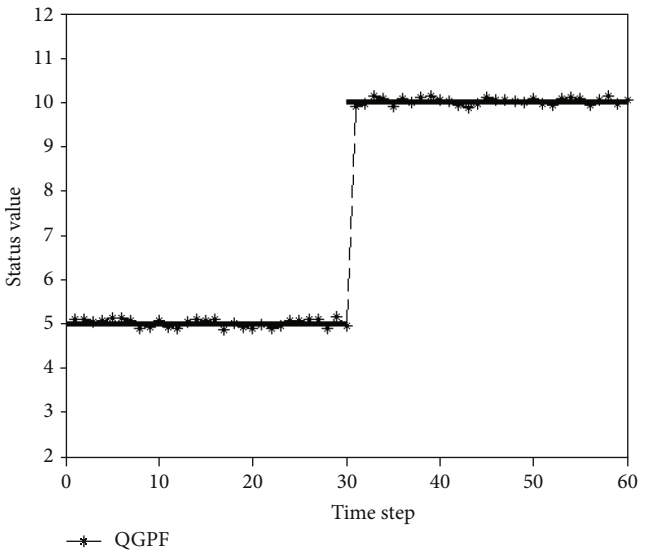


Figure 6: Comparison of time-constant target tracking track.

Figure 6 shows the comparison of real state and state estimation trajectory. The results show that the optimization of PF with QG can keep the diversity of particles, prevent the degradation of particles, and achieve stable and accurate target tracking.

PF algorithm has good performance in dealing with the system state problems of nonlinear and non-Gaussian distribution, but the effect of conventional particle filter algorithm is not robust under complex conditions, mainly including how to design and select the importance sampling density function and how to solve the problem of particle degradation along with the iteration process and the increase of the calculation amount caused by a large number of particle sampling. The simulation results show that the algorithm proposed in this paper has solved the above problems well and has satisfactory stability and accuracy.

6. Conclusion

Particle filter, which is commonly used in target tracking, is a method to realize the posterior probability density estimation of target state through the prediction and update of particle state and weight. Particle degradation is an inevitable problem in standard particle filter. After updating the particle state, in order to solve the problem of particle degradation,

it is necessary to resample the particle set to remove the unimportant particles. However, such removal often makes the particle set lose its diversity, and quantum genetic algorithm can well maintain the diversity of the population. Therefore, the introduction of quantum genetic algorithm into particle filter can better solve the problem of particle filter degradation. At the same time, the quantum parallel operation also improves the real-time tracking. In the process of target tracking, complex background, partial or total occlusion of target and illumination change will affect the tracking accuracy and real-time performance. In this paper, the particle filtering algorithm combined with the quantum genetic algorithm (QGA) uses binary coding and feature fusion of particles through qubits and quantum superposition states to reduce the degradation of particles and improve the tracking accuracy and real-time performance.

Data Availability

We are not supposed to share the data due to project privacy policy.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This research is supported by the "cyan and blue project" of Jiangsu Higher Education Institutions of China.

References

[1] M. Nieto, A. Cortés, O. Otaegui, J. Arróspide, and L. Salgado, "Real-time lane tracking using Rao-Blackwellized particle filter," *Journal of Real-Time Image Processing*, vol. 11, no. 1, pp. 179–191, 2016.

[2] P. M. Djuric, J. H. Kotecha, J. Zhang et al., "Particle filtering," *IEEE Signal Processing Magazine*, vol. 20, no. 5, pp. 19–38, 2003.

[3] F. S. Wang, M. Y. Lu, Q. J. Zhao, and Z. J. Yuan, "Particle filtering algorithm," *Chinese Journal of Computers*, vol. 37, no. 8, pp. 1679–1694, 2014.

[4] A. Doucet, S. Godsill, and C. Andrieu, "On sequential Monte Carlo sampling methods for Bayesian filtering," *Statistics and computing*, vol. 10, no. 3, pp. 197–208, 2000.

[5] R. Van Der Merwe, A. Doucet, N. De Freitas, and E. A. Wan, "The unscented particle filter," *Advances in neural information processing systems*, pp. 584–590, 2001.

[6] J. Y. Zuo, Y. Z. Zhang, and Y. Liang, "Particle filter based on adaptive part resampling," *Acta Automatica Sinica*, vol. 38, no. 4, pp. 647–651, 2012.

[7] X. Fu and Y. Jia, "An improvement on resampling algorithm of particle filters," *IEEE Transactions on Signal Processing*, vol. 58, no. 10, pp. 5414–5420, 2010.

[8] T. Li, T. P. Sattar, and S. Sun, "Deterministic resampling: unbiased sampling to avoid sample impoverishment in particle filters," *Signal Processing*, vol. 92, no. 7, pp. 1637–1645, 2012.

[9] Y. Kao and E. A. Zahara, "A Hybrid genetic algorithm and particle swarm optimization for multimodal functions," *Applied Soft Computing*, vol. 8, no. 2, pp. 849–857, 2008.

[10] L. Zhou, X. H. Cheng, and Y. X. Zhu, "Terrain aided navigation for autonomous underwater vehicles with coarse maps," *Measurement Science and Technology*, vol. 27, no. 9, pp. 1–9, 2016.

[11] F. Zhao, S. S. Ge, J. Zhang, and W. He, "Celestial navigation in deep space exploration using spherical simplex unscented particle filter," *IET Signal Processing*, vol. 12, no. 4, pp. 463–470, 2018.

[12] B. Claus and R. Bachmayer, "Terrain-aided navigation for an underwater glider," *Journal of Field Robotics*, vol. 32, no. 7, pp. 935–951, 2015.

[13] W. Wei, Q. Xu, L. Wang et al., "GI/Geom/1 queue based on communication model for mesh networks," *International Journal of Communication Systems*, vol. 27, no. 11, pp. 3013–3029, 2013.

[14] C. Kwok, D. Fox, and M. Meila, "Adaptive real-time particle filters for robot localization," in *Proceedings of IEEE International Conference on Robotics and Automation*, pp. 2836–2841, Washington D. C., USA, 2009.

[15] O. Straka and M. Ŝimandl, "A survey of sample size adaptation techniques for particle filters," *IFAC Proceedings Volumes*, vol. 42, no. 10, pp. 1358–1363, 2009.

[16] T. Li, S. Sun, T. P. Sattar, and J. M. Corchado, "Fight sample degeneracy and impoverishment in particle filters: a review of intelligent approaches," *Expert Systems with Applications*, vol. 41, no. 8, pp. 3944–3954, 2014.

[17] S. Li, S. Zhao, B. Cheng, E. Zhao, and J. Chen, "Lightweight particle filter for robust visual tracking," *IEEE Access*, vol. 6, pp. 32310–32320, 2018.

[18] M. Firouznia, K. Faez, H. Amindavar, and J. A. Koupaei, "Chaotic particle filter for visual object tracking," *Journal of Visual Communication and Image Representation*, vol. 53, pp. 1–12, 2018.

[19] M. D. Jenkins, P. Barrie, T. Buggy, and G. Morison, "Selective sampling importance resampling particle filter tracking with multibag subspace restoration," *IEEE Transactions on Cybernetics*, vol. 48, no. 1, pp. 264–276, 2018.

[20] L. Jia-qiang, Z. Rong-hua, Z. Yan-ping, C. Jin-li, and Z. Chun-yan, "Target tracking algorithm based on adaptive strong tracking particle filter," *IET Science, Measurement & Technology*, vol. 10, no. 7, pp. 704–710, 2016.

[21] R. H. Zhang, Z. C. He, H. W. Wang, F. You, and K. N. Li, "Study on self-tuning tyre friction control for developing main-servo loop integrated chassis control system," *IEEE Access*, vol. 5, pp. 6649–6660, 2017.

[22] R. Ma, L. Zhang, G. Li, S. X. Du Jiang, and D. Chen, "Grasping force prediction based on sEMG signals," *Alexandria Engineering Journal*, vol. 59, 2020.

[23] D. Tang and Y. J. Zhang, "Combining mean-shift tracking algorithm based on multi-feature fusion," in *Proceedings of the 33rd Chinese Control Conference*, 2014.

[24] S. Yi, Z. He, X. You, and Y.-M. Cheung, "Single object tracking via robust combination of particle filter and sparse representation," *Signal Processing*, vol. 110, pp. 178–187, 2015.

[25] Y. F. Cai et al., "Visual vehicle tracking based on deep representation and semisupervised learning," *Journal of Sensors*, vol. 2017, Article ID 6471250, 6 pages, 2017.

[26] J. Xin, X. Du, and J. Zhang, "Deep learning for robust outdoor vehicle visual tracking," *2017 IEEE International Conference on Multimedia and Expo (ICME)*, 2017, Hong Kong, China, 2017.

[27] W. Cheng, Y. Sun, G. Li, G. Jiang, and H. Liu, "Jointly network: a network based on CNN and RBM for gesture recognition," *Neural Computing and Applications*, vol. 31, Supplement 1, pp. 309–323, 2019.

[28] Z. H. Huang, X. Xu, H. H. Zhu, and M. C. Zhou, "An efficient group recommendation model with multiattention-based neural networks," in *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–14, 2020.

[29] Y. He, G. Li, Y. Liao et al., "Gesture recognition based on an improved local sparse representation classification algorithm," *Cluster Computing*, vol. 22, Supplement 5, pp. 10935–10946, 2019.

[30] L. Dong, Q. Guo, and W. Wu, "Speech corpora subset selection based on time-continuous utterances features," *Journal of Combinatorial Optimization*, vol. 37, no. 4, pp. 1237–1248, 2019.

[31] Y. Sun, C. Xu, G. Li et al., "Intelligent human computer interaction based on non redundant EMG signal," *Alexandria Engineering Journal*, vol. 59, 2020.

[32] W. Wei, H. Song, W. Li, P. Shen, and A. Vasilakos, "Gradient-driven parking navigation using a continuous information potential field based on wireless sensor network," *Information Sciences*, vol. 408, no. 2, pp. 100–114, 2017.

[33] X. Sun, H. Zhang, W. Meng, R. Zhang, K. Li, and T. Peng, "Primary resonance analysis and vibration suppression for the harmonically excited nonlinear suspension system using a pair of symmetric viscoelastic buffers," *Nonlinear Dynamics*, vol. 94, no. 2, pp. 1243–1265, 2018.

Hindawi
Wireless Communications and Mobile Computing
Volume 2020, Article ID 8867157, 12 pages
https://doi.org/10.1155/2020/8867157

Research Article

User-Edge Collaborative Resource Allocation and Offloading Strategy in Edge Computing

Zhenquan Qin, Xueyan Qiu, Jin Ye, and Lei Wang

School of Software, Dalian University of Technology, 116620, China

Correspondence should be addressed to Zhenquan Qin; qzq@dlut.edu.cn

Received 16 March 2020; Revised 11 May 2020; Accepted 25 May 2020; Published 12 June 2020

Academic Editor: Wei Wang

The foundation of urban computing and smart technology is edge computing. Edge computing provides a new solution for large-scale computing and saves more energy while bringing a small amount of latency compared to local computing on mobile devices. To investigate the relationship between the cost of computing tasks and the consumption of time and energy, we propose a computation offloading scheme that achieves lower execution costs by cooperatively allocating computing resources by mobile devices and the edge server. For the mixed-integer nonlinear optimization problem of computing resource allocation and offloading strategy, we segment the problem and propose an iterative optimization algorithm to find the approximate optimal solution. The numerical results of the simulation experiment show that the algorithm can obtain a lower total cost than the baseline algorithm in most cases.

1. Introduction

The rapid development of the Internet of Things (IoT) and mobile devices has facilitated the development of emerging applications such as machine learning [1] and face recognition [2]. Relying on these rapidly developing new IoT technologies and the Internet of Everything scenarios, the process of urban informatization is greatly accelerated, and smart cities have become a vision for development [3–5]. Smart cities can perform real-time analysis based on urban big data, provide new models of efficient and sustainable urban governance, and establish effective communication between people and cities. In recent years, the topic of smart cities has attracted widespread attention. With the introduction of various learning technologies, people hope to make cities smarter by collecting, storing, and processing big data.

One problem is that most of the emerging IoT technologies, including various learning technologies, are computationally intensive. However, due to volume and heat considerations, sensor nodes or other mobile devices that collect big data for smart cities usually have limited processing power, battery power, and storage space, making it difficult to meet the needs of all computing tasks. Cloud computing provides a solution [6, 7]. Migrating computing tasks from the device to the cloud server eliminates the limitations of the device itself. However, the cloud server is remotely isolated from the device in terms of topology and geographical location, and multiple storage and forwarding, as well as long-distance communications, cause huge transmission and propagation delays [8]. Smart cities need to establish efficient communication with people, and high latency will reduce the user experience. And for applications with high real-time requirements, this high latency is intolerable, and even in extreme cases, high latency can endanger user safety, such as autonomous driving [9], automatic navigation boats [10], and health care [11].

As an emerging distributed cloud architecture, edge computing provides a new solution for smart cities [12, 13]. Deploy servers near the demanders of the service, avoiding the high latency of a centralized cloud. At the same time, offloading all tasks to the edge server saves more energy compared to placing all tasks locally on the device. Moreover, edge computing can effectively reduce the pressure on the core network. For example, smart transportation is

an important part of smart cities. The intelligent transportation system analyzes the videos and videos acquired by a large number of traffic cameras to generate highly efficient traffic strategies. Traditional video surveillance systems must upload all surveillance videos to the cloud before analyzed, which will increase the traffic load on the core network. If edge computing is used, you can analyze at the edge server and reduce network pressure and the energy consumption of the entire monitoring system [14], which gives new vitality to smart transportation.

However, edge servers often have limited computing resources [15], considering the economic benefits and scalability of deployment. One reason is that edge servers are deployed in large numbers and close to users, and the economic benefits of deployment need to be considered. So a single edge server does not need and cannot have as many resources as a cloud center. Then, the edge server cannot fully process all tasks. If all tasks are transmitted to the edge server without any difference, it will result in low processing efficiency and long processing time for the task. So how to offload tasks based on demand is a question worth studying.

The process of migrating computing tasks from a mobile device to an edge server or a cloud center is called offloading [16–18]. The development of an offloading strategy is a classic and important issue in the field of edge computing. Offloading tasks to the edge server will inevitably lead to higher latency than local execution while reducing power consumption. Therefore, mobile devices need to choose the appropriate offloading strategy to meet their needs, such as minimizing task execution time [19–21], minimizing power consumption [22, 23], balancing delays, and power consumption [24, 25].

In many of the past work, resource allocation problems have been discussed. Most of them focus on solving the allocation of computing resources or the allocation of communication resources for edge servers [15, 22–24]. Some discussed the computational resource allocation and transmission power control of mobile devices [19, 25]. In fact, the resource allocation of mobile devices and edge servers will affect each other and work together to generate an offloading policy. Therefore, it is necessary to study the relationship between computing resource allocation and offloading decision generation.

Based on the above, we summarize the motivation of the work as follows: (1) The offloading decision should be reasonable. The constraints of computing resources and energy need to be considered to achieve a better task assignment. (2) Delay and energy consumption are two important parameters of the task. The allocation of computing resources has an impact on latency and energy consumption. (3) A parameter should be developed based on the deadline to express the real-time requirements of the task. (4) In heterogeneous edge servers, the execution time of tasks is different from local execution. The task execution costs need to explicitly consider the heterogeneity of edge servers.

In this paper, we proposed a collaborative computing resource allocation scheme and an offloading strategy for mobile devices and the edge server. Our contributions are as follows:

(1) We propose an integrated framework for computing resource allocation and computational task offloading in a mobile edge computing network

(2) We propose a computation offloading scheme that achieves lower execution costs by cooperatively allocating computing resources by mobile devices and the edge server

(3) Compared to the weighting factors of execution delay and energy consumption based on user preferences, we add the deadline to the definition of the weighting factor to reflect the real-time requirements of the task

(4) We segment the complex problem and propose Genetic-Algorithm-Based-Iterative (GABI) algorithm to obtain an approximate optimal solution

(5) The paper is organized as follows. In Section 2, we review the related works. Section 3 introduces the system modeling and problem formulation. Section 4 gives the solution. The simulation results are given and discussed in Section 5. Section 6 concludes this paper

2. Related Works

For mobile edge computing, computing offloading is the key technology. A lot of work has been done on the computation offloading technology to achieve the purpose of reducing calculation delay [8, 19–21] or saving equipment energy [22, 23, 26]. In [8], Xiao and Krunz studied the trade-off between users' QoE and fog node power efficiency in fog computing networks, focusing on users' QoE measured by average service response time. The author proposed a novel cooperative strategy called offload forwarding, and further proposed a distributed optimization algorithm based on distributed alternating direction method of multipliers (ADMM) to maximize the user's QoE. Rodrigues et al. in [19] considered computing and communication at the same time. To minimize service latency, the author controlled processing delay through virtual machine migration and improved transmission delay through transmission power control. In [20], Kao et al. proposed a complete polynomial-time approximation scheme (FPTAS) Hermes to solve the formulated NP-hard problem, thereby, minimizing application delay while meeting the specified resource utilization constraints. Jia et al. in [21] present an online task offloading algorithm that minimizes the completion time of the application on the mobile device. In addition to the line topology task graphs, the authors further considered the general topology task graphs. For concurrent tasks, the author used a load-balancing heuristic to increase the parallelism between mobile devices and the cloud. Some literature focuses on the energy consumption of mobile devices. In [22], You and Huang minimized the weighted sum of mobile energy consumption by solving the convex optimization problem under the constraint of calculating the waiting time. The author discussed both the infinite and limited edge cloud computing capabilities. For the latter, the author proposed

a suboptimal resource allocation algorithm to reduce complexity. You et al. in [23] studied the resource allocation of multiuser mobile edge computing systems based on time division multiple access (TDMA) and orthogonal frequency division multiple access (OFDMA), and by solving the convex optimization problem and the mixed-integer problem, the weighted total mobile energy consumption under the constraint of computing the waiting time is minimized. In [26], Zhang et al. proposed an offloading scheme that optimizes energy with guaranteed delay. This scheme considers the link to the status of the fronthaul network and the backhaul network at the same time and uses an artificial fish swarm algorithm for global optimization.

Time delay and energy are the main indicators to evaluate the performance of offloading calculation. It is meaningful to study the trade-off between them. The author in [15] designed the QoE maximization framework. QoE is a cost reduction achieved by offloading tasks to fog nodes or remote cloud servers, where the cost of performing tasks includes computing energy and computing latency. In [24], Zhang et al. proposed an online dynamic task allocation plan to study the trade-off between energy consumption and execution delay of the MEC system with energy harvesting function. An online dynamic Lyapunov-optimized offloading algorithm (ODLOO) is proposed, which could determine task allocation, reduce the execution delay by increasing the running frequency of the local CPU, and save energy by selecting a suitable data transmission channel. [25] proposed an energy-aware offloading scheme, which can jointly optimize the allocation of communication and computing resources with limited energy and sensitive delay. To calculate the mixed-integer nonlinear problem of shunting and resource allocation, Zhang et al. proposed an iterative search algorithm that combines internal penalty functions with DC programming to find the best solution. [27] studied the joint design of local execution and calculation offloading strategies in a multiuser MEC system. Mao et al. proposed an online algorithm based on Lyapunov optimization, which determines the CPU cycle frequency of local execution and the transmit power and bandwidth of the computation load distribution. The author used simulation experiments to verify that the proposed algorithm can balance the power consumption of mobile devices and the quality of computing experience. However, with the exception of [25], most of the work that studied the trade-off between energy consumption and delay has not clearly defined the weighting factors that weigh the two. Most of them described the weighting factors as a task attribute that changes according to user needs or verified the effectiveness of the proposed algorithm by adjusting the weighting factors. In [25], the author defined the weighting factor as the current residual energy ratio of mobile devices, which is an innovative idea and also gives some inspiration to this article.

The deadline is an important attribute for computing tasks [23, 24, 27–30]. The deadline of the task represents the delay tolerance of the task. In the work focusing on task scheduling, the deadline is used as an evaluation index. As in [31], the author formulated a task scheduling and offloading strategy so that more tasks can meet the deadline requirements, and the number of tasks that have exceeded the deadline, that is, the error rate, is used as the evaluation index. In most of the tasks that focus on calculating the offload, the deadline of the task is only a constraint, which does not give full play to the attribute of deadline and cannot express the real-time requirements of the task. In this paper, inspired by the literature [25], the deadline is introduced into the definition of the weighting factor.

Limited resources are a characteristic of edge computing systems, which prompted us to study the problem of resource allocation. Allocable resources include the communication power, channel bandwidth, and computing ability of user equipment and edge servers. Considering the allocation of multiple resources at the same time complicates the problem. The authors in [15] studied computing resource allocation, formulated a computational offloading game to model competition among IoT users, and effectively allocated the limited processing power of fog nodes. In [22], You and Huang formulated the optimal resource allocation problem as a convex optimization problem, considering the situation of unlimited and limited cloud computing capabilities. The authors in [23] further considered the OFDMA system, whose optimal resource allocation is formulated as a mixed-integer problem. Zhang et al. considered the allocation of communication and computing resources in [25], including channel selection, user equipment computing resource allocation, and communication power allocation. In [32], Yu et al. considered the allocation of radio and computing resources. They proposed a near-optimal algorithm for scheduling subcarriers and CPU time slots, respectively, and further proposed a joint scheduling algorithm to manage subcarriers and CPUs coordinately, thereby, achieving the goal of reducing energy consumption. In [33], Sun et al. studied the joint problem of network economics and resource allocation in MEC and considered the incentive framework that can maximize system efficiency in the Industrial Internet of Things (IIoT). They proposed two types of dual auction schemes with dynamic pricing, namely, a breakeven-based double auction (BDA) and a more efficient dynamic pricing based double auction (DPDA) to determine the matching pair between the IIoT MD and the edge server, and Pricing mechanism to achieve high system efficiency under local constraints.

3. System Model and Problem Formulation

3.1. System Model. We consider a network system with N mobile devices and an edge server as shown in Figure 1. The channel bandwidth between the mobile device $u_i (i \in \{1, 2, \cdots, N\})$ and the edge server is w. The propagation delay of edge server and mobile device communication is negligible because the distance between the two is often only a single hop or a limited number of hops.

In this network, the mobile device u_i with limited energy $E_i^{\max}$ will generate M tasks that need to be completed in each time slot. The jth $(j \in \{1, 2, \cdots, M\})$ task generated by device u_i in time slot t $(t \in \{1, 2, \cdots, T\})$ is represented as $\tau_{i,j,t} = (d_{i,j,t}, c_{i,j,t}^U, c_{i,j,t}^E, t_{i,j,t}^{\text{deadline}}, s_{i,j,t})$, where $d_{i,j,t}$ represents the

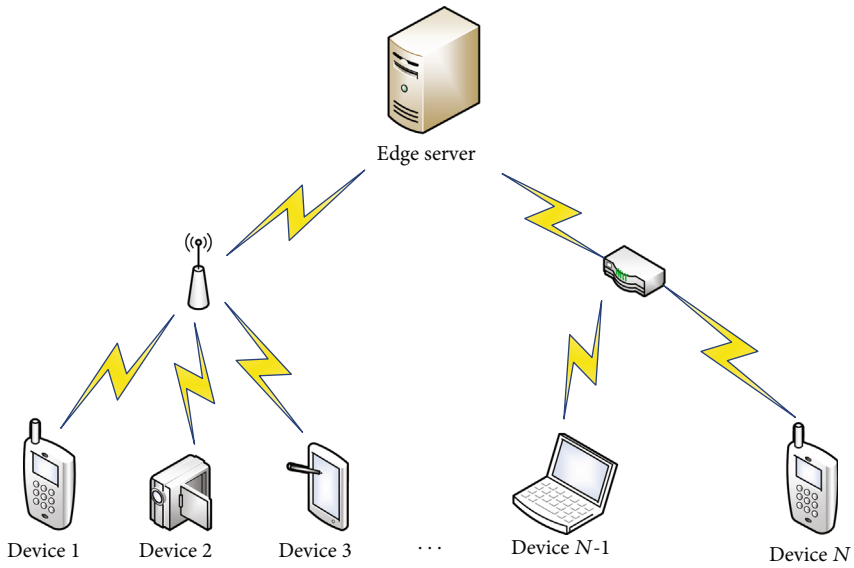


Figure 1: Network system model.

amount of data for the task, $c^U_{i,j,t}$ denotes the number of CPU cycles required for the task if it is executed locally, $c^E_{i,j,t}$ means the number of CPU cycles required for the task to execute on the heterogeneous edge server, and $t^{\text{deadline}}_{i,j,t}$ is the tolerable execution time. $s_{i,j,t}$ is the offloading strategies of task $\tau_{i,j,t}$. $s_{i,j,t} = 1$ when the task is performed locally, if not, $s_{i,j,t} = 0$.

In each time slot, after the mobile device generates the tasks and allocates the local computing resources, the information of the task and the local computing resource allocation scheme are transmitted to the edge server. The edge server calculates the offloading policy and returns it to the mobile device. The task is executed locally or offloaded according to the offloading policy.

3.2. Local Computing. We define $f^U_{i,j,t}$ as the CPU frequency of the mobile device u_i, which represents the local computation ability of the device. Note that this value could be changed to get a better offloading strategy. Then, the computation execution time $t^{\text{execute}^-U}_{i,j,t}$ for the locally performed task $\tau_{i,j,t}$ is

$$t^{\text{execute}^-U}_{i,j,t} = \frac{c^U_{i,j,t}}{f^U_{i,j,t}}. \tag{1}$$

We ignore the queue delay of local execution. Then, the time consumption $t^U_{i,j,t}$ of task $\tau_{i,j,t}$ local execution is

$$t^U_{i,j,t} = t^{\text{execute}^-U}_{i,j,t}. \tag{2}$$

The energy consumption can be expressed as

$$e^U_{i,j,t} = \kappa \left(f^U_{i,j,t}\right)^2 c^U_{i,j,t}, \tag{3}$$

where κ is the energy efficiency coefficient. It is related to the chip architecture, and we assume that this value is constant.

3.3. Edge Computing. When the mobile device u_i decides to offload the task $\tau_{i,j,t}$ to the edge server, it needs to obtain the data uplink transmission rate

$$r_{i,j,t} = w \log_2 \left(1 + \frac{p_i h_{i,j,t}}{\sigma + \sum_{k \in N, k \neq i} \left(1 - s_{k,j,t}\right) p_k h_{k,j,t}}\right), \tag{4}$$

where w is the channel bindwidth and p_i is the transmission power of the mobile device u_i. h_i and σ are channel gain and noise power. The summation term in the denominator represents interference between channels. Therefore, the transmission time of the uplink is

$$t^{\text{trans}}_{i,j,t} = \frac{d_{i,j,t}}{r_{i,j,t}}. \tag{5}$$

The time task $\tau_{i,j,t}$ consumes to execute on a heterogeneous edge server is

$$t^{\text{execute}^-E}_{i,j,t} = \frac{c^E_{i,j,t}}{f^E_{i,t}}, \tag{6}$$

where $f^E_{i,t}$ is the computing resource allocated by the edge server to the task from u_i in the time slot t. As with communication resources, computing resources are also allocated periodically.

So, we get the total time consumption of edge computing

$$t^E_{i,j,t} = t^{\text{trans}}_{i,j,t} + t^{\text{execute}^-E}_{i,j,t}. \tag{7}$$

We ignore the return time because many computationally intensive applications have a much smaller amount of data than the input data.

Energy consumption is

$$e_{i,j,t}^{E} = p_i t_{i,j,t}^{\text{trans}}. \tag{8}$$

In general, the time consumption and energy consumption constitute the execution cost of the task. The cost of locally executed task $G_{i,j,t}^{U}$ and offloaded task $G_{i,j,t}^{E}$ are expressed as the weighted sum of time and energy consumption

$$G_{i,j,t}^{U} = \alpha_{i,j,t} t_{i,j,t}^{U} + \left(1 - \alpha_{i,j,t}\right) \beta e_{i,j,t}^{U}, \tag{9}$$

$$G_{i,j,t}^{E} = \alpha_{i,j,t} t_{i,j,t}^{E} + \left(1 - \alpha_{i,j,t}\right) \beta e_{i,j,t}^{E}. \tag{10}$$

β is a normalization factor used to eliminate the unit difference between time and energy consumption, which can be equal to the ratio of average time to average energy consumption.

The weight factor $\alpha_{i,j,t}$ ($\alpha_{i,j,t} \in [0, 1]$) is the task's trade-off between time and energy consumption. By changing the weighting factor, the user can achieve faster execution or more energy-efficient execution. In this paper, we define the weighting factor by the time consumption of the task

$$\alpha_{i,j,t} = \frac{t_{i,j,t}^{\text{execute}^{-}U}}{t_{i,j,t}^{\text{deadline}}}. \tag{11}$$

The task is generated on the mobile device, and the deadline is established with the execution time on the mobile device, so it is reasonable to perform the weighting factor with the local execution time.

After adding the offloading decision $s_{i,j,t}$, the generalization of the cost of task $\tau_{i,j,t}$ is expressed as:

$$G_{i,j,t} = s_{i,j,t} G_{i,j,t}^{U} + \left(1 - s_{i,j,t}\right) G_{i,j,t}^{E}. \tag{12}$$

3.4. Problem Formulation. From what has been discussed above, we formulate problems as follows:

$$\min_{s, f^U, f^E} \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{j=1}^{M} \left[s_{i,j,t} G_{i,j,t}^{U} + \left(1 - s_{i,j,t}\right) G_{i,j,t}^{E} \right], \tag{13}$$

$$\text{s.t.} \quad f_i^{\min} \le f_{i,j,t}^{U} \le f_i^{\max}, \quad \forall i, j, t, \tag{14}$$

$$0 \le \sum_{i=1}^{N} f_{i,t}^{E} \le F^{E}, \quad \forall t, \tag{15}$$

$$s_{i,j,t} t_{i,j,t}^{U} + \left(1 - s_{i,j,t}\right) t_{i,j,t}^{E} \le t_{i,j,t}^{\text{deadline}}, \quad \forall i, j, t, \tag{16}$$

$$\sum_{t=1}^{T} \sum_{j=1}^{M} \left[s_{i,j,t} e_{i,j,t}^{U} + \left(1 - s_{i,j,t}\right) e_{i,j,t}^{E} \right] \le E_i^{\max}, \quad \forall i, j, t, \tag{17}$$

$$s_{i,j,t} \in \{0, 1\}, \quad \forall i, j, t. \tag{18}$$

Constraint (14) indicates that the computing resource allocated to a local task cannot exceed the maximum CPU frequency, and constraint (15) indicates that the computing resources allocated by an edge server in each time slot cannot exceed its maximum capacity. Constraint (17) ensures that the mobile device is able to complete computing tasks before all energy is consumed. Constraint (18) indicates that the offloading decision is a 0-1 variable.

The objective function is nonlinear and contains the multiplication of variables, so this is a complex mixed-integer nonlinear optimization problem with binary variables. This kind of problem is difficult to solve, so we will divide the problem in the following paper, and iteratively use genetic algorithm method to find the approximate solution and carry out simulation experiments to prove the validity of the solution.

4. Solution

4.1. Local Execution Cost. The cost $G_{i,j,t}^{U}$ of the local execution calculation task is split, and the specific expressions of the parameter weight factor, time consumption, and energy consumption are brought in.

$$\begin{aligned} G_{i,j,t}^{U} &= \alpha_{i,j,t} t_{i,j,t}^{U} + \left(1 - \alpha_{i,j,t}\right) \beta e_{i,j,t}^{U} \\ &= \frac{\left(c_{i,j,t}^{U}\right)^2}{\left(f_{i,j,t}^{U}\right)^2 t_{i,j,t}^{\text{deadline}}} + \left(1 - \frac{c_{i,j,t}^{U}}{f_{i,j,t}^{U} t_{i,j,t}^{\text{deadline}}}\right) \beta \kappa c_{i,j,t}^{U} \left(f_{i,j,t}^{U}\right)^2. \end{aligned} \tag{19}$$

After observation, we found that the value of $G_{i,j,t}^{U}$ depends only on $f_{i,j,t}^{U}$, so $G_{i,j,t}^{U}$ is a function of $f_{i,j,t}^{U}$.

Lemma 1. *The function $G_{i,j,t}^{U}(f_{i,j,t}^{U})$ is unimodal.*

Lemma 1 can be easily proved by the derivative and monotonicity of the function. Simplify the formula (19) and it looks like $y = ax^2 + bx + c(1/x^2)$, where $a > 0$, $b < 0$, and $c > 0$. The second derivative is $\mathrm{d}^2y/\mathrm{d}x^2 = 6c(1/x^4) + 2a$. In the positive interval, the second derivative is always positive, then, the first derivative is monotonically increasing. The first derivative $\mathrm{d}y/\mathrm{d}x = 2ax - 2c(1/x^3) + b$ is continuous when $x > 0$, and as x tends to zero and positive infinity, the derivative tends to be negative infinity and positive infinity. From the interval value theorem, there is a point that the first derivative is zero, and the point is a minimum point.

From Lemma 1, we get the CPU frequency $f_{i,j,t}^{*}$ that minimizes the local execution cost, which is the ideal best, but we still need to consider the constraints (14), (16), and (17).

In constraint (14), we have to consider the upper and lower bounds of $f_{i,j,t}^{U}$. And in constraint (16),

$$\begin{aligned} t_{i,j,t}^{U} = \frac{c_{i,j,t}^{U}}{f_{i,j,t}^{U}} \le t_{i,j,t}^{\text{deadline}}, \\ f_{i,j,t}^{U} \ge \frac{c_{i,j,t}^{U}}{t_{i,j,t}^{\text{deadline}}} \end{aligned}. \tag{20}$$

For a single task, we can override the constraint (17) to make it look simpler

$$\kappa\left(f_{i,j,t}^{U}\right)^{2} c_{i,j,t}^{U} \le E_{i,j,t}^{\max},$$
$$f_{i,j,t}^{U} \le \sqrt{\frac{E_{i,j,t}^{\max}}{\kappa c_{i,j,t}^{U}}}, \tag{21}$$

where $E_{i,j,t}^{\max}$ is a real-time value indicating the remaining battery power of the device while performing the task.

The CPU frequency of the mobile device $f_{i,j,t}^{U}$ fluctuates within $[f_i^{\min}, f_i^{\max}]$. Then, we get the new bound of $f_{i,j,t}^{U}$

$$f_{i,j,t}^{ub} = \min\left\{f_i^{\max}, \sqrt{\frac{E_{i,j,t}^{\max}}{\kappa c_{i,j,t}^{U}}}\right\},$$
$$f_{i,j,t}^{lb} = \max\left\{f_i^{\min}, \frac{c_{i,j,t}^{U}}{t_{i,j,t}^{\text{deadline}}}\right\}, \tag{22}$$

and the best value to get the lowest $G_{i,j,t}^{U}$

$$f_{i,j,t}^{U} = \begin{cases} f_{i,j,t}^{*}, & f_{i,j,t}^{lb} \le f_{i,j,t}^{*} \le f_{i,j,t}^{ub}, \\ f_{i,j,t}^{lb}, & f_{i,j,t}^{*} < f_{i,j,t}^{lb}, \\ f_{i,j,t}^{ub}, & f_{i,j,t}^{ub} < f_{i,j,t}^{*}. \end{cases} \tag{23}$$

4.2. Edge Execution Cost. We found that the problem of task execution cost on edge servers could not be considered separately like that on mobile devices.

$$G_{i,j,t}^{E} = \alpha_{i,j,t} t_{i,j,t}^{E} + \left(1-\alpha_{i,j,t}\right)\beta e_{i,j,t}^{E},$$
$$= \alpha_{i,j,t}\left(t_{i,j,t}^{trans} + \frac{c_{i,j,t}^{E}}{f_{i,t}^{E}}\right) + \left(1-\alpha_{i,j,t}\right)\beta e_{i,j,t}^{E}. \tag{24}$$

For locally executed computing tasks, different tasks are independent in terms of computing resource allocation. But it is different at the edge, where all tasks share the computing resources of one edge server. The task execution cost and the computing resources allocated by the edge server are inversely related. Therefore, it is meaningless to study the minimum cost of a computing task separately, and the resource allocation for all tasks needs to be considered.

We can consider the minimum cost of all tasks in a time slot. The edge server allocates computing resources for newly arrived tasks in each time slot, and tasks in different time slots are independent of each other.

$$\min_{f^E} \sum_{i=1}^{N}\sum_{j=1}^{M} \alpha_{i,j,t}\left(t_{i,j,t}^{\text{trans}} + \frac{c_{i,j,t}^{E}}{f_{i,t}^{E}}\right) + \left(1-\alpha_{i,j,t}\right)\beta e_{i,j,t}^{E}, \tag{25}$$

$$\text{s.t.} \quad 0 \le f_{i,t}^{E} \le F^{E}, \quad \forall i, \tag{26}$$

$$0 \le \sum_{i=1}^{N} f_{i,t}^{E} \le F^{E}. \tag{27}$$

Note that we have not defined time constraints and energy constraints here. We will mention them in the next section and use them as conditions for developing an offload strategy.

Problem (25) is a nonlinear optimization problem. Our goal is to obtain a global optimal solution. Here, we use the genetic algorithm to get an approximate optimal solution. Then, we can get the minimum edge computing cost $G_{i,j,t}^{E}$ according to $f_{i,t}^{E}$, the calculation result of question (25).

4.3. Offloading Decision and Resource Allocation. The initial allocation of computing resources is for all tasks in this time slot. That is, before the initial offload policy, the default offload policy is to offload all tasks, $\boldsymbol{s} = 0$. An obvious problem is that due to deadline constraints and energy constraints, offloading all tasks cannot be taken as the final offloading decision, and the allocation of computing resources with the default offloading condition is not reasonable. Therefore, in this section, we propose a Genetic-Algorithms-Based-Iterative (GABI) algorithm to obtain reasonable offloading decisions and resource allocation methods through iterative calculations.

The first step is to filter based on constraints. We denote the new offloading strategy $\boldsymbol{s}'$ and assign initial values to it $\boldsymbol{s}' = \boldsymbol{s}$. In the previous section, we mentioned that time constraint is one of the conditions for filtering. After obtaining the calculation resource allocation result $f_{i,t}^{E}$, we can get the total time of edge execution $t_{i,j,t}^{E}$. If $t_{i,j,t}^{E} > t_{i,j,t}^{\text{deadline}}$, then $s'_{i,j,t} = 1$.

For energy constraints, the energy consumption of the edge calculation is independent of the calculation resource allocation, so the energy constraint is not added in the problem (25). Moreover, we have obtained local computational resource allocations that satisfy energy constraints in the previous section, while the energy consumption of edge computations is generally less than local computation. So $s'_{i,j,t} = 1$ if $e_{i,j,t}^{E} > e_{i,j,t}^{U}$. We use the matrix $\boldsymbol{s}^T$ to record the current value of $\boldsymbol{s}'$, $\boldsymbol{s}^T = \boldsymbol{s}'$. Then, the process of filtering can be expressed as Algorithm 1.

$\boldsymbol{s}^T$ will be used later as one of the conditions for stopping the iteration. The other condition is G_1, which is the total task execution cost calculated by bringing $\boldsymbol{s}^T$ and $\boldsymbol{f}^E$ into (12).

The second step is to filter tasks based on the execution costs. After the computing resource allocation of the edge server is obtained, the execution cost of the task on the edge server can be calculated and further compared with the task's local execution cost. If a task is found to cost more to execute on the edge server, we think it is better to perform it locally on the user device. The second filtering is based on the first filtering, so we define a new offloading decision matrix $\boldsymbol{s}''$ and initialize it, $\boldsymbol{s}'' = \boldsymbol{s}'$. This step is shown in Algorithm 2.

After filtering, we get a new offloading strategy $\boldsymbol{s}''$. If $\boldsymbol{s}'' = \boldsymbol{s}$, then, we think that a stable offloading strategy has

Input: task set τ , offloading strategy $\boldsymbol{s}$
Output: new decision $\boldsymbol{s}'$
1: $\boldsymbol{s}' \longleftarrow \boldsymbol{s}$
2: **for** every task $\tau_{i,j,t}$ **do**
3: Get $t^E_{i,j,t}$ using $f^E_{i,t}$ from (25). $t^E_{i,j,t} = t^{trans}_{i,j,t} + t^{execute^-E}_{i,j,t} = (d_{i,j,t}/r_{i,j,t}) + (c^E_{i,j,t}/f^E_{i,t})$.
4: Get $e^E_{i,j,t}$. $e^E_{i,j,t} = p_i t^{trans}_{i,j,t} = p_i(d_{i,j,t}/r_{i,j,t})$.
5: **if** $t^E_{i,j,t} > t^{deadline}_{i,j,t}$ **or** $e^E_{i,j,t} > e^U_{i,j,t}$ **then**
6: $s'_{i,j,t} = 1$.
7: **end if**
8: **end for**

Algorithm 1: Filtering Based on Constraints

Input: task set τ , offloading strategy $\boldsymbol{s}'$
Output: new decision $\boldsymbol{s}''$
1: $\boldsymbol{s}'' \longleftarrow \boldsymbol{s}'$
2: **for** every task $\tau_{i,j,t}$ **do**
3: Get $G^E_{i,j,t}$. $G^E_{i,j,t} = \alpha_{i,j,t} t^E_{i,j,t} + (1-\alpha_{i,j,t})\beta e^E_{i,j,t} = (c^U_{i,j,t}/f^U_{i,j,t} t^{deadline}_{i,j,t}) t^E_{i,j,t} + (1 - c^U_{i,j,t}/f^U_{i,j,t} t^{deadline}_{i,j,t})\beta e^E_{i,j,t}$.
4: Get $G^U_{i,j,t}$. $G^U_{i,j,t} = ((c^U_{i,j,t})^2/(f^U_{i,j,t})^2 t^{deadline}_{i,j,t}) + (1 - c^U_{i,j,t}/f^U_{i,j,t} t^{deadline}_{i,j,t})\beta\kappa c^U_{i,j,t}(f^U_{i,j,t})^2$.
5: **if** $G^E_{i,j,t} \geqslant G^U_{i,j,t}$ **then**
6: $s''_{i,j,t} = 1$.
7: **end if**
8: **end for**

Algorithm 2: Filtering Based on Cost

been obtained, and the GABI algorithm can return directly. So we get the final offloading strategy $\boldsymbol{s}$ and the computing resource allocation method $\boldsymbol{f}^E$.

But if the offloading strategy $\boldsymbol{s}''$ is different from the default offloading strategy $\boldsymbol{s}$, the obtained computing resource allocation strategy cannot minimize the edge execution cost because some allocated resources are not used. Therefore, computing resources need to be reallocated.

Assign the value of $\boldsymbol{s}''$ to $\boldsymbol{s}$. Once again, the optimization problem of minimizing the edge execution cost is solved. This time, $\boldsymbol{s}$ is added to remove the locally executed task:

$$\min_{\boldsymbol{f}^E} \sum_{i=1}^{N}\sum_{j=1}^{M}\left(1 - s_{i,j,t}\right)G^E_{i,j,t}, \tag{28}$$

$$\text{s.t.} \quad 0 \leq f^E_{i,t} \leq F^E, \quad \forall i, \tag{29}$$

$$0 \leq \sum_{i=1}^{N} f^E_{i,t} \leq F^E. \tag{30}$$

After the new resource allocation scheme f^E is obtained, we get $\boldsymbol{s}'$ using Algorithm 1, and get G_2 using $\boldsymbol{s}'$ and $f^{E'}$ by (12). At this time, compare G_2 and G_1. If $G_2 > G_1$, then stop iteration, and the final offloading strategy is determined to be $\boldsymbol{s}^T$. Assign $\boldsymbol{s}^T$ to $\boldsymbol{s}$, then get the final computing resource allocation method $\boldsymbol{f}^E$ through (28), and the GABI algorithm returns.

If $G_2 \leqslant G_1$, the iteration continues: recording $\boldsymbol{s}'$ with $\boldsymbol{s}^T$, filtering based on cost, judging the iteration termination conditions, iterating again, and so on.

The computing resource allocation and the offloading decision is given in Algorithm 3. Algorithm 3 shows the whole process of the GABI algorithm.

5. Simulation Result

In this section, we design simulation experiments to verify the effectiveness of our proposed algorithm. We compare the performance of our proposed method with the baseline scheme, the probabilistic offloading scheme, and the energy-aware offloading scheme. We consider an edge computing network with one edge server and N mobile devices, where N ranges from 3 to 9. We set the maximum computing power of the edge server to 4 GHz, and the computing power of mobile devices ranges from 0.2 GHz to 1 GHz. The initial CPU frequency of the mobile device is randomly generated between 0.5 and 1 GHz, which is used to calculate the normalization factor β and the deadline of tasks. Mobile device transmission power is 20 dBm.

The simulation process consists of 10 time slots in which the mobile device generates 5 computational tasks. The amount of task data is randomly generated between [300,

Input: task set τ
Output: offloading decision $\boldsymbol{s}$, computing resource allocation scheme $\boldsymbol{f}^E$
1: **for all** $t \in T$ **do**
2: Initialization: $\boldsymbol{s} \longleftarrow \mathbf{0}$, $\boldsymbol{s}' \longleftarrow \boldsymbol{s}$, $\boldsymbol{s}^T \longleftarrow \boldsymbol{s}$, $\boldsymbol{s}'' \longleftarrow \boldsymbol{s}$.
3: Get G_1 according to (24)(12).
4: **repeat**
5: $\boldsymbol{s} \longleftarrow \boldsymbol{s}''$.
6: Calculate $\boldsymbol{f}^E$ using genetic algorithms through (28).
7: Get $\boldsymbol{s}'$ by calling Algorithm 1.
8: Get G_2 using $\boldsymbol{s}'$ through (24)(9)(10)(12).
9: **if** $G_2 > G_1$ **then**
10: $\boldsymbol{s} \longleftarrow \boldsymbol{s}^T$.
11: Calculate $\boldsymbol{f}^E$ using genetic algorithms through (28).
12: **return** $\boldsymbol{s}, \boldsymbol{f}^E$.
13: **end if**
14: $\boldsymbol{s}^T \longleftarrow \boldsymbol{s}$.
15: Calculate G_1 using $\boldsymbol{s}^T$ through (24)(9)(10)(12).
16: Get $\boldsymbol{s}''$ by calling Algorithm 2.
17: **until** $\boldsymbol{s}'' = \boldsymbol{s}$
18: **return** $\boldsymbol{s}, \boldsymbol{f}^E$.
19:**end for**

Algorithm 3: Genetic-Algorithms-Based-Iterative(GABI) algorithm

1200] KB. The local computing resources required to complete the task are randomly generated between [0.1, 1] GHz. Considering the heterogeneity of devices, the required computing resources performed on the edge server is [0.2, 1.2] times than that of local execution. The task deadline is randomly generated in [0.5, 5] s, and the value is guaranteed to be greater than the local task execution time calculated by the initial CPU frequency of the mobile device.

The total bandwidth of the network is 2 MHz, which is evenly distributed by all mobile devices. The path loss model between the mobile device and the edge server is considered as the lognormal distribution.

In Figure 2, we consider the impact of offloading strategy on total cost, time, and energy consumption under different mobile device quantities. "All local" and "All MEC" in Figure 2 represent all tasks executed locally and all offloading schemes. As shown, compared with two baselines, our solution maintains lower cost, medium execution time, and power consumption in most cases. When there are fewer mobile devices, the edge server allocates more resources to each user, so the task execution time can be even lower than the local execution, and the total task cost is lower. As the number of users increases, the edge execution time becomes longer, so the total cost of edge execution approaches or even exceeds local execution, even though the energy consumed by edge execution is still much lower than local execution. The energy consumption of local execution tasks is several times or even ten times that of edge execution, so even if the time consumption is slightly lower, there is still a higher total cost.

In Figure3, we compare the performance of our scheme and the probabilistic offloading scheme. The idea of the probabilistic offloading scheme is that the computing server has a certain probability of being offloaded by the edge server. Whether to offload is determined by probability, without comparing the cost of tasks performed locally and edge. We compared the cases with unloading probabilities of 0.2, 0.5, and 0.8. The experimental results show that our proposed scheme can always achieve the lowest task execution cost. When the number of users is small, the computing resources of the edge server are less competitive, and the total cost of the solution with a high probability of offloading is lower. This is consistent with the "ALL MEC" baseline solution in Figure2 when the number of users is low. As the number of users increases, the low-probability offloading solution gradually achieves lower costs. In terms of time and energy consumption, our solution is also at a relatively low level.

In Figure4, we compare the impact of the definition and treatment of weighting factors in this article and in the literature [25]. The document [25] defines the weighting factor $\alpha = \alpha * r^E$, where r^E is the ratio of the remaining energy of the current user equipment to the maximum battery capacity. In order to reflect the effect of the method in [25], we have selected a special piece of data. In this simulation, the remaining power of the user equipment is small. As a result, at the beginning of the third time slot, the weighting factor of the task calculated by the [25] method decreases rapidly, as shown by the polyline on the way with the legend "Rest Energy Defines α". In order to save energy, most tasks choose to offload, even if the cost of edge execution is much higher than local execution. The method in this paper does not focus on energy perception, so in most cases, the strategy of keeping the lowest cost is selected. When the local execution cost of sending some tasks is low, choose not to offload.

In Figure 5, we consider the impact of local computing resource allocation on total cost, time, and energy

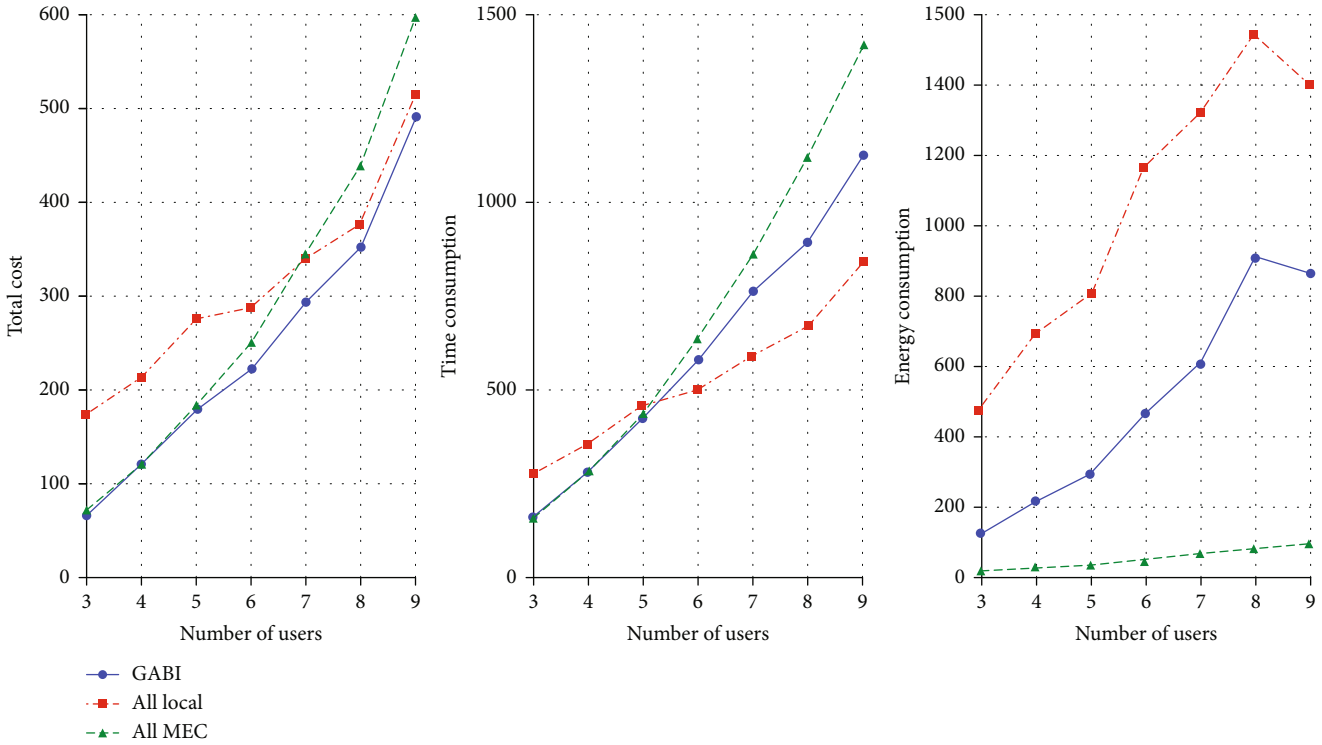


Figure 2: The impact of offloading strategy on total cost, time, and energy consumption under different mobile device quantities.

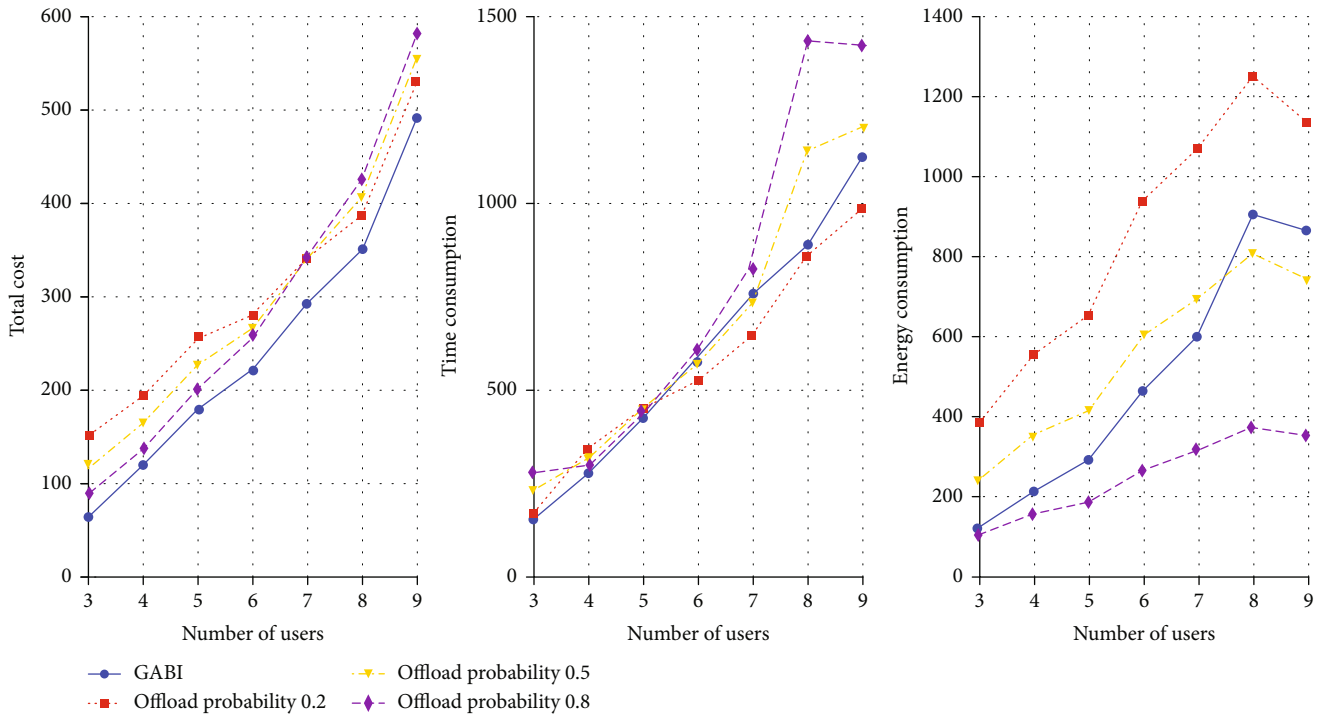


Figure 3: The impact of offloading strategy on total cost, time, and energy consumption under different mobile device quantities.

consumption under different user quantities. "fUmax", "fUmin", and "fUmid", respectively, represent local computing resource allocation with $f_i^{\max}$, $f_i^{\min}$, and $(f_i^{\max} + f_i^{\min})/2$. "fUoriginal" stands for the CPU frequency preset by the user equipment. Our approach to finding the best local resource allocation scheme is feasible and effective. Compared to the for baselines, our proposed method always finds the lowest cost. Higher CPU frequencies result in faster

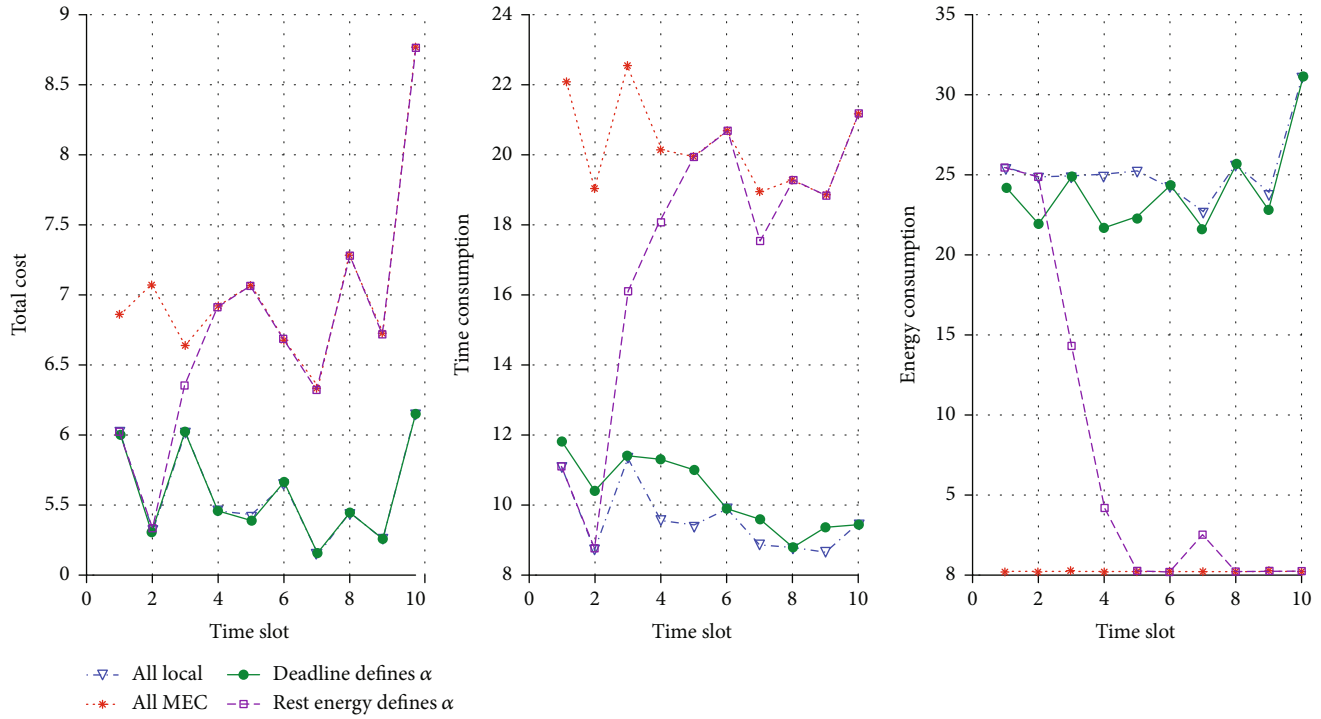


Figure 4: In multiple consecutive time slots, compare the impact of the proposed scheme and the energy-aware offloading scheme on the total cost, delay, and energy consumption.

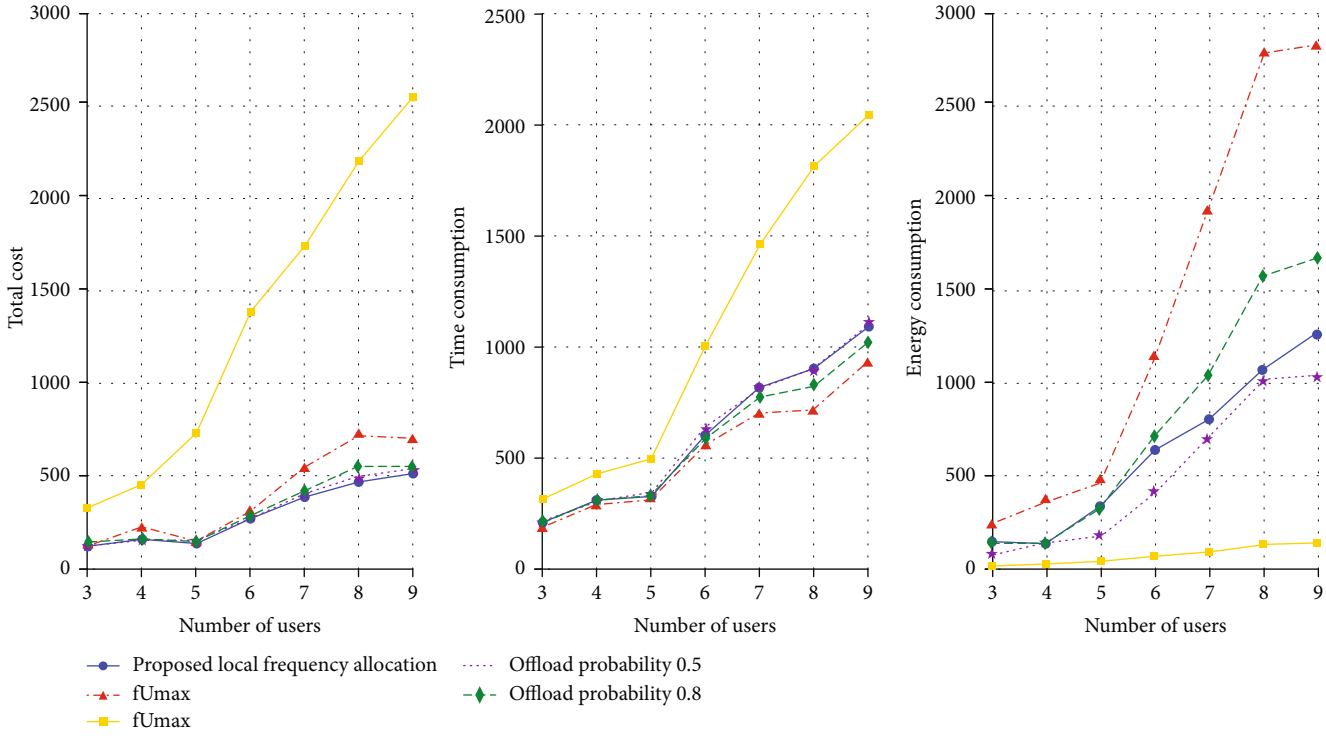


Figure 5: The impact of local computing resource allocation on total cost, time, and energy consumption under different mobile device quantities.

execution speeds, but at the same time result in higher energy consumption, as evidenced by the figure. Our algorithm achieves less energy consumption than "fUmax" and gets lower execution time than "fUmin", fully weighing the impact of time and energy on mission cost. It is noted that the average allocation "fUmid" can approach the optimal

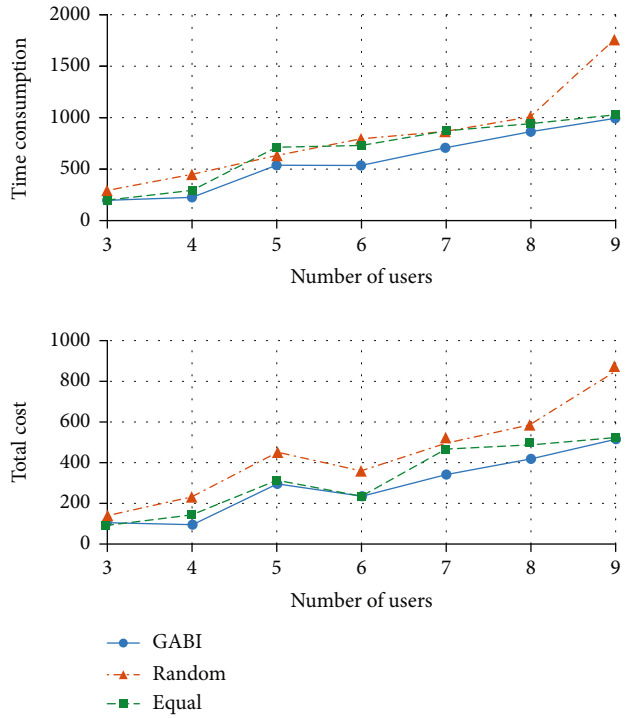


Figure 6: The impact of edge computing resource allocation on total cost and time consumption under different mobile device quantities.

solution in many cases. The reason is that in most cases, the value range of the user CPU is not extreme. The calculated local optimal resource allocation plan is a value that can be obtained in most cases and is very close to $(f_i^{\max} + f_i^{\min})/2$. Although the preset CPU frequency of the user equipment can also achieve better results, it is unstable.

In Figure 6, we compare the impact of edge server computing resource allocation schemes on time consumption and total cost under different user numbers. "Equal" and "Random", respectively, indicate that the edge server evenly allocates and randomly allocates computing resources to users. The edge server resource allocation scheme does not affect the energy consumption of edge execution, but it indirectly affects the offload decision and total cost. As shown in the figure, it is obvious that the random allocation scheme is costly and unstable, and the method of the average allocation is close to the optimal solution in many cases, but the proposed method can achieve the lowest cost and time consumption.

6. Conclusions

In this paper, we investigate the resource allocation and offloading decisions of mobile devices and edge servers in edge computing. To minimize the total cost of task execution, we define weighting factors based on deadlines to consider the impact of time and energy consumption on costs. We considered an edge computing network consisting of edge servers and users and built mathematical models. For difficult MINLP problems, we split the problem and iteratively optimize it. The result is that an approximate optimal solution is not a globally optimal solution. In future work, we will consider designing heuristic algorithms to reduce complexity and improve performance.

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 there is no conflict of interest regarding the publication of this paper.

Acknowledgments

The work was supported by "National Natural Science Foundation of China" with No. 61842601 and 61902052, "National Key Research and Development Plan" with No. 2017YFC0821003-2, "Dalian Science and Technology Innovation Fund" with No. 2019J11CY004, and "the Fundamental Research Funds for the Central Universities" with No. DUT19RC(3)003.

References

[1] S. Wang, T. Tuor, T. Salonidis et al., "When edge meets learning: adaptive control for resource-constrained distributed machine learning," in *IEEE INFOCOM 2018 - IEEE Conference on Computer Communications*, pp. 63–71, Honolulu, HI, USA, April 2018.

[2] P. Hu, H. Ning, T. Qiu, Y. Zhang, and X. Luo, "Fog computing based face identification and resolution scheme in internet of things," *IEEE Transactions on Industrial Informatics*, vol. 13, no. 4, pp. 1910–1920, 2017.

[3] R. Kitchin, "The real-time city? Big data and smart urbanism," *GeoJournal*, vol. 79, no. 1, pp. 1–14, 2014.

[4] K. Su, J. Li, and H. Fu, "Smart city and the applications," in *2011 International Conference on Electronics, Communications and Control (ICECC)*, pp. 1028–1031, Ningbo, China, September 2011.

[5] R. E. Hall, B. Bowerman, J. Braverman, J. Taylor, H. Todosow, and U. Von Wimmersperg, "The vision of a smart city," Tech. Rep., Brookhaven National Lab., Upton, NY, USA, 2000.

[6] K. Nowicka, "Smart city logistics on cloud computing model," *Procedia-Social and Behavioral Sciences*, vol. 151, pp. 266–281, 2014.

[7] T. Clohessy, T. Acton, and L. Morgan, "Smart City as a Service (SCaaS): a future roadmap for E-Government smart city cloud computing initiatives," in *2014 IEEE/ACM 7th International Conference on Utility and Cloud Computing*, pp. 836–841, London, UK, December 2014.

[8] Y. Xiao and M. Krunz, "Qoe and power efficiency tradeoff for fog computing networks with fog node cooperation," in *IEEE INFOCOM 2017 - IEEE Conference on Computer Communications*, pp. 1–9, Atlanta, GA, USA, May 2017.

[9] S. Liu, L. Liu, J. Tang, B. Yu, Y. Wang, and W. Shi, "Edge computing for autonomous driving: opportunities and challenges," *Proceedings of the IEEE*, vol. 107, no. 8, pp. 1697–1716, 2019.

[10] T. Yang, H. Feng, C. Yang, Y. Wang, J. Dong, and M. Xia, "Multivessel computation offloading in maritime mobile edge computing network," *IEEE Internet of Things Journal*, vol. 6, no. 3, pp. 4063–4073, 2019.

[11] A. M. Rahmani, T. N. Gia, B. Negash et al., "Exploiting smart e-Health gateways at the edge of healthcare Internet-of-Things: A fog computing approach," *Future Generation Computer Systems*, vol. 78, pp. 641–658, 2018.

[12] A. Giordano, G. Spezzano, and A. Vinci, "Smart agents and fog computing for smart city applications," in *Smart Cities. Smart-CT 2016. Lecture Notes in Computer Science, vol 9704*, E. Alba, F. Chicano, and G. Luque, Eds., pp. 137–146, Springer, Cham, 2016.

[13] Y. C. Hu, M. Patel, D. Sabella, N. Sprecher, and V. Young, "Mobile edge computing: a key technology towards 5g," *ETSI white paper*, vol. 11, no. 11, pp. 1–16, 2015.

[14] N. Chen, Y. Chen, Y. You, H. Ling, P. Liang, and R. Zimmermann, "Dynamic urban surveillance video stream processing using fog computing," in *2016 IEEE Second International Conference on Multimedia Big Data (BigMM)*, pp. 105–112, Taipei, Taiwan, April 2016.

[15] H. Shah-Mansouri and V. W. S. Wong, "Hierarchical fog-cloud computing for iot systems: a computation offloading game," *IEEE Internet of Things Journal*, vol. 5, no. 4, pp. 3246–3257, 2018.

[16] M. Aazam, S. Zeadally, and K. A. Harras, "Offloading in fog computing for iot: review, enabling technologies, and research opportunities," *Future Generation Computer Systems*, vol. 87, pp. 278–289, 2018.

[17] 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.

[18] X. Wang, Z. Ning, and L. Wang, "Offloading in internet of vehicles: a fog-enabled real-time traffic management system," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 10, pp. 4568–4578, 2018.

[19] T. G. Rodrigues, K. Suto, H. Nishiyama, and N. Kato, "Hybrid method for minimizing service delay in edge cloud computing through vm migration and transmission power control," *IEEE Transactions on Computers*, vol. 66, no. 5, pp. 810–819, 2017.

[20] Y.-H. Kao, B. Krishnamachari, M.-R. Ra, and F. Bai, "Hermes: latency optimal task assignment for resource-constrained mobile computing," *IEEE Transactions on Mobile Computing*, vol. 16, no. 11, pp. 3056–3069, 2017.

[21] M. Jia, J. Cao, and L. Yang, "Heuristic offloading of concurrent tasks for computation-intensive applications in mobile cloud computing," in *2014 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, pp. 352–357, Toronto, ON, Canada, April-May 2014.

[22] C. You and K. Huang, "Multiuser resource allocation for mobile-edge computation offloading," in *2016 IEEE Global Communications Conference (GLOBECOM)*, pp. 1–6, Washington, DC, USA, December 2016.

[23] C. You, K. Huang, H. Chae, and B. H. Kim, "Energy-efficient resource allocation for mobile-edge computation offloading," *IEEE Transactions on Wireless Communications*, vol. 16, no. 3, pp. 1397–1411, 2017.

[24] G. Zhang, W. Zhang, Y. Cao, D. Li, and L. Wang, "Energy-delay tradeoff for dynamic offloading in mobile-edge computing system with energy harvesting devices," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 10, pp. 4642–4655, 2018.

[25] J. Zhang, X. Hu, Z. Ning et al., "Energy-latency tradeoff for energy-aware offloading in mobile edge computing networks," *IEEE Internet of Things Journal*, vol. 5, no. 4, pp. 2633–2645, 2018.

[26] H. Zhang, J. Guo, L. Yang, X. Li, and H. Ji, "Computation offloading considering fronthaul and backhaul in small-cell networks integrated with mec," in *2017 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, pp. 115–120, Atlanta, GA, USA, May 2017.

[27] Y. Mao, J. Zhang, S. H. Song, and K. B. Letaief, "Power-delay tradeoff in multi-user mobile-edge computing systems," in *2016 IEEE Global Communications Conference (GLOBECOM)*, pp. 1–6, Washington, DC, USA, December 2016.

[28] T. Zhu, T. Shi, J. Li, Z. Cai, and X. Zhou, "Task scheduling in deadline-aware mobile edge computing systems," *IEEE Internet of Things Journal*, vol. 6, no. 3, pp. 4854–4866, 2019.

[29] Y. Xing and H. Seferoglu, "Predictive edge computing with hard deadlines," in *2018 IEEE International Symposium on Local and Metropolitan Area Networks (LANMAN)*, pp. 13–18, Washington, DC, USA, June 2018.

[30] Y. Wang, K. Wang, H. Huang, T. Miyazaki, and S. Guo, "Traffic and computation co-offloading with reinforcement learning in fog computing for industrial applications," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 2, pp. 976–986, 2019.

[31] J. Meng, H. Tan, C. Xu, W. Cao, L. Liu, and B. Li, "Dedas: Online task dispatching and scheduling with bandwidth constraint in edge computing," in *IEEE INFOCOM 2019 - IEEE Conference on Computer Communications*, pp. 2287–2295, Paris, France, April-May 2019.

[32] Y. Yu, J. Zhang, and K. B. Letaief, "Joint subcarrier and cpu time allocation for mobile edge computing," in *2016 IEEE Global Communications Conference (GLOBECOM)*, pp. 1–6, Washington, DC, USA, December 2016.

[33] W. Sun, J. Liu, Y. Yue, and H. Zhang, "Double auction-based resource allocation for mobile edge computing in industrial internet of things," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 10, pp. 4692–4701, 2018.

Hindawi
Wireless Communications and Mobile Computing
Volume 2020, Article ID 6896579, 15 pages
https://doi.org/10.1155/2020/6896579

Research Article

Capsules TCN Network for Urban Computing and Intelligence in Urban Traffic Prediction

Dazhou Li[1], **Chuan Lin**[2], **Wei Gao**[1], **Zeying Chen**[1], **Zeshen Wang**[3] and **Guangqi Liu**[4,5]

[1]*College of Computer Science and Technology, Shenyang University of Chemical Technology, Shenyang 110016, China*
[2]*Key Laboratory for Ubiquitous Network and Service Software of Liaoning Province, Dalian University of Technology, Dalian 116024, China*
[3]*School of Computer Science and Engineering, Northeastern University, Shenyang 110819, China*
[4]*Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang 110016, China*
[5]*University of Chinese Academy of Sciences, Beijing 100049, China*

Correspondence should be addressed to Chuan Lin; chuanlin1988@gmail.com

Received 15 January 2020; Revised 16 February 2020; Accepted 25 February 2020; Published 4 June 2020

Guest Editor: Wei Wang

Predicting urban traffic is of great importance to smart city systems and public security; however, it is a very challenging task because of several dynamic and complex factors, such as patterns of urban geographical location, weather, seasons, and holidays. To tackle these challenges, we are stimulated by the deep-learning method proposed to unlock the power of knowledge from urban computing and proposed a deep-learning model based on neural network, entitled Capsules TCN Network, to predict the traffic flow in local areas of the city at once. Capsules TCN Network employs a Capsules Network and Temporal Convolutional Network as the basic unit to learn the spatial dependence, time dependence, and external factors of traffic flow prediction. In specific, we consider some particular scenarios that require accurate traffic flow prediction (e.g., smart transportation, business circle analysis, and traffic flow assessment) and propose a GAN-based superresolution reconstruction model. Extensive experiments were conducted based on real-world datasets to demonstrate the superiority of Capsules TCN Network beyond several state-of-the-art methods. Compared with HA, ARIMA, RNN, and LSTM classic methods, respectively, the method proposed in the paper achieved better results in the experimental verification.

1. Introduction

Empowered by Internet of Things (IoTs) technologies and advanced algorithms that can collect and handle massive traffic datasets, urban computing and intelligence can make more informed decisions and create feedback loops between actual traffic situation and management department in the urban environment [1]. It can bridge the gaps between ubiquitous sensing, intelligent computing, cooperative communication, and big data management technologies to create novel solutions which can improve urban traffic environments, quality of life, and smart city systems [2]. In these urban computing methods, the huge datasets used by the scientists are all from various sources, such as geographic information, taxi GPS, and online weather web sites [3].

Urban traffic prediction has become a challenging urgent task for the development of a smart urban city, as it can afford visions for urban planning and traffic administration to improve the performance of urban transportation, as well as provide warnings for public security emergency message as timely [4]. Moreover, urban traffic prediction has been an important research issue with highly social shock [5]. When some emergences happen such as traffic accidents, an earthquake, tornado, and national holiday, urban traffic prediction becomes the top priority for authority (e.g., law enforcement) and traffic management operators (e.g., bus/ferry/subway) to protect people's safety and keep the work of social infrastructures [6]. Particularly for an enormous population city such as New York and London, the urban traffic is very heavy, which commonly leads to

more probability for different traffic collisions and accident situations [6].

To meet this challenge, we are with the purpose of deriving the urban traffic prediction from period, trend, geospatial, and external influences and generate an accurate prediction for the urban traffic in the next time window, which is considered to be an available way to dispose the urban computing. We propose a neural network-based method called Capsules TCN Network based on collected big traffic mobility data and two deep-learning architecture TCN and Capsules Network. For real time, we also proposed a further improvement method for spatial-temporal data processing to achieve supervision of urban area vehicle density.

2. Related Work

Traffic flow prediction has been considered as a key functional component of intelligent transportation systems. Meanwhile, artificial intelligence technology is rapidly growing and the fifth-generation communication technology is approaching [7–15]. Massive traffic data are being continuously collected through all kinds of sources, some of which can be treated and utilized as streaming data for understanding and predicting urban traffic [6]. All these stimulate us to take new efforts and achieve new success on this social issue by using such streaming mobility data and advanced artificial intelligence technologies [6].

The evolution of traffic flow can be considered to be a spatiotemporal process. As early as the 1970s, the autoregressive integrated moving average (ARIMA) model was used to predict the short-term traffic flow of expressways [16]. Traffic flow prediction based on a time series method is a widely used traffic flow prediction technology. Levin and Tsao applied Box-Jenkins time series analysis to predict highway traffic flow and found that the ARIMA (0, 1, 1) model was useful in the prediction of the most statistically significant [17]. Hamed et al. used the ARIMA model to predict the traffic volume of urban arterial roads [18]. In order to improve the prediction accuracy, many variants of ARIMA were proposed, such as Kohonen-ARIMA [19], subset ARIMA [20], ARIMAX [21], space-time ARIMA [22], and seasonal ARIMA [23]. In addition to ARIMA-type time series models, other types of time series models are also used for traffic flow prediction [24].

On account of the random and nonlinear nature of traffic flow, nonparametric methods have received widespread attention in the field of traffic flow prediction. Davis and Nihan used the KNN method for short-term traffic prediction on expressways [25]. Chang et al. proposed a dynamic multi-interval traffic forecasting model based on KNN nonparametric regression [26]. Faouzi developed an autoregressive function with a smooth kernel function for short-term traffic flow prediction, in which a function estimation technique was applied [27]. Sun et al. used a local linear regression model for short-term traffic prediction [28]. A traffic flow prediction method based on Bayesian network was also proposed [29]. It proposed an online learning weighted support vector regression (SVR) for short-term traffic flow prediction. Various artificial neural network models for predicting traffic flow have been established [30–32]. The MA, ES, and ARIMA models are used to obtain three related time series, which are the basis of the nature in the aggregation phase [33]. Zargari et al. developed different linear programming, multilayer perceptron, and fuzzy logic models to estimate 5- and 30-minute traffic flows [34]. Cetin and Comert combined the ARIMA model with the expectation maximization and cumulative sum algorithm [35].

Yao et al. proposed to combine the principal component analysis method with SVR and select urban multisection data to establish a road network short-term prediction model that took into account the relationship between time and space of multiple sections [36]. Li et al. used the wavelet decomposition and wavelet reconstruction of the traffic flow sequence data and then the use of Kalman filtering for dynamic data prediction [37]; Sun et al. proposed the application of the gray system theory to intersection traffic volume prediction [38]. Xiong et al. combined traditional linear models with artificial intelligence prediction models and proposed a short-term traffic flow prediction method based on artificial neural networks and Kalman filtering [39].

This article is divided into 6 sections: The first section describes the research background, significance, and purpose of the traffic forecast of urban vehicle traffic. The second section introduces the current situation and the structure of this article. The third section models the traffic forecast in urban areas and introduces the structure of Capsules TCN Network, which has two main technologies: Capsules Network and Temporal Convolutional Network. At the same time, the Capsules TCN Network model results are superresolution reconstructed to obtain a regional traffic flow forecast map with higher accuracy. The fourth section introduces the dataset used in the experiments and the data preprocesses, the experimental criteria, and the comparative baselines. Moreover, in the experimental environment, platform construction is introduced and the experimental results are demonstrated and analyzed. The fifth section summarizes the whole research.

3. Analytical Model of Regional Traffic

3.1. Regional Flow Prediction Problem. In urban areas, the indicator of vehicle flow can be used to indicate the vehicle flow in an area. This indicator can well reflect the traffic, population density, and public safety of a region. This article predicts two types of vehicle group traffic: inflow and outflow, as shown in Figure 1(a). Inflow refers to the total volume of vehicles entering a certain area from other areas within a given time interval. Outflow represents the total flow of vehicles leaving the area in a given time interval. Both types of traffic are used to indicate the movement patterns of vehicle traffic in urban areas. Understanding them can be of great help in risk assessment and traffic management. Inflow and outflow can be measured by the number of cars driving near the road, the number of cars driving on public transportation systems (e.g., subway and buses), the number of taxis, or all available data. Figure 1(b) shows an example of using the GPS trajectory of a rental car to measure the amount of traffic. The results show that the inflow

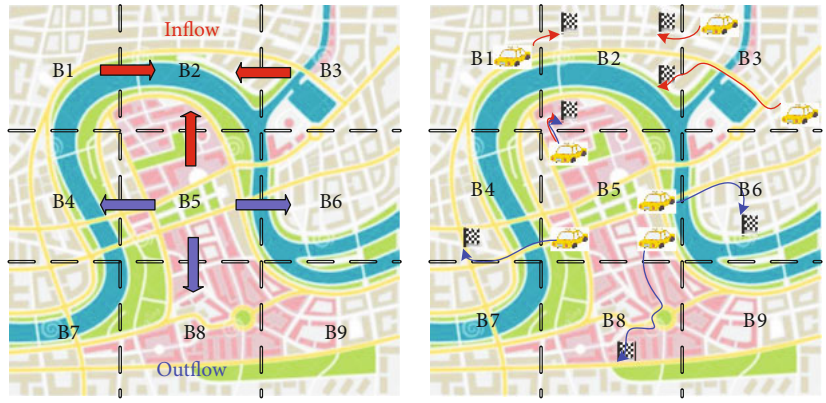


Figure 1: (a) Inflow and outflow. (b) Using the GPS trajectory of a rental car to measure the amount of traffic.

in area B2 is 4 and the outflow in B5 is also 4. Obviously, predicting traffic flow can be regarded as a spatial-temporal prediction problem.

There are three complex factors in the spatial-temporal prediction problem:

3.1.1. Space dependence. As shown in Figure 1(a), the inflow in the B2 area is affected by the outflow in its vicinity (such as B5). Similarly, the outflow of B5 will affect the inflow of other regions (such as B2). The inflow of the B2 region will affect its own outflow. Urban traffic flow may be even affected by distant areas. For example, people who live far away from the office always take the car or taxi to work, which means that the outflow of long-distance residential areas directly affects the inflow of office areas.

3.1.2. Time dependence. The change of the traffic flow in any area is generally continuous from the perspective of time. It means the traffic flow at the next moment and the traffic flow at the previous moment have the strongest correlation. With the increase of the time interval, the correlation of traffic flow will gradually decrease. Figure 2 shows the time-varying curves of the traffic flow in a typical residential area and a typical working area from our dataset. It can be seen that both curves are relatively smooth, reflecting the continuous change characteristics described above. At the same time, it can be seen from Figure 2 that the change curve of the traffic flow in the living area is different from the change curve of the traffic flow in the working area, which reflects the regional differences.

Different regions have different numbers of population densities. Residential areas are suitable for living and resting. In a residential area, each person has a larger unit space that is more suitable for living and resting. Therefore, the lower the population density of a residential area, the better the residential area. In the work area, the closer the workers are, the more convenient the communication is and the work is more efficient. Therefore, the population density in the work area is much larger than that in residential areas. Different population densities determine different needs for public transportation. It can be seen from Figure 2 that although the trend of the number of taxis in the residential

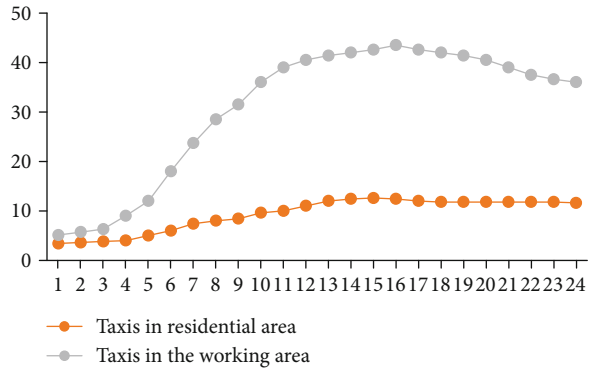


Figure 2: Time continuity and difference of the number of taxis in the residential area and the working area.

area and the working area is basically the same over time, there are obvious differences in the magnitude of the two.

As shown in Figure 3, whether it is a change in the traffic flow in the work area or a change in the traffic flow in the residential area, there are obvious characteristics of periodic changes. To further complicate matters, this periodicity will also be different under different time scales. When you observe in days, you can see the daily fluctuations of vehicles from morning to night. When you observe in weeks, you can see fluctuations of vehicles from work. If you look at the unit of year, you can see the impact of the climate and holidays on the traffic flow in the four seasons.

This paper divides time dependence into period and tendency. Period: traffic during the morning rush hour is similar on consecutive working days. The morning rush hour usually occurs from 8 AM to 10 AM, and the evening rush hour is usually from 17 to 21 PM, repeated every 24 hours. Tendency: there is a cyclical difference between traffic between a working day and a nonworking day, with a time interval of one week.

3.1.3. External factors. Some external factors, such as weather conditions and holidays, can drastically change traffic flow in different areas of the city. As shown in Figure 4, a rainstorm affects the speed of traffic on the road and further

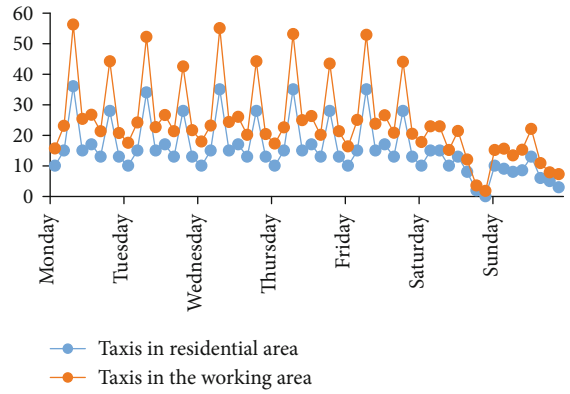


Figure 3: Time periodicity of the number of taxis in the residential area and the working area.

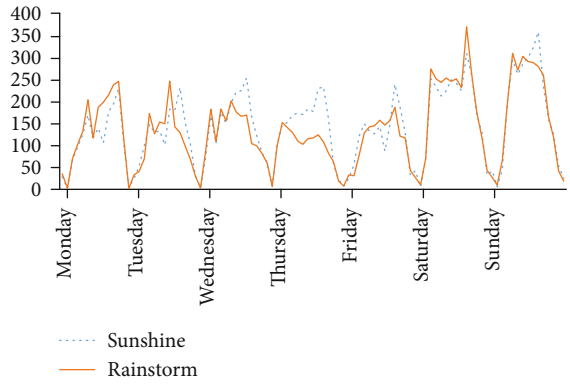


Figure 4: Impact of extreme weather on the number of taxis in the working area.

changes the area's traffic volume. Figure 5 shows the impact of holidays on a regional traffic.

There are many ways to divide the city area. According to the function, it can be divided into working area, residential area, mixed area, etc. It can also be divided according to the structure of the urban road network, and the city can be divided into main roads by using the map division method. The division method is introduced in the paper. We use grids to divide cities according to latitude and longitude. As shown in Figure 6(a), j and k represent the number of rows and columns in the area, respectively. In actual life, the values of j and k can be adjusted according to different city sizes and different application scenarios. In this paper, the scenario is divided into 16×16 grids.

Let R be the trajectory set of the ith time interval. For the grid (j, k) located in the jth row and the kth column, the inflow and outflow at the time interval i are defined as

$$\alpha_i^{j,k} = \sum_R \text{Count}(L_{i-1} \notin (j,k) \text{ and } L_i \in (j,k)), \quad (1)$$

$$\beta_i^{j,k} = \sum_R \text{Count}(L_i \in (j,k) \text{ and } L_{i+1} \notin (j,k)). \quad (2)$$

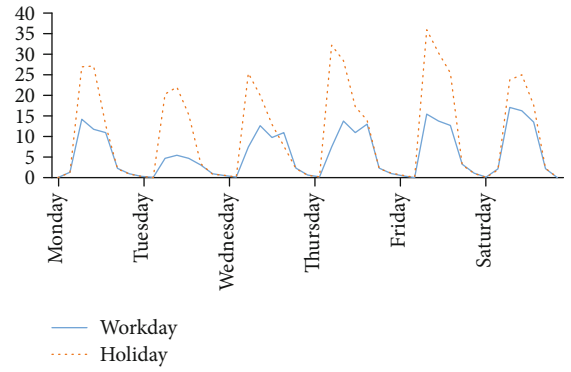


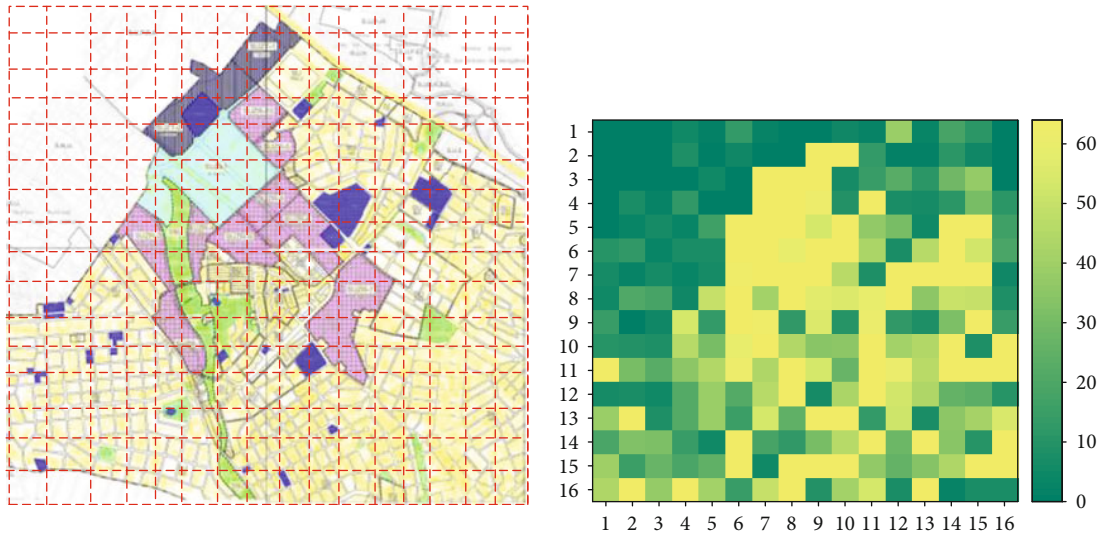
Figure 5: Impact of social event on the number of taxis in the working area.

Among them, L_i is the trajectory of all the vehicles in R at ith time interval. Here, the trajectory is determined according to the GPS coordinates in the dataset and the grids divided by the map. In the ith time interval, inflow of 16×16 grids in the entire area can be represented by a matrix composed of $\alpha_i^{j,k}$ as shown Figure 6(b). The traffic prediction problem is transformed into known historical data $\alpha_i^{j,k}$ and $\beta_i^{j,k}$ to predict $\alpha_{i+1}^{j,k}$ and $\beta_{i+1}^{j,k}$ in the next moment.

3.2. Algorithm Model of Capsules TCN. Both recurrent neural networks (RNN) and long-term short-term memory (LSTM) are capable of learning remote time dependence. However, if RNN or LSTM is used to simulate time periods and trends, it requires very long input sequences, which make the entire training process very complicated. According to the knowledge of space-time domain, only a few previous key frames will affect the next key frame. Therefore, we use time period, tendency, and geographic space to select key frames for modeling. Figure 7 shows the architecture of Capsules TCN Network proposed in the paper. It consists of four primary parts, which model time period, tendency, geospatial, and external influences.

As shown Figure 7, first, the methods introduced in formulas (1) and (2) are used to convert the inflow and outflow of the entire city at each time interval into a 2-channel matrix. The spaced 2-channel stream matrix in each time segment is sent to the first two parts, respectively, and the same network structure of proposed Capsules TCN Network is used for modeling. This structure also captures the spatial dependence between nearby and distant areas. They are provided to the same neural network structure in the external factors. The output of the four parts is fused in the way of fully convolutional networks. Finally, the result is mapped to the range [-1, 1] by the Sigmoid function, which produces faster convergence than the standard logic function during the backpropagation learning process. The entire neural network structure consists of two important methods: Capsules Network [40] (CapsulesNet) and Temporal Convolutional Network [41] (TCN).

3.2.1. CapsulesNet in Capsules TCN Network. Sabour et al. published a paper on Google Brain entitled "Dynamic



(a) (b)

Figure 6: (a) Grid division of urban areas. (b) Inflow matrix in the ith time slice of the urban areas in the (a).

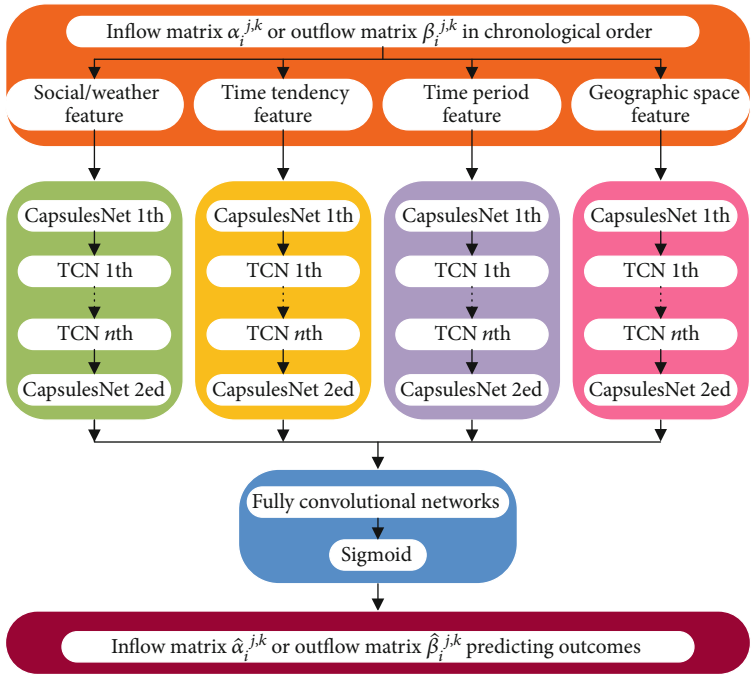


Figure 7: Capsules TCN architecture.

routing between capsules" [40]. We use the ideas from the reference when designing our Capsules TCN Network. Figure 8 shows the architecture of Capsules network (CapsulesNet). CapsulesNet, like ordinary neural networks, consists of many layers. The lowest capsule layer is called the primary capsule layer: each capsule unit in them receives a region of a matrix as input and detects the presence and posture of a specific object, and higher layers can detect larger and more complex objects.

Capsules are a group of neurons whose input and output vectors represent instantiation parameters of a specific entity type (that is, the probability of certain objects, conceptual entities, etc. appearing and certain attributes). The capsules at the same level use the transformation matrix to predict the instantiation parameters of higher-level capsules. When multiple predictions are consistent (this paper uses dynamic routing to make predictions consistent), higher-level capsules become active. The activation of the neurons in the capsule represents the various properties of the specific entities present in the matrix. These properties can include many different parameters, such as pose (position, size, and orientation), deformation, speed, reflectivity, color, texture, and more.

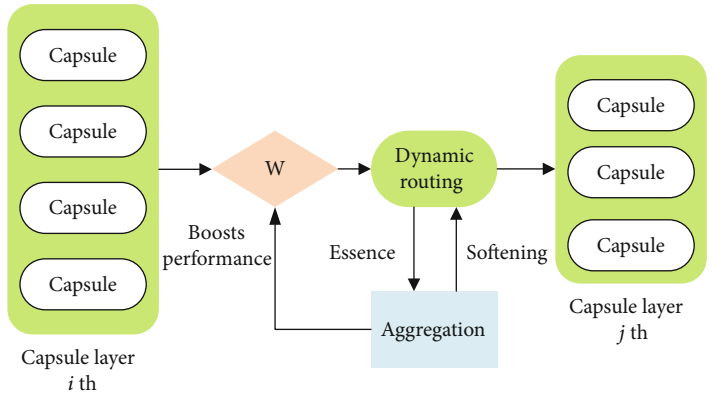


Figure 8: Capsule network architecture.

The length of the input-output vector represents the probability of an entity appearing, so its value must be between 0 and 1. To achieve this compression and complete capsule level activation, Sabour et al. used a nonlinear function called "squashing." This nonlinear function ensures that the length of the short vector can be shortened to almost zero, and the length of the long vector is compressed to close to but not more than 1 [40]. Here is the expression for this nonlinear function [40]:

$$V_j = \frac{\|S_j\|^2}{1+\|S_j\|^2}\frac{S_j}{\|S_j\|}, \tag{3}$$

where V_j is the output vector of capsule j, which S_j is the weighted sum of the vector output by all capsules in the previous layer to capsule j in the current layer, which S_j is simply the input vector of capsule j. The nonlinear function can be divided into two parts [40], namely,

$$\begin{matrix}\frac{\|S_j\|^2}{1+\|S_j\|^2},\\ \frac{S_j}{\|S_j\|},\end{matrix} \tag{4}$$

the first part is the scaling of the input vector S_j, and the second part is the unit vector of the input vector S_j. This nonlinear function not only retains the direction of the input vector but also compresses the length of the input vector to the interval [0,1). When S_j is zero, V_j can take 0, and when S_j is infinity, V_j approaches 1 infinitely. This nonlinear function can be seen as a kind of compression and reallocation of the vector length, so it can also be seen as a way to "activate" the output vector after the input vector.

Then, as mentioned above, the input vector of capsule is equivalent to the scalar input of a classic neural network, and the calculation of this vector is equivalent to the way of propagation and connection between two layers of capsules. The calculation of the input vector is divided into two phases, namely, linear combination and routing. This process can be expressed by the following formula [40]:

$$S_j = \sum_i c_{ij}\hat{u}_{j|i}, \hat{u}_{j|i} = W_{ij}u_i, \tag{5}$$

where $\hat{u}_{j|i}$ is a linear combination of u_i, which can be seen as a general neuron in the previous layer outputs with different strengths to a neuron in the next layer [40]. Just that capsule has a set of neurons (to generate a vector) at each node compared to a general neural network, which $\hat{u}_{j|i}$ means that the output vector of the ith capsule in the previous layer is multiplied by the corresponding weight vector (W_{ij} representing a vector). The resulting prediction vector $\hat{u}_{j|i}$ can also be understood as the strength of connecting to the jth capsule in the latter layer if the previous layer is the ith capsule.

After $\hat{u}_{j|i}$ decision is made, routing needs to be used for the second stage of allocation to calculate S_j in the output nodes. This process involves iterative updates c_{ij} using dynamic routing. We can get the S_j of the next layer of capsule through routing and then put S_j into the "squashing" nonlinear function to get the output of the next layer. The entire capsule layer and the process of propagation between them have been completed.

Coupling coefficient c_{ij} is updated and determined iteratively by a dynamic routing process. The sum of the coupling coefficients between capsule i and all capsules in the next level is 1. In addition, c_{ij} is determined by "routing softmax," and b_{ij} in the softmax function is initialized to 0. The softmax of c_{ij} is calculated as [40]

$$c_{ij} = \frac{\exp\left(b_{ij}\right)}{\sum_k \exp(b_{ik})}. \tag{6}$$

b_{ij} depends on the position and type of the two capsules but does not depend on the current input matrix. The consistency between the current output V_j of each capsule j in the subsequent hierarchy can be measured. The prediction vector of the previous capsule i iteratively update the

coupling coefficient with the consistency of the measurement. This paper simply measures this consistency by the inner product as

$$a_{ij} = V_j \cdot \widehat{u}_{j|i}. \tag{7}$$

This part also involves using routing to update the coupling coefficient [40]. The routing process is the update process. It calculates the product of V_j, and $\widehat{u}_{j|i}$ updates b_{ij} by adding it to the original b_{ij} and then uses softmax (b_{ij}, j) to update c_{ij}. When the output V_j is new, it can be updated c_{ij} iteratively, so that the parameters are updated directly by calculating the consistency of the input and output without back propagation.

For all capsule i and capsule j, initialize b_{ij} to equal to zero. The routing algorithm is very easy to converge; basically, it can have a good effect in 3 iterations. c_{ij} is updated through consistent routing. It does not need to be updated according to the loss function, but other convolution parameters and W_{ij} in the entire network need to be updated according to the loss function. In general, these parameters can be updated directly for the loss function using standard back propagation. The expression of this loss function is [40]

$$L_c = T_c \max(0, m^+ - \|v_c\|)^2 + \lambda(1 - T_c) \max(0, \|v_c\| - m^-)^2, \tag{8}$$

where c is the classification category, T_c is the indication function of classification (c exists as 1, and c does not exist as 0), m^+ is the upper boundary, and m^- is the lower boundary. In addition, v_c modulus is the L_2 distance of the vector.

3.2.2. TCN in Capsules TCN Network. TCN has better performance than a baseline recursive architecture in a wide range of sequence modeling tasks. Because these tasks include various benchmarks that are often used to evaluate recurrent network designs, it shows that the recent success of convolutional architectures in applications such as sequence processing is not limited to these areas [41].

TCN is based on two principles: the network produces an output of the same length as the input, and it cannot leak from the future to the past. To complete the first point, TCN uses a one-dimensional full convolutional network architecture, where each hidden layer is the same length as the input layer, and a zero-padding length (kernel size-1) is added to keep the subsequent layers from the previous layers. To achieve the second point, TCN uses causal convolution, and the output at time t is only transformed with elements from current time and earlier layers from the previous layer. It can be found by careful observation that TCN = 1D FCN causal convolution.

The major difference between TCN convolution and ordinary 1D convolution is the use of dilated convolutions. The higher the level, the larger the convolution window, and the more "holes" in the convolution window. More formally, for a 1D sequence input $X \in R^n$ and a filter $f : \{0, \cdots, k-1\} \to R$, the dilated convolution operation F on element s of the sequence is defined as

$$F(s) = (x * df)(s) \sum_{i=0}^{k-1} f(i) * x_s - d * i, \tag{9}$$

where d is the expansion factor, k is the size of the filter, and $x_s - d * i$ represents the past direction [28]. Therefore, expansion is equivalent to introducing a fixed step between every two adjacent filter faucets.

A primitive timing sequence convolution is just able to run back over at a point in time with size linear in the depth of timing sequence of the network. It makes a challenge to put in the mentioned causal convolution for time series, in which a longer history is critical. To acquire an exponentially large receptive field, a good part of the solution is dilated convolution. As illustrated in formula (9), d is the expansion factor. When $d = 1$, the expansion convolution is reduced to regular convolution. In order to figure a broad range of inputs, a larger dilation can be applied at the top level of the output. This ensures that there is a wider scale that expand the receptive field of a convolution within the effective history, meanwhile also extending for a long effective history using deep networks.

Every two such convolutional layers and identity mapping are encapsulated into a residual module (the residual module here is different from ResNet). The residual module contains RelU function, and a fully convolutional layer is used instead of a fully connected layer in the last few layers, as shown in Figure 9(a).

Generally, when using expanded convolution, we will increase d exponentially as the depth of the network increases. When the expansion factor is 1, as shown in Figure 9(b), the expansion convolution degenerates into causal convolution with a receptive field of 2. When the expansion factor is 2, the convolution kernel of the expanded convolution becomes 4. The final output contains all input information. By controlling the expansion factor, the size of the convolution kernel is increased to achieve the purpose of increasing the receptive field.

There are two disadvantages in large-scale neural networks: (1) it is too time-consuming; (2) it is easy to be overfitting. The dropout layer prevents overfitting of the network. Dropout is the process of training the network during deep learning. First, a part of the neural network units is temporarily dropped from the network with a certain probability, which is equivalent to finding a more streamlined network from the original network.

3.3. Superresolution Matrix of Inflow and Outflow Based on GAN. Unlike traditional time series prediction, the result of urban traffic flow prediction is a matrix rather than a simple value. When a high-resolution prediction result is needed, for example, the city is divided into 32×32. Through the two neural network models of the Capsules Network and Temporal Convolutional Network within a minute time level to obtain the final predicted 32×32 time results, it cannot be achieved by the hardware conditions at this stage. Therefore,

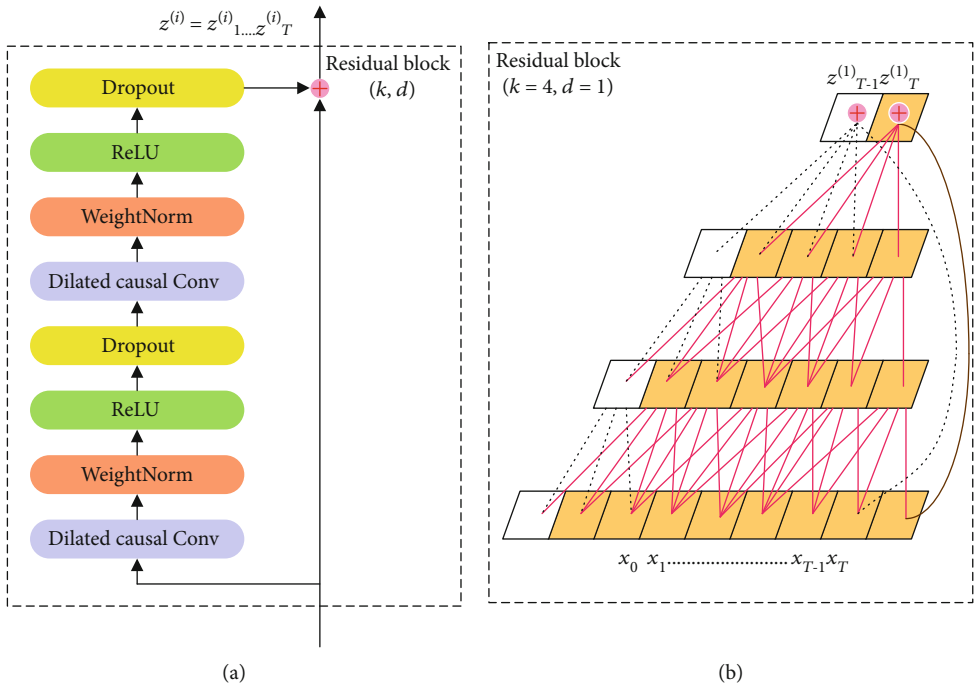


Figure 9: (a) TCN residual block in proposed architectural. (b) TCN residual connection in proposed architectural.

by reducing the resolution (matrix dimension) of the input data, it can achieve an exponential time-saving effect. Using a GAN-based superresolution reconstruction model is reasonable to reconstruct the high-resolution prediction results. Although the accuracy of the prediction result is sacrificed, it can obtain a minute-level high-resolution prediction result under the available hardware conditions. Under the urban traffic command and public safety guarantee scenarios, it is vital to obtain near-accurate results faster, which can provide better support for decision makers to make timely and effective judgments.

Because the amount of data is huge and the calculation is complicated, only the vehicle scene prediction of the experimental scene city in the 16×16 grid is calculated. However, in actual life applications, the experimental scene city is divided into 16×16 grids which is not inadequate. Dividing the city into a finer-grained grid is undoubtedly the solution to this problem. However, the traffic flow at the next interval cannot be predicted in time for more data to be computed.

The superresolution reconstruction reconstructs the city traffic flow a 16×16 experimental scene and obtains a 32×32 traffic flow prediction result. When we want to get better 32×32 fine results, the input data that needs to be processed increases by 4 times, and the overall calculation volume will also increase exponentially. We directly predict 32×32 results based on the predicted 16×16 results based on Generative Adversarial Network (GAN). The overall structure and workflow of traffic superresolution reconstruction of GAN are shown in Figure 10.

Figure 10 shows the structure of the superresolution reconstruction process based on GAN. The 16×16 experimental scene of urban vehicle traffic is used as a low-resolution matrix sequence after convolution layers to form a set of arranged matrixes. This set of matrixes output a 32×32 high-resolution matrix after passing through the GAN.

The process of inputting a convolution layer of a low-resolution matrix is based on the input of a low-resolution matrix of a frame and then convolving the matrix. The training process of the convolutional layer network is the optimization process of the parameters. The spatial transformation can be expressed as

$$I'_{t+k} = T_{\theta_i}(I_{t+k}). \tag{10}$$

The matrix I'_{t+k} represents the high-resolution matrix obtained by transforming $T_{\theta_i}(I_{t+k})$, and the transformation is $T(.)$ [42]. Regarding the loss function of the convolutional layer network, we utilize a regularization method to express it. The optimal parameter estimation process can be expressed as [42]

$$\theta_i^* = \arg_{\theta_i} \min \left\| I_t - I'_{t+k} \right\|_2^2 + \lambda \left\| Q I'_{t+k} \right\|_2^2, \tag{11}$$

among them, θ_i^* represents the parameters of the optimization estimation, λ is a regularization parameter, and Q is a Laplacian. Differentiate the parameters θ_i^* on the right side of formula (11), and make the differentiated result equal to 0. Use the fastest gradient descent method to iteratively solve the equation until the error is less than a preset threshold. The output parameter θ_i^* is the estimated optimal parameter.

The weight representation of the reconstruction network refers to defining a weight for each input low-resolution matrix, then performing weight representation on the input

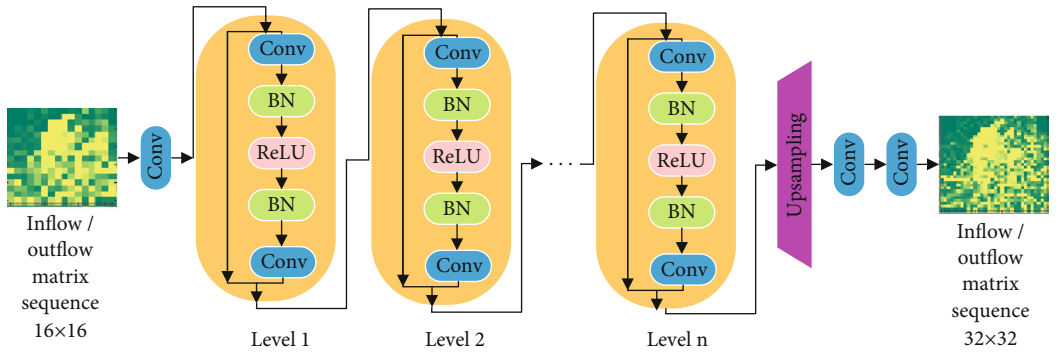


Figure 10: The architecture of superresolution matrix of inflow and outflow based on GAN.

low-resolution matrix to obtain a frame of high-frequency detail information. We add a convolution layer before the generative adversarial reconstruction network to complete the weight representation of the low-resolution matrix after the convolution layer. The mathematical expression of the weight representation can be expressed as [42]

$$X(m,n) = \sum_{k=0}^{K-1} \omega_k(m,n) I_{t+k}(m,n), \tag{12}$$

where $\omega_k(m,n)$ represents the weight value corresponding to the matrix block of the low-resolution matrix sequence. Generally, the same weight is defined for the matrix block. K represents the number of input low-resolution matrixes, (m,n) represents the serial number corresponding to the matrix block, and $(m \in 0, \rightarrow, M-1\,; n \in 0, \rightarrow, N-1)$.

Table 1: Table of dataset.

Dataset	Taxi GPS
Type of data	Taxi GPS track data
Time span	2018/6/10-2019/10/4
Time interval	30 minutes
Map grid size	(16, 16)
Track data	
Number of taxis	5000+
Available time interval	31,724
External factor data	
Holiday	26
The weather	12 types
Temperature	[-14.2, 38.6]
Wind speed	[0, 31.7]

4. Numerical Evaluation and Discussion

4.1. Experimental Data and Preprocessing. In the experimental verification part, the urban taxi dataset (taxi GPS) of the experimental scenario is used, and the data is shown in Table 1.

This article uses the reserve method: (1) the dataset is divided into two disjoint parts, one is the training set and the other is test set; (2) keep the data distribution roughly consistent, similar to stratified sampling; (3) in this paper, the amount of data for one year is used as the training set, and the amount of data for 4 months is used as the validation set. The amount of training set data should account for 75%.

We mainly use historical taxi traffic data prediction to refer to the forecast of rental vehicle traffic data at the future moment. The experiment selects the urban taxi GPS track data from the experimental scenes from June 10, 2018, to June 10, 2019, as the training set, and the remaining data as the test set. In order to facilitate the display and calculation of the results, we select the period from 8:00 to 10:00 AM for analysis.

The grid is divided into 16×16 grids, as shown in Figure 6(a). The GPS trajectory of the taxi is then mapped to the grid area, and a grid area map is developed, as shown in Figure 6(b). The grids represent regions, and the line segments connect the two regions (connected by taxi in this article). The area map actually combines data from the road network and taxi trajectory.

In Keras, learnable parameters are initialized with a uniform distribution with default parameters. The convolution of CapsulesNet *1*st and all TCNs uses 32 filters of size 3×3, and CapsulesNet *2*ed uses the convolution of 2 filters of size 3×3. Each Capsules TCN Network unit consists of 4 TCNs and 2 CapsulesNets. Table 2 for details, there are five additional hyperparameters in Capsules TCN Network.

In our superresolution reconstruction experiment, the 16×16 grid map is also an input as a low-frame image, and a 32×32 grid map is obtained through calculation by a GAN-based traffic prediction network. The magnification of the reconstruction experimental resolution is 2×2. The initial learning rate is set to 10^{-4}, and with each 10,000 iterations, the learning rate drops by 5%. In order to balance the convergence and training time of the network, the maximum number of iterations for superresolution reconstruction is set to 106.

4.2. Experimental Environment and Evaluation Criteria. The experimental verification in our research mainly runs on

Table 2: Table of detail in Capsules TCN architecture.

Network layer	Size	Social/weather feature	Time tendency feature	Time period feature	Geographic space feature
CapsulesNet *1*st	16×16	$3 \times 3, 32$	$3 \times 3, 32$	$3 \times 3, 32$	$3 \times 3, 32$
TCN *1*st	16×16	$\begin{bmatrix} 3 \times 3, 32 \\ 3 \times 3, 32 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 32 \\ 3 \times 3, 32 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 32 \\ 3 \times 3, 32 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 32 \\ 3 \times 3, 32 \end{bmatrix} \times 2$
TCN *2*ed	16×16	$\begin{bmatrix} 3 \times 3, 32 \\ 3 \times 3, 32 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 32 \\ 3 \times 3, 32 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 32 \\ 3 \times 3, 32 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 32 \\ 3 \times 3, 32 \end{bmatrix} \times 2$
TCN *3*rd	16×16	$\begin{bmatrix} 3 \times 3, 32 \\ 3 \times 3, 32 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 32 \\ 3 \times 3, 32 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 32 \\ 3 \times 3, 32 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 32 \\ 3 \times 3, 32 \end{bmatrix} \times 2$
TCN *4*th	16×16	$\begin{bmatrix} 3 \times 3, 32 \\ 3 \times 3, 32 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 32 \\ 3 \times 3, 32 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 32 \\ 3 \times 3, 32 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 32 \\ 3 \times 3, 32 \end{bmatrix} \times 2$
CapsulesNet *2*ed	16×16	$3 \times 3, 32$	$3 \times 3, 32$	$3 \times 3, 32$	$3 \times 3, 32$

the GPU server, and its detailed information is shown in Table 3.

We use Root Mean Square Error (RMSE) to evaluate the model [43].

$$\mathrm{RMSE} = \sqrt{\frac{1}{Z}\sum_{i}(x_i - x\wedge_i)^2}, \tag{13}$$

where x is the real value and $\hat{x}$ is the corresponding predicted value; Z is the number of all available true values. The RMSE is used to measure the deviation between the observed value and the true value, which is more suitable in this experiment.

Furthermore, in order to measure the quality of the superresolution reconstruction algorithm, evaluation indicators need to be used. The requirements for reconstruction results are different in different application scenarios, so the evaluation standards used are also different. Evaluation methods are generally divided into two categories, one is subjective evaluation and the other is objective evaluation. In objective evaluation, the two most commonly used evaluation indicators are Peak Signal-to-Noise Ratio (PSNR) [44] and Structural Similarity (SSIM) [45].

The specific calculation formula of PSNR [44] is described as follows:

$$\begin{aligned} \mathrm{PSNR} &= 10\log_{10}\frac{\left(2^l - 1\right)^2}{\mathrm{MSE}}, \\ \mathrm{MSE} &= \frac{1}{mn}\sum_{x=1}^{m}\sum_{y=1}^{n}[f\wedge(x,y) - f(x,y)]^2. \end{aligned} \tag{14}$$

MSE is the mean square error, $f(x, y)$ represents the reference matrix. In the experiment, 32×32 grids represent the matrix. It can be known from the formula that when the PSNR of the matrix to be evaluated is larger, the reconstruction result is better.

Table 3: Table of detail in experimental environment.

OS	Windows7
Memory	32 GB
CPU	Intel Core i5-6500 3.20 GHz4 cores
GPU	Nvdia 1080Ti∗2
Software	
CUDA ver.	8.0
CUDNN ver.	8.0
Keras ver.	1.1.1
TensorFlow ver.	1.1.0

The specific calculation formula of SSIM [45] is given as follows:

$$\mathrm{SSIM} = \frac{\left(2\mu_f\mu_{\hat{f}} + C_1\right)\left(2\sigma_{f,\hat{f}} + C_2\right)}{\left(\mu_f^2 + \mu_{\hat{f}}^2 + C_1\right)\left(\sigma_f^2 + \sigma_{\hat{f}}^2 + C_2\right)}, \tag{15}$$

where μ_f is the average value of the reference matrix, $\mu_{\hat{f}}$ represents the average value of the matrix to be evaluated, σ_f is the variance of the reference matrix, and $\sigma_{\hat{f}}$ is the variance of the matrix to be evaluated.

4.3. Effect of Hyperparameters on Experimental Results. The number of CapsulesNet has an effect on the taxi GPS dataset experiments, as shown in Figure 11(a). The network depth also greatly affects the experimental results. As shown in Figure 11(b), the number of TCN increases; the RMSE of the model fluctuates. It indicates that the network is not the deeper the better, because it captures not only close-space dependencies but also far-space dependencies. When the network is very deep (such as when the number is 15), training becomes very difficult. Based on the above comparison, the number of CapsulesNet is two, and the number of TCNs is set to four.

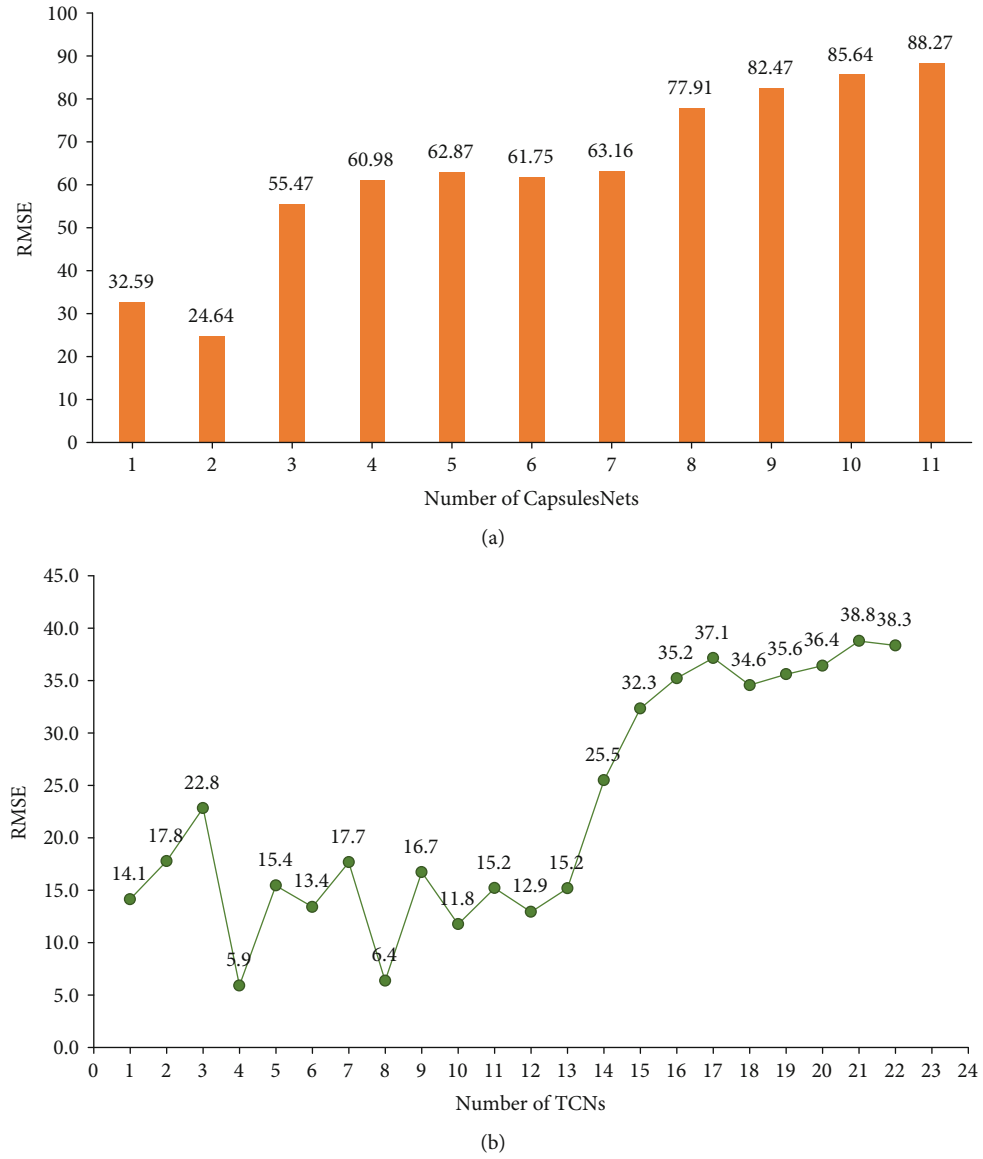


Figure 11: (a) Effect of the number of CapsulesNets on experimental results. (b) Effect of the number of TCNs on experimental results.

4.4. Experimental Results and Analysis. By comparing with historical average (HA), autoregressive integrated moving average (ARIMA), recurrent neural network (RNN), and long-term short-term memory (LSTM) network, the validity of the Capsules TCN Network model for urban area traffic flow prediction is verified.

HA predict the inflow and outflow of people based on the historical average of inflow and outflow at the same time and area in the past. For example, to predict the inflow of a region from 10:00 to 10:30 AM this Thursday morning, calculate the average of the inflow from 10:00 to 10:30 AM every Thursday morning in this region.

ARIMA is a well-known model for understanding and predicting future values in a time series. In the traditional linear model, the autoregressive integrated moving average model has been widely used in passenger flow prediction. It is a general formula for autoregressive (AR) models, integral (I) models, or moving average (MA) models.

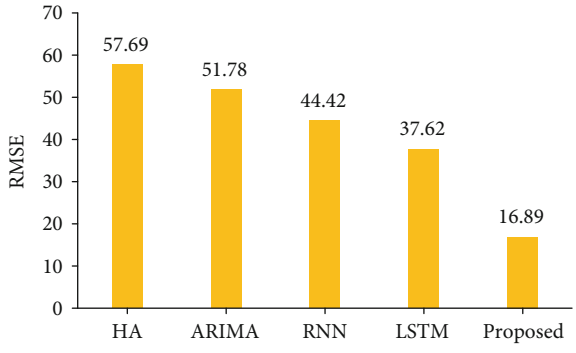


Figure 12: Comparison of the proposed algorithm with traditional existing algorithms.

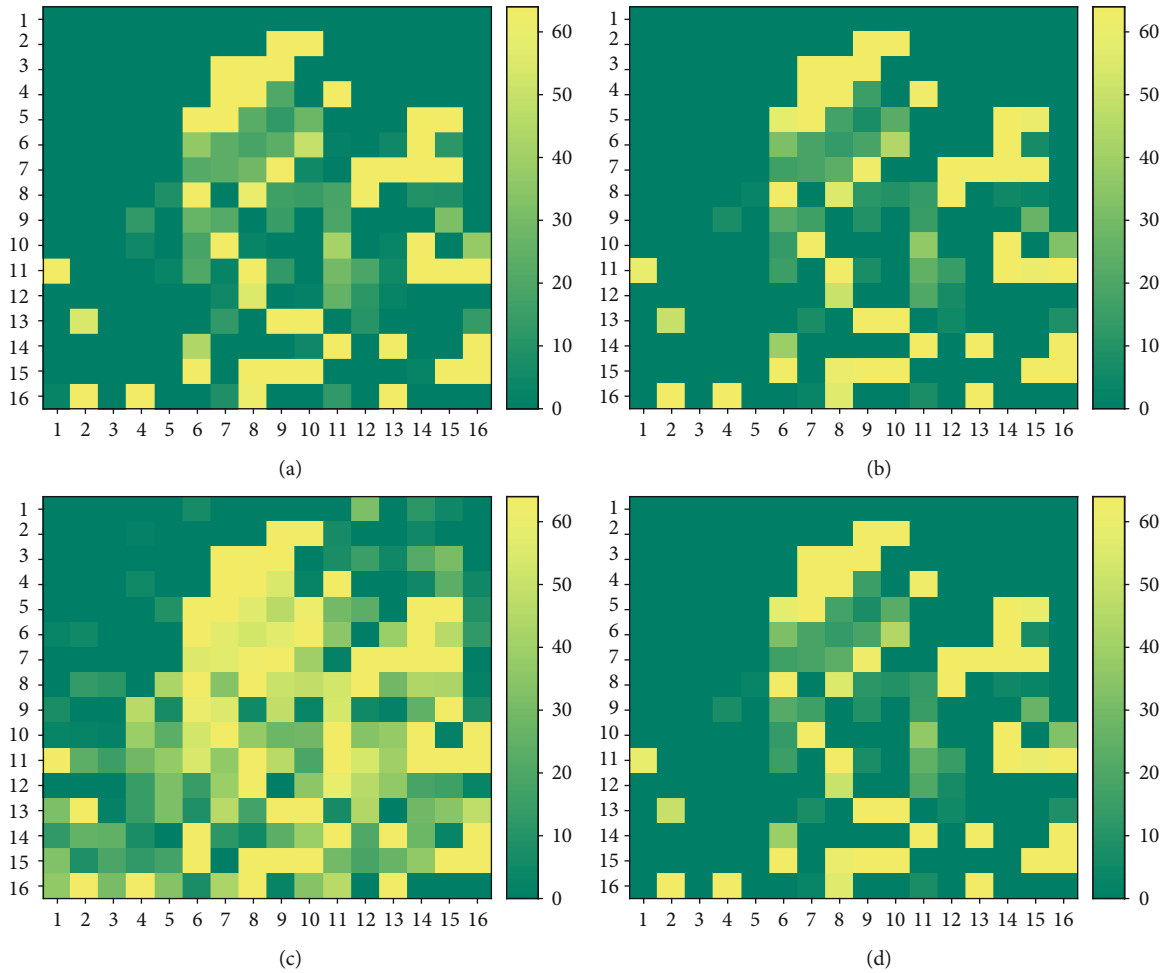


Figure 13: The prediction results of the algorithm in 4 different times (a) PM 8:00. (b) PM 8:30. (c) PM 9:00. (d) PM 10:00.

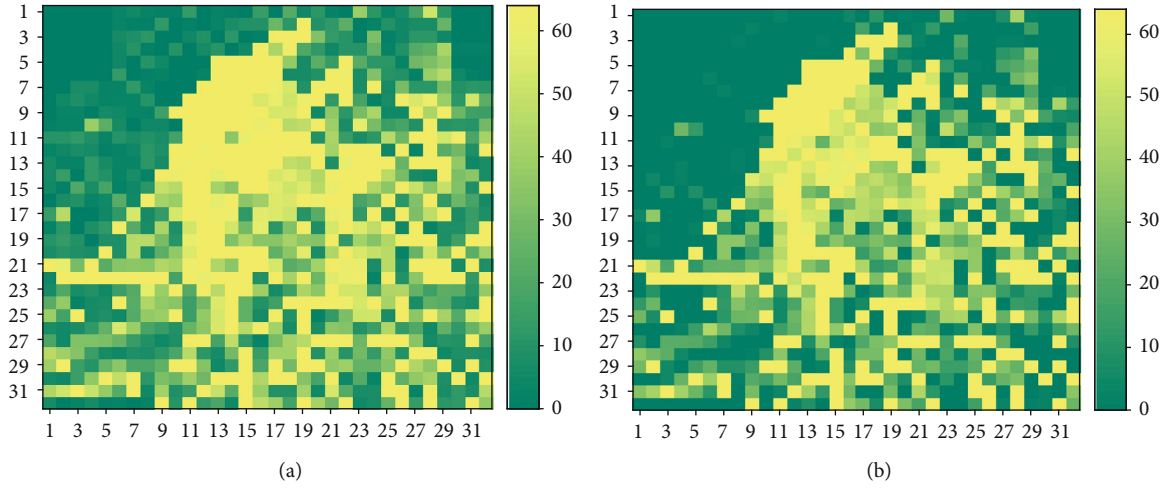


Figure 14: (a) Results of 32×32 matrix based on superresolution GAN reconstruction according to 16×16 matrix of Figure 6(b). (b) Real 32×32 matrix based on input data according to 16×16 matrix of Figure 6(b).

RNN is a deep-learning model that captures time sequence dependencies. Formally, RNNs can train sequences of any length [46].

LSTM is a special RNN that can learn long-term time dependencies [47].

We compare the RMSE between the Capsules TCN Network and the true value and then compare it with other prediction models to verify the validity of Capsules TCN Network. The results are shown in Figure 12. According to the comparison results, it can be seen that the proposed Capsules TCN Network has smaller RMSE. It has higher prediction accuracy than other models, indicating the effectiveness of the proposed Capsules TCN Network for traffic prediction tasks.

Figure 13 shows the spatial-temporal distribution of taxi traffic during the morning rush hour at 8:30-10:00 AM on September 30, 2018. From the results in Figure 13, it can be seen that the proposed Capsules TCN Network better grasps the spatiotemporal characteristics of the changes in taxi traffic and makes predictions with sufficient accuracy.

The experimental results of superresolution matrix of inflow and outflow based on GAN are also demonstrated. Figure 14(a) is the result of superresolution reconstruction based on GAN, and Figure 14(b) is the real value, when the input low-resolution matrix is used from Figure 6(b). According to experiments, it can be seen subjectively that GAN-based superresolution reconstruction has achieved good reconstruction results and is close to the real value visually. The objective assessment is as follows: PSNR is 33.844 and SSIM is 0.93. We also obtain the PSNR and SSIM of superresolution matrix of inflow and outflow based on GAN in 64×64 and 128×128, respectively. In 64×64 scene, PSNR is 28.94 and SSIM is 0.88. In 128×128 scene, PSNR is 22.75 and SSIM is 0.79.

5. Conclusions

Traffic forecasting has been a core issue in transportation planning and management, and it has also been a major issue in urban computing. The prediction of traffic volume can help the development of urban traffic safety, and traffic flow will be more order. We propose a method based on the Capsules Network and Temporal Convolutional Network to predict traffic flow in local areas of the city. This method is called Capsules TCN Network. The Capsules TCN Network model can learn the spatial dependence, time dependence, and external factors of traffic flow prediction. We evaluated the GPS track data of urban taxis in the experimental scenarios and verified that the model has a good applicability in vehicle traffic prediction. Because the accuracy of regional traffic flow is different in different scenarios, we propose a GAN-based superresolution reconstruction model of traffic flow to improve the accuracy of Capsules TCN Network model results. The experimental results show that the GAN-based traffic superresolution reconstruction model not only has a better subjective visual effect but also has more prominent objective evaluation indicators.

Data Availability

The dataset used in this article is from a commercial company. If you need the dataset used in this study, you can send a usage request to centaureacyanus@foxmail.com. After being authorized by the company, the dataset will be transmitted to the applicant in the form of an email attachment.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This project is supported by The Science and Technology Funds from Liaoning Education Department (No. LQ2017 008), the Doctoral Research Startup Fund Project of Liaoning Province (No. 2016011968), and the China Postdoctoral Science Foundation (No. 2019M661096).

References

[1] W. Wang, J. Chen, J. Wang, J. Chen, and Z. Gong, "Geography-aware inductive matrix completion for personalized Point-of-Interest recommendation in smart cities," *IEEE Internet of Things Journal*, vol. 7, no. 5, pp. 4361–4370, 2020.

[2] W. Wang, J. Chen, J. Wang, J. Chen, J. Liu, and Z. Gong, "Trust-enhanced collaborative filtering for personalized point of interests recommendation," *IEEE Transactions on Industrial Informatics*, p. 1, 2019.

[3] J. Liu, T. Li, P. Xie, S. Du, F. Teng, and X. Yang, "Urban big data fusion based on deep learning: an overview," *Information Fusion*, vol. 53, pp. 123–133, 2020.

[4] Z. Pan, Y. Liang, W. Wang, Y. Yu, Y. Zheng, and J. Zhang, "Urban traffic prediction from spatio-temporal data using deep meta learning," in *KDD '19: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 1720–1730, Anchorage, AK, USA, July 2019.

[5] R. Jiang, X. Song, D. Huang et al., "DeepUrbanEvent: a system for predicting citywide crowd dynamics at big events," in *KDD '19: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 2114–2122, Anchorage, AK, USA, July 2019.

[6] P. Xie, T. Li, J. Liu, S. Du, X. Yang, and J. Zhang, "Urban flow prediction from spatiotemporal data using machine learning: a survey," *Information Fusion*, vol. 59, pp. 1–12, 2020.

[7] R. Li, Z. Zhao, X. Zhou et al., "Intelligent 5G: when cellular networks meet artificial intelligence," *IEEE Wireless Communications*, vol. 24, no. 5, pp. 175–183, 2017.

[8] J. Pérez-Romero, O. Sallent, R. Ferrús, and R. Agustí, "Artificial intelligence-based 5G network capacity planning and operation," in *2015 International Symposium on Wireless Communication Systems (ISWCS)*, pp. 246–250, Brussels, Belgium, August 2015.

[9] Y. Bi, C. Lin, H. Zhou, P. Yang, X. Shen, and H. Zhao, "Time-constrained big data transfer for SDN-enabled smart city," *IEEE Communications Magazine*, vol. 55, no. 12, pp. 44–50, 2017.

[10] D. Zou, S. Li, X. Kong, H. Ouyang, and Z. Li, "Solving the combined heat and power economic dispatch problems by an improved genetic algorithm and a new constraint handling strategy," *Applied Energy*, vol. 237, pp. 646–670, 2019.

[11] X. Shen, D. Zou, N. Duan, and Q. Zhang, "An efficient fitness-based differential evolution algorithm and a constraint handling technique for dynamic economic emission dispatch," *Energy*, vol. 186, p. 115801, 2019.

[12] M. Yao, M. Sohul, V. Marojevic, and J. H. Reed, "Artificial intelligence defined 5G radio access networks," *IEEE Communications Magazine*, vol. 57, no. 3, pp. 14–20, 2019.

[13] C. Lin, Y. Bi, H. Zhao, Z. Liu, S. Jia, and J. Zhu, "DTE-SDN: a dynamic traffic engineering engine for delay-sensitive transfer," *IEEE Internet of Things Journal*, vol. 5, no. 6, pp. 5240–5253, 2018.

[14] Y. Fu, S. Wang, C. X. Wang, X. Hong, and S. McLaughlin, "Artificial intelligence to manage network traffic of 5G wireless networks," *IEEE Network*, vol. 32, no. 6, pp. 58–64, 2018.

[15] D. Zou, S. Li, X. Kong, H. Ouyang, and Z. Li, "Solving the dynamic economic dispatch by a memory-based global differential evolution and a repair technique of constraint handling," *Energy*, vol. 147, pp. 59–80, 2018.

[16] M. S. Ahmed and A. R. Cook, *Analysis of freeway traffic time-series data by using Box-Jenkins techniques (No. 722)*, Transportation Research Board, 1979.

[17] M. Levin and Y. D. Tsao, "On forecasting freeway occupancies and volumes (abridgment)," *Transportation Research Record*, no. 773, pp. 47–49, 1980.

[18] M. M. Hamed, H. R. Al-Masaeid, and Z. M. B. Said, "Short-term prediction of traffic volume in urban arterials," *Transportation Engineering*, vol. 121, no. 3, pp. 249–254, 1995.

[19] M. Van Der Voort, M. Dougherty, and S. Watson, "Combining Kohonen maps with ARIMA time series models to forecast traffic flow," *Transportation Research Part C: Emerging Technologies*, vol. 4, no. 5, pp. 307–318, 1996.

[20] S. Lee and D. B. Fambro, "Application of subset autoregressive integrated moving average model for short-term freeway traffic volume forecasting," *Transportation Research Record*, vol. 1678, no. 1, pp. 179–188, 1999.

[21] B. M. Williams, "Multivariate vehicular traffic flow prediction: evaluation of ARIMAX modeling," *Transportation Research Record*, vol. 1776, no. 1, pp. 194–200, 2001.

[22] Y. Kamarianakis and P. Prastacos, "Forecasting traffic flow conditions in an urban network: comparison of multivariate and univariate approaches," *Transportation Research Record*, vol. 1857, no. 1, pp. 74–84, 2003.

[23] B. M. Williams and L. A. Hoel, "Modeling and forecasting vehicular traffic flow as a seasonal ARIMA process: theoretical basis and empirical results," *Journal of Transportation Engineering*, vol. 129, no. 6, pp. 664–672, 2003.

[24] B. Ghosh, B. Basu, and M. O'Mahony, "Multivariate short-term traffic flow forecasting using time-series analysis," *IEEE Transactions on Intelligent Transportation Systems*, vol. 10, no. 2, pp. 246–254, 2009.

[25] G. A. Davis and N. L. Nihan, "Nonparametric regression and short-term freeway traffic forecasting," *Journal of Transportation Engineering*, vol. 117, no. 2, pp. 178–188, 1991.

[26] H. Chang, Y. Lee, B. Yoon, and S. Baek, "Dynamic near-term traffic flow prediction: system-oriented approach based on past experiences," *IET Intelligent Transport Systems*, vol. 6, no. 3, pp. 292–305, 2012.

[27] N. E. El Faouzi, "Nonparametric traffic flow prediction using kernel estimator," in *Transportation and traffic theory: proceedings of the 13th International Symposium on Transportation and Traffic Theory*, Lyon, France, July 1996.

[28] H. Sun, H. X. Liu, H. Xiao, R. R. He, and B. Ran, "Use of local linear regression model for short-term traffic forecasting," *Transportation Research Record*, vol. 1836, no. 1, pp. 143–150, 2003.

[29] Y. S. Jeong, Y. J. Byon, M. M. Castro-Neto, and S. M. Easa, "Supervised weighting-online learning algorithm for short-term traffic flow prediction," *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, no. 4, pp. 1700–1707, 2013.

[30] K. Y. Chan, T. S. Dillon, J. Singh, and E. Chang, "Neural-network-based models for short-term traffic flow forecasting using a hybrid exponential smoothing and Levenberg–Marquardt algorithm," *IEEE Transactions on Intelligent Transportation Systems*, vol. 13, no. 2, pp. 644–654, 2012.

[31] K. Kumar, M. Parida, and V. K. Katiyar, "Short term traffic flow prediction for a non urban highway using artificial neural network," *Procedia-Social and Behavioral Sciences*, vol. 104, pp. 755–764, 2013.

[32] W. Zheng, D. H. Lee, and Q. Shi, "Short-term freeway traffic flow prediction: Bayesian combined neural network approach," *Journal of Transportation Engineering*, vol. 132, no. 2, pp. 114–121, 2006.

[33] M.-C. Tan, S. C. Wong, J.-M. Xu, Z.-R. Guan, and P. Zhang, "An aggregation approach to short-term traffic flow prediction," *IEEE Transactions on Intelligent Transportation Systems*, vol. 10, no. 1, pp. 60–69, 2009.

[34] S. A. Zargari, S. Z. Siabil, A. H. Alavi, and A. H. Gandomi, "A computational intelligence-based approach for short-term traffic flow prediction," *Expert Systems*, vol. 29, no. 2, pp. 124–142, 2012.

[35] M. Cetin and G. Comert, "Short-term traffic flow prediction with regime switching models," *Transportation Research Record*, vol. 1965, no. 1, pp. 23–31, 2006.

[36] Y. Zhisheng, S. Chunfu, X. Zhihua, and Y. Hao, "Short-term traffic flow prediction of road networks based on principal component analysis and support vector machines," *Journal of Jilin University*, vol. 38, no. 1, pp. 48–52, 2008.

[37] L. Cunjun, R. Yang, and Z. Jiashu, "Traffic flow forecasting method based on wavelet analysis," *Computer Application*, vol. 23, no. 12, pp. 7-8, 2003.

[38] S. Yan, C. Senfa, and Z. Zhenguo, "Application of grey system theory to traffic flow prediction at detector-free intersections," *Journal of Southeast University: Natural Science Edition*, vol. 32, no. 2, pp. 256–258, 2002.

[39] X. Weiqing, Y. Xiaobo, J. Shouxu, and L. Zhijun, "Short-term traffic flow prediction based on BP neural network and fuzzy inference system," *Intelligent Computers and Applications*, vol. 5, no. 2, pp. 43–46, 2015.

[40] S. Sabour, N. Frosst, and G. E. Hinton, "Dynamic routing between capsules," in *Advances in neural information processing systems*, pp. 3856–3866, Curran Associates, Inc, 2017.

[41] S. Bai, J. Z. Kolter, and V. Koltun, "An empirical evaluation of generic convolutional and recurrent networks for sequence modeling," 2018, http://arxiv.org/abs/1803.01271.

[42] S. Lian, H. Zhou, and Y. Sun, "FG-SRGAN: a feature-guided super-resolution generative adversarial network for unpaired image super-resolution," in *Advances in Neural Networks* –

ISNN 2019. ISNN 2019. Lecture Notes in Computer Science, vol 11554, H. Lu, H. Tang, and Z. Wang, Eds., Springer, Cham, 2019.

[43] T. Chai and R. R. Draxler, "Root mean square error (RMSE) or mean absolute error (MAE)?–arguments against avoiding RMSE in the literature," *Geoscientific Model Development*, vol. 7, no. 3, pp. 1247–1250, 2014.

[44] A. Hore and D. Ziou, "Image quality metrics: PSNR vs. SSIM," in *2010 20th International Conference on Pattern Recognition*, pp. 2366–2369, Istanbul, Turkey, August 2010.

[45] S. S. Channappayya, A. C. Bovik, and R. W. Heath Jr., "Rate bounds on SSIM index of quantized images," *IEEE Transactions on Image Processing*, vol. 17, no. 9, pp. 1624–1639, 2008.

[46] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning. Book in preparation for MIT Press*, 2016, http://www.deeplearningbook.org.

[47] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

Hindawi
Wireless Communications and Mobile Computing
Volume 2020, Article ID 3240675, 11 pages
https://doi.org/10.1155/2020/3240675

Research Article

Research on Phase Combination and Signal Timing Based on Improved K-Medoids Algorithm for Intersection Signal Control

Guojiang Shen, Xiangyu Zhu, Wei Xu, Longfeng Tang, and Xiangjie Kong

College of Computer Science and Technology, Zhejiang University of Technology, Hangzhou, China

Correspondence should be addressed to Xiangjie Kong; xjkong@ieee.org

Received 11 February 2020; Accepted 17 April 2020; Published 9 May 2020

Academic Editor: Bingxian Lu

Aiming at the problem of intersection signal control, a method of traffic phase combination and signal timing optimization based on the improved K-medoids algorithm is proposed. Firstly, the improvement of the traditional K-medoids algorithm embodies in two aspects, namely, the selection of the initial medoids and the parameter k, which will be applied to the cluster analysis of historical saturation data. The algorithm determines the initial medoids based on a set of probabilities calculated from the distance and determines the number of clusters k based on an exponential function, weight adjustment, and elbow ideas. Secondly, a phase combination model is established based on the saturation and green split data, and the signal timing is optimized through a bilevel programming model. Finally, the algorithm is evaluated over a certain intersection in Hangzhou, and results show that this algorithm can reduce the average vehicle delay and queue length and improve the traffic capacity of the intersection in the peak hour.

1. Introduction

With the rapid development of urban construction and socioeconomy, traffic congestion, one of China's urban diseases, not only brings tremendous pressure to urban traffic management but also seriously affects the harmonious development of cities. Many modern transportation facilities and applications can benefit from better performance of signal timing schemes [1–4]. For example, space-time road resources can be allocated more reasonably, the accuracy of traffic speed prediction can be improved [2], and the optimized signal cycle time and green split scheme can help make better-coordinated control [4]. In [5, 6], the authors studied the application of mobile crowdsourcing (MCS) in smart cities. In [7, 8], the authors integrate geographic and temporal influences into points of interest (POI) recommendations to help people find points of interest.

In recent years, several algorithms have been presented in the literature for traffic signal phase combination and timing optimization. In [9], the authors studied the dynamic prediction traffic signal control framework for a single intersection and optimized the signal timing according to the predicted arrival flow. In [10], a queuing and dissipation model of the intersection traffic flow was presented, which provided a theoretical basis for optimizing the intersection phase and timing. In [11], the authors considered an adaptive traffic signal control method based on fuzzy logic. This method optimized the phase duration and phase sequence. The results showed that the average queue length, the maximum queue length, and the parking rate were significantly shortened, but only lower queue lengths were considered. In [12], fog calculation was used to process traffic data, and a phase combination method based on a genetic algorithm was presented. The authors in [13] studied dynamic programming algorithms to optimize signal timing and phase, thereby, reducing average vehicle latency. In [14], the Artificial Bee Colony algorithm was adopted to optimize the signal cycle time and the green split, reducing the average vehicle delay and the average queue length, but the algorithm needed to obtain the vehicle speed online and calculate it. In [15], the authors considered a dynamic phase control method based on traffic flow, but it needed real-time detection and calculation of road conditions, resulting in poor practical application effect. In [16], the clustering algorithm was applied to

process vehicle motion information, which was the basis for subsequent optimization, but only optimized the signal timing, excluding phase combination. In [17], a traffic signal segmentation algorithm based on the two-dimensional clustering was presented. It matched the best timing scheme for the current traffic conditions through the clustering analysis. However, the intersection traffic flow model cannot distinguish between a left turn and straight vehicles.

In [18], the authors studied the interval data-based K-means clustering method, and the clustering results can accurately describe the trend of traffic state evolution at an urban intersection. In [19], the K-means clustering algorithm was used to group traffic flows and divide the traffic condition level and provides a theoretical basis for matching the most suitable traffic signal control scheme in different situations. In [20], the author studies a dynamic traffic control method that predicts congestion by the clustering thought. In [21], a traffic signal control method based on the K-means clustering algorithm was presented, and the number of clusters was defined as two. The authors in [22] studied the improved affinity propagation (AP) clustering algorithm, which provided efficient and accurate traffic state information for traffic signal control. The average waiting time was effectively reduced. In [23], the authors studied the K-means clustering method to optimize the best switching time of time-of-day (TOD) control scheme, but the number of clusters needed to be specified in advance, which largely affected the effectiveness of the method. Similarly, the authors in [24] used the Kohonen cluster and K-means cluster to optimize TOD breakpoints and proved that K-means had a better performance. However, it was still necessary to specify the number of clusters and the initial cluster centers in advance, which was easy to fall into local optimum.

The existing researches mainly have the following shortcomings:

(1) the intersection traffic flow model is established without considering all of the flow directions

(2) the practical value of online data acquisition and frequent signal switching solutions is not high

(3) the number of clusters depends heavily on prior or empirical knowledge

To solve the problems above, this paper proposes a traffic phase combination and signal timing optimization method based on the improved K-medoids algorithm. Firstly, the improved K-medoids algorithm is used to cluster the historical saturation data, which can select the number of schemes k more quickly and accurately. Then, the phase combination model is established since K-medoids correspond to k pairs of saturation and green split data, which can combine the flow direction with similar traffic demand to improve the utilization of green time. Finally, the bilevel programming model is used to optimize the signal cycle time and green split of each phase, so that the timing scheme can be further optimized based on the phase combination. After clustering, each medoid composing a scheme library corresponds to a traffic scheme. In experiments, we choose an appropriate traffic scheme according to the Euclidean distance between the actual traffic saturation and medoids.

The paper is organized as follows: Section 2 introduces the traditional K-medoids clustering algorithm and its improvement. Section 3 designs the phase combination and signal timing optimization algorithm. Section 4 provides experimental results and comparisons with the traditional K-medoids algorithm. Section 5 provides conclusions and describes directions for future research.

2. Improved K-Medoids Algorithm

In this section, we first introduce the traditional k-medoids algorithm, then, to find better initial medoids and the appropriate parameter k, an improvement is introduced. Finally, we apply the improved k-medoids algorithm to the traffic saturation dataset into k clusters, and each cluster corresponds to one set of traffic scheme.

2.1. Traditional K-Medoids Algorithm. Clustering is an unsupervised learning algorithm that partitions the origin data into several clusters, where the data in the same cluster are similar to each other but different from the data in other clusters. K-medoids algorithm is a partition-based clustering algorithm. Compared with K-means clustering, it is less sensitive to outliers. Among many k-medoids algorithms, partitioning around medoids (PAM) is one of the most classical and powerful [25].

K-medoids algorithm first randomly selects k representative data points as the initial medoids, each medoid corresponds to one cluster. Secondly, Euclidean distance is applied to calculate the distance between all data and the chosen medoid, each data point will be assigned to the most similar medoid. Thirdly, such a new medoid in each cluster is found to minimize the criterion function within the cluster. The algorithm will stop until all of the medoids are equal to the previous ones, otherwise, assign each data to the nearest medoid and generate k new clusters. The Euclidean distance $d_{(x_i,y_i)}$ is used to measure the similarity between all of the data points and the medoids, which can be calculated as follows:

$$d_{(x_i,y_j)} = \sqrt{\left(x_{i1} - y_{j1}\right)^2 + \left(x_{i2} - y_{j2}\right)^2 + \cdots + \left(x_{\text{in}} - y_{jn}\right)^2}, \tag{1}$$

where x_i and y_i are both n-dimensional data objects.

The criterion function in within-cluster can be calculated as:

$$E_i = \sum_{b_j \in B_i} d_{\left(b_j, c_i\right)}^2, \tag{2}$$

where B_i is the cluster after clustering, b_j is the data point in the cluster B_i, and c_i is the medoid of the cluster B_i.

The criterion function is described as follows:

$$E = \sum_{i=1}^{k} \sum_{b_j \in B_i} d_{(b_j, c_i)}{}^2. \tag{3}$$

where k is the number of clusters.

2.2. The Improvement of K-Medoids Algorithm. For the K-medoids clustering algorithm, the number of clusters and the initialization have a great influence on the clustering process and results. In [26], a density peak clustering algorithm is proposed. This algorithm can select medoids and confirm the correct number of clusters. In [27], the author studied the K-medoids clustering algorithm based on a subset of candidate medoids and gradually increasing the number of clusters, thereby, improving the clustering performance of the algorithm. In order to reduce the negative impact when the initial medoids have a low dispersion degree, this paper proposes an initial point probability selection method based on the Euclidean distance. In addition, in order to reduce the artificial dependence for selecting initial medoids and avoid the excessive gap between each cluster, this paper proposes an optimization for selecting an optimal parameter k based on exponential function, weight adjustment, and elbow idea.

2.2.1. Improved Method for Selecting Initial Medoids. After selecting a point in sample data as the first medoid c_1 randomly, the Euclidean distance $d_{(b_h, c_i)}$ is applied to calculate the distance between each point b_h and the nearest medoid c_i, and the probability p_h that point b_h will be selected as the next cluster medoid can be calculated as:

$$p_h = \frac{d_{(b_h, c_i)}{}^2}{\sum_{b_j \in B} d_{(b_j, c_i)}{}^2}, \tag{4}$$

where B is the dataset, and the probability set P can be obtained as follows:

$$\mathrm{P} = [0, p_1, p_1 + p_2, \cdots, p_1 + p_2 + \cdots + p_{n-1}, 1], \tag{5}$$

where n is the number of samples in the dataset.

The roulette wheel method is used to select the cluster medoid $c_i (i \geq 2)$ (see Figure 1):

Step 1. We generate a random number r between $[0, 1)$, if r belongs to the interval $[p_1 + p_2 + \cdots + p_{i-1}, p_1 + p_2 + \cdots + p_{i-1} + p_i)$ in P, point b_i will be the second cluster medoid c_2.

Step 2. We recalculate the probabilities that each point in the dataset will be selected as the next medoid.

Step 3. We select the next medoid according to the probability set P and the roulette wheel method.

The steps mentioned above will be repeated until k centers are selected. The purpose is to make the initial medoids more discrete, which are closer to the real cluster centers.

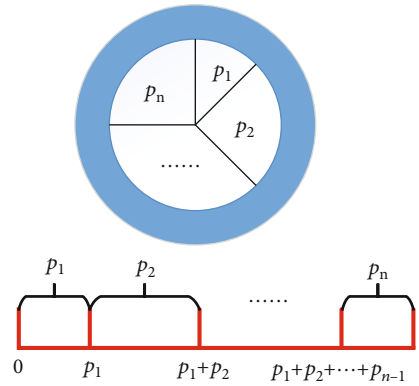


Figure 1: Roulette wheel method.

The number of iterations can be reduced, but settle the problem of trapping in a local optimum.

2.2.2. Improved Method for Selecting the Number of Clusters. The traditional criterion function in each cluster is the sum of all data within the cluster, which will make a big difference among clusters, and the classification will also be uneven. To settle the problem, this paper uses the exponential function e^x to optimize the criterion function calculation method. The criterion function in within-cluster can be calculated as:

$$S_i = e^{\sum_{b_j \in B_i} d_{(b_j, c_i)}{}^2} \tag{6}$$

In order to avoid exponential explosion, the weight coefficient t is employed, and the criterion function S can be calculated as follows:

$$S = \sum_{i=1}^{k} e^{\frac{\sum_{b_j \in B_i} d_{(b_j, c_i)}{}^2}{e}}. \tag{7}$$

With the optimization, the criterion function S can be calculated for different k. Following the increasement of parameter k, S will decrease. According to the elbow idea, S drops dramatically at the beginning, then, S reaches an elbow, finally, the curve of S turns to a plateau. The value k corresponding to the elbow is regarded as the optimal number of clusters.

2.3. Clustering with Saturation Data. Traffic saturation data is a collection of saturation at intersections, a single piece of data can be described as:

$$\mathrm{X}_i = [x_{i1}, x_{i2}, \cdots, x_{\mathrm{in}}], \tag{8}$$

where n is the number of intersections.

The improved K-medoids algorithm described in Section 2.2 is then applied to the traffic saturation data, which divides the data into k clusters, and the initial cluster medoids are selected according to the distance probability p_h. The phase and timing optimization can be performed according to the cluster medoids, and each cluster corresponds to one set of

traffic scheme, which means there will be k sets of initial traffic schemes.

3. Phase Combination and Signal Timing

In order to improve the adaptability of the traffic schemes for maching different traffic conditions, we establish the phase combination model and optimize the signal timing using the bilevel programming model.

3.1. Phase Combination Model. Signal phase refers to one or more flow directions displayed by the same signal lamp in a signal cycle time. The phase combination model mainly analyzes the conflicts of traffic flows with different directions, and use clustering ideas to merge nonconflicting flows with similar traffic characteristics into one phase. A reasonable combination can effectively reduce the release time of the green light, improve the utilization of the green light, and ensure transportation safety.

Two traffic flows are conflicting if there is a collision point of the vehicle travel path in these two directions. For example, the traffic flow in the east-west direction and the south-north direction are conflict, while the traffic flow in the east-west direction and the west-east direction are compatible. The conflict matrix can be constructed as follows:

$$\mathbf{C} = \begin{bmatrix} 0 & \varphi_{12} & \varphi_{13} & \cdots & \varphi_{1n} \\ \varphi_{21} & 0 & \varphi_{23} & \cdots & \varphi_{2n} \\ \varphi_{31} & \varphi_{32} & 0 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \varphi_{(n-1)n} \\ \varphi_{n1} & \varphi_{n2} & \cdots & \varphi_{n(n-1)} & 0 \end{bmatrix}, \quad (9)$$

where φ_{ij} indicates whether the flow direction and j is conflict. If not, the value is 0, otherwise, 1.

The distance matrix is used to represent the difference between traffic flows, which can be constructed based on the saturation of flow directions, green signal split data, and the conflict matrix:

$$\mathbf{D} = \begin{bmatrix} 0 & d_{12} & d_{13} & \cdots & d_{1n} \\ d_{21} & 0 & d_{23} & \cdots & d_{2n} \\ d_{31} & d_{32} & 0 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & d_{(n-1)n} \\ d_{n1} & d_{n2} & \cdots & d_{n(n-1)} & 0 \end{bmatrix}, \quad (10)$$

where the element d_{ij} in the matrix can be calculated as follows:

$$d_{ij} = \left(y_i - y_j\right)^2 + \varphi_{ij} = \left(x_i\lambda_i - x_j\lambda_j\right)^2 + \varphi_{ij}, \quad (11)$$

where y_i is the traffic flow ratio of the flow direction i, which reflecting the traffic demand not affected by the signal control scheme. x_i is the saturation of the flow direction i, and λ_i is the initial green split of the flow direction i.

Input:
The distance matrix **D**
Output:
The final phase combination scheme D_{final}
Begin
1. for $i = 1$ to n, do
2, For $j = 1$ to $(i - 1)$, do
3. If $d_{ij} < 1$ Then
4. D_{first}.add(d_{ij}) ;
5. End If
6. End For
7. End For
8. While $c_1, c_2, c_3, \cdots, c_n!$ = each other, do
9. $D_{each} \longleftarrow [d_{c_1c_2}, d_{c_3c_4}, d_{c_5c_6}, \cdots, d_{c_{n-1}c_n}]$;
10. D_{all}.add(D_{each}) ;
11. $S_{each} \longleftarrow$ sum(D_{each}) ;
12. S_{all}.add(S_{each}) ;
13. End While
14. $S_{\min} \longleftarrow$ Selectmin(S_{all}) ;
15. $z \longleftarrow$ The index of $S_{\min}$ in S_{all} ;
16. $D_{final} \longleftarrow D_{all}[z]$;
End

Algorithm 1: Phase combination.

Since the distance between the flow direction i and j is the same as the distance between the flow direction j and i, the distance matrix is symmetric, that is, $d_{ij} = d_{ji}$. To ensure the balance of traffic flows in each phase, we optimize the phase combination according to the distance matrix between flow directions to make the combination more rational. For a typical crossroad, four-phase schemes are usually used, each phase consists of two flow directions, and the same flow direction traffic must be released only once in one cycle. Considering the symmetry of the distance matrix and all-zero values on the main diagonal, only the lower triangle needs to be processed. Algorithm 1 shows the optimization of the phase combination. If the distance between two flows is equal to or greater than 1, these two flows are physically conflicting. Hence, we select all the flow pairs with their distances less than 1 to form the D_{first} vector. If one scheme in the D_{first} contains all flow direction and each direction c_i only appears once, it will be saved as D_{each} to D_{all}. Then, we calculate the sum of the distances in D_{each} and insert it into the S_{all} as S_{each}, and the index z of the minimum $S_{\min}$ in S_{all} is selected. Finally, we choose the optimal scheme D_{final} according to z in D_{all}.

For example, there are two schemes here (see Figure 2): Scheme A takes east left movement and east through movement as one phase, west left movement and west through movement as another phase. Scheme B takes east left movement and west left movement as one phase, east through movement and west through movement as another phase, the distances of above four combinations are 0.2, 0.1, 0.3, and 0.4, respectively. The scheme A is chosen because the sum of the first two values is smaller than that of the last two values.

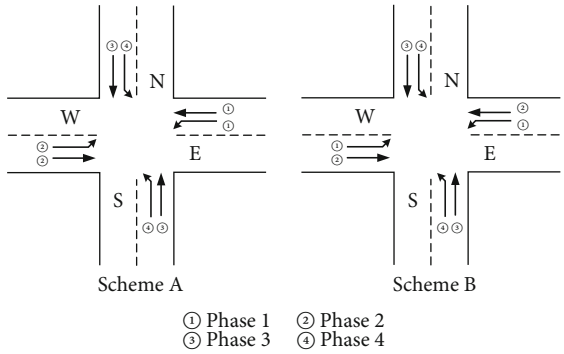


Figure 2: Schematic diagram of phase scheme comparison.

3.2. Traffic Signal Timing Bi-Level Programming Model. The bilevel programming model is a system optimization model with a two-tier hierarchical structure. The upper and lower levels have their own objective functions and constraints [28, 29]. The objective functions and constraints of the upper-level problem are not only related to the upper decision variables but also depending on the optimal solution of the lower level problem, while the optimal solution of the lower level problem is affected by the upper decision variables.

We establish a traffic signal timing optimization algorithm based on the bilevel programming model. The framework of the traffic signal timing optimization algorithm is shown in Figure 3.

3.2.1. Establishment of the Bilevel Programming Model. The signal cycle time is the key control parameter that determines the quality of traffic signal control in traffic signal timing, and the saturation can reflect the rationality of the signal cycle time to some extent. We establish the upper-level programming model with saturation as the decision target, which can be calculated as:

$$J = \min\left(|\bar{x} - \tilde{x}|\right), \tag{12}$$

where $\bar{x}$ is the average saturation of each phase, and $\tilde{x}$ is the target average saturation.

Under the condition of fixed signal cycle time constraints, the mean square error (MSE) of the saturation is used to evaluate the rationality of green split distribution. With the MSE, the lower-level programming model can be established as:

$$\sigma = \min\left(\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(x_i - \bar{x}\right)^2}\right), \tag{13}$$

where N is the number of signal phases. The saturation of each phase can be calculated as:

$$x_i = \frac{f_i}{q_i \lambda_i}, \tag{14}$$

where f_i is the arrival traffic flow for phase i, q_i is the average of each flow direction saturated flow in phase i, and λ_i is the initial green split of phase i.

3.2.2. Solution of the Bilevel Programming Model. The single-step action set with signal cycle time changes is designed to obtain the optimal signal cycle time of the upper-level programming model, the action set can be expressed as follows:

$$\boldsymbol{action}1 = [a_1, -a_1, 0], \tag{15}$$

where a_1, in seconds, is the adjustment step size for cycle time.

The three elements in action1 represent three operations, including addition, subtraction and invariance, respectively. For example, if the initial signal cycle time is T, the action1 is $[a_1, -a_1, 0]$, and the signal cycle time after each adjustment according to action1 will be $[T + a_1, T - a_1, T]$.

Algorithm 2 shows the optimization for signal cycle time. Each action of Equation (15) is executed in the initial signal cycle time T_0, and J is calculated by Equation (12) and (14), which will be inserted into J_{all}. The minimum $J_{\min}$ in J_{all} is selected, if its corresponding action is nonzero, the action will be taken, and the signal cycle time after execution will be updated as the initial scheme T_0 for the next iteration. The algorithm will loop until the action corresponding to $J_{\min}$ is zero, and the signal cycle time at this time $\mathbf{Now}[w]$ is regarded as the optimal signal cycle time T_f.

Similar to the upper level, to solve the optimal green split of the lower-level programming model, we design a set of single-step changes in the green time of each signal phase, the action set is

$$\begin{aligned}\text{action2} = [&[a_2, a_2, -a_2, -a_2], [a_2, -a_2, a_2, -a_2], [a_2, -a_2, -a_2, a_2],\\ &[-a_2, a_2, a_2, -a_2], [-a_2, a_2, -a_2, a_2], [-a_2, -a_2, a_2, a_2],\\ &[a_2, -a_2, 0, 0], [a_2, 0, -a_2, 0], [a_2, 0, 0, -a_2],\\ &[-a_2, a_2, 0, 0], [0, a_2, -a_2, 0], [0, a_2, 0, -a_2],\\ &[-a_2, 0, a_2, 0], [0, -a_2, a_2, 0], [0, 0, a_2, -a_2],\\ &[-a_2, 0, 0, a_2], [0, -a_2, 0, a_2], [0, 0, -a_2, a_2], [0, 0, 0, 0]],\end{aligned} \tag{16}$$

where a_2, in seconds, is the adjustment step size for green time.

The four elements in action2 represent the adjustment of green time of each phase in the four-phase scheme. For example, if the initial green time is $[g_1, g_2, g_3, g_4]$, the action2 is $[[a_2, a_2, -a_2, -a_2], \cdots, [0, 0, 0, 0]]$ and the green time of each phase after each adjustment according to action2 will be $[[g_1 + a_2, g_2 + a_2, g_3 - a_2, g_4 - a_2], \cdots, [g_1, g_2, g_3, g_4]]$.

Algorithm 3 shows the process for green split optimization. Considered the premise of the green split optimization algorithm that the signal cycle time is fixed, the sum of all elements in the action matrix is zero. According to the initial scheme of green split, the initial timing scheme is obtained by multiplying the signal cycle time. Each action of Equation (16) is executed, respectively, and then σ value of the corresponding action can be saved into $\boldsymbol{\sigma}_{\text{all}}$ according to Equation

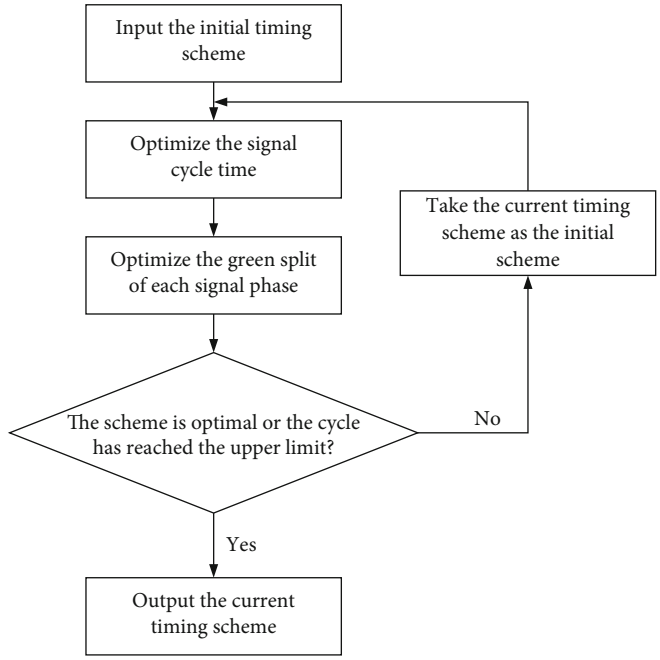


Figure 3: The framework of the traffic signal timing optimization algorithm.

Input:
The initial signal cycle time T_0, the average arrival traffic flow for each phase $\mathbf{f} = [f_1, f_2, \cdots, f_n]$, the average of each flow direction saturated flow in each phase $\mathbf{q} = [q_1, q_2, \cdots, q_n]$, the initial green split of each phase $\boldsymbol{\lambda} = [\lambda_1, \lambda_2, \cdots, \lambda_n]$ and the action set action1
Output:
The optimal signal cycle time T_f
Begin
1. $w_0 \longleftarrow$ The index of 0 in action1 ;
2. $w \longleftarrow$ A number! = w_0 ;
3. While w! = w_0, do
4. For each a_i in $action1$, do
5. $TNow[i] \longleftarrow$ Calculate Adjust_T(T_0, a_i) ;
6. End For
7. $J_{all} \longleftarrow$ [];
8. For each TN_i in $TNow$, do
9. J_{all}.add(Calculate J($TN_i, f_i, q_i, \lambda_i$)) ;
10. End For
11. $J_{\min} \longleftarrow$ Selectmin(J_{all}) ;
12. $w \longleftarrow$ The index of $J_{\min}$ in J_{all} ;
13. If w = = w_0 Then
14. $T_f \longleftarrow TNow[w]$;
15. Else
16. $T_0 \longleftarrow TNow[w]$;
17. End If
18. End While
End

Algorithm 2: Signal cycle time optimization.

(13) and (14). Then we select the minimum $\sigma_{\min}$ in $\boldsymbol{\sigma}_{\text{all}}$, if its corresponding action is not $[0, 0, 0, 0]$, the action will be taken, and the green timing scheme after execution is updated as the initial scheme $\boldsymbol{g}_0$ for the next iteration. The algorithm will loop until the action corresponding to $\sigma_{\min}$ is $[0, 0, 0, 0]$, and the green time of each signal phase at this time is converted into green split, and the optimal green split scheme λ_f is output.

We complete the green split optimization in the lower-level programming model, which will be fed back to the upper level. While in the upper level, the signal cycle time is optimized heuristically and iteratively under the restriction

Input:
The signal cycle time T, the arrival traffic flow for each phase $\mathbf{f} = [f_1, f_2, \cdots, f_n]$, the average of each flow direction saturated flow in each phase $\mathbf{q} = [q_1, q_2, \cdots, q_n]$, the initial green split of each phase $\boldsymbol{\lambda}_0 = [\lambda_{01}, \lambda_{02}, \cdots, \lambda_{0n}]$, and the action set action2
Output:
The optimal green split scheme$\boldsymbol{\lambda}_f = [\lambda_{f1}, \lambda_{f2}, \cdots, \lambda_{fn}]$

$I_0 \longleftarrow$ The index of [0, 0, 0, 0] in action2 ;
$I \longleftarrow$ A number! = I_0 ;
For i in λ_0, do
    g_0.add($\lambda_{0i} T$) ;
While I! = I_0, do
    For each a_i in $action2$, do
        $GNow[i] \longleftarrow$ Calculate Adjust_g(g_0, a_i) ;
    End For
    $\sigma_{all} \longleftarrow$ [];
    For each GN_i in $GNow$, do
        σ_{all}.add(Calculate $\sigma(GN_i, f_i, q_i)$) ;
    End For
    $\sigma_{\min} \longleftarrow$ Selectmin(σ_{all}) ;
    $I \longleftarrow$ The index of $\sigma_{\min}$ in σ_{all} ;
    If I = = I_0, do
        $g_f \longleftarrow GNow[I]$;
    Else
        $g_0 \longleftarrow GNow[I]$;
    End If
End While
For i in g_f, do
    λ_f.add(g_{fi}/T) ;
End

Algorithm 3: Green split optimization.

Figure 4: The actual layout of the intersection.

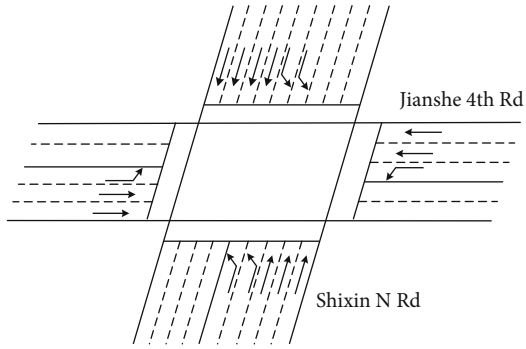


Figure 5: The simulation layout of the intersection.

of the green split, until the scheme is optimal or the cycle reaches the upper limit.

4. Simulation Experiment and Result Analysis

4.1. Experimental Methods and Experimental Data. The experiment is simulated in SUMO (Simulation of Urban Mobility), which is an open-source, highly portable, microscopic, and continuous traffic simulation software. The real-world intersection, Jianshe 4th Rd and Shixin N Rd in Xiaoshan District, Hangzhou, China (see Figure 4) is chosen as the operating environment.

According to the traffic laws and regulations in our country, the right turn movement can pass the intersection at any time without being controlled by the signal light; thus, only the left turn and the straight vehicles are considered in the simulation. Figure 5 shows the simulation structure of the intersection.

The traffic flow data were provided by the traffic control department of Xiaoshan District, Hangzhou, from 7 : 00 a.m. to 9 : 00 a.m. on November 20th, 2018. The original data was the traffic flow data of the signal cycle time and the timing scheme of the corresponding time, which was processed into saturation data set for clustering, and then, timing

Table 1: Traffic flow data of each flow direction in each time period of the intersection (veh/h).

Time	LE	SE	LS	SS	LW	SW	LN	SN
7 : 00	128	242	168	476	92	186	266	368
7 : 30	202	364	150	980	96	198	320	798
8 : 00	124	240	184	758	88	164	238	662
8 : 30	138	275	143	752	98	150	282	760
9 : 00	118	220	186	576	102	148	224	448
Saturation flow	1529	1641	1347	2360	1286	1606	1722	2228

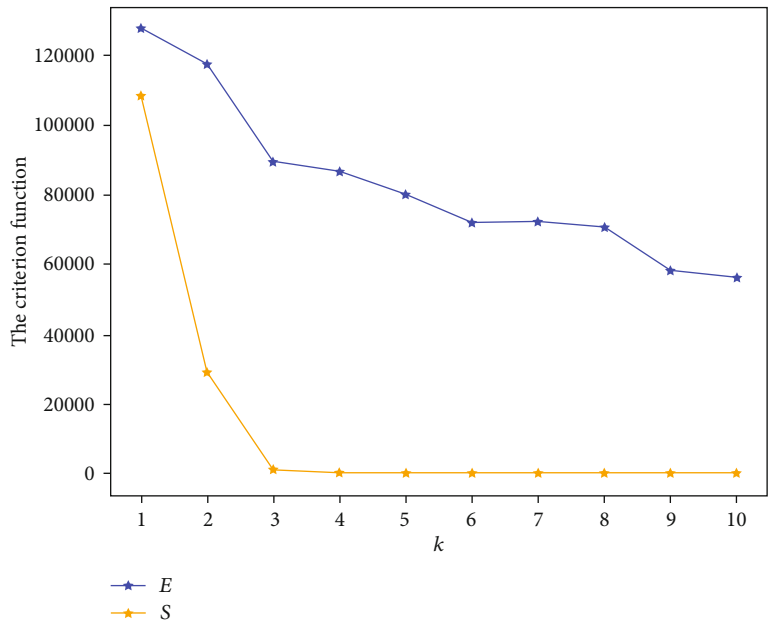


Figure 6: Algorithm comparison on the saturation dataset.

optimization was carried out based on the original timing scheme. Then, the data were divided into 5 periods, and flows of each direction were calculated every half an hour. The traffic flow data at a certain point in time is the average flow of the adjacent 15 minutes, that is, the traffic flow at 8 : 00 a.m. is the average traffic flow from 7 : 45 a.m. to 8 : 15 a.m. The flow data were divided into eight flow directions, such as the left turn for eastbound movement, the through for eastbound and so on. The average traffic flow of all lanes in each flow direction is recorded in Table 1, which has been converted into the hourly traffic flow to the inlet, and the through flow of each flow direction is also recorded. In this table, "E," "S," "W," and "N" refer to eastbound, southbound, westbound, and northbound, respectively. "L" and "S" mean left turn and straight vehicles. For example, "LE" represents the traffic flow of the left turn in the eastbound movement.

The signal timing scheme generated by the improved K-medoids clustering algorithm is compared with the scheme generated by the traditional one to ensure the fairness of the experiment. In order to avoid the exponential explosion and make the criterion function E and S be in the same order of magnitude, the weight coefficient t is set as 11000. Additionally, we set the target average saturation δ_x to 70 according to the actual intersection traffic demand. In order to avoid missing the optimal timing scheme due to overlarge step size, the signal cycle time adjustment step a_1 and the green time adjustment step a_2 are both set as 1.

Table 2: Performance comparison of the traditional and improved K-medoids algorithm.

Algorithm	Number of clusters	Average execution time(s)
Traditional K-medoids	[3–6]	1.764
Improved K-medoids	3	1.623

In addition, the proposed algorithm is compared to the fixed phase scheme and the traffic flow and vector angle based on the optimization scheme [17].

4.2. Analysis of Results. The criterion functions of different k using traditional and improved K-medoids algorithm are shown, expressed by E and S, respectively. As k increases, the criterion functions decrease, and the rate of decline also stabilizes. In both cases, the optimal k is 3, while using the

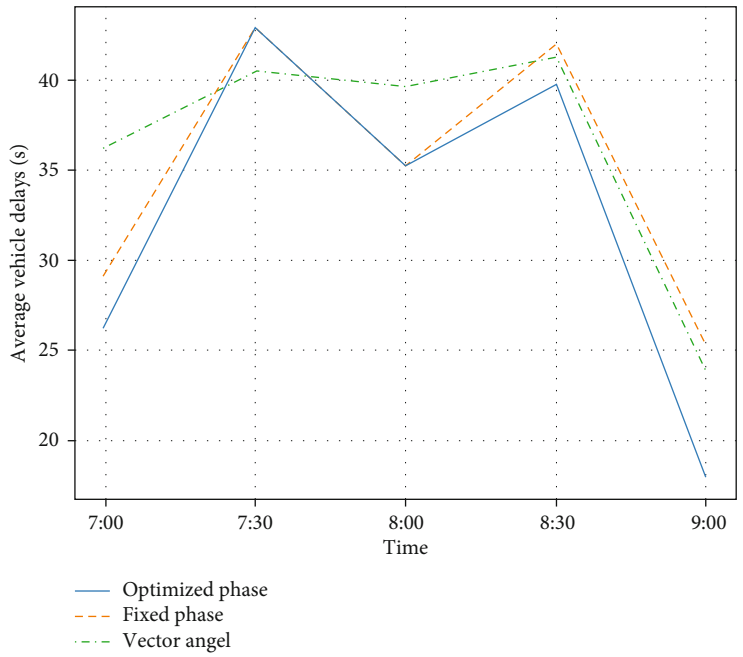


Figure 7: Comparison of average vehicle delays at intersection.

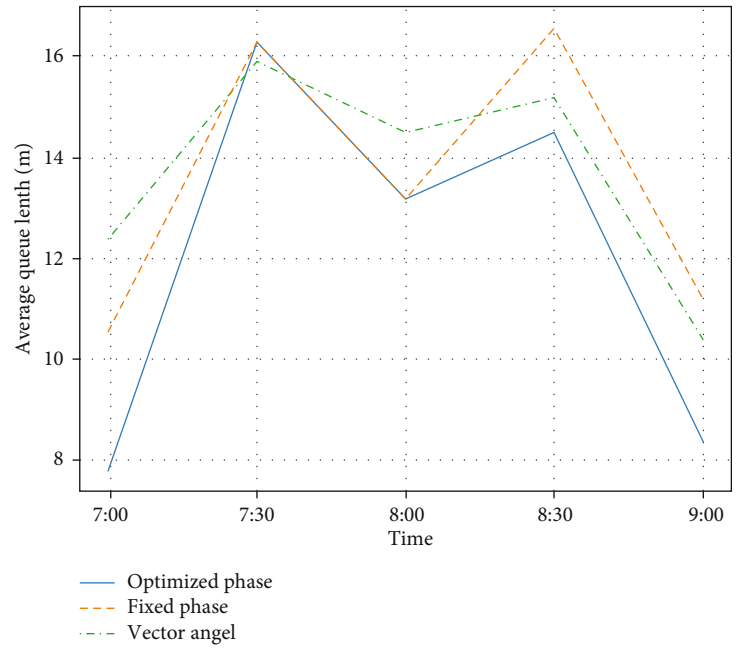


Figure 8: Comparison of average queue length at intersection.

improved K-medoids, it is easier to reach the result, and the elbow point can be identified more unambiguously. Figure 6 shows the curves of both algorithms, which is more intuitive.

Table 2 shows the different performances of the traditional and improved K-medoids algorithm. As for the number of clusters, in different iterations, the traditional K-medoids may reach the elbow when k is in range of 3 and 6, which is ambiguous to identify, while the improved K-medoids can

Table 3: Comparison of three optimization schemes at intersection.

	Average vehicle delays (s)	Average queue length (m)
Fixed phase	34.842	13.554
Vector angle	36.304	13.668
Optimized phase	32.380	12.012

Table 4: Delay comparison of three optimization schemes in flow direction.

Time	Fixed phase			Vector angle			Optimized phase		
	SS	SW	LE	SS	SW	LE	SS	SW	LE
7 : 00	14.44	23.89	35.57	20.66	45.82	37.62	15.91	28.88	27.28
7 : 30	29.81	24.04	40.20	21.35	53.38	41.17	29.82	24.05	40.19
8 : 00	22.74	23.61	32.67	29.84	55.28	44.68	22.73	23.62	32.67
8 : 30	35.35	20.92	39.16	27.55	55.99	50.72	29.18	26.45	30.31
9 : 00	12.21	25.20	34.72	11.79	31.85	29.78	14.92	30.91	21.41
Average	22.91	23.53	36.46	22.24	48.46	40.79	22.51	26.78	30.37

always reach the elbow when $k = 3$. In addition, the improved K-medoids runs faster than the traditional version, and that may because we optimize the selection of initial optimizing, which reduces the number of interactions.

Average vehicle delay and average queue length are used to evaluate the performance of the proposed algorithm.

Figures 7 and 8 show the curves of optimized phase and timing schemes under different conditions compared to fixed schemes that optimize only timing and vector angle-based schemes. The outperformance of our proposed method can be seen in all time periods. Table 3 shows the averaged values of the above two evaluation indexes, we can see that the proposed method outperforms the fixed phase method with improvements of 2.462 s (7.07%), and 1.542 m (11.38%) on the vehicle delay and the queue length, and also shows improvements of 3.924 s (10.81%) and 1.656 m (12.16%) compared to the traffic flow and vector angle-based optimization scheme.

Table 4 shows the delay comparison of three optimization schemes in SS, SW, and LE. We can see that the method proposed by us has a great improvement on the average vehicle delays in each flow direction compared to the traffic flow and vector angle-based optimization scheme. In our proposed method, the average vehicle delays of SW and LE is different from that of the fixed phase method, this is because the phase of SW and LE has changed. Compared with the fixed phase method, the average vehicle delays of LE in our method are reduced, but the average vehicle delays of SW are increased. The main reason is that our method improves the overall traffic capacity of the intersection rather than the single flow direction.

5. Conclusions

In this paper, we optimize the traditional K-medoids clustering algorithm in terms of the clustering number and initial medoids selection. In order to match the changes of traffic flow in different time periods adaptively, the phase combination optimization model is established to optimize the phase, and the bilevel programming model aims to optimize the signal timing, which can maximize the utilization of green time. The proposed algorithm is optimized for each flow direction. Whereas the flow saturation may be different when the overall situation is similar, we will study the difference of different flow saturation to achieve the optimal control effect of the intersection.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Acknowledgments

The authors express their sincere thanks to the researchers of Institute of Computer Intelligent System, Zhejiang University of Technology, China, for their help. This work was partially supported by the National Natural Science Foundation of China under Grant No. 61701443.

References

[1] G. Shen, L. Zhu, J. Lou, S. Shen, Z. Liu, and L. Tang, "Infrared multi-pedestrian tracking in vertical view via Siamese convolution network," *IEEE Access*, vol. 7, pp. 42718–42725, 2019.

[2] G. Shen, C. Chen, Q. Pan, S. Shen, and Z. Liu, "Research on traffic speed prediction by temporal clustering analysis and convolutional neural network with deformable kernels (May, 2018)," *IEEE Access*, vol. 6, pp. 51756–51765, 2018.

[3] G. Shen, X. Han, J. Zhou, Z. Ruan, and Q. Pan, "Research on intelligent analysis and depth fusion of multi-source traffic data," *IEEE Access*, vol. 6, pp. 59329–59335, 2018.

[4] G. Shen and Y. Yang, "A dynamic signal coordination control method for urban arterial roads and its application," *Frontiers of Information Technology & Electronic Engineering*, vol. 17, no. 9, pp. 907–918, 2016.

[5] X. Kong, X. Liu, B. Jedari, M. Li, L. Wan, and F. Xia, "Mobile crowdsourcing in smart cities : technologies , applications , and future challenges," *IEEE Internet of Things Journal*, vol. 6, no. 5, pp. 8095–8113, 2019.

[6] X. Kong, F. Xia, J. Li, M. Hou, M. Li, and Y. Xiang, "A shared bus profiling scheme for smart cities based on heterogeneous mobile crowdsourced data," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 2, pp. 1436–1444, 2020.

[7] W. Wang, J. Chen, J. Wang, J. Chen, and Z. Gong, "Geography-aware inductive matrix completion for personalized point of interest recommendation in smart cities," *IEEE Internet of Things Journal*, 2019.

[8] W. Wang, J. Chen, J. Wang, J. Chen, J. Liu, and Z. Gong, "Trust-enhanced collaborative filtering for personalized point of interests recommendation," *IEEE Transactions on Industrial Informatics*, 2019.

[9] Z. Yao, L. Shen, R. Liu, Y. Jiang, and X. Yang, "A dynamic predictive traffic signal control framework in a cross-sectional vehicle infrastructure integration environment," *IEEE Transactions*

on Intelligent Transportation Systems, vol. 21, no. 4, pp. 1455–1466, 2020.

[10] S. Le Vine, X. Liu, F. Zheng, and J. Polak, "Automated cars: Queue discharge at signalized intersections with 'Assured-Clear-Distance-Ahead' driving strategies," *Transportation Research Part C: Emerging Technologies*, vol. 62, pp. 35–54, 2016.

[11] A. Vogel, L. Oremovic, and R. Simic, "Fuzzy Traffic Light Control Based on Phase Urgency," in *2019 International Symposium ELMAR*, Zadar, Croatia, September 2019.

[12] C. Tang, S. Xia, C. Zhu, and X. Wei, "Phase timing optimization for smart traffic control based on fog computing," *IEEE Access*, vol. 7, no. 8, pp. 84217–84228, 2019.

[13] Y. Guo, J. Ma, C. Xiong, X. Li, F. Zhou, and W. Hao, "Joint optimization of vehicle trajectories and intersection controllers with connected automated vehicles: combined dynamic programming and shooting heuristic approach," *Transportation Research Part C: Emerging Technologies*, vol. 98, no. 1, pp. 54–72, 2019.

[14] C. Ma, W. Hao, A. Wang, and H. Zhao, "Developing a coordinated signal control system for urban ring road under the vehicle-infrastructure connected environment," *IEEE Access*, vol. 6, no. 5, pp. 52471–52478, 2018.

[15] Y. Wang and G. Zheng, "Research on polling based traffic signal control strategy with fuzzy control," in *2018 IEEE 4th International Conference on Computer and Communications (ICCC)*, Chengdu, China, December 2018.

[16] H. Rashid, "Intelligent traffic light control based on clustering using vehicular ad-hoc networks," in *2015 7th Conference on Information and Knowledge Technology (IKT)*, pp. 1–6, Urmia, Iran, May 2015.

[17] W. Hao, "Traffic signal segmentation algorithm based on two-dimensional clustering of traffic volume and vector angles," in *2017 International Smart Cities Conference (ISC2)*, Wuxi, China, September 2017.

[18] W. Rao, J. Xia, W. Lyu, and Z. Lu, "Interval data-based k -means clustering method for traffic state identification at urban intersections," *IET Intelligent Transport Systems*, vol. 13, no. 7, pp. 1106–1115, 2019.

[19] H. Chu and C. Wang, "Using K-means algorithm for the road junction time period analysis," in *2017 IEEE 8th International Conference on Awareness Science and Technology (iCAST)*, pp. 441–446, Taichung, Taiwan, November 2017.

[20] A. Rao, A. Phadnis, A. Patil, T. Rajput, and P. Futane, "Dynamic traffic system based on real time detection of traffic congestion," in *2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA)*, pp. 1–5, Pune, India, August 2018.

[21] J. Datesh, W. T. Scherer, and B. L. Smith, "Using k-means clustering to improve traffic signal efficacy in an IntelliDrive^{SM} environment," in *2011 IEEE Forum on Integrated and Sustainable Transportation Systems*, pp. 122–127, Vienna, Austria, June-July 2011.

[22] W. Liu, G. Qin, Y. He, and F. Jiang, "Distributed cooperative reinforcement learning-based traffic signal control that integrates V2X networks' dynamic clustering," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 10, pp. 8667–8681, 2017.

[23] X. Wang, W. Cottrell, and S. Mu, "Using k-means clustering to identify time-of-day break points for traffic signal timing plans," in *Proceedings 2005 IEEE Intelligent Transportation Systems, 2005*, pp. 586–591, Vienna, Austria, September 2005.

[24] Y. Jun and Y. Yang, "Optimization of time of day break points of intersection using cluster method," *International Journal of Advancements in Computing Technology*, vol. 5, no. 6, pp. 479–487, 2013.

[25] S. Theodoridis and K. Koutroumbas, *Pattern Recognition, Fourth Edition*, Academic Press, 2008.

[26] A. Rodriguez and A. Laio, "Machine learning. Clustering by fast search and find of density peaks," *Science*, vol. 344, no. 6191, pp. 1492–1496, 2014.

[27] D. Yu, G. Liu, M. Guo, and X. Liu, "An improved K-medoids algorithm based on step increasing and optimizing medoids," *Expert Systems with Applications*, vol. 92, pp. 464–473, 2018.

[28] X. Wang, Y. Chang, and P. Zhang, "Traffic signal optimization based on system equilibrium and bi-level multi-objective programming model," in *Green Intelligent Transportation Systems. GITSS 2016. Lecture Notes in Electrical Engineering, vol 419*, W. Wang, K. Bengler, and X. Jiang, Eds., pp. 429–438, Springer, Singapore, 2018.

[29] G. Zhang, J. Han, and J. Lu, "Fuzzy bi-level decision-making techniques : a survey," *International Journal of Computational Intelligence Systems*, vol. 9, pp. 25–34, 2016.

Hindawi
Wireless Communications and Mobile Computing
Volume 2020, Article ID 7917021, 14 pages
https://doi.org/10.1155/2020/7917021

Research Article

A Deep Multiscale Fusion Method via Low-Rank Sparse Decomposition for Object Saliency Detection Based on Urban Data in Optical Remote Sensing Images

Cheng Zhang[1] **and Dan He**[2]

[1]*City Institute, Dalian University of Technology, China*
[2]*Dalian University of Finance and Economics, China*

Correspondence should be addressed to Cheng Zhang; zhangc@dlut.edu.cn

Received 6 February 2020; Accepted 16 April 2020; Published 8 May 2020

Academic Editor: Qingchen Zhang

The urban data provides a wealth of information that can support the life and work for people. In this work, we research the object saliency detection in optical remote sensing images, which is conducive to the interpretation of urban scenes. Saliency detection selects the regions with important information in the remote sensing images, which severely imitates the human visual system. It plays a powerful role in other image processing. It has successfully made great achievements in change detection, object tracking, temperature reversal, and other tasks. The traditional method has some disadvantages such as poor robustness and high computational complexity. Therefore, this paper proposes a deep multiscale fusion method via low-rank sparse decomposition for object saliency detection in optical remote sensing images. First, we execute multiscale segmentation for remote sensing images. Then, we calculate the saliency value, and the proposal region is generated. The superpixel blocks of the remaining proposal regions of the segmentation map are input into the convolutional neural network. By extracting the depth feature, the saliency value is calculated and the proposal regions are updated. The feature transformation matrix is obtained based on the gradient descent method, and the high-level semantic prior knowledge is obtained by using the convolutional neural network. The process is iterated continuously to obtain the saliency map at each scale. The low-rank sparse decomposition of the transformed matrix is carried out by robust principal component analysis. Finally, the weight cellular automata method is utilized to fuse the multiscale saliency graphs and the saliency map calculated according to the sparse noise obtained by decomposition. Meanwhile, the object priors knowledge can filter most of the background information, reduce unnecessary depth feature extraction, and meaningfully improve the saliency detection rate. The experiment results show that the proposed method can effectively improve the detection effect compared to other deep learning methods.

1. Introduction

With the rapid promotion of information technology, urban data has become one of the important information sources for human beings. And the amount of information received by people has increased exponentially [1, 2]. How to select the object regions of human interest from the mass of image information in urban becomes a significant research. Studies have found that under a complex scene, the human visual processing system will focus on several objects, named region of interest (ROI) [3]. ROI is relatively close to human visual perception. Saliency, as the image pretreatment process, can be widely applied in remote sensing areas such as visual tracking, image classification, image segmentation, and target relocation.

The saliency detection method mainly contains two aspects: top-down and bottom-up. The top-down-based saliency detection method [4–6] is a task-driven process. The ground-truth images are labeled manually for supervised training. It integrates more perceptions of humans to obtain the salient map. However, the bottom-up method is a data-driven process and pays more attention to the images'

Figure 1: Saliency detection instance.

features such as contrast, position, and texture to compute the saliency map (SM). Itti et al. [7] proposed a spatial visual model taking full advantage of local contrast and obtained the saliency map via the image differences from the center to the surrounding. Hou and Zhang [8] put forward a saliency detection algorithm based on Spectral Residual (SR). Achanta et al. [9] proposed a frequency-tuned (FT) method based on the image frequency domain to calculate saliency. A detection method combining histogram was presented to calculate global contrast [10]. Furthermore, other relevant methods were raised and showed better effect [11–15]. But they do not analyze the image from the dimensions.

Yan et al. [16] treated the saliency region of the image as sparse noise and the background as a low-rank matrix. It calculated the saliency of the image by using the sparse representation and robust principal component analysis algorithm. Firstly, the image was decomposed into 8×8 blocks. Every image block was sparsely encoded and merged into a coding matrix. Then, the coding matrix was decomposed by robust principal component analysis. Finally, the sparse matrix obtained by decomposition was devoted to establish the saliency factor of the corresponding image block. However, because the large-size saliency object contained many image blocks, the saliency object in each image block no longer satisfied the sparse feature; thus, it greatly affected the detection effect. Lang et al. [17] utilized a multitask low-rank recovery approach for saliency detection. The multitask low-rank representation algorithm was used to decompose the feature matrix and constrained the consistency of all feature sparse components in the same image blocks. The algorithm used the consistency information of multifeature description, and its effect was improved. However, since the large-size target contained a large number of feature descriptions, the feature was no longer sparse. The reconstruction error could not solve this problem, so this method could not completely detect the saliency object with a large size. To perfect the result of the above method, Shen and Wu [18] proposed a low-rank matrix recovery (LRMR) algorithm combining bottom-up and top-down algorithm (providing high-level and low-level information, respectively). First, it performed the superpixel segment in the image and several features were extracted. Then, the feature transformation matrix and a priori knowledge, including size, texture, and color, were obtained by network learning to transform the feature matrix. Finally, the low-rank and sparse decomposition of the transformed matrix were carried out by using the robust principal component analysis algorithm. This method improved the deficiency to some extent. However, due to the limitation of center prior and the failure of color prior to complex scenes, this algorithm was not ideal for detecting images with complex backgrounds.

The saliency detection method using different low-level features is usually only effective for a specific type of image, which is not suitable for multiobject images in complex scenes [19–21]. Figure 1 is the instance of saliency detection. The low-level features of visual stimuli lack an understanding of the nature of saliency objects and cannot represent the features at a deeper level. For noisy objects in the image, if they are similar to the low-level features but do not belong to the same category, they are often wrongly detected as saliency objects. Yang et al. [22] showed a bag of word model to detect saliency. Firstly, the prior probability saliency map could be obtained through the object feature, and a word bag model representing the middle semantic features was established to calculate the conditional probability saliency graph. Finally, two saliency images were synthesized by Bayesian inference. The middle semantic features could represent the image content more accurately than the bottom features. Therefore, the detection effect was more accurate. Jiang et al. [23] took saliency detection as a regression problem and integrated regional attributes, contrast, and feature vectors of regional background knowledge at multiscale segmentation conditions. The saliency map was obtained by supervised learning. Due to the introduction of background knowledge features, the algorithm had a better ability to identify background objects, and thus obtained more accurate foreground detection results.

Deep learning (DL) combines low-level features to form more abstract high-level features, a typical representative is a convolutional neural network (CNN). Many saliency detection methods have adopted CNN to optimize the result. Li et al. [24] proposed deep CNN to detect saliency. Firstly, region and edge information were obtained by using the hyper-pixel algorithm and bilateral filtering. DCNN was utilized to extract the regions and edge features in raw images. Finally, the region confidence graph and edge confidence graph generated by CNN were integrated into the conditional random field to judge the saliency. Wang et al. [25] proposed recurrent fully CNN (i.e., RFCNN) for saliency detection, which mainly included two steps: pretraining and fine-tuning. RFCN was used to train the original image to correct the saliency prior image. Then, the traditional algorithm was used to further optimize the modified saliency graph.

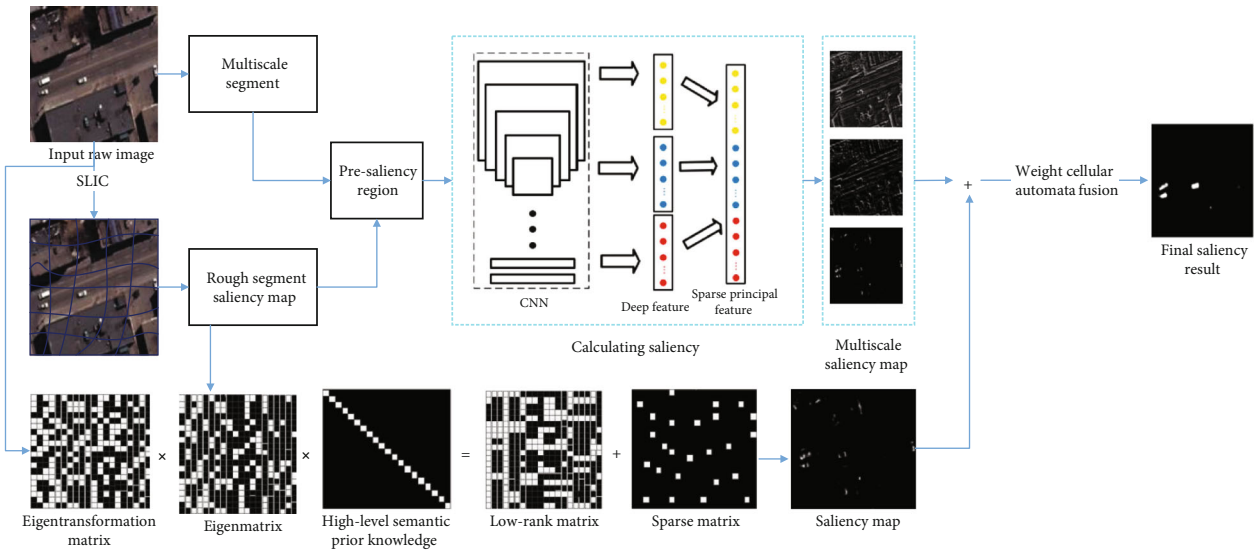


Figure 2: The framework of proposed saliency detection.

Lee et al. [26] proposed a deep saliency (DS) algorithm for saliency detection using low and high-level information in a unified CNN framework. VGG-Net was used to extract the advanced features. It mainly extracted the low-level features. Then, the CNN was used to encode the distance graph. Finally, the coded low-level distance graph was connected with higher features. A full-connected CNN classifier was adopted to evaluate the features' information and obtain the saliency graph [27]. The above DL methods show the excellent performance in terms of saliency detection rate. But there are still some disadvantages such as slow speed and highly complex calculations.

In this paper, we propose a deep multiscale fusion method via low-rank sparse decomposition for object saliency detection in optical remote sensing images. The main contributions are as follows.

(a) First, multiscale segmentation is executed for remote sensing images. For the first segmentation graph, the depth features of all the superpixel blocks are extracted by CNN

(b) Then, we calculate the saliency value, and the proposal region is generated. The superpixel blocks of the remaining proposal regions of the segmentation graph are input into the CNN network. By extracting the depth feature, the saliency value is calculated and the proposal regions are updated. Meanwhile, the color, texture, and edge feature mean values of all the pixels in each superpixel are calculated to construct the feature matrix. In order to make the image background facilitate low-rank sparse decomposition, the above feature matrices need to be transformed so that the background can be represented as a low-rank matrix in the new feature space

(c) To make use of the high-level information and improve the detection effect of the ROI, the fully convolutional neural network is used for learning features, and the high-level semantic prior knowledge matrix is obtained. The feature matrix is transformed by using the feature transformation matrix and the high-level semantic prior knowledge. The robust principal component analysis algorithm is used to decompose the transformed matrix into a low-rank sparse decomposition to obtain a saliency map. The process is iterated continuously to obtain the saliency map on each scale

(d) Finally, the weight cellular automata method fuses the multiscale saliency graphs. It is shown that the proposed method can effectively improve the detection effect compared to other DL methods

The remainder of the paper is organized as follows. The proposed deep multiscale fusion method for saliency detection is analyzed in section II. Section III introduces the saliency region extraction based on multiscale segmentation. Saliency is calculated based on the deep features in section IV. The performance and robustness are evaluated in section V. Conclusion is drawn in section VI.

2. Deep Multiscale Fusion for Saliency Detection

The proposed deep multiscale fusion method for saliency detection in optical remote sensing images is shown in Figure 2.

Firstly, the image I is segmented into a small number of superpixel blocks by using the superpixel segmentation algorithm. The deep feature is extracted from all the superpixel blocks. The color, texture, and edge feature mean value of all the pixels in each superpixel are calculated to construct the feature matrix. In order to make the image background facilitate low-rank sparse decomposition, the above feature matrix needs to be transformed so that the background can be represented as a low-rank matrix in the new feature space. And the multidimensional feature containing the key

information of the image is extracted by PCA (principal component analysis). The rough segmentation saliency graph is obtained based on the calculation of key features, where we can extract the initial saliency region to obtain the superpixel set *Suppix*. Then, we adopt *Suppix* to centralize the similarity degree between superpixel and the nonobject region. The input image is segmented at different scales. The region containing the superpixel block in the *Su ppix* set is selected for depth feature extraction. Saliency maps and *Suppix* sets at the next scale are obtained based on the same method. The robust PCA is used to decompose the transformed matrix into a low-rank sparse decomposition to obtain a saliency map. Weight cellular automata fusion is used to obtain the final SM M_{final}.

3. Saliency Region Extraction Based on Multiscale Segmentation

Superpixel segmentation is to gather adjacent similar pixel points into image regions with different sizes according to the low-level features such as brightness, thus reducing the complexity of significance calculation. The superpixel segmentation algorithm mainly includes watershed [28] and simple linear iterative clustering (SLIC) [25] method. We combine their respective characteristics, SLIC method is used to obtain the segmentation results with regular shape and uniform size during rough segmentation, and the watershed algorithm is used to obtain better object contour during fine segmentation in this study.

For N segmentation scales $(s_1,\cdots,s_n)$. $Sup_j = \{Sp_i^j\}_{i=1}^{N_j}$ denotes the obtained superpixel set at a certain segmentation scale. N_j denotes the superpixel number at scale s_j. $Sp_i^j(v) = \{R, G, B, L, a, b\}$ is pixel's color feature vector in the superpixel.

For the input image, we extract color, texture, and edge features to construct the feature matrix.

(i) Color feature. The gray value of R, G, B, hue, and saturation are extracted to describe the color feature of the image

(ii) Edge feature. Steerable pyramid filter is used to decompose the image in multiple scales and directions. Filters with 3 scales and 4 directions are selected to obtain 12 responses as the edge features of the image

(iii) Texture feature. Gabor filter is used to extract texture features at different scales and directions. Here, 3 scales and 12 directions are selected to obtain 36 responses as the texture features

It calculates the mean value of all pixel features in each superpixel to represent the eigenvalue f_i. All the eigenvalues constitute the eigenmatrix $F = [f_1, f_2,\cdots,f_N]$, $F \in R^{d\times N}$.

The saliency region of the image is regarded as sparse noise and the background as a low-rank matrix. In the complex background, the image background similarity degree after clustering is still not high. Therefore, the features in the original image are not conducive to low-rank sparse decomposition. In order to find a suitable feature space, most image backgrounds can be represented as low-rank matrices; in this paper, the eigentransformation matrix is obtained based on the gradient descent method. The process of obtaining the eigentransformation matrix is as follows:

(a) Construct marker matrix $Q = \text{diag}\{q_1, q_2,\cdots,q_N\}$. If the superpixel p_i is within the marked saliency region manually, $q_i = 0$. Otherwise, $q_i = 1$

(b) According to the following formula, the optimal model of transformation matrix T is utilized to learn the features of raw image

$$T_{optimal} = \arg\min_T O(T) = \frac{1}{K}\sum_{k=1}^{K} \|TF_kQ_k\|_\circ - \gamma\|T\|_\circ \quad (1)$$

Where $F_k \in R^{d\times N_k}$ is the feature matrix of kth image. N_k represents the superpixel number of kth image. $Q_k \in R^{N_k\times N_k}$ is the labeled matrix of the kth image. $\|\cdot\|_\circ$ represents the kernel norm of the matrix, that is, the sum of all singular values of the matrix. γ is the weight coefficient. $\|T\|_2$ denotes the ℓ_2 norm of the matrix T. c is a constant to prevent T from arbitrarily increasing or decreasing. If the eigentransformation matrix T is appropriate, then TFQ is low rank. $-\gamma\|T\|_\circ$ is to avoid obtaining the general solution when the rank of T is arbitrarily small.

(c) Find the $T_{optimal}$ gradient descent direction, that is

$$\frac{\partial O(T)}{\partial T} = \frac{1}{K}\sum_k \frac{\partial\|TF_kQ_k\|_\circ}{\partial T} - \gamma\frac{\partial\|T\|_\circ}{\partial T} \quad (2)$$

(d) Adopt the following formula to update the eigentransformation matrix T until the algorithm converges to the local optimal. α is the step size

$$T_{t+1} = T_t - \alpha\frac{\partial O(T)}{\partial T} \quad (3)$$

3.1. Extracting Proposal Region. The segmentation graph of a rough segmentation scale s_j is taken as input. The saliency map Map_j is obtained by depth feature extraction and saliency value calculation. The Map_j, as the object prior knowledge in the next segmentation, is used to guide the proposal region extraction. The saliency Map_j is binarized. The value of Map_j is divided into K channels by the adaptive threshold strategy. $p(i)$ is used to represent the number of pixels in the channel i. The channel k with

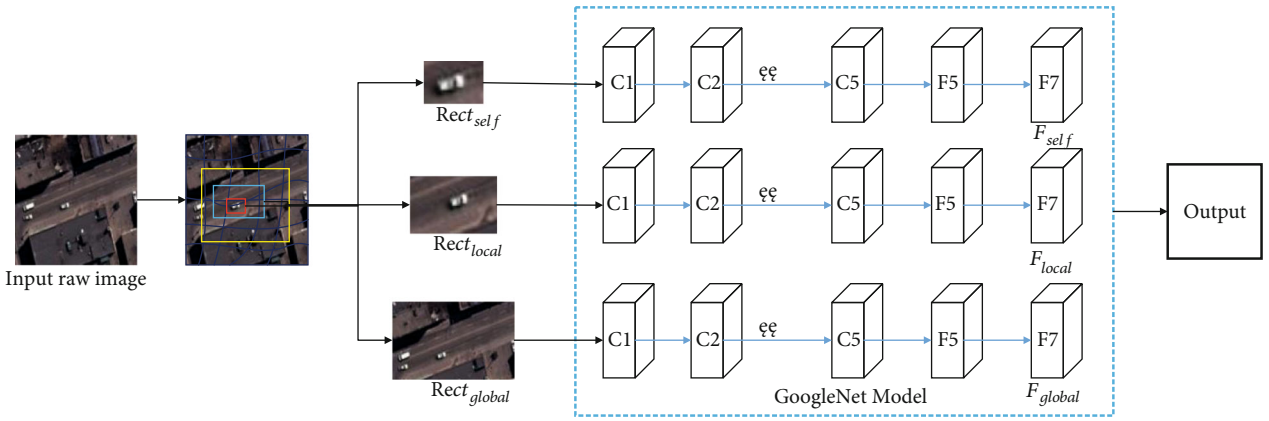


Figure 3: Deep features extraction based on CNN.

the largest number of pixels in all channels is determined. The threshold value T is calculated by the formula (4).

$$T = (k+1)/K \tag{4}$$

In order to prevent T from getting larger, the significant pixel will not be binarized to 0 when the saliency object occupies the most space in the image. The pixel number in each channel must satisfy $p(i)/area(I) < \varepsilon$. Where $area(I)$ is the pixel number of image I. $\varepsilon \in [0.6, 0.9]$ is an experience value. The binarization object a priori map is denoted as $MapB_j$. We adopt $MapB_j$ as the prior knowledge. The corresponding superpixel area of superpixel set $Sup_{j+1} = \{Sp_i^{j+1}\}_{i=1}^{N_{j+1}}$ in the next scale s_{j+1} constitutes the proposal saliency superpixel set $Suppix_{j+1} = \{Sp_i^{j+1}\}_{i=1}^{M_{j+1}}$. M_{j+1} is the number of proposal saliency superpixel at the scale s_{j+1}, $M_{j+1} < N_{j+1}$. Assume that Num_i is the total number of the superpixel Sp_i^{j+1}. num is the pixel number with a value of 1 at the corresponding position of the binary map $MapB_j$. If $num/Num_j > 0.5$, the superpixel at the corresponding position is considered to belong to $Suppix_{j+1}$.

3.2. Region Optimization. The proposal object superpixel set may contain some background areas or missing saliency areas. It needs to optimize the proposal object area. It removes the possible background area in $Suppix_{j+1}$ and adds the possible saliency area in the background area. According to the Euclidean distance between the two color spaces, the difference matrix is $Difmat$. It is a symmetric matrix with N_{j+1} order.

$$Difmat(i,j) = Difmat\left(Sp_i, Sp_j\right) = \sqrt{\sum_{k=1}^{6}\left(F_{i,k} - F_{j,k}\right)^2} \tag{5}$$

Where $F_{i,k}$ is the kth feature of superpixel region Sp_i. $k = [1,\cdots,6]$ corresponds to R, G, B, L, a, and b, respectively. For $Sp_k \in Suppix_{j+1}$, it calculates the local average dissimilarity degree through equation (6),

$$MavDif(Sp_k) = \frac{\sqrt{\sum_{l=1,l\neq k}^{M_{j+1}} Difmat(Sp_k, Sp_l)^2}}{M_{j+1}} \tag{6}$$

Where $Sp_k, Sp_l \in Suppix_{j+1}$, M_{j+1} is superpixel number in the proposal saliency region set $Suppix_{j+1}$. We calculate the average dissimilarity degree of each superpixel Sp_k in $Suppix_{j+1}$ and its adjacent background region:

$$MavDif(Sp_k)' = \frac{\sqrt{\sum_{l=1,l\neq k}^{M'_{j+1}} Difmat(Sp_k, Sp_l)^2}}{M'_{j+1}} \tag{7}$$

Where $Sp_k \in Suppix_{j+1}$, $Sp_l \notin Suppix_{j+1}$ and Sp_k is adjacent to Sp_l. M'_{j+1} represents the number of superpixels adjacent to Sp_k in the background area. If $MavDif(Sp_k)' > MavDif(Sp_k)$, it indicates that Sp_k is more similar to the adjacent background area, then Sp_k will be removed from $Suppix_{j+1}$.

Similarly, for any $Sp_h \notin Suppix_{j+1}$, the average dissimilarity $MavDif(Sp_h)'$ between Sp_h and adjacent background region, and average dissimilarity $MavDif(Sp_h)$ between Sp_h and adjacent proposal saliency region can be calculated. If the condition $MavDif(Sp_h)' > MavDif(Sp_h)$ is satisfied, then the similarity between Sp_h and the adjacent saliency region is higher than that of other background regions. Therefore, Sp_h is added to $Suppix_{j+1}$. $Suppix_{j+1}$ is constantly updated by comparing the superpixel in $Suppix_{j+1}$ with other saliency regions and background regions. Until the superpixel in $Suppix_{j+1}$ is no longer changed.

3.3. Deep Feature Extraction of Proposal Region. This deep feature extraction method based on CNN is as shown in Figure 3. In the first superpixel segmentation, the deep

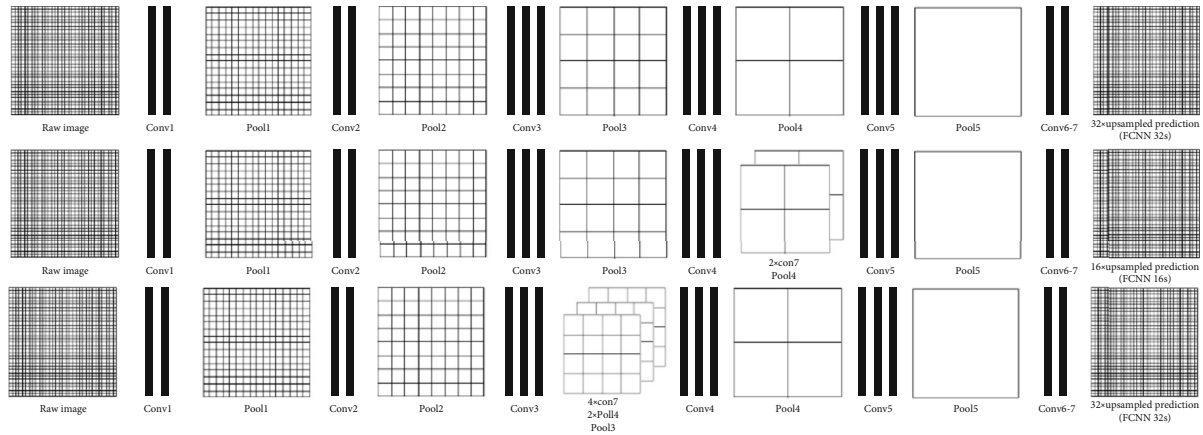


Figure 4: FCNN model.

features of all superpixels are extracted. In the subsequent deep feature extraction process, only the superpixel in *Su ppix* set is extracted. Under a certain segmentation strategy, the computation is greatly reduced and the computation speed is increased.

Assuming it is not the first time to segment the superpixel, the local and global features are extracted for superpixel Sp_i. The local features of the superpixel include two parts: (1) the deep feature F_{self} containing its own region; (2) deep feature F_{local} containing itself and the adjacent superpixel region.

First, according to *Suppix* set, it extracts the minimum rectangular region Re ct_{self} of each superpixel Sp_i. Since most superpixels are not regular rectangles, the extracted rectangles must contain other pixels. These pixels are represented by the average value of the superpixel. The depth feature F_{self} only containing its own region can be obtained through the deep CNN.

If we only adopt the saliency calculation of F_{self} to acquire saliency detection value is meaningless. It is impossible to determine whether it is saliency without comparing it with the saliency of other adjacent superpixels. Therefore, it still needs to extract Re ct_{local} to further obtain F_{local} of the deep local feature. The location of the region in the image is an important factor to judge whether it is saliency or not. It is generally believed that the area in the center of the image is more likely to be saliency than the region at the edge. Therefore, the whole image is taken as the input, and the deep feature F_{global} of the global region is extracted.

If it only uses the bottom feature to extract the saliency map, due to many interference objects, the final saliency map is not ideal. Therefore, the high-level information needs to be added to improve the detection effect. The adopted high-level semantic prior knowledge is mainly to predict the most likely ROI based on previous experience (i.e., training samples). The FCNN is used to train the high-level semantic prior knowledge, which is integrated into the feature transformation process to optimize the final saliency map. Higher-order features can be learned from the primitive data without preprocessing in the multi-stage global training process of CNN.

FCNN can accept input images with any size. The difference between FCNN and CNN is that the deconvolution layer replaces the full connection layer. Finally, pixel classification is carried out on the feature map of the upsampling. A binary prediction is produced for each pixel, and a classification result at the pixel level is output. Thus, the problem of image segmentation at the semantic level is solved. Semantic a priori is an important high-level information in the detection of the ROI, which can assist the detection of the ROI. Therefore, this paper adopts FCNN to obtain high-level semantic prior knowledge and applies it to the detection of the ROI.

The network structure of FCNN is shown in Figure 4. Based on the original classifier, this paper utilizes the back propagation algorithm to fine-tune the parameters in all FCNN layers. In the network structure, the first row gets the feature map after alternately seven convolutional layers and five pooling layers. The last step of the deconvolution layer is to conduct the upsampling of the feature map with a step size 32 pixels. The network structure in this paper is denoted as FCNN-32s. It is found that the precision decreases because of the maximum pool operation. It directly executes upsampling for the feature map of downsampling, which will result in very rough output and details loss. Therefore, in this paper, the features with step size 32 pixels obtained from the upsampling are extended by 2 times, which is summed with the feature with step size 16 pixels. Then, the obtained feature is recovered to the original image for training, and the FCNN-16s model is obtained. So more accurate detailed information is obtained than that of FCNN-32s. We adopt the same method to train the network to obtain the FCNN-8s model, the prediction of detailed information is more accurate. Experiments show that although lower-level feature fusion for training networks can make detailed information prediction more accurate, the effect of low-rank sparse decomposition on the result is not significantly improved. Since the training time will increase sharply, this paper adopts FCNN-8s model to acquire the high-level priori knowledge of images.

The deep CNN model comprises an input layer, multiple convolution layers, downsampling layer, full connection layer, and output layer. The downsampling layer and

convolution layer form the intermediate structure of the neural network. The former is used for feature extraction, and the latter is for feature calculation. The fully connected layer is connected with the downsampling layer, which can output the feature. The output of the convolution layer is:

$$d_n^l = f\left(\sum_{\forall m}\left(d_m^{i-1} \cdot k_{m,n}^l\right) + b_m^l\right) \tag{8}$$

Where d_n^l and d_m^{i-1} are the feature maps of the current layer and the previous layer. $k_{m,n}^l$ is the convolution kernel of the model. $f(x) = 1/[1 + e^{-x}]$ is the neuron activation function. b_n^l is neuron bias. The feature extraction result of the downsampling layer is:

$$d_n^l = f\left(k_n^l \times \frac{1}{s^2}\sum_{s\times s} d_n^{l-1} + b_n^l\right) \tag{9}$$

Where $s \times s$ is the downsampling template scale. k_n^l is the template weight. In this paper, the trained GoogleNet model is used to extract the depth features of the proposal object region. On the strength of this model, the labeled output layer is removed to obtain a depth feature. The convolution layer C1 uses 96 filters with $11 \times 11 \times 3$ size to filter the input image with size $224 \times 224 \times 3$. The convolution layers C2, C3, C4, and C5 take the output of the downsampling layer as their input, respectively. The convolution processing is carried out by using the self-filter, and several output feature graphs are obtained and transmitted to the next layer. The full connection layers F6 and F7 have 4096 features. The output of each full connection layer can be denoted as:

$$d_n^{\text{out}} = f\left(\sum\left(d_n^{\text{out}-1} \times k_{m,n}^{\text{out}}\right) + b_n^{\text{out}}\right) \tag{10}$$

3.4. Saliency Calculation Based on Deep Feature. PCA [28] is the common method for dimension reduction of high-dimensional data, which can replace p high-dimensional features with a smaller number of m features. For n superpixels, the output features can constitute a sample matrix W with $n \times p$ dimension. The correlation coefficient matrix $R = (r_{ij})_{p\times p}$ of the sample is calculated by the formula (11):

$$r_{ij} = \frac{\sum_{k=1}^n (x_{ki} - \bar{x}_i)(x_{kj} - \bar{x}_j)}{\sqrt{\sum_{k=1}^n (x_{kj} - \bar{x}_j)^2 \sum_{k=1}^n (x_{ki} - \bar{x}_i)^2}}, i, j = 1, 2, \cdots, p. \tag{11}$$

Where $\bar{x}_i = 1/n\sum_{i=1}^n x_{ij}$. By solving the equation $|\lambda I - R| = 0$, we find the eigenvalues and order them. Then, we calculate the contribution rate and cumulative contribution rate of each eigenvalue λ_i:

$$con_{rate} = \lambda_i / \sum_{k=1}^p \lambda_k, cum_{rate} = \sum_{k=1}^i \lambda_k / \sum_{k=1}^p \lambda_k, i, j = 1, 2, \cdots, p. \tag{12}$$

We calculate the corresponding orthogonal unit vector $z_i = [z_{i1}, z_{i2}, \cdots, z_{ip}]^T$ of each eigenvalue λ_i. The unit vector corresponding to the first m features with a cumulative contribution rate 95% is selected to form the transformation matrix $Z = [z_1, z_2, \cdots, z_m]_{p\times m}$. The high-dimensional matrix m is reduced by formula (13). $Sp_i(df) = (f_{i,1}, f_{i,2}, \cdots, f_{i,m})$ denotes the m-dimension principal component feature. The principal component features are extracted by the same transformation matrix in the segmentation maps with different scales.

$$Sp_i(df) = W_{n\times p} Z_{p\times m} \tag{13}$$

3.5. Contrast Feature. The contrast feature reflects the difference degree between the region and its adjacent region. The contrast feature $w^c(Sp_i)$ of the superpixel Sp_i is defined by its distance from other superpixels features, as given in equation (14):

$$w^c(Sp_i) = \frac{1}{n-1}\sum_{i=1, i\neq k}^n \left\| Sp(df)_i - Sp(df)_k \right\|_2 \tag{14}$$

Where n denotes the number of superpixel. $\|\cdot\|_2$ is 2-norm.

3.6. Spatial Feature. In the human visual system, we pay different attentions in different spatial positions. The distance between the pixel at different positions and the image center satisfies the Gaussian distribution. For any superpixel Sp_i, its spatial feature $w^s(Sp_i)$ is calculated as:

$$w^s(Sp_i) = e^{\frac{-d(Sp_{i,x}, c)}{\sigma^2}} \tag{15}$$

Where $Sp_{i,x}$ is the central coordinate of superpixel Sp_i. c is the central region. If the average distance from the image center is smaller, the spatial feature is larger. The saliency value of the superpixel Sp_i is denoted as:

$$Map(Sp_i) = w^c(Sp_i) \times w^s(Sp_i) \tag{16}$$

We obtain the SM of the first segmented image and use it as the object prior knowledge to guide the proposal region extraction and optimization.

3.7. Saliency Detection Based on Low-Rank Sparse Decomposition. The background in the image can be expressed as a low-rank matrix. The saliency region can be regarded as sparse noise. For an original image, the eigenmatrix $F = [f_1, f_2, \cdots, f_N] \in R^{d\times N}$ and the eigentransformation matrix T are obtained. Then, we use the FCN to obtain the high-level prior knowledge P. The low-rank sparse decomposition of the transformed matrix is carried out by robust PCA.

$$(L^*, S^*) = \arg\min_{L,S}(\|L\|_\circ + \lambda\|S\|_1) \quad s.t.\ TFP = L + S \tag{17}$$

Input: Raw image I, multiscale segment number N and segment parameter in each scale.
Output: Saliency map.
for $i = 1 : N$
{
if i=1 then
(1) According to the determined parameters, we use SLIC to segment image I;
(2) Determine the input region Re ct_{self}, Re ct_{local}, Re ct_{global} of each superpixel;
(3) The above is input GoogleNet to extract deep feature F_{self}, F_{local}, F_{global};
(4) The deep features of all superpixels constitute a matrix W, and the transformation matrix A of W is calculated by using PCA to obtain the principal component features;
(5) According to the principal component features, saliency values without object priors are calculated to obtain the first segmentation saliency map Map_1;
else
(6) According to the determined parameters, we use Watershed algorithm to segment image;
(7) The saliency map Map^{i-1} is taken as object priori map. Then it extracts and optimizes proposal object set *Suppix*;
(8) Determine the input region Re ct_{self}, Re ct_{local}, Re ct_{global} in *Suppix*;
(9) The above is input GoogleNet to extract deep feature F_{self}, F_{local}, F_{global};
(10) The deep features of all superpixels constitute a matrix W, and the transformation matrix A of W is calculated by using PCA to obtain the principal component features;
(11) According to the principal component features, saliency values with object priors are calculated. And we obtain the saliency map Map_i;
end if
}
(12) Calculate the saliency map weight w_i at each scale;
(13) Adopt weight cellular automata to fuse the obtained N saliency maps and get final SM.

Algorithm 1: Proposed saliency detection method.

Where F is the eigenmatrix. T is the learned eigentransformation matrix. P is a high-level prior knowledge matrix. L is a low-rank matrix. S represents the sparse matrix. $\|\cdot\|_\circ$ represents the kernel norm of the matrix, that is, the sum of all singular values of the matrix. $\|\cdot\|_1$ represents the ℓ_1-norm of the matrix, the sum of the absolute values of all the elements in the matrix. Supposing that $S*$ is the optimal solution for the sparse matrix. The saliency map can be calculated by the following equation.

$$Sal(p_i) = \|S^*(:,i)\|_1 \tag{18}$$

Where $Sal(p_i)$ represents the saliency value of superpixel p_i. $\|S^*(:,i)\|_1$ represents the ℓ_1-norm of the ith column vector of $S*$, that is, the sum of the absolute values of all the elements in the vector.

3.8. Saliency Map Fusion Based on Weight Cellular Automata. Wang and Wang [29] adopted the multilayer cellular automata (MCA) for object fusion. Each pixel represents a cell. In the m-layer cellular automata, the cellular of the saliency map has m-1 neighbors. They are at the same positions in other saliency maps.

If cellular i is labeled as foreground, the foreground probability of its neighbor j at the same position in other SMs is $\lambda = P(\eta_i = +1 \mid i \in F)$. Saliency maps obtained by different methods are considered to be independent. When updating synchronously, all saliency maps are considered to have the same weight. There are guiding and refining relationships between the saliency maps at different segmentation scales. The weights cannot be considered as equally during the fusion process. In different segmentation scales, it is assumed that the weight of the SM obtained by the first segmentation scale is λ_1, represented by $w_i = \lambda_1$. The SM weight with different scale is expressed as:

$$w_i = \lambda_{i-1} + (1 - o_i/O_i), i = 1, 2, \cdots, 6 \tag{19}$$

Where O_i denotes the total pixel number in the proposal object set. o_i is the superpixel number in the ith saliency map. Set $\lambda_1 = 1$. Synchronous updating mechanism $f : Map^{M-1} \longrightarrow Map$ is defined as:

$$l\left(Map_m^{t+1}\right) = w_m \sum_{k=1,k\neq m}^{M} \operatorname{sign}\left(Map_k^t - \gamma_k \cdot I\right) \cdot \ln\left(\frac{\lambda}{1-\lambda}\right) + l\left(Map_m^t\right) \tag{20}$$

Where $Map_m^t = [Map_{m,1}^t, \cdots, Map_{m,H}^t]^T$ represents the saliency value of all the cellular of the mth SM at time t. Matrix I is a matrix with H elements. If the neighbor of cellular is judged as foreground, then the saliency value should be increased. We obtain the final saliency map by formula (21). $T2$ is the next time.

$$Map_{final} = \frac{1}{N} \sum_{m=1}^{M} \left(Map_m^{T_2} + Sal(p_i)\right) \tag{21}$$

The proposed deep multiscale fusion method for object saliency detection is summarized as depicted in Algorithm 1.

4. Experiments and Analysis

In this section, we obtain the experiment data from Google Earth. The remote sensing image size is from 512×512 pixel to 2048×2048 pixel. The spatial resolution is 1 m. The experiment environment is Intel(R), Core(TM), i7-8750, CPU2.2 Hz, Geforce GTX1060 with MATLAB 2017a platform.

4.1. Evaluation Index and Parameter Setting. In the experiment, the PR curve, F-measure, and mean absolute error (MAE) of the saliency map are compared to evaluate the effect of saliency detection to select a better segmentation scale.

Precision and Recall are the two most commonly used evaluation criteria in image saliency detection. If the PR curve is higher, the effect of saliency detection is better. Otherwise, it is poor. For the given manual labeled Ground Truth G and the saliency map S, the definition of Precision and Recall is given in equation (22):

$$\Pr ecision = \frac{sum(S, G)}{sum(S)}, \operatorname{Re} call = \frac{sum(S, G)}{sum(G)} \tag{22}$$

Where $sum(S, G)$ represents the sum of the value after the pixels of visual feature graph S multiplying that of G. $sum(S)$ is the sum of all pixels in the visual feature graph S. $sum(G)$ represents the sum of all pixels in G.

When calculating F-measure, the adaptive threshold T of each image is used to segment the image.

$$T = \frac{2}{W \cdot H} \sum_{x=1}^{W} \sum_{y=1}^{H} S(x, y) \tag{23}$$

Where the W and H denote the width and height of the image, respectively. It calculates the average precision and recall of the SM. The average F-measure value is calculated according to equation (24). The effect of saliency is better if the F-measure value is excellent. F-measure value is used for the comprehensive evaluation of accuracy and recall. β^2 is often set to 1.

$$F = \frac{2Pecision \cdot \operatorname{Re} call}{Pecision + \operatorname{Re} call} \tag{24}$$

MAE is used to evaluate the saliency model by comparing the difference between the SM and the GT. We use formula (25) to compute the MAE value of each input image. The calculated MAE value can be used to draw a histogram. If the MAE value is lower, the proposed algorithm is better.

$$MAE = \frac{1}{W \cdot H} \sum_{x=1}^{W} \sum_{y=1}^{H} |S(x, y) - G(x, y)| \tag{25}$$

4.2. Segment Scale Determination. The main parameter of this algorithm is the segmentation scale. Many segmentation scales can increase the computational complexity. Few scales will affect the accuracy of saliency detection. Therefore, 15 segmentation scales are set according to experience. We conduct experiments on randomly selected remote sensing image data. Then, we extract the depth features of all superpixels in the segmentation graphs and calculate the saliency map. The histogram of the PR curve with different segmentation scales is shown in Figure 5. Three segmentation scales with better effects are selected from them. Through comparative analysis, it is found that the three segmentation scales 10, 11, 12 have a relatively better saliency detection effect. The three segmentation scales are selected as the final segmentation scales of the proposed method.

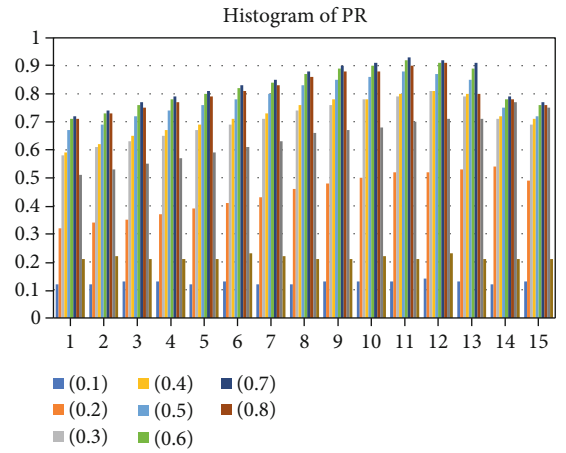


Figure 5: Histogram of PR curve with different segmentation scales.

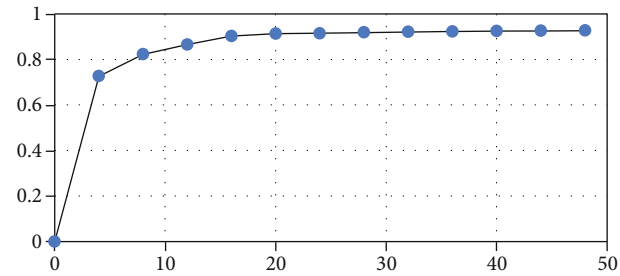
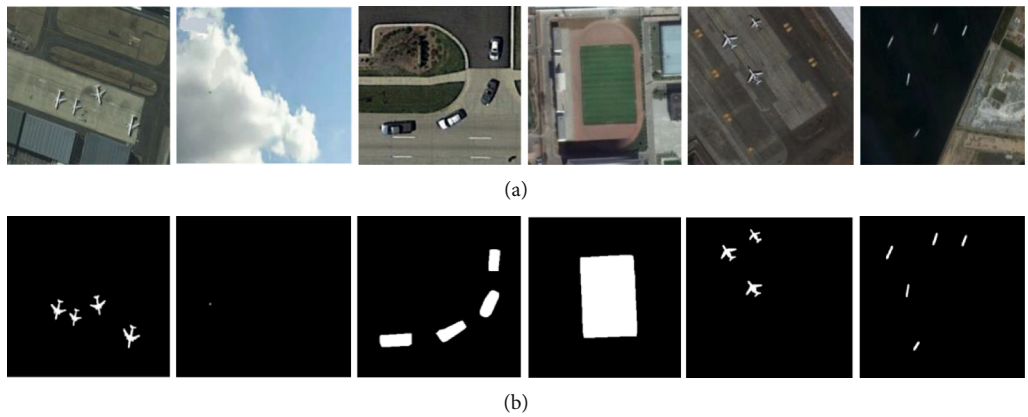


Figure 6: Relation between PEV and the principal components.

4.3. PCA Parameter Determination. To verify the effectiveness of PCA on selecting principal components from depth features, this section adopts the depth features extracted from each superpixel block as the data set. The percentage of explained variance (PEV) is used to measure the importance of the principal component in the overall data as formula (26). PEV is a main index to describe the distortion rate of data.

$$PEV = \sum_{i=1}^{m} R_{ii}^2 / tr\left(\sum\right) \tag{26}$$

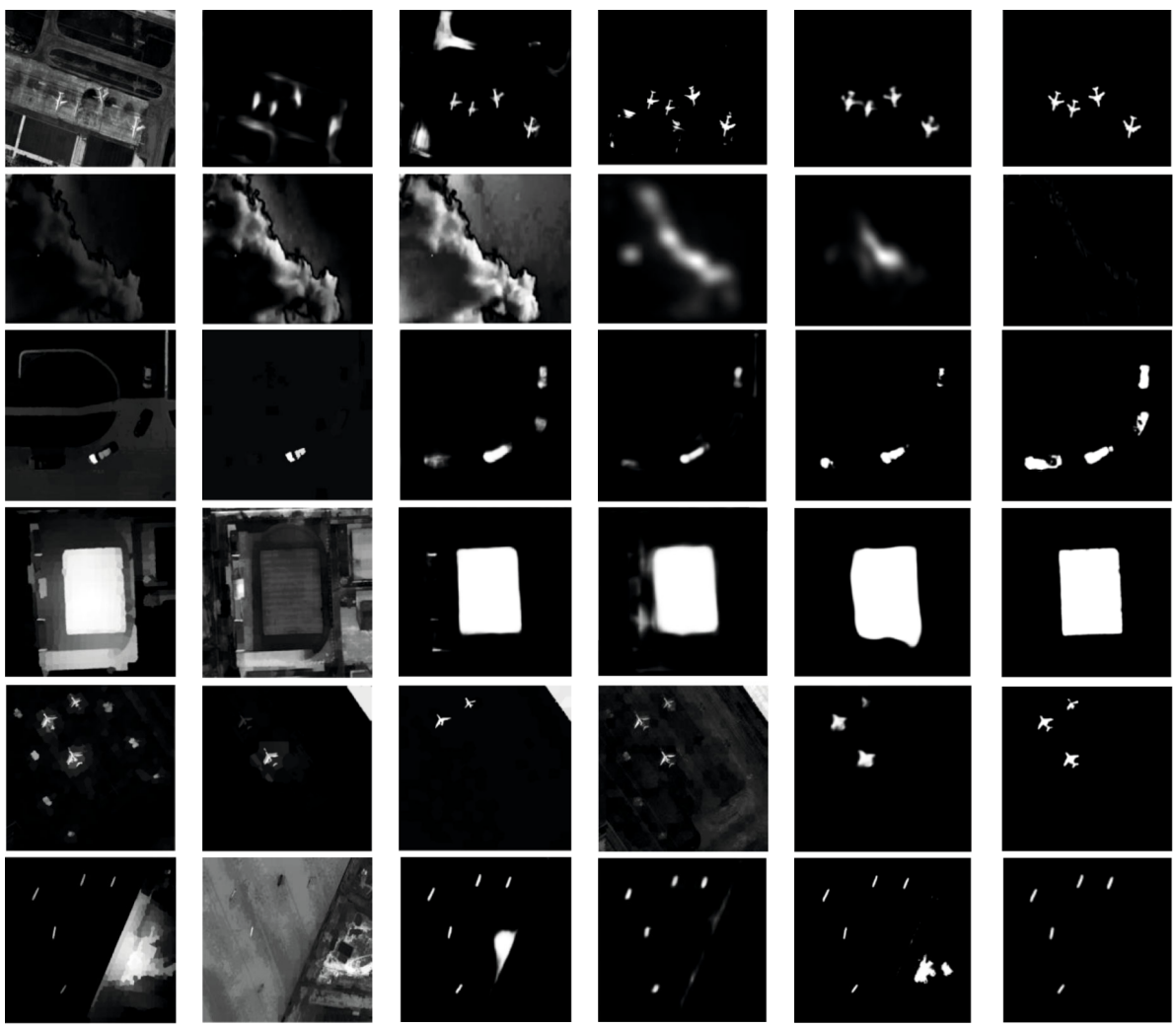
Where R_{ii}^2 is the right matrix of the main component matrix M' after singular value decomposition. $\sum$ denotes the covariance matrix. Figure 6 shows the relation between PEV and the top 50 principal components. It reveals that



(a)

(b)

Figure 7: Test images: airplane1, cloud, vehicle, playground, airplane2, boat. (a) Raw remote sensing images; (b) Ground Truth.



(a) (b) (c) (d) (e) (f)

Figure 8: Comparison of saliency images. (a) RA. (b) RB. (c) SC. (d) RAD. (e) SCLR. (f) Proposed.

with the increase of principal component number, PEV shows an upward trend. But the trend grows slowly. When the number of principal components exceeds 20, the PEV reaches to 90%, which is considered to represent the overall information of the data. In this paper, the top 20 principal components are selected for saliency calculation.

Table 1: Different indexes with different methods on different objects.

Object	Method	Precision	Recall	F-measure	MAE
Airplane1	RA	79.9%	73.5%	72.8%	19.3%
	RB	81.1%	75.8%	76.4%	17.2%
	SC	81.8%	74.6%	77.2%	15.7%
	RAD	87.4%	77.4%	79.5%	14.6%
	SCLR	91.7%	75.4%	81.8%	12.5%
	Proposed	95.6%	65.3%	82.5%	9.8%
Cloud	RA	84.6%	68.9%	73.8%	21.2%
	RB	89.1%	71.8%	75.4%	17.8%
	SC	90.7%	74.1%	76.2%	14.1%
	RAD	91.6%	73.7%	78.4%	12.6%
	SCLR	93.6%	72.8%	80.9%	11.3%
	Proposed	97.4%	71.5%	83.6%	7.6%
Vehicle	RA	89.2%	77.4%	78.7%	19.5%
	RB	91.5%	79.9%	80.8%	17.6%
	SC	93.1%	79.3%	82.1%	13.8%
	RAD	93.6%	79.5%	82.5%	13.1%
	SCLR	94.3%	78.6%	83.7%	11.7%
	Proposed	98.2%	72.4%	89.1%	8.7%
Playground	RA	85.7%	77.8%	81.9%	16.5%
	RB	87.8%	74.2%	83.7%	14.8%
	SC	89.9%	74.6%	84.1%	13.1%
	RAD	91.8%	73.3%	84.5%	12.5%
	SCLR	92.4%	71.8%	86.7%	10.2%
	Proposed	97.2%	59.6%	89.7%	9.4%
Airplane2	RA	86.4%	78.9%	77.1%	15.8%
	RB	87.6%	78.3%	78.6%	14.6%
	SC	88.2%	77.1%	79.4%	13.5%
	RAD	88.3%	76.6%	80.8%	12.9%
	SCLR	91.3%	75.8%	81.6%	11.9%
	Proposed	96.3%	62.4%	83.9%	8.1%
Boat	RA	85.4%	84.1%	72.4	24.5%
	RB	88.6%	78.2%	74.1%	21.2%
	SC	89.7%	76.1%	76.4%	19.4%
	RAD	91.6%	74.8%	79.2%	15.7%
	SCLR	92.7%	73.4%	82.5%	11.4%
	Proposed	95.2%	68.3%	89.7%	7.4%

4.4. Saliency Detection Comparison with Other State-of-the-Art Methods. In this section, five state-of-the-art methods including RA [30], RB [31], SC [32], RAD [33], and SCLR [34] are conducted comparison with the proposed deep multiscale fusion method. And we conduct experiments on some optical remote sensing images based on urban data, namely airplane (512×512 pixel), playground (1024×1024 pixel), boat (1024×1024 pixel), vehicle (512×512 pixel), and cloud (2048×2048 pixel). Due to the limited space, we only display the results of several remote sensing objects. The test images

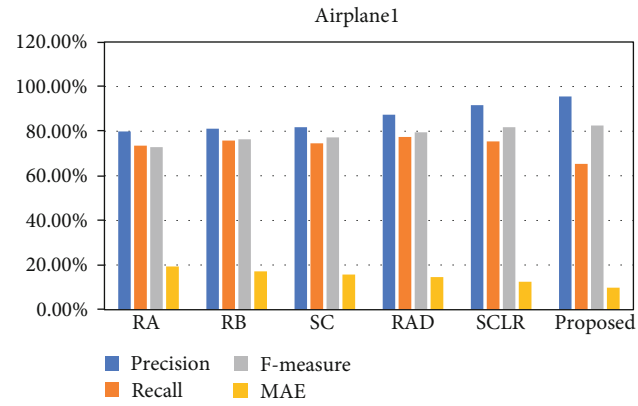


Figure 9: Airplane1 comparison with different methods.

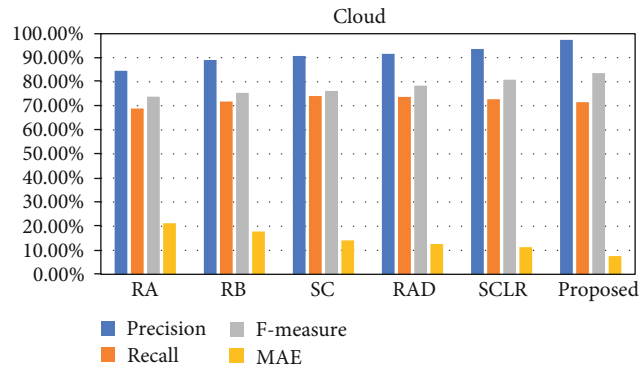


Figure 10: Cloud comparison with different methods.

along with their relevant GT maps are shown in Figure 7. Figure 8 displays the saliency results with different methods.

Figure 8 shows the comparison of saliency detection results with different methods. It can be seen that the detection effect of this algorithm is obviously better than other algorithms.

Table 1 is the F-measure result. With the change of Recall, the Precision of the method in this paper has better value and keeps a high level. However, in terms of F-measure value, our method is 7.18% higher than the second better method. Under the condition of complex background information, both the PR curve value and F-measure value of the proposed method are significantly higher than other algorithms. It fully demonstrates the advantages of the proposed algorithm in relatively complex image information. Similarly, the MAE of this proposed algorithm is lower than that of other algorithms. Figures 9–14 are the subjective evaluation results for the six objects.

We also adopt IoU (Intersection-Over-Union) to illustrate the effectiveness of the proposed method [35, 36]. The IoU is calculated as follows:

$$IoU = \frac{Area\ of\ Overlap}{Area\ of\ Union} \tag{27}$$

The greater IoU shows a better effect. The results are shown in Table 2.

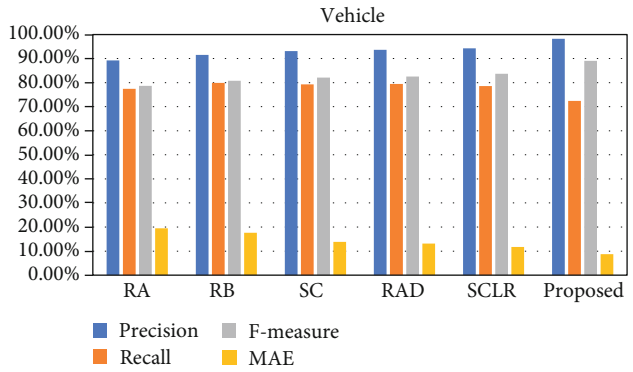


Figure 11: Vehicle comparison with different methods.

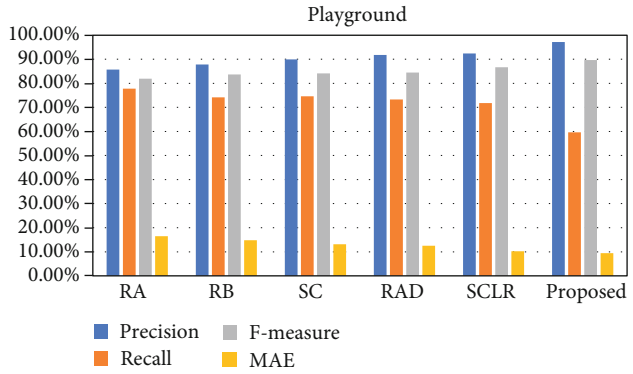


Figure 12: Playground comparison with different methods.

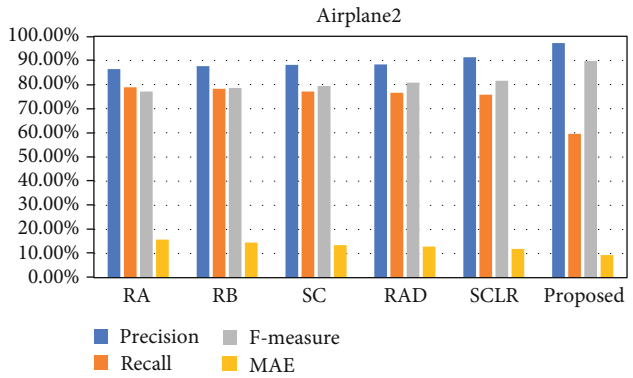


Figure 13: Airplane2 comparison with different methods.

From Table 2, we can see that our proposed method has a better saliency detection effect than other methods.

There are also apparent differences in the detection time among different algorithms. In terms of the speed of saliency detection, the proposed method is faster than other methods as given in Figure 15. Though deep learning-based algorithms need to train many samples, compared with other deep learning methods, the processing efficiency is improved by about 4%. Overall, the deep multiscale fusion method has a better effect on saliency detection for remote sensing images.

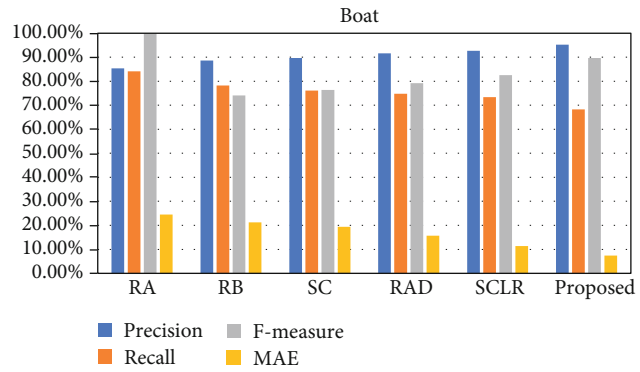


Figure 14: Boat comparison with different methods.

Table 2: IoU comparison.

Method	RA	RB	SC	RAD	SCLR	Proposed
Airplane1	69.3%	72.5%	76.7%	77.5%	79.8%	82.4%
Cloud	68.4%	72.7%	77.1%	79.2%	81.6%	83.5%
Vehicle	69.7%	73.8%	76.5%	78.5%	79.4%	81.6%
Playground	66.4%	72.9%	80.1%	81.4%	83.7%	86.4%
Airplane2	63.9%	71.3%	74.9%	76.4%	78.2%	81.9%
Boat	65.4%	70.8%	73.2%	75.9%	77.2%	79.5%

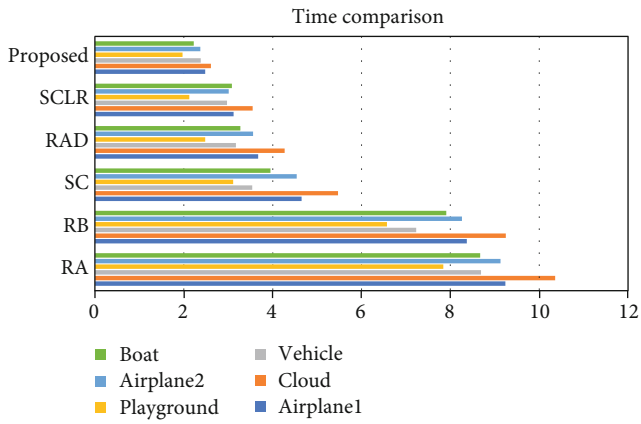


Figure 15: Time comparison with different methods.

5. Conclusions

The saliency detection algorithm based on DL can overcome the shortcomings of the traditional saliency detection algorithms. However, the detection efficiency is obviously insufficient. Therefore, we present a deep multiscale fusion method for object saliency detection in optical remote sensing images based on urban data. Through the deep feature extraction, we calculate the saliency value and use the weight cellular automata to integrate and optimize the scale saliency map. Results reveal that the proposed method can efficiently acquire the saliency detection results than other methods. In the future, some new models based on deep learning will be researched. And the new methods will be applied to practical engineerings.

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 is no conflict of interest regarding the publication of this paper.

References

[1] Q. Zhang, L. Yang, Z. Chen, and P. Li, "Incremental Deep Computation Model for Wireless Big Data Feature Learning," in *IEEE Transactions on Big Data*, no. article 1, 2019.

[2] P. Li, Z. Chen, L. T. Yang, Q. Zhang, and M. J. Deen, "Deep Convolutional Computation Model for Feature Learning on Big Data in Internet of Things," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 2, pp. 790–798, 2018.

[3] S. Yin, Y. Zhang, and K. Shahid, "Large Scale Remote Sensing Image Segmentation Based on Fuzzy Region Competition and Gaussian Mixture Model," *IEEE Access*, vol. 6, pp. 26069–26080, 2018.

[4] H. Quan, S. Feng, and B. Chen, "Two Birds With One Stone: A Unified Approach to Saliency and Co-Saliency Detection via Multi-Instance Learning," *IEEE Access*, vol. 5, pp. 23519–23531, 2017.

[5] C. Lang, J. Feng, S. Feng, J. Wang, and S. Yan, "Dual Low-Rank Pursuit: Learning Salient Features for Saliency Detection," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 27, no. 6, pp. 1190–1200, 2016.

[6] J. Zhu, Y. Qiu, R. Zhang, J. Huang, and W. Zhang, "Top-Down Saliency Detection via Contextual Pooling," *Journal of Signal Processing Systems*, vol. 74, no. 1, pp. 33–46, 2014.

[7] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 11, pp. 1254–1259, 1998.

[8] X. Hou and L. Zhang, "Saliency detection: a spectral residual approach," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–8, IEEE, Minneapolis, MN, USA, 2007.

[9] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk, "Frequency-tuned salient region detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1597–1604, IEEE, Miami, FL, USA, 2009.

[10] M. Cheng, G. Zhang, N. Mitra, X. Huang, and S. Hu, "Global contrast based salient region detection," in *Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 409–416, IEEE, Providence, RI, 2011.

[11] Q. Zhang, C. Bai, T. Yang, Z. Chen, P. Li, and H. Yu, *A Unified Smart Chinese Medicine Framework for Healthcare and Medical Services*, IEEE/ACM Transactions on Computational Biology and Bioinformatics, 2019.

[12] P. Li, Z. Chen, L. T. Yang, J. Gao, Q. Zhang, and M. J. Deen, "An Incremental Deep Convolutional Computation Model for Feature Learning on Industrial Big Data," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 3, pp. 1341–1349, 2019.

[13] G. Lee, Y. W. Tai, J. Kim et al., "ELD-Net: An Efficient Deep Learning Architecture for Accurate Saliency Detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 7, pp. 1599–1610, 2018.

[14] Q. Zhang, C. Bai, Z. Chen et al., "Deep Learning Models for Diagnosing Spleen and Stomach Diseases in Smart Chinese Medicine with Cloud Computing," in *Concurrency and Computation: Practice and Experience*, p. e5252, 2019.

[15] W. Wang, J. Chen, J. Wang, J. Chen, and Z. Gong, "Geography-Aware Inductive Matrix Completion for Personalized Point of Interest Recommendation in Smart Cities," *IEEE Internet of Things Journal*, 2019.

[16] J. Yan, M. Zhu, H. Liu, and Y. Liu, "Visual saliency detection via sparsity pursuit," *IEEE Signal Processing Letters*, vol. 17, no. 8, pp. 739–742, 2010.

[17] C. Lang, G. Liu, J. Yu, and S. Yan, "Saliency detection by multitask sparsity pursuit," *IEEE Transactions on Image Processing*, vol. 21, no. 3, pp. 1327–1338, 2012.

[18] X. Shen and Y. Wu, "A unified approach to salient object detection via low rank matrix recovery," in *Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 853–860, IEEE, Providence RI, USA, 2012.

[19] C. Dai, X. Liu, J. Lai, P. Li, and H. Chao, "Human Behavior Deep Recognition Architecture for Smart City Applications in the 5G Environment," *IEEE Network*, vol. 33, no. 5, pp. 206–211, 2019.

[20] W. Wang, J. Chen, J. Wang, J. Chen, J. Liu, and Z. Gong, "Trust-Enhanced Collaborative Filtering for Personalized Point of Interests Recommendation," *IEEE Transactions on Industrial Informatics*, p. 1, 2019.

[21] C. Dai, X. Liu, and J. Lai, "Human action recognition using two-stream attention based LSTM networks," *Applied soft computing*, vol. 86, p. 105820, 2020.

[22] S. Yang, C. Zhao, and W. Xu, "A novel salient object detection method using bag-of-features," *Acta Automatica Sinica*, vol. 42, pp. 1259–1273, 2016.

[23] H. Jiang, J. Wang, Z. Yuan, Y. Wu, N. Zheng, and S. Li, "Salient object detection: a discriminative regional feature integration approach," in *Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2083–2090, IEEE, Portland, OR, USA, 2013.

[24] Y. Li, Y. Xu, S. Ma, and H. Shi, "Saliency detection based on deep convolutional neural network," *Journal of Image and Graphics*, vol. 21, pp. 53–59, 2016.

[25] L. Wang, L. Wang, H. Lu, P. Zhang, and X. Ruan, "Saliency detection with recurrent fully convolutional networks," in *Proceedings of the Computer Vision-ECCV 2016. Lecture Notes in Computer Science, vol. 9908*, pp. 825–841, Springer, Amsterdam, Netherlands, 2016.

[26] G. Lee, Y. Tai, and J. Kim, "Deep saliency with encoded low level distance map and high level features," in *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 660–668, IEEE, LasVegas, NV, USA, 2016.

[27] P. Li, Z. Chen, L. T. Yang, J. Gao, Q. Zhang, and M. J. Deen, "An Improved Stacked Auto-Encoder for Network Traffic Flow Classification," *IEEE Network*, vol. 32, no. 6, pp. 22–27, 2018.

[28] C. Yang, J. Pu, Y. Dong, G. S. Xie, Y. Si, and Z. Liu, "Scene classification-oriented saliency detection via the modularized prescription," *The Visual Computer*, vol. 35, no. 4, pp. 473–488, 2019.

[29] A. Wang and M. Wang, "RGB-D Salient Object Detection via Minimum Barrier Distance Transform and Saliency Fusion,"

IEEE Signal Processing Letters, vol. 24, no. 5, pp. 663–667, 2017.

[30] S. Chen, X. Tan, B. Wang et al., "Reverse attention for salient object detection," in *European Conference on Computer Vision*, Springer, Cham, 2018.

[31] W. Zhu, S. Liang, Y. Wei, and J. Sun, "Saliency optimization from robust background detection," *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2814–2821, 2014.

[32] Q. Hou, M.-M. Cheng, X. Hu, A. Borji, Z. Tu, and P. Torr, "Deeply supervised salient object detection with short connections," *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5300–5309, 2017.

[33] X. Hu, L. Zhu, J. Qin, C.-W. Fu, and P.-A. Heng, "Recurrently aggregating deep features for salient object detection," *Thirty-Second AAAI Conference on Artificial Intelligence*, pp. 6943–6950, 2018.

[34] N. Liu and J. Han, "A deep spatial contextual long-term recurrent convolutional network for saliency detection," *IEEE Transactions on Image Processing*, vol. 27, no. 7, pp. 3264–3274, 2018.

[35] J. Gao, P. Li, Z. Chen, and J. Zhang, "A Survey on Deep Learning for Multimodal Data Fusion," *Neural Computation*, vol. 32, no. 5, pp. 829–864, 2020.

[36] J. Gao, P. Li, and Z. Chen, "A canonical polyadic deep convolutional computation model for big data feature learning in Internet of Things," *Future Generation Computer Systems*, vol. 99, pp. 508–516, 2019.

Hindawi
Wireless Communications and Mobile Computing
Volume 2020, Article ID 8817419, 9 pages
https://doi.org/10.1155/2020/8817419

Research Article

A New Kinect-Based Posture Recognition Method in Physical Sports Training Based on Urban Data

Dianchen He[1,2] **and Li Li**[1]

[1]*College of Sports Science, Woosuk University, Jeonju, Republic of Korea*
[2]*College of Sports Science, Shenyang Normal University, Shenyang, China*

Correspondence should be addressed to Li Li; 48241194@qq.com

Received 11 March 2020; Accepted 15 April 2020; Published 30 April 2020

Academic Editor: Qingchen Zhang

Physical data is an important aspect of urban data, which provides a guarantee for the healthy development of smart cities. Students' physical health evaluation is an important part of school physical education, and postural recognition plays a significant role in physical sports. Traditional posture recognition methods are with low accuracy and high error rate due to the influence of environmental factors. Therefore, we propose a new Kinect-based posture recognition method in a physical sports training system based on urban data. First, Kinect is used to obtain the spatial coordinates of human body joints. Then, the angle is calculated by the two-point method and the body posture library is defined. Finally, angle matching with posture library is used to analyze posture recognition. We adopt this method to automatically test the effect of physical sports training, and it can be applied to the pull-up of students' sports. The position of the crossbar is determined according to the depth sensor information, and the position of the mandible is determined by using bone tracking. The bending degree of the arm is determined through the three key joints of the arm. The distance from the jaw to the bar and the length of the arm are used to score and count the movements. Meanwhile, the user can adjust his position by playing back the action video and scoring, so as to achieve a better training effect.

1. Introduction

Urban big data is a massive amount of dynamic and static data generated from the subjects and objects including various urban facilities, organizations, and individuals, which have been collected and collated by city governments, public institutions, enterprises, and individuals using a new generation information technologies. Big data can be shared, integrated, analyzed, and mined to give people a deeper understanding of the status of urban operations and help them make more informed decisions on urban administration with a more scientific approach, thereby, optimizing the allocation of urban resources, reducing the operating costs of the urban system, and promoting the safe, efficient, green, harmonious, and intelligent development of the cities as a whole.

Nowadays, the sports hardware facilities are also perfect with the development of the smart cities [1–3]. The quality of people's life has been guaranteed and improved. The human body has a rich and variety of movements [4]. In many applications, a more comprehensive analysis for human movement is needed, such as behavior monitoring, movement analysis, and medical rehabilitation. If the human body can be identified and tracked in real time, then the posture of the human body can be identified accurately, this process can make it more convenient to observe and learn human behavior [5]. Therefore, it is necessary to find a good way to recognize human posture.

In recent years, human motion recognition based on Kinect has shown great significance in the field of medicine, and many institutions are carrying out relevant researches. Enea et al. [6] installed a Kinect on a walker to extract information about the legs for medical gait analysis. Chen et al. [7] used depth information extracted from Kinect to detect the body's joints and used a random forest classifier to classify depth image pixels into multiple parts of the body. Thang

et al. [8] adopted human anatomy marks and the human body skeleton model to obtain a depth map and estimate the posture of the human body. The geodesic distance was used to measure the distance between the different parts of the body. Xu et al. [9] used the Kinect sensor to obtain human body images and identify 3D human body posture. Yang et al. [10] utilized a Kinect device to capture the scene and estimate the body's limb posture. Moreover, some other researchers had done the related works to improve the posture recognition [11–13]. These methods only take physical characteristics into consideration; it ignores the global features, which shows a poor effect on physical sports training. Although deep learning methods have been applied to many fields, it is not mature in physical training and rehabilitation. In this paper, we focus on the study of improving the Kinect technology for posture recognition.

Many medical experts have brought Kinect to medical rehabilitation because it is cheap and useful such as using Kinect for rehabilitation treatment. The basic idea is to use depth information and skeleton tracking technology to track the body limb and determine the position of the body. At the end, it can identify the movement of the body. In reference [14], Kinect rehabilitation training could effectively enhance the quality of rehabilitation. It not only assisted patients to recover motor function but also improved their psychological quality and reduced their negative emotions. Wang et al. [15] designed a rehabilitation system for the human shoulder that required the patient to touch his or her hand to a set point. However, it was impossible to measure the specific location of the joint in real time. Rehabilitation training usually does not require patients to carry out rapid and large movements, but requires high tracking accuracy of Kinect's human skeleton. If accurate tracking of the human hand and leg skeleton can be observed, then, the movements of patients can be identified more accurately, so as to achieve better recovery effect.

Our main contributions are as follows:

(a) We propose a new Kinect-based posture recognition method in a physical sports training system based on urban data

(b) First, Kinect is used to obtain the spatial coordinates of the human body joints

(c) Then, the angle is calculated by the two-point method and the body posture library is defined

(d) Finally, angle matching with posture library is used to analyze posture recognition

(e) We adopt this method to automatically test the effect of physical sports training and it can be applied to the pull-up of students' sports. This method can measure the angle between the skeleton in real time, improve the accuracy of posture matching, and can accurately identify the human posture. The new algorithm is simple with high efficiency

The remainder of the paper is organized as follows. In Section 2, we give an outline of the Kinect imaging principle. Section 3 describes the posture recognition method in detail. The performance and robustness are evaluated in Section 4. The conclusion is drawn in Section 5.

2. Kinect Imaging Principle

Kinect emits a near-infrared light to get a depth map. Kinect actively tracks large objects regardless of the amount of light. The depth image of Kinect is captured by an infrared projector and camera. The projection and reception are overlapped. There are similar processes for transmitting, capturing, and computing visual representations.

Structural light is with specific patterns such as points, lines, and surfaces. The principle of depth image acquisition based on structured light is to project the structured light into the scene and the image sensor to obtain the pattern corresponding to the structured light [16]. Because the structure light will be changed according to the shape of the object and the depth information of each point in the image can be calculated by using the triangle principle and the obtained pattern.

The Microsoft Kinect depth concept uses the light coding technology, which is different from the traditional two-dimensional pattern projection method of structured light. Kinect's light-coding infrared transmitter emits three-dimensional depth coding. Laser speckle is the source of light coding and the result of diffuse reflection of the laser. That is to form random spots, and the spots are not the same anywhere in the space. So every time the light source is labeled and all random spots are saved. When an object is placed in this area, if a random speckle of the object's appearance is obtained, the object's position can be found. Thus, it can obtain a depth image of the scene.

3. Coordinate Transformation

Camera space refers to the 3D spatial coordinates used by the Kinect. The origin coordinate ($x = 0$, $y = 0$, and $z = 0$) is in the center of the Kinect's infrared camera. The x-axis is to the left of the Kinect irradiation direction. The y-axis is the upward direction along the Kinect irradiation direction. The z-axis is along the Kinect irradiation direction.

The coordinate system of the depth image is the origin of an infrared camera. The positive x-axis and y-axis directions are horizontal to the right and vertical downward, respectively. The z-axis is the spatial coordinate system of the camera axis direction and meets the right-hand spiral criterion, and its type is *DepthImagePoint* (x, y, and z). The point type of the bone tracking coordinate system is *SkeletonPoint* (x, y, and z). Where (x, y, and z) is the spatial coordinate system. Only the x and y values are used in the 3D coordinates of the bone joint points. The z value is related to the distance from the object to the Kinect. If the z value is smaller and the distance is closer, then, the bone image is larger. Because Kinect does not use the same camera to collect depth images and color images, because the corresponding coordinate systems are different, so they need to be transformed into coordinate systems. The KinectSensor the SDK provides *MapDepthToSkeletonPoint*, *MapSkeletonPointToDepth*, *MapDepthToColorImagePoint*, and *MapSkeletonPointToColor* to transform coordinate system.

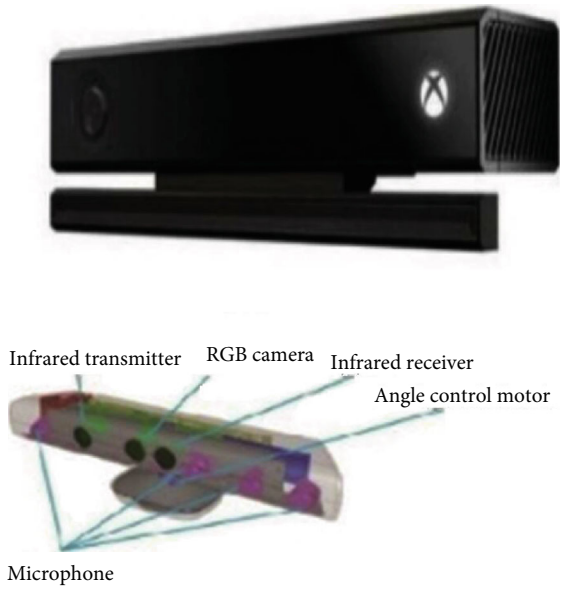


Figure 1: Kinect structure.

4. Proposed Posture Recognition Method

The Kinect camera integrates devices such as infrared transmitter, RGB camera, and infrared receiver, as shown in Figure 1. The best working range is 1.2-3.5 m, the horizontal angle of the RGB camera is 57°, the vertical angle is 43°, and the shooting frequency is 30 Hz, which can ensure high accuracy in a fast scanning moment.

The human posture recognition algorithm is mainly composed of skeleton acquisition, angle measurement, angle matching, and posture recognition. The flow of the algorithm is shown in Figure 2. Firstly, the human skeleton is obtained and the spatial coordinates of the skeleton joints are calculated. Then, it calculates the distance between the joints and the angle between the joints. Finally, the calculated angle is matched with the angle template in the posture library to evaluate the posture recognition.

Kinect can provide the three-dimensional coordinates with 20 bone joints of the human body. Figure 3 is the skeleton diagram of the human body.

5. Calculating the Distance between the Joints

In the above analysis, 20 key points of the human body have been obtained. Next, the distance between the two key joints is computed. Firstly, the scene depth information obtained by Kinect is used to calculate the actual distance between the person and the camera. In reference [17], the obtained depth value was used to calculate the actual distance d from the target to the Kinect sensor, i.e.,

$$d = K \tan\left(Hd_{\text{raw}} + L\right) - O, \tag{1}$$

where d_{raw} is the depth value. $H = 3.4 \times 10^{-4}$rad, $K = 12.35$ cm, $L = 1.18$rad, and $O = 3.7$cm. The transformation formula

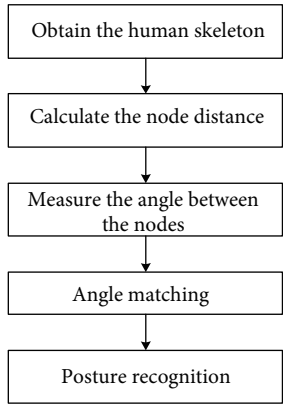


Figure 2: Flowchart of the proposed algorithm for posture recognition.

from pixel coordinate ($x_{\text{image}}, y_{\text{image}}$, and$z_{\text{image}}$) of depth image to actual coordinate ($x_{\text{real}}, y_{\text{real}}$, and$z_{\text{real}}$) is

$$\begin{cases} x_{\text{real}} = \left(x_{\text{image}} - \dfrac{w}{2}\right)\left(z_{\text{real}} + D'\right)F \cdot \dfrac{w}{h}, \\ y_{\text{real}} = \left(y_{\text{image}} - \dfrac{h}{2}\right)\left(z_{\text{real}} + D'\right)F, \\ z_{\text{real}} = d, \end{cases} \tag{2}$$

where $D' = -10$, and $F = 0.0021$ according to the abundant experiments. The resolution of Kinect is $w \times h = 640 \times 480$. $X(x_1, x_2, x_3)$ and $Y(y_1, y_2, y_3)$ are two points in the spatial coordinate system. Combining equations (1) and (2), the actual coordinate of the joint can be obtained. Then, we use the Euclidean distance to get the distance between the joints.

$$D(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + (x_3 - y_3)^2} \tag{3}$$

6. Calculating the Angle

The three-point method (three joints) is mainly used to solve the angles between human body connection points. The coordinates of the actual position of the key nodes calculated by the formula (2) are used to calculate the distance of the three key nodes related to the human body as shown in Figure 4. By using the cosine law (equation (5)), the angles between the connection points are calculated. The main disadvantage of this method in the recognition of human posture is that the instability of the closed joint has a great impact on the angle measurement during the measurement process resulting in inaccurate posture recognition. Figure 5(a) shows the angle measurement effect of the three-point method.

$$\begin{cases} a = D(B, C), \\ b = D(A, C), \\ c = D(A, B), \end{cases} \tag{4}$$

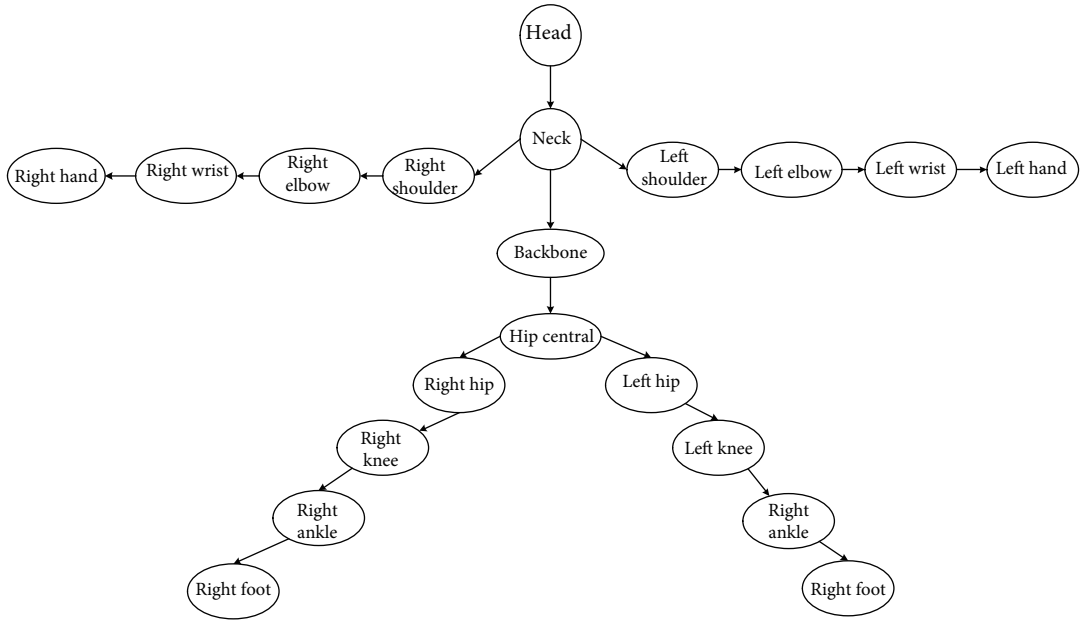


Figure 3: Joints of the human skeleton.

$$\theta = \cos^{-1}\frac{a^2 + c^2 - b^2}{2ac}. \quad (5)$$

7. Posture Definition

Equation (6) is used to define the angle condition of the joint.

$$P_A = \{P_1, P_2, \theta, \tau\}. \quad (6)$$

So it is centered on P_1. The angle between the joint P_2 and the x-axis is θ. τ is the set angle threshold value. The definition of more postures only needs to determine the angle relationship between the joints, and different thresholds can be set to meet different precision requirements. Set $\theta_i (i = 1, 2, 3, 4)$ as the angle of the joint. θ_1 = (left shoulder, left elbow), θ_2 = (left elbow, left wrist), θ_3 = (right shoulder, right elbow), and θ_4 = (right elbow, right wrist). τ is the threshold value. Then, the posture definition should satisfy the angle condition as follows: $\Delta = \{\theta_1, \theta_2, \theta_3, \theta_4, \tau\}$.

T- type (Posture starting): $\Delta T = (180, 180, 0, 0, 15)$;

Hands up: $\Delta D = (180, 90, 0, 90, 15)$;

Put down hands: $\Delta_0 = (270, 270, 270, 270, 15)$;

Raise left hand, flat right hand: $\Delta_{L_0} = (180, 90, 0, 0, 15)$;

Raise left hand, put down right hand: $\Delta_{L_1} = (180, 90, 270, 270, 15)$;

Raise right hand, flat left hand: $\Delta_{R_0} = (180, 180, 0, 90, 15)$;

Raise right hand, put down left hand: $\Delta_{R_1} = (270, 270, 0, 90, 15)$;

8. Human Body Posture Matching

In this paper, the threshold range of angle is set up when building a posture library. First, it traverses all the angles, then, it determines if the four angles are within the specified threshold. If YES, then the posture matching is successful. That is, all angles satisfy equation (7). If one of the angles is not satisfied, the matching is not successful and the match is resumed.

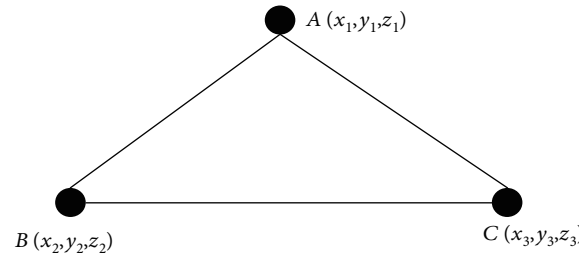


Figure 4: The calculation of joints angle.

$$\max_{i \in N} |\theta_i - a_i| < T, \quad (7)$$

where θ_i is the measured angle, a_i is the set expected angle, and T is the threshold value.

9. Experiments and Analysis

In this section, the algorithm tests seven actions in physical education class from Shenyang Normal University. It can be seen from Table 1 that when the threshold value is set to 15°. The T-type, the raised hands are recognized with 100%. The recognition rate of putting down hands is recognized with 96%.

Participants make corresponding actions according to the prompts. It starts with a T-type with multiple stages. There is one action in link 1 and two consecutive actions in link 2. The rest can be done in the same manner. Only by

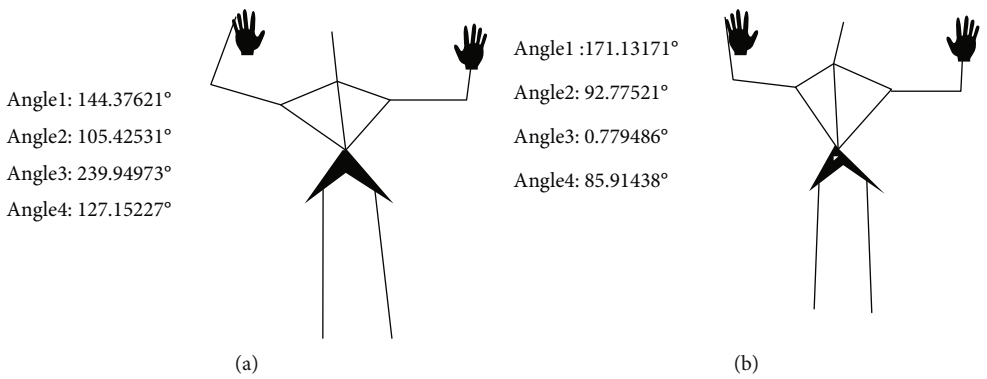


Figure 5: Results of angle measurement. (a) Three-point method. (b) Two-point method.

Table 1: Statistical analysis of posture recognition experiments.

Action	Number of people	Test times/per capita	Correct recognition number	Recognition rate (%)
T-type	10	10	100	100
Raise hands	10	10	100	100
Put down hands	10	10	96	96
Raise left hand (flat right hand)	10	10	100	100
Raise left hand (put down right hand)	10	10	100	100
Raise right hand (flat left hand)	10	10	100	100
Raise right hand (put down left hand)	10	10	100	100

completing the low link, we can enter the high link. If this link is completed, and the next link does not meet the requirements, it will start from the first link again.

In this section, we test student sports pull-ups using our method. Kinect's core technology is bone tracking, which allows the device to better capture human motion and extract deep information. Microsoft Kinect adopts the depth measurement technology of Light Coding. To obtain the spatial position of the key joints, the human body in the depth image needs to be segmented through the machine learning method. Finally, the depth image is transformed into the bone image [18–20].

The conversion process from depth image to bone image requires three steps: body recognition, body part recognition, and joint recognition. The Kinect SDK can track the 3D coordinates of the 25 bone points for 30 frames in real time. Kinect 2.0 can identify the location of six people at the same time and give complete skeletal information of two people at the same time. With each joint of the three states: TRACKED, NOT TRACKED, and INFERRED can get a complete skeleton of the human body connected by 25 nodes.

The scoring module firstly extracts the user's motion information by using the Kinect module and then extracts the angle features and position information of the data [21, 22]. Since there is no time limit during the pull-up test, the score of each movement is calculated by presetting the scoring criteria and rules to judge whether the user's arm is straight and the position relationship between the jaw and the bar. There are two indicators to evaluate the pull-up, namely, the distance between the lower jaw and the bar when the body is at the highest point and the bending angle of the arm at the end of the movement.

By calculating the distance difference between the height (hi_i) of the mandible and the height of the actual crossbar ($\Delta\mathrm{hi}_i$) when the human body is at the highest point, we can calculate the score $A\mathrm{score}_i$, namely,

$$A\mathrm{score}_i = \Delta\mathrm{hi}_i \times \delta_i, \tag{8}$$

where δ_i is the variable used to determine the threshold of scoring within each distance interval. It is similar to calculate the bending angle of the arm and the distance between the lower jaw and the bar. The score $B\mathrm{score}_i$ is set by measuring the angle difference $\Delta\mathrm{Angle}_i$ between the bending angle of the i-th elbow joint and the angle at full extension, and the angle of the left and right elbow is averaged to set the score $B\mathrm{score}_i$. The relation between the two angles can be expressed by the function $f(x)$, namely,

$$B\mathrm{score}_i = f(\Delta\mathrm{Angle}_i). \tag{9}$$

Since in the actual test, it is impossible to ask everyone to fully extend their arms, so we let the function $f(x)$ be a segmented function to set a certain threshold range. In this system

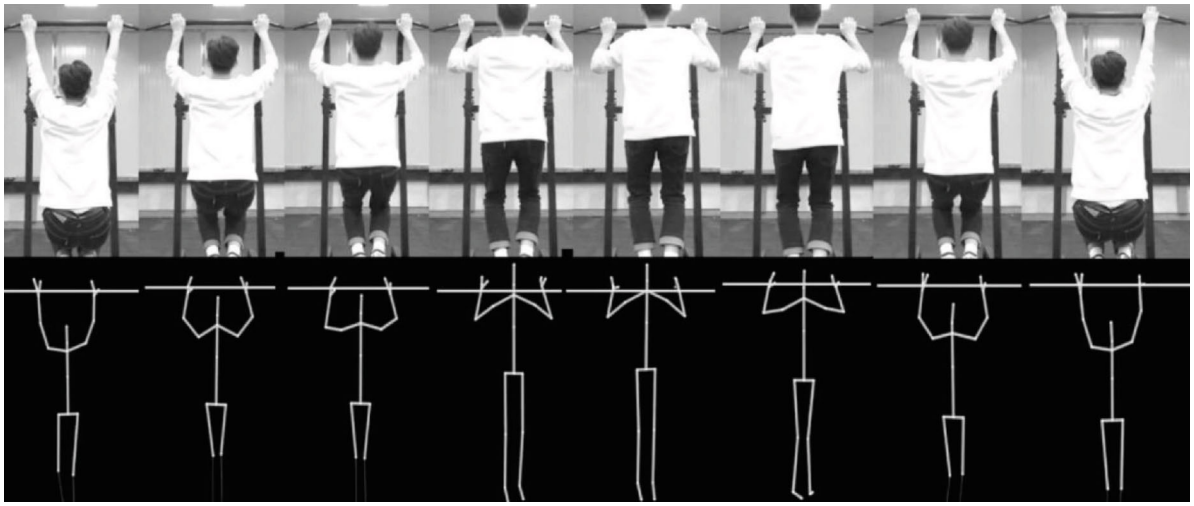


Figure 6: Lower crossing the jaw with arms straight.

$$f(x) = \begin{cases} 10, x \in [0, 20), \\ 5, x \in [20, 40), \\ 0, x \in [40, 90). \end{cases} \quad (10)$$

According to the national physical test standards, it is necessary to take the distance between the lower jaw and the crossbar and the straightening degree of the arm into comprehensive consideration to score the pull-up. Therefore, the weighted value of both should be taken to obtain the final score, i.e.,

$$\text{Score}_i = \alpha A\text{score}_i + \beta B\text{score}_i, \quad (11)$$

where α and β are the weight coefficients of $A\text{score}_i$ and $B\text{score}_i$, respectively. The weights of the two indexes can be changed by selecting different coefficients. In this system, $\alpha = \beta = 1$.

This project counts and scores according to the position relationship between the user's jaw and the crossbar and the bending angle of the arm. When the lower jaw crossing the bar is detected and the arm bending angle is within a certain threshold range then a count is made and the corresponding score is given. Other cases are not counted, and the corresponding score will be given. The following observations provide the counting and scoring results in several situations (The full score is 10).

The system detects that when the user's body is at the highest point, the lower jaw crosses the crossbar, the bending angle of the left arm is 172°, and the bending angle of the right arm is 163°. It counts once and determines the scoring interval according to the bending angle of the arm. We synthesize the $A\text{score}_i$ and $B\text{score}_i$, it obtains ten points as shown in Figure 6.

The system detects that when the user's body is at the highest point, the lower jaw crosses the crossbar, the bending angle of the left arm is 151°, and the bending angle of the right arm is 148°. It does not count and determines the scoring interval according to the bending angle of the arm. We synthesize the $A\text{score}_i$ and $B\text{score}_i$, it obtains five points as shown in Figure 7.

The system detects that when the user's body is at the highest point, the lower jaw does not cross the crossbar, the bending angle of the left arm is 171°, and the bending angle of the right arm is 160°. It does not count and determines the scoring interval according to the bending angle of the arm. We synthesize the $A\text{score}_i$ and $B\text{score}_i$, it obtains six points as shown in Figure 8.

The system detects that when the user's body is at the highest point, the lower jaw does not cross the crossbar, the bending angle of the left arm is 158°, and the bending angle of the right arm is 161°. It does not count and determines the scoring interval according to the bending angle of the arm. We synthesize the $A\text{score}_i$ and $B\text{score}_i$, it obtains two points as shown in Figure 9.

Under natural conditions in the laboratory, 100 groups of experimental data are selected for the experiment. In this paper, the accuracy and real time of body recognition are tested, respectively. The experimental results are shown in Tables 2 and 3. It can be seen from the table that the recognition accuracy of the proposed method in this paper is over 88%.

This proposed method is compared with the other three body recognition methods containing DTW [23], IKS [24], and ConvNets [25]. Indicators include accuracy and time. Accuracy refers to the proportion of the correct sample in the total test samples.

10. Conclusions

This algorithm is developed by combining Microsoft Visual Studio 2010 with Microsoft Kinect SDK 1.7. The experiment shows that this method can measure the angle between the skeleton in real time and identify the posture of the human body accurately. The algorithm is simple and accurate. Moreover, different angle ranges can be set according to the requirements of different postures, so the reusability is

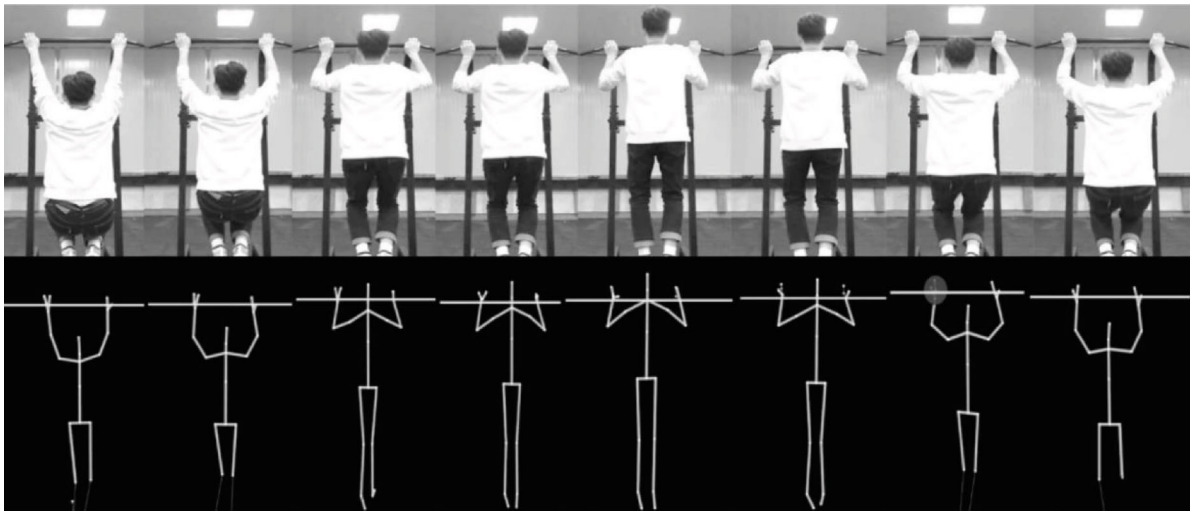


Figure 7: Lower crossing the jaw with arms not straight.

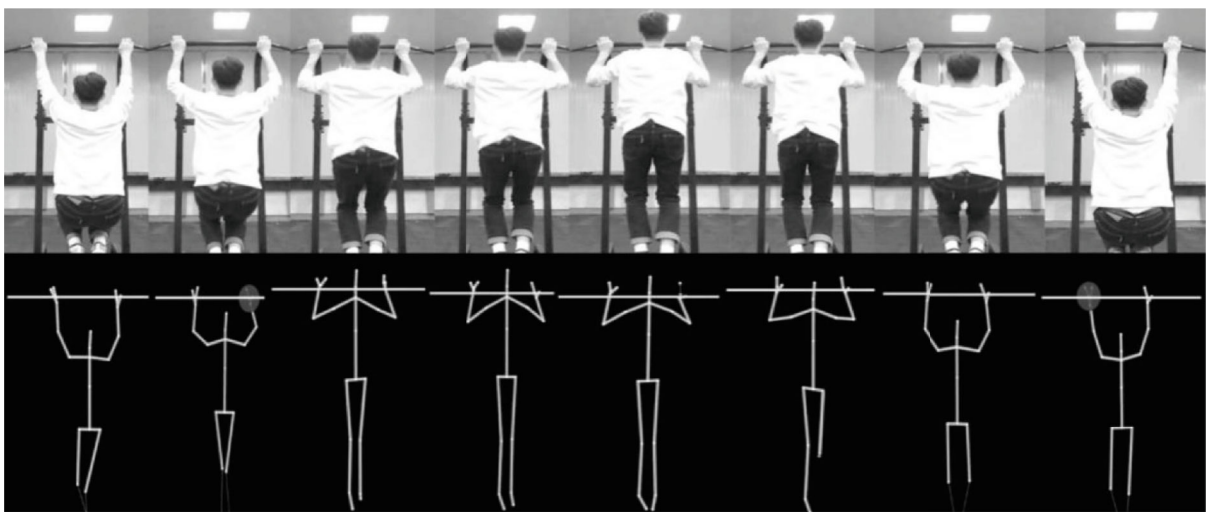


Figure 8: Lower not crossing the jaw with arms straight.

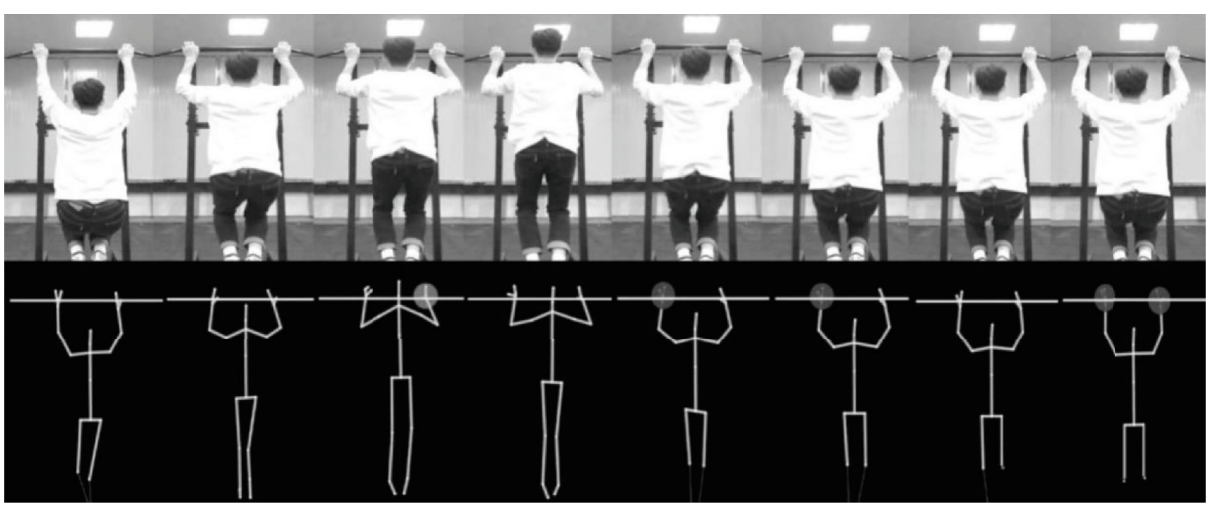


Figure 9: Lower not crossing the jaw with arms not straight.

Table 2: Body recognition results.

Posture	Recognition number			
	Wave	Bow	Walk	Squat
Wave	98	92	96	98
Bow	—	90	—	—
Walk	—	—	89	—
Squat	—	—	—	97

Table 3: Results comparison with different methods.

Method	DTW	IKS	ConvNets	Proposed
Accuracy (%)	85.4	89.6	92.7	94.6
Time (s)	0.208	0.124	0.088	0.032

strong. Although the Kinect sensor can obtain the depth information of the human body and calculate the spatial position of the human body, it is not accurate enough to identify such problems as the coincidence of the joints. Therefore, while paying attention to the development of human behavior analysis, we should study problems such as skeleton correction to further improve the accuracy of skeleton. In the future, we will adopt deep learning and artificial intelligence method to perfect the quality of physical for all national persons.

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 is no conflict of interest regarding the publication of this paper.

References

[1] P. Li, Z. Chen, L. T. Yang, Q. Zhang, and M. J. Deen, "Deep convolutional computation model for feature learning on big data in internet of things," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 2, pp. 790–798, 2018.

[2] E. Gomede, F. Gaffo, G. Briganó, R. de Barros, and L. Mendes, "Application of computational intelligence to improve education in smart cities," *Sensors*, vol. 18, no. 1, p. 267, 2018.

[3] Q. Zhang, L. T. Yang, Z. Chen, and P. Li, "Incremental deep computation model for wireless big data feature learning," *IEEE Transactions on Big Data*, 2019.

[4] P. Li, Z. Chen, L. T. Yang, J. Gao, Q. Zhang, and M. J. Deen, "An incremental deep convolutional computation model for feature learning on industrial big data," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 3, pp. 1341–1349, 2019.

[5] G. Jim, "Inside the race to hack the Kinect," *New Scientist*, vol. 208, no. 2789, pp. 22-23, 2010.

[6] E. Cippitelli, S. Gasparrini, S. Spinsante, and E. Gambi, "Kinect as a tool for gait analysis: validation of a real-time joint extraction algorithm working in side view," *Sensors*, vol. 15, no. 1, pp. 1417–1434, 2015.

[7] X. Chen, Z. Cao, Y. Xiao, and Z. Fang, "Hand pose estimation in depth image using CNN and random forest," in *Tenth International Symposium on Multispectral Image Processing and Pattern Recognition (MIPPR2017)*, Xiangyang, China, 2017.

[8] N. D. Thang, T. S. Kim, Y. K. Lee, and S. Lee, "Estimation of 3-D human body posture via co-registration of 3-D human model and sequential stereo information," *Applied Intelligence*, vol. 35, no. 2, pp. 163–177, 2011.

[9] D. Xu, X. Xiao, X. Wang, and J. Wang, "Human action recognition based on Kinect and PSO-SVM by representing 3D skeletons as points in lie group," in *2016 International Conference on Audio, Language and Image Processing (ICALIP)*, pp. 568–573, Shanghai, China, July 2016.

[10] Y. Yang, F. Pu, Y. Li, S. Li, Y. Fan, and D. Li, "Reliability and validity of Kinect RGB-D sensor for assessing standing balance," *IEEE Sensors Journal*, vol. 14, no. 5, pp. 1633–1638, 2014.

[11] W. Wang, J. Chen, J. Wang, J. Chen, and Z. Gong, "Geography-aware inductive matrix completion for personalized point of interest recommendation in smart cities," *IEEE Internet of Things Journal*, 2019.

[12] W. Wang, J. Chen, J. Wang, J. Chen, J. Liu, and Z. Gong, "Trust-enhanced collaborative filtering for personalized point of interests recommendation," *IEEE Transactions on Industrial Informatics*, p. 1, 2019.

[13] Q. Zhang, C. Bai, L. T. Yang, Z. Chen, P. Li, and H. Yu, "A unified smart Chinese medicine framework for healthcare and medical services," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 2019.

[14] T. Hoang, H. Dang, and V. Nguyen, "Kinect-based virtual training system for rehabilitation," in *2017 International Conference on System Science and Engineering (ICSSE)*, pp. 53–56, Ho Chi Minh City, Vietnam, July 2017.

[15] Q. Wang, P. Markopoulos, B. Yu, W. Chen, and A. Timmermans, "Interactive wearable systems for upper body rehabilitation: a systematic review," *Journal of Neuroengineering and Rehabilitation*, vol. 14, no. 1, p. 20, 2017.

[16] Q. Zhang, C. Bai, Z. Chen et al., "Deep learning models for diagnosing spleen and stomach diseases in smart Chinese medicine with cloud computing," *Concurrency and Computation: Practice and Experience*, 2019.

[17] S. Samoil and S. N. Yanushkevich, "Depth assisted palm region extraction using the Kinect v2 sensor," in *2015 Sixth International Conference on Emerging Security Technologies (EST)*, pp. 74–79, Braunschweig, Germany, September 2015.

[18] L. Xie and H. J. Liao, "A posture recognition method based on skeletal node and geometric relation using Kinect," *Applied Mechanics and Materials*, vol. 543-547, pp. 2879–2883, 2014.

[19] B. Li, B. Bai, C. Han, H. Long, and L. Zhao, "Novel hybrid method for human posture recognition based on Kinect V2," in *Computer Vision. CCCV 2017. Communications in Computer and Information Science, vol 771*, Springer, Singapore, 2017.

[20] Z. Zhang, Y. Liu, A. Li, and M. Wang, "A novel method for user-defined human posture recognition using Kinect," in *2014 7th International Congress on Image and Signal Processing*, pp. 736–740, Dalian, China, October 2014.

[21] J. Gao, P. Li, Z. Chen, and J. Zhang, "A survey on deep learning for multimodal data fusion," *Neural Computation*, vol. 32, no. 5, pp. 829–864, 2020.

[22] J. Gao, P. Li, and Z. Chen, "A canonical polyadic deep convolutional computation model for big data feature learning in internet of things," *Future Generation Computer Systems*, vol. 99, article S0167739X19307393, pp. 508–516, 2019.

[23] N. Li, Y. Dai, R. Wang, and Y. Shao, "Study on action recognition based on kinect and its application in rehabilitation training," in *2015 IEEE Fifth International Conference on Big Data and Cloud Computing*, pp. 265–269, Dalian, China, August 2015.

[24] M. Eltoukhy, J. Oh, C. Kuenze, and J. Signorile, "Improved kinect-based spatiotemporal and kinematic treadmill gait assessment," *Gait & Posture*, vol. 51, pp. 77–83, 2017.

[25] S. Neili, S. Gazzah, M. A. El Yacoubi, and N. E. Ben Amara, "Human posture recognition approach based on ConvNets and SVM classifier," in *2017 International Conference on Advanced Technologies for Signal and Image Processing (ATSIP)*, pp. 1–6, Fez, Morocco, May 2017.

Hindawi
Wireless Communications and Mobile Computing
Volume 2020, Article ID 3714879, 13 pages
https://doi.org/10.1155/2020/3714879

Research Article

Research on Intelligent Guidance Optimal Path of Shared Car Charging in the IOT Environment

Yuefang Sun,[1] Kangkang Jin,[2] Zhaozhuang Guo,[1] Chen Zhang,[2] and Hao Wang[2]

[1]*School of Liberal Arts and Sciences, North China Institute of Aerospace Engineering, Langfang, Hebei, China*
[2]*Engineering Computing and Simulation Innovation Lab, North China University of Science and Technology, Tangshan, Hebei, China*

Correspondence should be addressed to Kangkang Jin; jinkangkang@stu.ncst.edu.cn

Received 15 December 2019; Revised 18 February 2020; Accepted 27 March 2020; Published 24 April 2020

Academic Editor: Bingxian Lu

In recent years, with the improvement of Internet of Things (IOT) technology, a "shared" service concept has appeared in people's life. In the limited available resources, it is of great value to study the optimal path of charging pile selection for shared cars. With the help of Internet of Things technology and through analyzing the collected data, this paper introduces three path optimization methods, the Dijkstra algorithm, heuristic algorithm A^*, and improved particle swarm optimization (PSO) algorithm; establishes relevant convergence conditions; and takes the actual path cost as the criterion to judge the optimal path. In addition, this paper studies the optimal path from the shared car to the charging pile. Through the simulation experiment, the results show that compared with the traditional optimal path algorithm, the improved particle swarm optimization algorithm has strong parallelism and better search effect for optimal path selection in the case of large number of traffic path nodes and complex paths, which fully reflects the performance advantage of the algorithm.

1. Introduction

As a long-term problem, air pollution has not been fully improved, and automobile exhaust is an important cause of air pollution. Based on this, with the support of IOT technology, it is of great significance to study the optimal path from shared cars to charging piles.

In order to distribute traffic network reasonably and improve traffic operation efficiency, the United States began to study the Intelligent Transportation IOT in the 1960s. In 1991, the United States Congress passed the Integrated Land Transportation Efficiency Act (ISTEA) [1]. The simulation results showed that once the ISTRA scheme was adopted, the capacity of the US road network would be increased by 20%~30% [2]. The intelligent transportation system (ITS) was first used in the parking lot vehicle guidance system. The earliest foreign parking guidance system appeared in Aachen, Germany, in 1971. Then it was first widely used in Britain and other European countries, Japan, and other developed countries [3, 4]. In China, Beijing [5, 6], Shanghai, Guangzhou [7], and other cities take the lead in introducing the parking guidance system, effectively alleviating traffic congestion. More and more cities have begun to use the parking guidance system [8]. More and more cities are applying parking guidance systems. The rise of domestic shared electric vehicles, combined with the development of domestic and international urban parking guidance systems, has pointed out the development goals and directions for the study of the optimal path from shared electric cars to charging piles.

The core of DRGS is the path guidance algorithm, but the current path guidance is mostly based on static information. In order to obtain the shortest path in the former N trips and reduce the complexity to the polynomial order, Wei et al. [10] proposed the generalized shortest path by using the improved Dijkstra algorithm. Xiaolei [11] combined the Dijkstra algorithm with genetic algorithm and studied the method of obtaining the optimal path quickly without the restriction of the network condition. In 2009, in order to avoid a single algorithm falling into the local optimal defect and improving the optimization efficiency, van der Zijpp and Catalano [12] proposed a shortest path

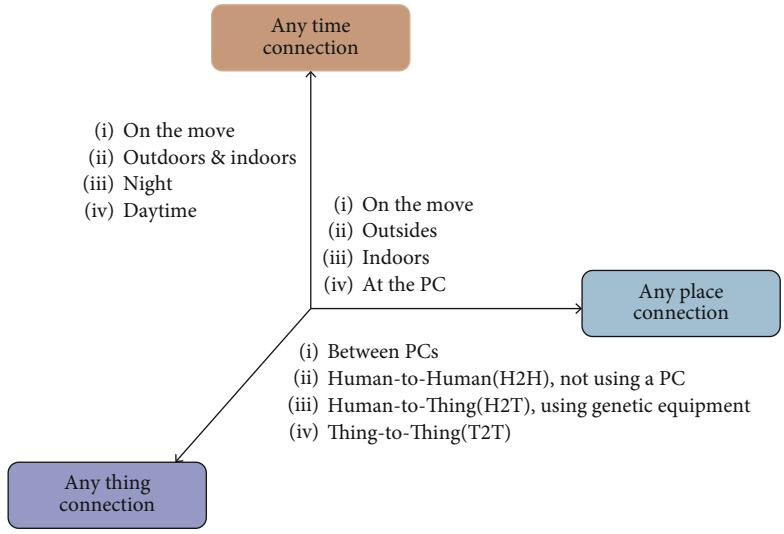


Figure 1: Connection dimension in the IOT.

optimization algorithm based on simulated annealing algorithm and genetic algorithm. The traditional path guidance method was only limited to the optimal path for one individual to find the starting point and the end point, which could not take into account the overall road impact factors and realize the path dynamic induction [13]. Therefore, it was necessary to design an optimal path in the sharing of electric car charging, so as to improve the utilization ratio of the shared car and alleviate the traffic pressure of traffic congestion [14].

Dynamic route guidance system (DRGS) was first studied in the United States, the United Kingdom, Japan, and other developed countries. Other path guidance systems based on DRGS have achieved some results [9]. The core of DRGS is path guidance algorithm, but the current path guidance is mostly based on static information. Wei et al. [10] proposed the generalized shortest path, using the improved Dijkstra algorithm to find the shortest path in the front N routes and reduce the complexity to a polynomial order. Xiaolei [11] used the method of combining the Dijkstra algorithm with genetic algorithm to find the optimal path without the restriction of road network conditions. van der Zijpp and Catalano [12] proposed a shortest path optimization algorithm based on simulated annealing algorithm and genetic algorithm in 2009, which avoided the defect that a single algorithm was easy to fall into local optimum and improved the optimization efficiency. The traditional route guidance method was limited to an individual to find the optimal path of the starting point and the end point and could not take into account the overall impact of road factors to achieve dynamic route guidance [13]. Therefore, it is necessary to design an optimal path when sharing electric car needs to charge, which can not only improve the utilization rate of the shared car but also alleviate traffic congestion and other traffic pressures [14].

The above researches on parking guidance system and urban route guidance based on IOT have certain limitations, and there are few researches on the path planning of guiding shared cars to charging piles at home and abroad, because of the complexity of urban traffic network, the uncertain situation of vehicles in road network, the real time of vehicles, and roads situation. When we use real-time updated traffic data to search the path, higher requirements are put forward for the optimal path selection.

2. IOT Technology

IOT is an important part of the new generation of information technology, which is called the third wave of development of the world information industry after the computer and the Internet. The early IOT was put forward based on the background of a logistics system, and the radio frequency identification (RFID) technology was used as the substitute of bar code identification to realize the intelligent management of the logistics system. The concept of "Internet of Things" was formally proposed at the 2005 ITU World Summit on the Information Society in Tunis. The report noted that the IOT was a new dimension of communication in the world of information and communication technology (shown in Figure 1), which extended any time, any place, and connection any person to connect any object. The connection of all things forms the IOT [15].

The introduction of the ITU report makes the IOT more and more important in the world. Workers in all related fields have studied the IOT from different aspects. However, different workers have different points of view on the IOT, and the description of the IOT is different. Therefore, the concept of IOT is not reaching a consensus.

Here, we only introduce a concept, which is widely recognized. The definition of "Internet of Things Technology" is [16] a network technology through the RFID, infrared sensors, the global positioning system (GPS), laser scanners, and other information-sensing equipment. In order to achieve intelligent identification, positioning, tracking, monitoring, and management, any items connected with the Internet for information exchange and communication can

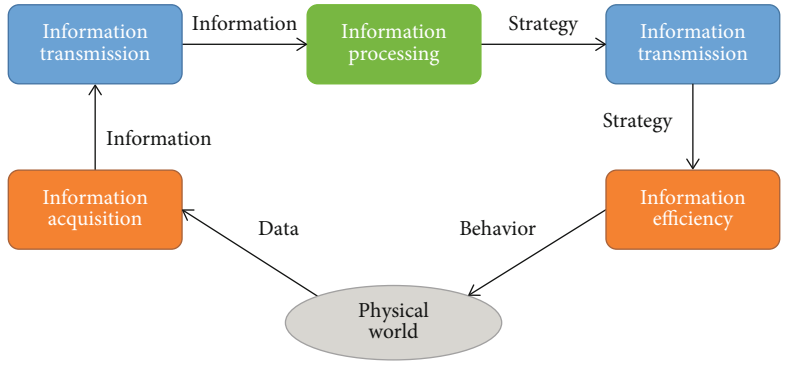


Figure 2: The information function model of the IOT.

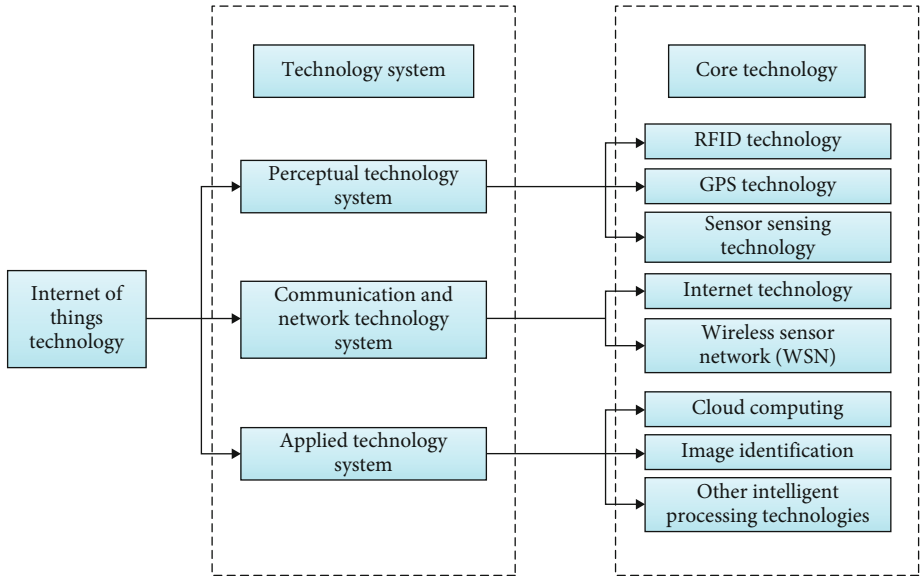


Figure 3: Technology system of IOT.

be called Internet technology [17] according to the agreement. Its core and foundation is still "Internet technology"; it is a kind of network technology that extends and expands on the basis of Internet technology. The clients expand and extend to any goods and objects, exchange information, and communicate with each other [18].

For a clearer description of the key links in the IOT, according to the perspective of information science, the information function model of the IOT is abstracted around the flow process of information, as shown in Figure 2.

2.1. Three Key Technologies in IOT Applications. There are four levels of general IOT architecture: the perceptive layer, the network layer, the data intelligent processing layer, and the application layer. The perceptive layer is like the human skin and the senses, which is used to identify objects and collect information. The network layer is like a human neural network that transmits information to the brain for processing. The data intelligent processing layer is the core technology, which is realized data-centric. The application layer includes the application support layer and various specific object-networking applications. The IOT includes three key technologies in the application, and its technical architecture is shown in Figure 3.

(1) Sensor technology: the collected signal is analog signal, which can be processed by computer only if the analog signals are converted into digital signal

(2) RFID tag technology: the integration of radio frequency identification and embedded technology, which has a broad application prospects in automatic identification, goods logistics management

(3) Embedded system technology: a composite technologies, which integrates computer hardware and software technology, sensor technology, integrated circuit technology, electronic application technology

2.2. Features of the IOT. The IOT is based on the Internet, which not only has many characteristics of the Internet but also extends its own more prominent features. The basic characteristics of the network of things are perception, reliability, and intelligence, which are mainly reflected in the following three aspects:

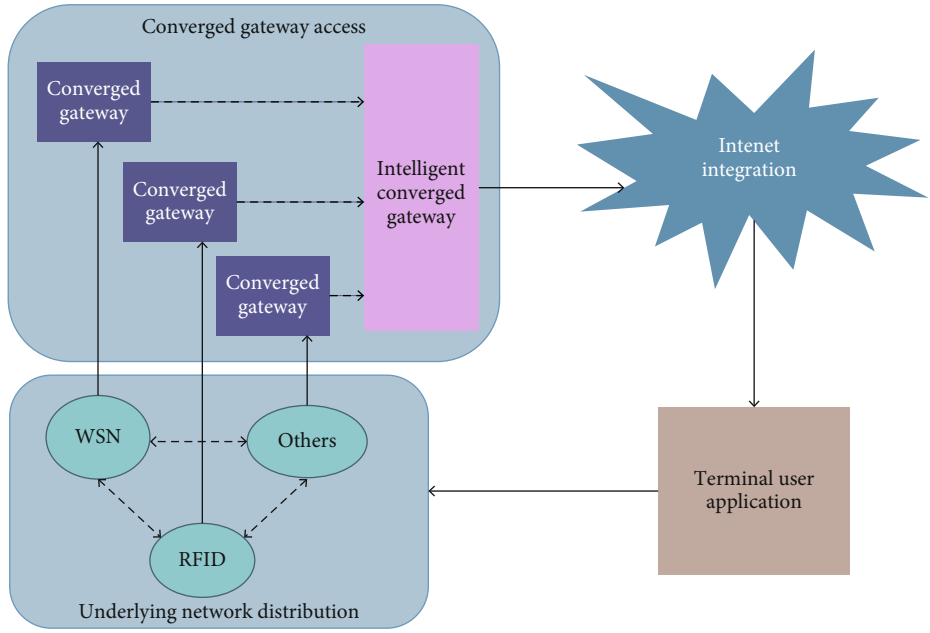


Figure 4: The architecture of IOT.

2.2.1. Perception. This is mainly embodied in the identification and communication of the IOT. In the entire IOT, there is a large number of sensors. Each sensor is a source of information. The sensor receives the information and identifies the information through a specific protocol. Today, with the rapid development of our science and technology information, information data is constantly being updated, so that the sensors must constantly perceive the data and update the information.

2.2.2. Reliability. Reliability is mainly reflected in the reliability of information transmission, through the Cloud computing; fuzzy recognition technology, corresponding to the change of various things; reliable transmission; and sending of instructions in time.

2.2.3. Intelligence. The IOT is the communication between things and things, without the interference of human beings. In other words, this is not through the authorization of the human, but the person is only a kind of consciousness control. The IOT can transmit, analyze, and process information through various technologies. Finally, intelligent decisions can be made through intelligent control.

Thus, comprehensive perception, reliable transmission, and intelligent processing are the three important features of the IOT. In the "smart planet," the importance of the IOT has been involved, which embodies the profound perception, comprehensive interconnection, and the powerful intelligence of the IOT.

2.3. The System Architecture of the IOT. Here, we introduce one of the architectures of the IOT, which includes the underlying network distribution, converged gateway access, Internet convergence, and terminal-user application. The structure is showed in Figure 4.

In Figure 4, a large number of the underlying network systems are selectively distributed in the physical space, according to their respective characteristics, which forms the network distribution. The underlying network collects the exchange information of barter and transfers it to intelligent converged gateway via RFID, WSNs, WLAN, and so on [19]. Through the intelligent convergence gateway, the network is connected to the network fusion system. Finally, the network approach, which includes radio and television network, Internet, and telecommunication network, is used to reach the terminal user application system. At the same time, the terminal user can influence the underlying network to different applications through subjective behavior, so as to realize material association information interaction between human and things, things to things, and things to human. The underlying network distribution including heterogeneous networks, such as WSNs, RFID system, and WLAN, allows the system to identify the properties of the objects and collects and captures the information, through the information interaction of heterogeneous network, which can realize the object's perception of the external physical environment. From the view of network function, the underlying network should have the dual functions of information acquisition and routing, while the underlying heterogeneous networks need to collaborate with each other to accomplish specific tasks.

Converged gateway access mainly completes the information collected from the underlying network smoothly access to the transmission network. The access technology includes the wired access methods and the wireless access methods. The intelligent convergence gateway usually has the powerful storage, processing, and the communication capabilities. The key is combined with the underlying network and smoothly access to the integrated network upward. The optimized network system includes the radio

and television, Internet, and telecommunication network, which mainly completes the long-distance transmission of the information [20]. The terminal user application system mainly completes the information related service discovery and the application function.

3. Question-Making and Analysis

3.1. Propose the Problem of Optimizing Guide Path. With the development of urbanization and industrialization, the problem of urban traffic congestion is becoming more and more serious. The role of shared cars in relieving traffic congestion pressure is more and more obvious. But as the scope of the application of shared cars increases, how to guide it to find the optimal charging pile routing problem has plagued the managers of electric vehicle charging. With the development of the IOT, the traffic monitoring and control equipment in the city can do the illegal snapping, the flow control, and so on. The Traffic Management Center can use the sensor or monitor the video processing technology, so that the real-time monitoring of each vehicle. The path guidance becomes possible. The real-time traffic data of the city is updated rapidly, and in front of the dynamic traffic flow information, a higher requirement is put forward on how to find the optimal route of charging pile quickly.

3.2. Analysis of Path Influencing Factors. The optimal path problem in real-time traffic is not equal to the shortest path research in theory. Combined with traffic laws and regulations, one-way road traffic, turn restrictions, and so on, we guide the optimal path of the electric vehicle to charge pile by taking the shortest driving distance of electric vehicle to the charging pile, the least driving time and the lowest power consumption and other optimization criteria as the goal. The related influence factors mainly come from several aspects.

(1) Real-time traffic conditions

In real traffic, if multiple electric cars choose the same optimal path, exceeding the traffic capacity of the road, it will definitely result in the worst-case scenario. Therefore, when choosing the optimal path, according to the traffic flow on the channel, driving speed and other factors on the road congestion degree quantification, which can be expressed as unimpeded, general congestion, congestion, and serious congestion, the real-time traffic information collected through the vehicle traveling recorder, the intersection electronic surveillance, and so on can be used to analyze and quantify the traffic conditions.

(2) Intersection delay time

In the theoretical research, the delay time between nodes are ignored when the optimization algorithm is used to study the optimal path, and the delay time of the subjective judgment section has no effect on the time optimization. In actual traffic, traffic signal (traffic light) is the most obvious time delay for intersection and has great influence on optimization time.

(3) The impact of road one-way street

In the actual road, the traffic management department will set up a one-way street on some specific roads, while the traditional path optimization methods are both two-way alley at the default node. This optimization mode does not conform to the actual traffic path.

3.3. The Intelligent Transportation System Model of IOT. According to the system structure of the IOT and the functional requirements of the intelligent transportation system, the model of intelligent traffic system based on IOT is proposed. The intelligent traffic system is layered to realize the analysis and description of the function and properties of the model.

In the perceptual layer, the application of intelligent traffic based on the IOT is built on vehicles, personnel, environment, and other related infrastructure. The information collection is completed by the beacon, the roadmap sensor, and other basic sensing tools. In the network layer, the intelligent traffic system based on IOT emphasizes the sharing and utilization of information, through satellite, computer network, wireless mobile communication, and other related technologies to complete the transmission of traffic information in the network layer. In addition, the storage and processing of information can be provided different information according to different service. The application layer, which can realize the service of adjusting signal lights according to traffic flow and providing traffic information in real time, and carry out different traffic management and control according to different service fields, feedbacks the realization in time and constantly expands more intelligent service functions.

In a word, the intelligent traffic system based on the IOT is a network system that realizes the collection, transmission, control, and application of traffic information. With the development of the technology of IOT, the service level of the intelligent transportation system will continue to expand and serve the traveler better.

4. Research Scheme of Optimal Path in Urban Network

The urban traffic route map is abstracted as a direction graph. The electric car parking point is compared to the starting point, and the destination is compared to the endpoint. According to the conditions such as path length, traffic signal, and two-way traffic and other conditions, the starting point, the end point, and the intersection are defined as the nodes in the traffic network, and the road between the two nodes is defined as the path between two points. Give weight to these information to determine the degree of traffic congestion, then we find the best path to guide the electric vehicle to the charging pile.

The traditional optimal path selection can be divided into the following kinds of problems according to the node, the path characteristic, and the weight value assigned by the path, as shown in Figure 5.

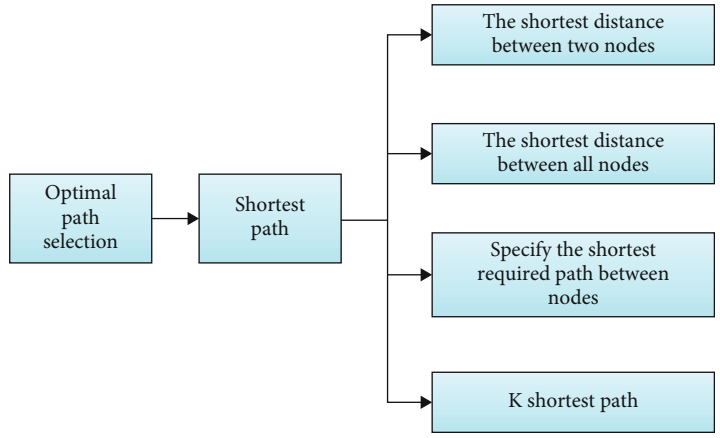


Figure 5: The classification of the optimal path.

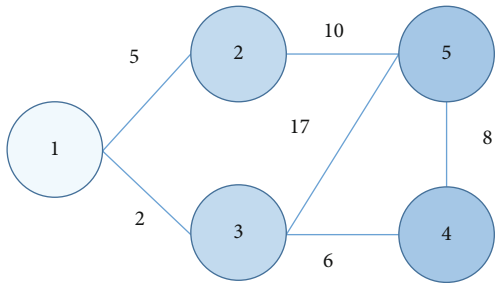


Figure 6: Network structure.

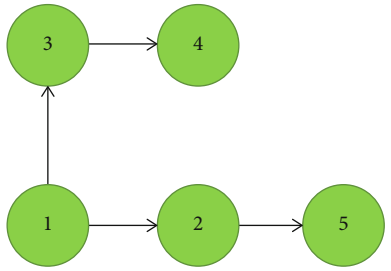


Figure 7: The tree diagram.

4.1. The Dijkstra Algorithm [21]. The Dijkstra algorithm is an algorithm that solves the weights of all paths in network graph greater than zero. At present, the research of this algorithm is more mature, and its biggest disadvantage is that it ignores the characteristic of the independent individuals in the network topology model, so that the algorithm runs for a long time. There is a certain effect on the selection of the optimal path of local small samples; however, its high time complexity has great limitation in the selection of the optimal path in a large scale.

Basic implementation steps:

The Dijkstra algorithm is an algorithm that solves the weights of all paths greater than zero in network graph and calculates the optimal path in the network which is the shortest path. Taking Figure 6 as an example, we briefly introduce the steps to solve the shortest path. Figure 7 is a tree structure of the network structure.

Table 1: Algorithm implementation process.

Order	V	S	i	d_i	Optimal path
1	{2,3,4,5}	{1}	3	2	$1 \longrightarrow 3$
2	{2,4,5}	{1,3}	2	5	$1 \longrightarrow 2$
3	{4,5}	{1,2,3}	4	8	$1 \longrightarrow 3 \longrightarrow 4$
4	{5}	{1,2,3,4}	5	15	$1 \longrightarrow 2 \longrightarrow 5$

(1) Set V: store the optimal path node not found; set S: store the shortest path node found; i represents the node in the set V which is the shortest distance from the set S; d_i represents the distance of the shortest path. (2) Store point 1 in set S; store the points 2, 3, 4, and 5 in set V. (3) Find the nearest point in set V, that is, point 3; move point 3 from V to S, and get $d3 = 2$. (4) Iterate to guide all points to set S. At this point, the shortest distance between point 1 and other points can be obtained.

The implementation process of the algorithm is as shown in Table 1:

4.2. Heuristic AlgorithmA^*. The shortest path problem has many meanings. In the heuristic algorithm, the optimal path selection refers to the function $h(n)$ defined on the node of a search tree, which is used to evaluate the optimal selection of the path from this node to the target node. Heuristic is a searching algorithms with the most abundant resource data. A^* is a classical heuristic search algorithm, which can effectively find the optimal path in the static road network. The formula (1) is the basis for selecting the lowest cost node:

$$f(n) = g(n) + h(n), \tag{1}$$

where $f(n)$ is the objective function of the optimal path from the starting node to the current node, $h(n)$ is the price within acceptable estimates, and $g(n)$ represents the actual path of the initial node to node n.

4.3. The Floyd Algorithm. The Floyd algorithm is an algorithm [22] to solve the shortest path between any two points, which can correctly deal with the shortest path problem of the directed graph or negative power (but not the negative power loop), and is also used to compute the transitive closure of the directed graph.

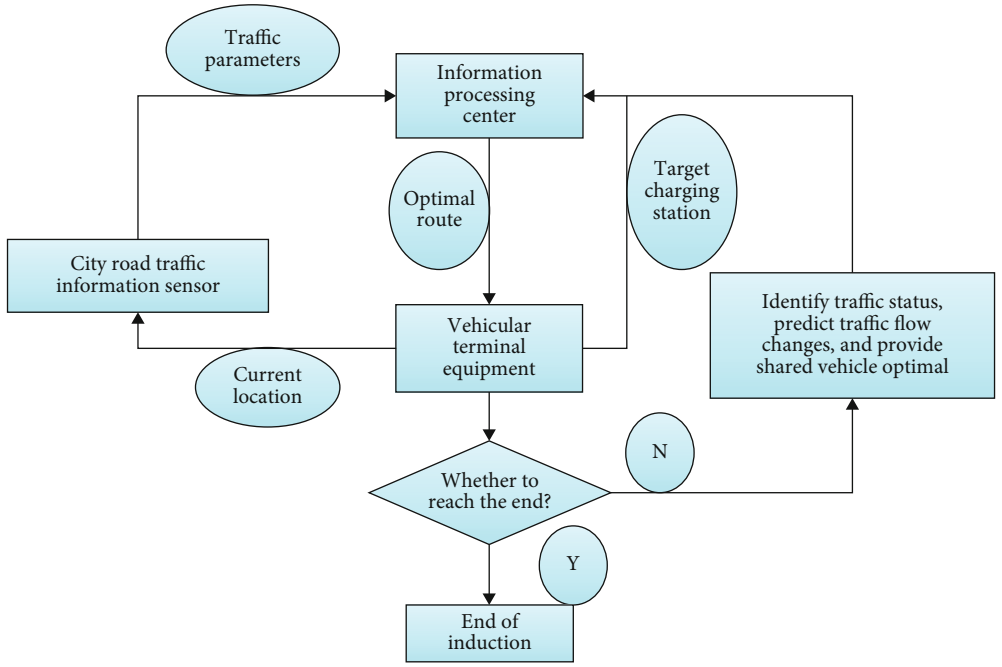


Figure 8: Traffic guidance flow chart.

When calculating the shortest path of each vertex in graph $G = (V, E)$ by the Floyd algorithm, two matrices need to be introduced. The element $a[i, j]$ in matrix S represents the distance from vertex i ($i-$th *vertex*) to vertex j ($j-$th *vertex*). The element $b[ij]$ in the matrix p represents the vertices of the vertex i to the vertex j after the value recorded by $b[i, j]$.

Assuming that the number of vertices in graph G is N, then the matrix D and the matrix p need to be updated N times. At first, the distance of vertex a in matrix D is the weight of vertex i to vertex j, and if i and j are not adjacent, then $a[i, j] = 0$, the value of the matrix p is the value of j of vertex $b[i, j]$. The next step is the N times update of the matrix D. In the first update, if the "the distance of $a[i, j]$" is greater than "$a[i, 0] + a[0, j]$, which represents the distance through the first vertex between i and j," update $a[i, j]$ to "$a[i, 0] + a[0, j]$" and update $b[i, j]$ to $b[i, 0]$. Similarly, when the $k^{-\text{th}}$ update, if "the distance of $a[i, j]$" is greater than "$a[i, k-1] + a[k-1, j]$," then update $a[i, j]$ to "$a[i, k-1] + a[k-1, j]$" and update $b[i, j]$ to $b[i, k-1]$. Iterative operation N times and the program ends.

The Floyd algorithm is a kind of exhaustive algorithm. When calculating the shortest path between any two points, the Floyd algorithm needs to calculate the point with connectivity on the way. However, there are many nodes in the urban traffic road network, so the time complexity is high and it is not suitable for this algorithm.

5. Research on Optimal Induction Path Based on PSO Algorithm

5.1. Optimal Path Calculation Process. When the electric vehicle charging administrator starts the car, the destination information is entered into the vehicle terminal device. The technology of IOT can perceive traffic congestion degree, traffic management information, and other real-time traffic information. The transmission layer transmission network is sent to the data processing center, and the received information is processed and analyzed through the network layer. The most reasonable selection scheme of charging pile and the optimal induction path are calculated, and then the real-time traffic information and system decision scheme are sent to the administrator through the transmission layer to achieve the goal of the optimal induction path. The overall traffic guidance process is shown in Figure 8.

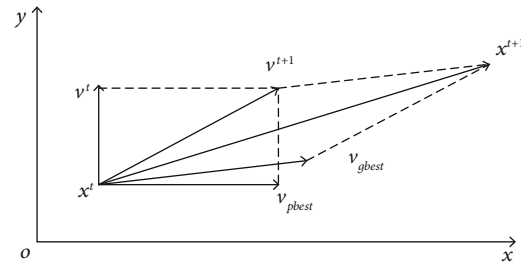


Figure 9: The principle of particle motion.

5.2. Identification of Traffic State. The traffic status of cities can be divided into fluent, basic smooth, mild congestion, moderate congestion, and serious congestion. The selection of the optimal guide path depends on the status of the traffic status of each section and predicts the traffic status of each possible road section.

The identification steps of traffic condition include the extraction and selection of traffic features, the classification of traffic status, and the prediction of traffic status. Traffic status recognition is mainly achieved the traffic volume, line length, road saturation, average speed, vehicle density, and

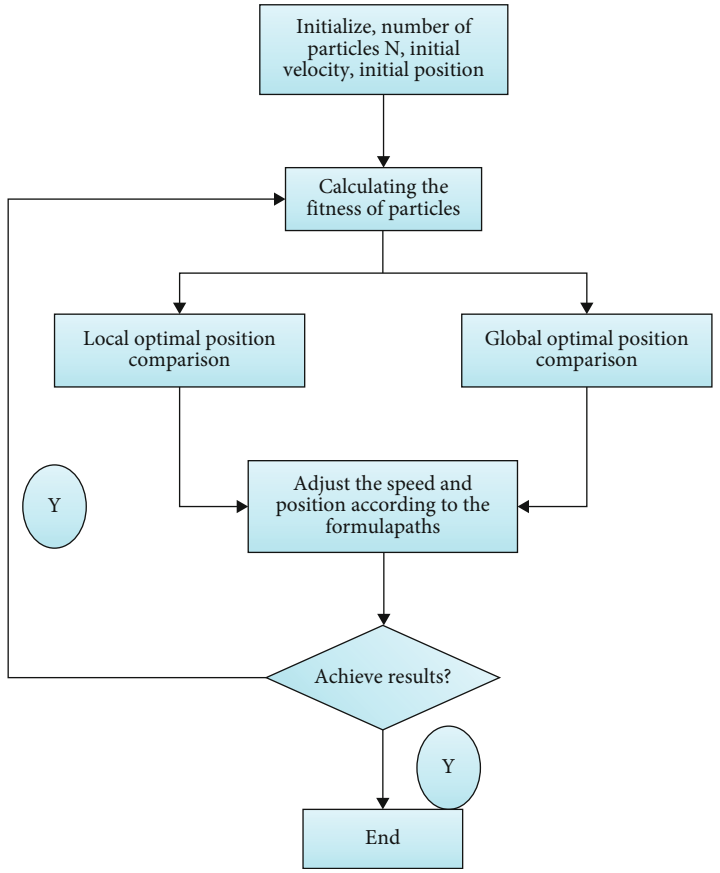


Figure 10: Algorithm flow chart.

other detectable characteristics by sensing equipment. Then we can classify the traffic state based on the traffic characteristics. Finally, according to the real-time traffic data, we can predict the traffic state of the road that may be selected in the near future.

5.3. The Calculation of the Optimal Path. According to the purpose of the profit of the shared car managers and the storage of electric vehicles after using, the main content of this research is to use the particle swarm optimization algorithm to select the optimal path from the parking point to the charging pile for the shared car need to be charged.

5.3.1. The Basic Idea of PSO Algorithm. The particle swarm algorithm was proposed by Eberhart, an expert in computational intelligence research in the United States, and Dr. Kennedy, a psychologist research specialist in 1995 [23–25]. The algorithm is a random search algorithm based on the regularity of bird activity and the swarm intelligence. In this algorithm, the bird is regarded as a particle and only retains the bird's flight function and the group behavior. Therefore, the particles in the algorithm are abstracted into the birds flying at a certain speed and dynamically adjusts according to the flight experience. The position of the particle can be used to represent the solution of the optimization problem, the direction and distance of particle movement can be controlled by velocity, and the pros and cons of each particle can be evaluated by a fitness function.

In solving the optimization problem, the particle swarm algorithm assumed that the size of the particle swarm is M and extends to the N-dimensional space, where the particle i is represented as a vector $X_i = (x_1, x_2, \cdots, x_N)$ in the N-dimensional space, and the flight velocity is $V_i = (v_1, v_2, \cdots, v_N)$. Each particle has a fitness value determined by the objective function and knows the best position (pbest) and the present position of X_i that it has found so far. This can be seen as the particle's own flight experience. In addition, each particle also knows the best location (gbest) of all particles found in this group so far (gbest is the optimal value in pbest). This can be seen as the experience of the peer group. Particles determine the next movement through their experience and the best experience of their peers. After finding (pbest, gbest) the two optimal values, the particle uses the following formula to update its speed and position.

$$v_{ij}(t+1) = \omega v_{ij}(t) + c_1 \text{ rand }() \left(p_{ij} - x_{ij}(t)\right) + c_2 \text{ rand }() \left(p_{gj} - x_{ij}(t)\right) \tag{2}$$

$$x_{ij}(t+1) = x_{ij}(t) + v_{ij}(t+1) \tag{3}$$

Among them, ω is called the inertia weight, which is the ability to search for new search space of particles, t is the

Table 2: Units for magnetic properties.

(a)

Path label R	Marking of starting and ending points	Length between starting and ending points (m)
$R1$	(1,5)	1192
$R2$	(2,5)	1145
$R3$	(2,21)	668
$R4$	(3,25)	242
$R5$	(4,5)	216
$R6$	(5,18)	676
$R7$	(18,6)	842
$R8$	(18,19)	511
$R9$	(6,7)	1900
$R10$	(6,8)	634
$R11$	(7,8)	639
$R12$	(8,10)	607
$R13$	(10,11)	672
$R14$	(10,9)	255
$R15$	(8,11)	902
$R16$	(11,17)	409
$R17$	(11,12)	675
$R18$	(11,14)	367
$R19$	(12,13)	140
$R20$	(14,15)	622
$R21$	(14,16)	825
$R22$	(17,19)	534
$R23$	(19,21)	695
$R24$	(19,20)	600
$R25$	(20,22)	1500
$R26$	(20,16)	651

(b)

Path label R	Credibility φ	Cost coefficient λ	Comprehensive cost
$R1$	0.92	0.90	136.22
$R2$	0.88	0.92	126.41
$R3$	0.72	0.94	175.82
$R4$	0.87	0.91	28.63
$R5$	0.91	0.96	18.66
$R6$	0.84	0.89	96.26
$R7$	0.93	0.85	50.10
$R8$	0.87	0.84	55.80
$R9$	0.91	0.87	148.77
$R10$	0.91	0.88	76.59
$R11$	0.93	0.93	41.60
$R12$	0.74	0.95	149.93
$R13$	0.68	0.97	208.59
$R14$	0.69	0.81	64.03
$R15$	0.74	0.82	192.31
$R16$	0.68	0.84	109.94
$R17$	0.86	0.82	77.49
$R18$	0.93	0.80	20.55
$R19$	0.93	0.83	8.13
$R20$	0.93	0.94	40.93
$R21$	0.92	0.95	62.70
$R22$	0.93	0.93	34.76
$R23$	0.87	0.92	83.12
$R24$	0.94	0.91	32.76
$R25$	0.94	0.94	84.60
$R26$	0.93	0.90	41.01

current iteration number, c_1 and c_2 are the learning factors, c_1 regulates the individual experience part, c_2 regulates the social cognition part and generally sets them to the same value, and rand() is a random number between (0, 1). In each dimension, the particle has a maximum limiting speed $V_{\max}$, and if one dimension speed exceeds $V_{\max}$, the velocity of this dimension is limited to V max. The motion principle of the particles is shown in Figure 9.

Its algorithm flow is shown in Figure 10.

5.3.2. Optimal Path Selection Based on Ant Colony Algorithm. The problem of optimal path selection is similar to the law of birds foraging. The shared car docking point is set as the starting point. The destination is that the charging pile is regarded as a food source for birds to find. The shared car need to pass through a section, node from the starting point to the location of the charging pile. In the path selection process, the shared car will choose the desired path according to its own judgment. The whole network is a directed graph, the vehicle is an artificial birds with intelligent behavior, and the artificial birds just find the path from the starting point to the point of termination and does not return to the original start point. Through the abstraction of the traffic process, the actual cost of traffic is taken as the heuristic information, so that an artificial bird swarm system with optimal path selection can be established.

In the actual traffic, we should not only consider the length of the path, but also take into account the real-time traffic, the delay time of intersection, and the influence of one-way street. Here we set up the comprehensive cost as the best reference standard. The formula is as follows:

$$R_n = \sum_{i-1}^{n} d_i \cdot \lambda / (1 - \varphi_i) \tag{4}$$

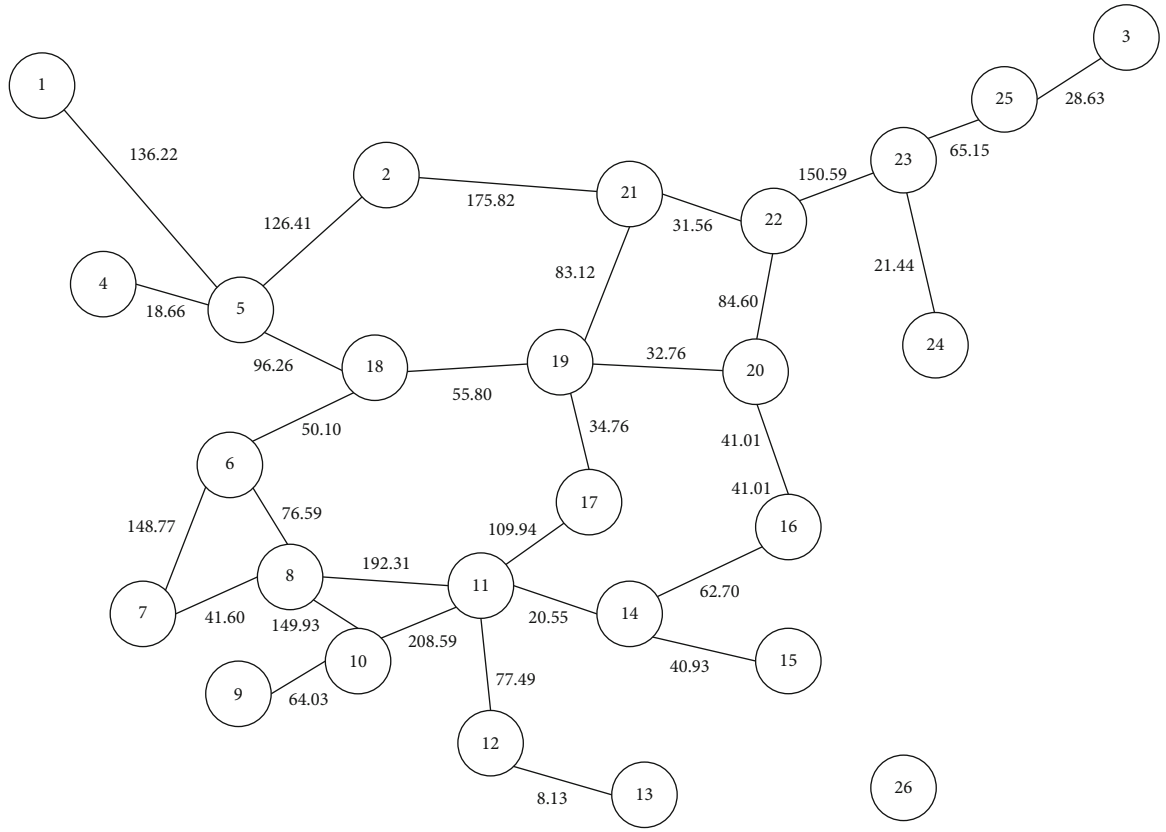


Figure 11: Location of each node.

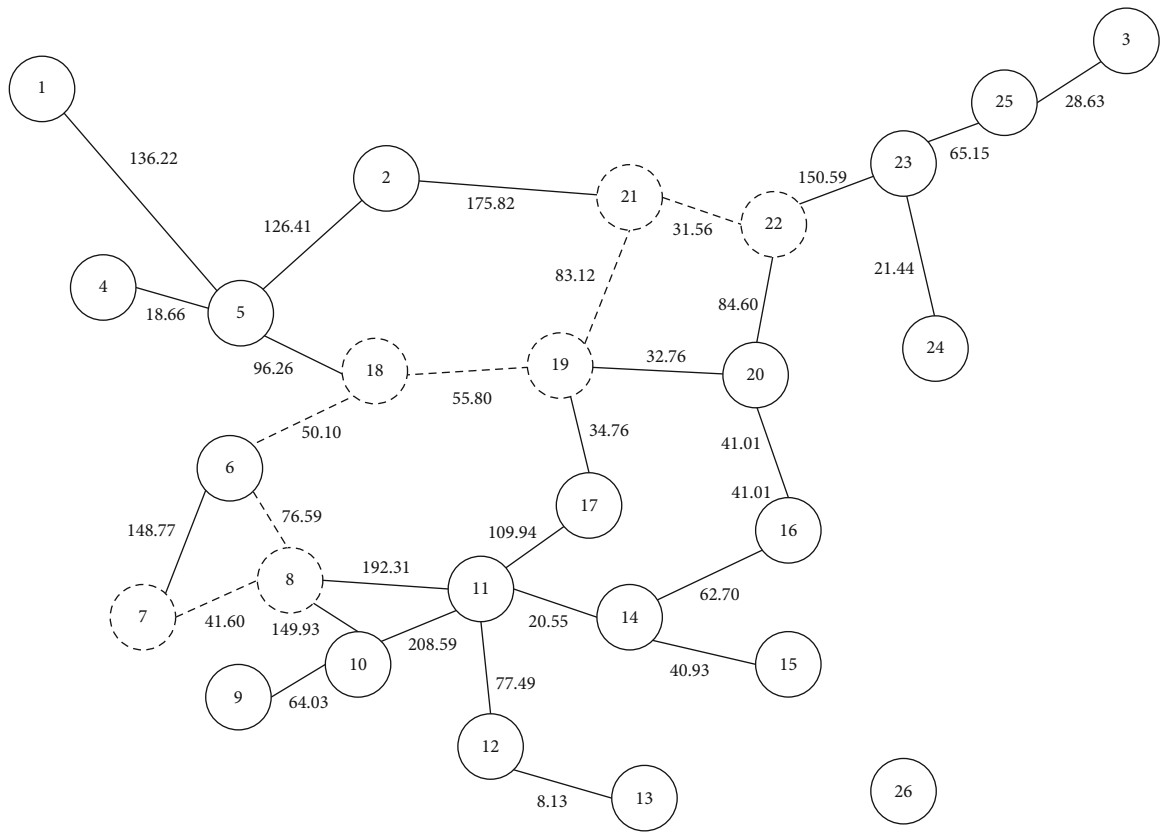


Figure 12: Dijkstra algorithm simulation results (dashed line).

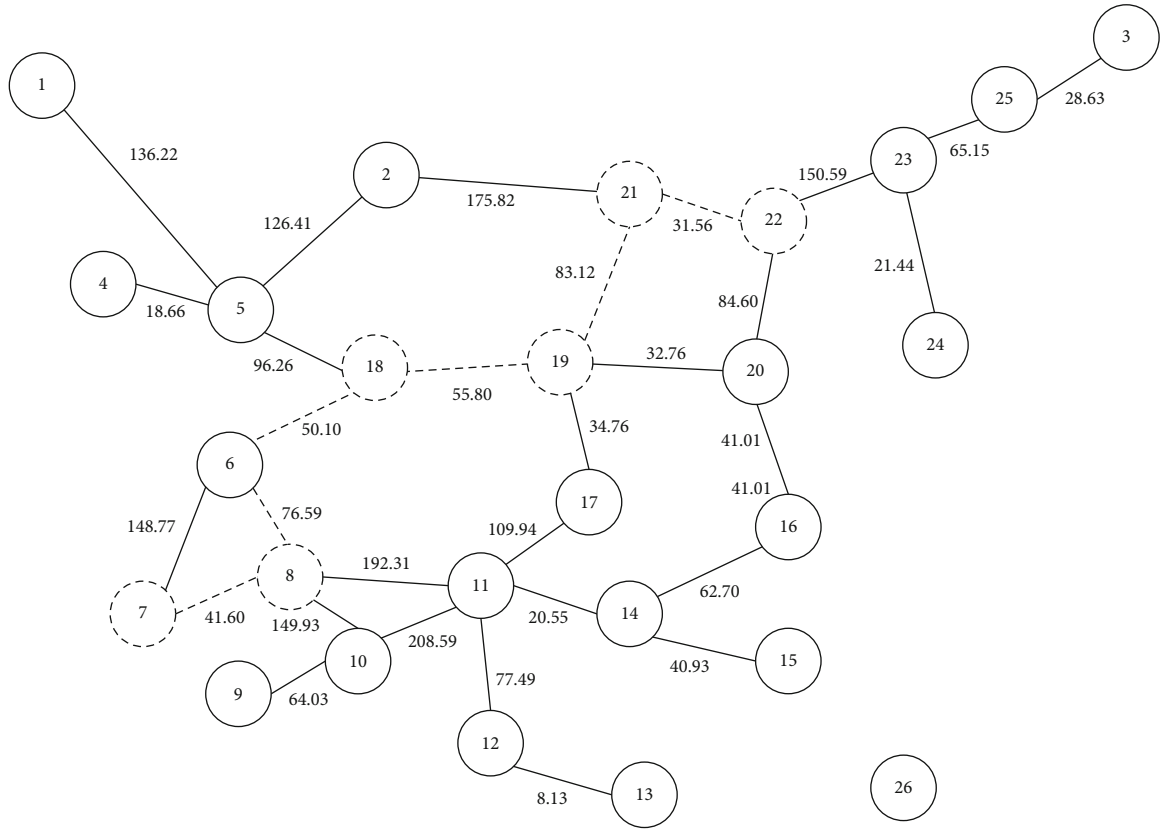


Figure 13: Heuristic algorithm simulation results (dashed line).

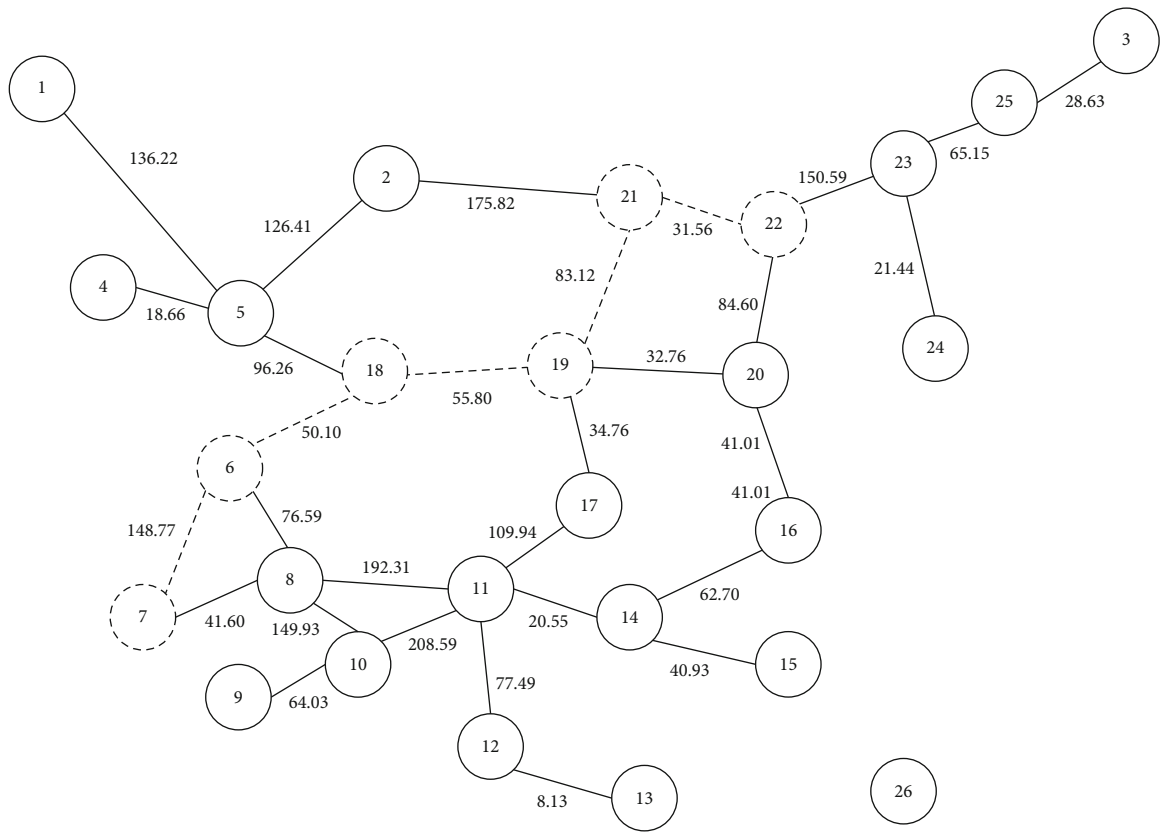


Figure 14: Particle swarm optimization algorithm simulation results (dashed line).

Table 3: Experimental results of the algorithm.

Algorithm class	Actual path consumption	Theory path consumption	Theory path consumption	Time spend
PSO	318.76	318.76	0	0.0586
Dijkstra	349.38	318.76	9.60%	0.0361
A^*	349.38	318.76	9.60%	0.0421

Where λ represents the cost parameters of the road section and φ indicates the reliability coefficient of the road section.

Through experimental simulation, the data of path label R, marking of starting and ending points, length between starting and ending points (m) are arranged as shown in Table 2. The Dijkstra algorithm, heuristic algorithm A^*, and particle swarm optimization method are used to calculate the optimal path, respectively, in which the cost parameter and the reliability coefficient among the section are calculated according to the actual situation. Assuming that there are 26 nodes in a city and numbered with the number of Arabia digital 1-26, and the distribution of nodes is shown in Figure 11; we should calculate the comprehensive cost of each path, respectively. Here, we use the Dijkstra algorithm, heuristic algorithm, and particle swarm algorithm to obtain the optimal path from node 7 (shared car initial point) to node 22 (shared car charging pile). Their shortest routes are 7⟶8⟶6⟶18⟶19⟶21→22, 7⟶8⟶6⟶18⟶19⟶21⟶22, and 7⟶6⟶18⟶19⟶21⟶22. The details are shown in dashed lines in Figure 12, Figure 13, and Figure 14. The remaining any two nodes can use the same method to solve the optimal path.

According to the data of Table 3, compared with the heuristic algorithm A^* and the Dijkstra algorithm, the actual path cost of particle swarm optimization algorithm is the lowest, and there is no deviation from the theoretical path cost. However, the actual path cost of heuristic algorithm A^*, and the Dijkstra algorithm is deviated from the theoretical value, indicating that these two algorithms are prone to fall into local optimization in the process of finding the optimal path. At the same time, because the network constructed by simulation is not large enough, it takes longer to adopt particle swarm optimization algorithm than heuristic algorithm A^* and Dijkstra.

The principle of the Dijkstra algorithm is to traverse the whole network, which has obvious advantages when the network size is small. Particle swarm optimization is a heuristic algorithm with simple principle. Particles can conduct current search based on previous search information, which greatly reduces the possibility of blind search. Therefore, with the increase of network construction, the advantage of particle swarm optimization algorithm will become more obvious.

Particle swarm optimization is a bionic optimization algorithm based on multiple agents. Therefore, the application of particle swarm optimization algorithm to the optimal path selection process can fully reflect the advantages of the algorithm, and the algorithm itself has great development prospects and research value. By selecting different starting and ending points, this paper uses particle swarm optimization algorithm to find the shortest path and calculates credibility, cost coefficient, and comprehensive cost. The sorting results are shown in Table 2.

Data Availability

The data sets analyzed in this study can be obtained from the corresponding author upon reasonable request.

Conflicts of Interest

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

Acknowledgments

This work was supported by the Natural Science Youth Foundation of Hebei Province (No. A2018409015), the Doctoral Science Foundation of North China Institute of Aerospace Engineering (No. BKY-2018-23), and the Research Project of Langfang Science and Technology Bureau (No. 2019029012).

References

[1] R. J. Dilger, "ISTEA: a new direction for transportation policy," *Publius: The Journal of Federalism*, vol. 22, no. 3, pp. 67–78, 1992.

[2] Transportation Research Board Special, *ISTEA and INTERMODAL PLANNING: Concept, Practice, Vision*, Transportation Research Board Special Report, 1993.

[3] Y. Sun, H. Qiang, J. Xu, and G. Lin, "Internet of Things-Based online condition monitor and improved adaptive fuzzy control for a medium-low-speed maglev train system," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 4, pp. 2629–2639, 2020.

[4] J. Verplanke, M. K. McCall, C. Uberhuaga, G. Rambaldi, and M. Haklay, "A shared perspective for PGIS and VGI," *The Cartographic Journal*, vol. 53, no. 4, pp. 308–317, 2016.

[5] Z. Y. Zhou, *Study on the Function Design and Key Model of Urban Stop-and-Click Guidance Information System*, South East Human Science transportation Planning and Management, 2004.

[6] B. Li, T. Liu, and H. Zhu, *Research on the Technical Scheme of Shanghai Public Parking Information System*, Third China Intelligent Transportation Conference, 2007.

[7] Z. Y. Zhou, J. Chen, and W. Wang, "Review on the development and application of urban parking guidance information system," *Traffic and Transportation Systems Engineering and Information*, vol. 2, 2006.

[8] Y. Wen, Y. Lu, J. Yan, Z. Zhou, K. M. von Deneen, and P. Shi, "An algorithm for license plate recognition applied to intelligent transportation system," *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 3, pp. 830–845, 2011.

[9] A. M. Yang, X. L. Yang, Y. Han, Y. K. Guo, J. M. Liu, and H. Q. Zhang, "Wireless channel optimization of Internet of things," *IEEE Access*, vol. 6, no. 1, pp. 54064–54074, 2018.

[10] W. Wei, X. Xia, M. Wozniak, X. Fan, R. Damaševičius, and Y. Li, "Multi-sink distributed power control algorithm for cyber-physical-systems in coal mine tunnels," *Computer Networks*, vol. 161, pp. 210–219, 2019.

[11] X. L. Wen, "The shortest path optimization algorithm based on hybrid algorithm," *Journal of Tianjin Polytechnic University*, vol. 25, no. 1, pp. 37–40, 2009.

[12] N. J. van der Zijpp and S. F. Catalano, "Path enumeration by finding the constrained K-shortest paths," *Transportation Research*, vol. 39, no. 6, pp. 545–563, 2005.

[13] Organisation for Economic Co-operation and Development, *Urban Mobility System Upgrade How Shared Self-Driving Cars Could Change City Traffic*, OECD Publishing, 2015.

[14] International Telecommunication Union, *Internet Reports 2005: The Internet of Things*, ITU, Geneva, 2005.

[15] J. Li, L. Zhang, X. Feng, K. Jia, and F. Kong, "Feature extraction and area identification of wireless channel in mobile communication," *Journal of Internet Technology*, vol. 20, no. 2, pp. 545–553, 2019.

[16] Y. Q. Mao and S. B. Shen, "Information model and capability analysis of Internet of things," *Journal of Software*, vol. 25, no. 8, pp. 1685–1695, 2014.

[17] H. Fu, G. Manogaran, K. Wu, M. Cao, S. Jiang, and A. Yang, "Intelligent decision-making of online shopping behavior based on Internet of things," *International Journal of Information Management*, vol. 50, pp. 515–525, 2020.

[18] A. Yang, Y. Li, C. Liu, J. Li, Y. Zhang, and J. Wang, "Research on logistics supply chain of iron and steel enterprises based on block chain technology," *Future Generation Computer Systems*, vol. 101, pp. 635–645, 2019.

[19] J. Wu, "Green wireless communications: from concept to reality [Industry Perspectives]," *IEEE Wireless Communications*, vol. 19, no. 4, pp. 4-5, 2012.

[20] W. Wei, J. Su, H. Song, H. Wang, and X. Fan, "CDMA-based anti-collision algorithm for EPC global C1 gen 2 systems," *Telecommunication Systems*, vol. 67, no. 1, pp. 63–71, 2018.

[21] J. Peng and Y. Shao, "Intelligent method for identifying driving risk based on V2V multisource big data," *Complexity*, vol. 2018, Article ID 1801273, 9 pages, 2018.

[22] J. Kennedy and R. C. Eberhart, "Particle swarm optimization," in *Proceedings of ICNN'95 - International Conference on Neural Networks*, pp. 1942–1948, Perth, WA, Australia, November-December 1995.

[23] A. Yang, C. Zhang, Y. Chen, Y. Zhuansun, and H. Liu, "Security and privacy of smart home systems based on the Internet of Things and stereo matching algorithms," *IEEE Internet of Things Journal*, vol. 2019, no. article 1, 2019.

[24] C. W. Reynolds, "Flocks, herds and schools: a distributed behavioral model," *ACM Siggraph Computer Graphics*, vol. 21, no. 4, pp. 25–34, 1987.

[25] J. Wu, L. Zhang, S. Yin, H. Wang, G. Wang, and J. Yuan, "Differential diagnosis model of hypocellular myelodysplastic syndrome and aplastic anemia based on the medical big data platform," *Complexity*, vol. 2018, Article ID 4824350, 12 pages, 2018.