

AI-Driven Anomaly Detection Technologies to Support Emerging Smart Applications

Lead Guest Editor: Fa Zhu

Guest Editors: Mian Ahmad Jan and Rajan Shankaran





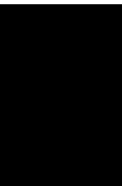
AI-Driven Anomaly Detection Technologies to Support Emerging Smart Applications

Wireless Communications and Mobile Computing

**AI-Driven Anomaly Detection
Technologies to Support Emerging
Smart Applications**

Lead Guest Editor: Fa Zhu


Guest Editors: Mian Ahmad Jan and Rajan
Shankaran




Copyright © 2022 Hindawi Limited. All rights reserved.

This is a special issue published in “Wireless Communications and Mobile Computing.” All articles are open access articles distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Chief Editor

































Zhipeng Cai , USA

Associate Editors

Ke Guan , China
Jaime Lloret , Spain
Maode Ma , Singapore

Academic Editors

Muhammad Inam Abbasi, Malaysia
Ghufran Ahmed , Pakistan
Hamza Mohammed Ridha Al-Khafaji ,
Iraq
Abdullah Alamoodi , Malaysia
Marica Amadeo, Italy
Sandhya Aneja, USA
Mohd Dilshad Ansari, India
Eva Antonino-Daviu , Spain
Mehmet Emin Aydin, United Kingdom
Parameshchhari B. D. , India
Kalapaveen Bagadi , India
Ashish Bagwari , India
Dr. Abdul Basit , Pakistan
Alessandro Bazzi , Italy
Zdenek Becvar , Czech Republic
Nabil Benamar , Morocco
Olivier Berder, France
Petros S. Bithas, Greece
Dario Bruneo , Italy
Jun Cai, Canada
Xuesong Cai, Denmark
Gerardo Canfora , Italy
Rolando Carrasco, United Kingdom
Vicente Casares-Giner , Spain
Brijesh Chaurasia, India
Lin Chen , France
Xianfu Chen , Finland
Hui Cheng , United Kingdom
Hsin-Hung Cho, Taiwan
Ernestina Cianca , Italy
Marta Cimitile , Italy
Riccardo Colella , Italy
Mario Collotta , Italy
Massimo Condoluci , Sweden
Antonino Crivello , Italy
Antonio De Domenico , France
Floriano De Rango , Italy


Antonio De la Oliva , Spain
Margot Deruyck, Belgium
Liang Dong , USA
Praveen Kumar Donta, Austria
Zhuojun Duan, USA
Mohammed El-Hajjar , United Kingdom
Oscar Esparza , Spain
Maria Fazio , Italy
Mauro Femminella , Italy
Manuel Fernandez-Veiga , Spain
Gianluigi Ferrari , Italy
Luca Foschini , Italy
Alexandros G. Fragkiadakis , Greece
Ivan Ganchev , Bulgaria
Óscar García, Spain
Manuel García Sánchez , Spain
L. J. García Villalba , Spain
Miguel Garcia-Pineda , Spain
Piedad Garrido , Spain
Michele Girolami, Italy
Mariusz Glabowski , Poland
Carles Gomez , Spain
Antonio Guerrieri , Italy
Barbara Guidi , Italy
Rami Hamdi, Qatar
Tao Han, USA
Sherief Hashima , Egypt
Mahmoud Hassaballah , Egypt
Yejun He , China
Yixin He, China
Andrej Hrovat , Slovenia
Chunqiang Hu , China
Xuexian Hu , China
Zhenghua Huang , China
Xiaohong Jiang , Japan
Vicente Julian , Spain
Rajesh Kaluri , India
Dimitrios Katsaros, Greece
Muhammad Asghar Khan, Pakistan
Rahim Khan , Pakistan
Ahmed Khattab, Egypt
Hasan Ali Khattak, Pakistan
Mario Kolberg , United Kingdom
Meet Kumari, India
Wen-Cheng Lai , Taiwan

Jose M. Lanza-Gutierrez, Spain
Pavlos I. Lazaridis , United Kingdom
Kim-Hung Le , Vietnam
Tuan Anh Le , United Kingdom
Xianfu Lei, China
Jianfeng Li , China
Xiangxue Li , China
Yaguang Lin , China
Zhi Lin , China
Liu Liu , China
Mingqian Liu , China
Zhi Liu, Japan
Miguel López-Benítez , United Kingdom
Chuanwen Luo , China
Lu Lv, China
Basem M. ElHalawany , Egypt
Imadeldin Mahgoub , USA
Rajesh Manoharan , India
Davide Mattera , Italy
Michael McGuire , Canada
Weizhi Meng , Denmark
Klaus Moessner , United Kingdom
Simone Morosi , Italy
Amrit Mukherjee, Czech Republic
Shahid Mumtaz , Portugal
Giovanni Nardini , Italy
Tuan M. Nguyen , Vietnam
Petros Nicolitidis , Greece
Rajendran Parthiban , Malaysia
Giovanni Pau , Italy
Matteo Petracca , Italy
Marco Picone , Italy
Daniele Pinchera , Italy
Giuseppe Piro , Italy
Javier Prieto , Spain
Umair Rafique, Finland
Maheswar Rajagopal , India
Sujan Rajbhandari , United Kingdom
Rajib Rana, Australia
Luca Reggiani , Italy
Daniel G. Reina , Spain
Bo Rong , Canada
Mangal Sain , Republic of Korea
Praneet Saurabh , India

Hans Schotten, Germany
Patrick Seeling , USA
Muhammad Shafiq , China
Zaffar Ahmed Shaikh , Pakistan
Vishal Sharma , United Kingdom
Kaize Shi , Australia
Chakchai So-In, Thailand
Enrique Stevens-Navarro , Mexico
Sangeetha Subbaraj , India
Tien-Wen Sung, Taiwan
Suhua Tang , Japan
Pan Tang , China
Pierre-Martin Tardif , Canada
Sreenath Reddy Thummaluru, India
Tran Trung Duy , Vietnam
Fan-Hsun Tseng, Taiwan
S Velliangiri , India
Quoc-Tuan Vien , United Kingdom
Enrico M. Vitucci , Italy
Shaohua Wan , China
Dawei Wang, China
Huaqun Wang , China
Pengfei Wang , China
Dapeng Wu , China
Huaming Wu , China
Ding Xu , China
YAN YAO , China
Jie Yang, USA
Long Yang , China
Qiang Ye , Canada
Changyan Yi , China
Ya-Ju Yu , Taiwan
Marat V. Yuldashev , Finland
Sherali Zeadally, USA
Hong-Hai Zhang, USA
Jiliang Zhang, China
Lei Zhang, Spain
Wence Zhang , China
Yushu Zhang, China
Kechen Zheng, China
Fuhui Zhou , USA
Meiling Zhu, United Kingdom
Zhengyu Zhu , China






Contents

Point Cloud Intensity Correction for 2D LiDAR Mobile Laser Scanning

Xu Liu, Qiuji Li , Youlin Xu, and Xuefeng Wei

Research Article (22 pages), Article ID 3707985, Volume 2022 (2022)

Efficient Semantic Enrichment Process for Spatiotemporal Trajectories

Bin Zhao , Mingyu Liu , Jingjing Han , Genlin Ji , and Xintao Liu 




Research Article (13 pages), Article ID 4488781, Volume 2021 (2021)

A Deep Learning-Based Inventory Management and Demand Prediction Optimization Method for Anomaly Detection

Chuning Deng  and Yongji Liu 




Research Article (14 pages), Article ID 9969357, Volume 2021 (2021)

Comparison Analysis of Different Time-Scale Heart Rate Variability Signals for Mental Workload Assessment in Human-Robot Interaction

Shiliang Shao , Ting Wang , Yawei Li, Chunhe Song , Yihan Jiang, and Chen Yao

Research Article (12 pages), Article ID 8371637, Volume 2021 (2021)

Unsupervised Anomaly Detection for Glaucoma Diagnosis

Wei Zhou , Yuan Gao, Jianhang Ji, Shicheng Li , and Yugen Yi 


Research Article (14 pages), Article ID 5978495, Volume 2021 (2021)

EAWNet: An Edge Attention-Wise Objecter for Real-Time Visual Internet of Things

Zhichao Zhang , Hui Chen, Xiaoqing Yin , and Jinsheng Deng

Research Article (15 pages), Article ID 7258649, Volume 2021 (2021)

An Optimized Fingerprinting-Based Indoor Positioning with Kalman Filter and Universal Kriging for 5G Internet of Things

Shuai Huang , Kun Zhao , Zhengqi Zheng, Wenqing Ji, Tianyi Li, and Xiaofei Liao

Research Article (10 pages), Article ID 9936706, Volume 2021 (2021)

Research Article

Point Cloud Intensity Correction for 2D LiDAR Mobile Laser Scanning

Xu Liu,¹ Qiujie Li¹ ,¹ Youlin Xu,¹ and Xuefeng Wei²

¹College of Mechanical and Electronic Engineering, Nanjing Forestry University, Nanjing 210037, China

²College of Intelligent Manufacturing, Huanghuai University, Zhumadian 463000, China

Correspondence should be addressed to Qiujie Li; liqiuje_1@163.com

Received 20 April 2021; Revised 13 September 2021; Accepted 21 December 2021; Published 20 January 2022

Academic Editor: Zhipeng Cai

Copyright © 2022 Xu Liu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The acquisition of point cloud data by mobile laser scanning (MLS) includes not only the information about the 3D geometry of the object but also the intensity from the scanned object. However, due to the influence of various factors, there is a large deviation between the intensity and the spectral reflection characteristics of the scanned object. Intensity correction should be carried out before this method is applied to object recognition. A new point cloud intensity correction method for 2D MLS that was developed by combining theoretical derivation with empirical correction is proposed in this paper. First, based on the LiDAR formula, the main factors influencing MLS intensity are investigated, and a distance piecewise polynomial and an incident angle cosine polynomial are adopted to obtain the intensity correction model of UTM-30LX 2D LiDAR on a diffuse reflector plate. Second, according to the scan pattern, a 2D scan grid is constructed to organize the MLS intensity, and a new method of spherical neighborhood search fitting plane is proposed to accurately calculate the cosine of the incident angle. Finally, the obtained intensity correction model is utilized to correct the MLS intensity of a wall. Two groups of verification experiments are carried out on single sites and multiple sites to test the effect of the intensity correction model. Overall, the improvements in intensity consistency range from 70% to 92.7% after correction within the tested ranges of distance and incident angles [0.52 m–5.34 m, 0°–74°]. The results indicate that the proposed intensity correction model yields highly accurate fitting and can effectively remove the deviation in MLS intensity caused by the distance and incident angle.

1. Introduction

Light detection and ranging (LiDAR), a type of noncontact active remote sensing sensor, can quickly scan high-resolution and high-precision 3D point cloud data on the target surface. LiDAR has been successfully applied in fields, such as global climate change research, smart cities, forest resource surveys, environmental monitoring, and basic mapping [1, 2]. LiDAR can not only obtain the 3D geometric information of the target surface, but it can also record the intensity information. The intensity corresponds to the coordinate information one-to-one without registration and has the feature of pixel-level fusion. It represents the reflection spectral characteristics of the object target to the laser and can be used as an important feature of target classification [3–5]. However, due to the influence of various factors, such as scanner characteristics, atmospheric

transmission characteristics, target surface parameters, and data acquisition parameters, there is a large deviation between the intensity and the spectral reflection characteristics of the object target. The phenomena of the same object target with different spectra and the same spectrum of different object targets are apt to occur. Therefore, it is necessary to eliminate the influence of various factors through correction [6–8].

Intensity correction methods can be divided into theoretical correction and empirical correction [9]. The theoretical correction method focuses on the analysis of the relationship between multiple factors causing the intensity change, and the regression model of each influencing factor and intensity is established according to the LiDAR ranging principle. Song et al. [10] systematically discussed and analyzed the radiation characteristics, as well as the key influencing factors of intensity data from the perspective of

the LiDAR radiation transmission mechanism. They also eliminated these factors through a theoretical correction model. Bolkas [11] studied the intensity correction method of the incident angle and evaluated the intensity correction effect of four kinds of light reflection models under different colors and glosses. Cheng et al. [12] proposed a method to eliminate the distance effect of laser intensity by using a piecewise polynomial model, which effectively eliminated the intensity deviation caused by distance. Fang et al. [13], based on the principle of laser imaging, deduced an intensity theoretical model that considered the defocus effect of the receiving optical system and applied the model to the intensity correction of fresco. The theoretical correction model is applicable to a wide range of scenarios, but the process of the established model is more complex. The parameters in the theoretical correction model are always set as the measured value under the ideal state, which produces large errors in the actual intensity correction. For example, the measurement of the atmospheric attenuation coefficient and transmission coefficient of a LiDAR optical system usually has fluctuation errors. The reflection characteristics of the scanned object's surface are very complex, and it is difficult to regard it as an ideal diffuse reflection target.

The empirical correction method does not rely on any theoretical model. However, it only establishes the intensity correction model in the form of elementary functions (such as polynomial, cosine, or exponential functions). It estimates the function parameters from the actual measured data, which is suitable for situations where the structure and parameters of LiDAR are poorly known. Li and Cheng [14] established two data-driven models to correct the effects of the distance and incident angle on intensity and then applied the corrected intensity to the damage detection of historic buildings. Coren and Sterzai [15] adopted the empirical correction method to establish an exponential function model between the intensity and atmospheric attenuation coefficient, which reduced the fluctuation of the intensity data of asphalt pavement. Vain et al. [16] and Korpela et al. [17] used the empirical correction method to compensate for the intensity change caused by the LiDAR internal automatic gain control system. For the intensity data of specific scenes, the empirical correction method has a higher accuracy than the theoretical correction method, but it is only applicable to this specific application scene; otherwise, it needs to be remodeled.

LiDAR measurements are often classified into three types: terrestrial laser scanning (TLS), mobile laser scanning (MLS), and airborne laser scanning (ALS). TLS can directly obtain high-precision 3D point cloud data from the surrounding natural environment by setting up fixed sites. MLS, on the other hand, can quickly scan 3D point cloud data on both sides of the road or on one side by carrying LiDAR on mobile devices, such as cars [18]. Compared with TLS, MLS is flexible and expands the scanning range of LiDAR. MLS also has mature applications in digital city 3D modeling, urban environment monitoring, and urban resource surveys [19, 20]. ALS can obtain the surface spatial information within a large area in a short time with high working efficiency by carrying LiDAR onto the aircraft to realize scanning. However, compared with MLS, point cloud

data obtained via ALS measurements are less accurate. Moreover, MLS has an advantage over ALS in obtaining vertical point cloud data. Yang et al. [21] showed that, compared with ALS, MLS has considerable advantages in street tree identification at the scale of a single tree, as well as a strong data acquisition ability to penetrate the inner canopy and trunk.

LiDAR sensors are usually divided into two kinds: 3D and 2D (also known as multithread and single-thread sensors). The 3D LiDAR can accurately obtain 3D geometric information of the surrounding environment, and it is mainly used in the field of unmanned driving, which requires high accuracy. However, 3D LiDAR is usually expensive. In contrast, 2D LiDAR can obtain high-precision 3D information about the environment throughout the process of moving through fan-shaped scanning frame by frame, which is advantageous because of the fast scanning speed, the small size, low power consumption, and low manufacturing cost. In addition, 2D LiDAR's acquisition of point cloud data has low redundancy and simple data fusion. Thus, it can be directly indexed according to the frame number and in-frame number of the measured points, avoiding accuracy loss caused by grid partitioning [22]. 2D LiDAR using the MLS measurement method is widely used in the field of tree parameter extraction [23, 24] and urban block ground object classification [25, 26]. The point cloud data collection process by 2D LiDAR is shown in Figure 1. Xu et al. [27] and Nan et al. [28] designed a real-time automatic target spray system based on MLS 2D LiDAR detection and compared the performance with that of infrared, ultrasonic sensors, and 3D LiDAR, proving that 2D LiDAR has great advantages in the identification accuracy and rapid measurement of distance.

To the author's knowledge, most previous intensity correction studies have focused on 3D LiDAR sensors for ALS and TLS, whereas 2D LiDAR sensors for MLS are still relatively rare. Based on the advantages of MLS over ALS and TLS analyzed above in application value, the important contribution of this work proposes a high-precision intensity correction method suitable for 2D MLS. In recent years, there have been some studies on intensity correction for TLS similar to this paper. Tan and Cheng [29] studied TLS intensity correction based on the laser ranging formula. The relationship between intensity and new variables was established by combining the cosine of the incident angle and the square of the distance. In the intensity correction for different areas of the white wall, the intensity variance-to-mean ratio ε was between 0.6 and 0.78, indicating that the improvements in intensity consistency ranged from 22% to 40%. Subsequently, Tan and Cheng [30] investigated the effects of incidence angle and distance on intensity data and corrected the intensity data of Faro Focus3D 120. For four reference targets with reflectances of 20%, 40%, 60%, and 80%, the linear interpolation method was used to fit the relationship between incidence angle versus intensity and distance versus intensity. A total of 20 small regions with a size of approximately 15 cm \times 15 cm in the white lime wall surface were randomly sampled to verify the intensity correction effect. The intensity variance-to-mean ratio ε

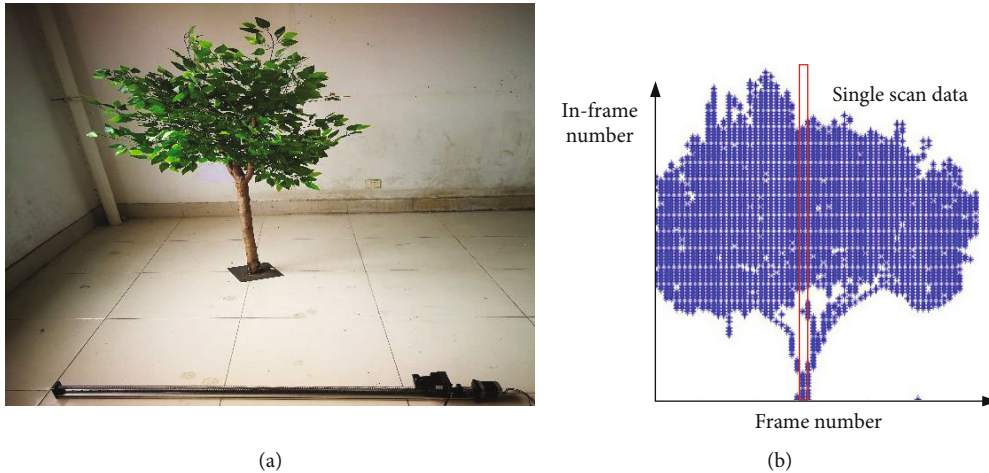


FIGURE 1: Data acquisition process of the MLS system based on a single 2D LiDAR. (a) A single frame of the 2D LiDAR. (b) Index structure for the 2D LiDAR data.

was 0.26, 0.14, 0.19, and 0.21. The results showed that the intensity consistency was improved by 74%~86% by using the intensity correction model established by the above four reference targets. For MLS intensity correction research, Teo and Yu [31] proposed a distance intensity correction workflow for MLS road point clouds using a data-driven approach. The relationship between distance and intensity was fitted by a piecewise polynomial. By comparing the differences between the scanners and the stripe intensities before and after correction, the improvements in intensity consistency ranged from 47% to 56%.

In contrast from the general data-driven intensity correction model, which directly uses various models to establish the approximate relationship between intensity and its influencing factors, lacking the necessary theoretical basis, a new point cloud intensity correction method for 2D MLS is proposed by combining theoretical derivation with empirical correction. In addition, according to the scan pattern, a 2D scan grid is constructed to organize the MLS intensity of the wall, and a new method of a spherical neighborhood search fitting plane is proposed to accurately calculate the cosine of the incident angle. In particular, the relationship between the neighborhood radius and the distance is discussed in the process of plane fitting. The accuracy of incident angle measurement is improved by selecting an appropriate neighborhood radius. Two groups of verification experiments are carried out on single sites and multiple sites to test the effect of the intensity correction model. A single-site experiment shows that the ϵ value of the five same area regions within the range of the distance and incident angle [0.52 m-1.55 m, 0° - 74°] is approximately 0.3, indicating that the consistency of intensity has been improved by 70% after correction. Multisite experiments concluded that the ϵ values of sites A, B, C, and D were 0.073, 0.079, 0.233, and 0.280, respectively, within the range of distance and incident angle [1.52 m-5.34 m, 0° - 62°]. This means that the improvements in intensity consistency range from 72% to 92.7% after correction. Overall, this approach is superior to the latest study mentioned above.

The proposed method of the intensity correction model is introduced in Section 2. The process of obtaining the intensity correction model is described in Section 3. To demonstrate the practical effectiveness of the intensity correction model, two different scenes of wall intensity correction are presented and analyzed in Section 4. Section 5 concludes with the findings.

2. Methods

2.1. MLS 2D LiDAR Intensity Correction Process. Figure 2 is the flow chart of intensity correction for the MLS 2D LiDAR point cloud. In this paper, the intensity correction process consists of two steps. The first step is for model establishment, and the second step is for model validation. Model establishment is when the intensity correction model is obtained. First, the influencing factors of MLS intensity (including target reflectivity, distance, and incident angle) are analyzed for the acquisition of the intensity correction model. Second, the intensity multiplicative model is established, and the intensity correction formula for the distance and incident angle is deduced. Finally, the parameters of the intensity correction model are estimated, including the order and coefficient of the polynomial. Model validation uses the intensity correction model to correct the intensity data of the wall point cloud. For the wall point cloud intensity data scanned by MLS 2D LiDAR, the distance intensity and incident angle intensity are extracted. Then, the obtained intensity correction model is used to correct the intensity data of the wall point cloud.

2.2. Influencing Factors of MLS Intensity. Assuming that the measured target surface is a Lambert surface (the surface with ideal diffuse reflection characteristics), according to the laser ranging formula [32], the echo power received by the laser detector is

$$P_r = \frac{P_t D_r^2 \eta_{\text{atm}} \eta_{\text{sys}}}{4} \cdot \frac{\lambda \cos \theta}{R^2}, \quad (1)$$

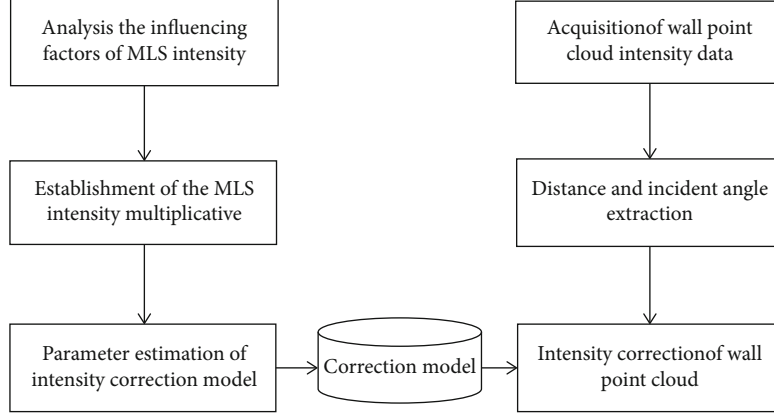


FIGURE 2: Flow chart of intensity correction for the MLS 2D LiDAR point cloud.

where the received laser power P_r is a function of the transmitting laser power P_t , the aperture of the receiving detector is denoted by D_r , the atmospheric one-way extinction coefficient is η_{atm} , the transmission coefficient of the optical system is η_{sys} , the target reflectivity is λ , the distance R is between the measuring point and the receiving end of the LiDAR laser (referred to as the distance), and the cosine of the laser incident angle is denoted by $\cos \theta$. For the MLS system with a short distance, it can be considered that P_t , D_r , η_{atm} , and η_{sys} are constants. Tan and Cheng [29] showed that intensity is the laser echo energy represented digitally. Additionally, there is a certain linear relationship between the received laser power and intensity. According to the above analysis, the main influencing factors of MLS intensity are target reflectivity λ , distance R , and incident angle θ .

2.3. The Multiplicative Model of Intensity by MLS. The objective of intensity correction in this paper is to remove the influence of the distance and incident angle on MLS intensity. Then, the intensity value is only related to the reflectivity of the target. Since the received laser power is nonlinearly processed internally by LiDAR, it is not possible to directly use Formula (1) for theoretical correction. In addition, considering that the effects of the reflectivity, distance, and incident angle are independent in theory, they can each be solved separately. The multiplicative model of intensity can be established as

$$I(\lambda, R, \theta) = f_\lambda(\lambda) f_R(R) f_\theta(\cos \theta). \quad (2)$$

$I(\lambda, R, \theta)$ is the original intensity, and $f_\lambda(\lambda)$, $f_R(R)$, and $f_\theta(\cos \theta)$ represent functions of the independent influence of the target reflectivity, distance, and incident angle as $I(\lambda, R, \theta)$. To eliminate the influence of the distance and incident angle on intensity, $I(\lambda, R, \theta)$ at any distance and incident angle should be transformed to the corrected intensity I_c at reference distance R_0 and reference incident angle θ_0 for a Lambert body with the same target reflectivity.

$$I_c = f_\lambda(\lambda) f_R(R_0) f_\theta(\cos \theta_0) = \frac{f_R(R_0) f_\theta(\cos \theta_0)}{f_R(R) f_\theta(\cos \theta)} \cdot I(\lambda, R, \theta). \quad (3)$$

Formula (3) shows that the transformation relationship between I_c and $I(\lambda, R, \theta)$ is established by $f_R(R)$ and $f_\theta(\cos \theta)$.

After satisfying the conditions of a target with reflectivity, λ is scanned at different distances R and at a constant incident angle θ_0 ; then, the intensity after the distance correction I_{rc} can be written as

$$I_{rc} = \frac{f_R(R_0) f_\theta(\cos \theta_0)}{f_R(R) f_\theta(\cos \theta_0)} \cdot I(\lambda, R, \theta) = \frac{f_R(R_0)}{f_R(R)} \cdot I(\lambda, R, \theta). \quad (4)$$

Under the above premise conditions, $R_{\min} \leq R \leq R_{\max}$ (R_{\min} and R_{\max} are the minimum value and maximum value of the distance measuring range, respectively). According to the Weierstrass theorem [33], a continuous function on a closed interval can be uniformly approximated by a polynomial series. According to the definition, $f_R(R)$ describes the relationship between distance R and $I(\lambda, R, \theta)$ in the distance-intensity data. The intensity regression value $I(\lambda, R, \theta_0)$ under reflectivity λ and reference incident angle θ_0 can be obtained by polynomial regression fitting. Tan and Cheng [30] showed that due to the short-distance effect of the LiDAR optical system, the intensity increases with an increasing distance when the distance is relatively close and decreases with an increasing distance when the distance is relatively far. According to the variable of distance R , piecewise polynomial modeling can be adopted here to illustrate the relationship between R and $I(\lambda, R, \theta_0)$ as

$$f_R(R) = I(\lambda, R, \theta_0) = \begin{cases} \sum_{k=0}^K a_k R^k, & R \leq R_t, \\ \sum_{l=0}^L b_l \left(\frac{1}{R}\right)^l, & R > R_t, \end{cases} \quad (5)$$

where a_k and b_l are the coefficients of the distance polynomial, K and L are the order of the distance polynomial, and

R_t is the distance cutoff point. Combined with Formula (4), I_{rc} can be expressed as follows:

$$I_{rc} = \frac{f_R(R_0)}{f_R(R)} \cdot I(\lambda, R, \theta) = \begin{cases} \frac{I(\lambda, R, \theta) \sum_{k=0}^K a_k R_0^k}{\sum_{k=0}^K a_k R^k}, & R \leq R_t, \\ \frac{I(\lambda, R, \theta) \sum_{l=0}^L b_l (1/R_0)^l}{\sum_{l=0}^L b_l (1/R)^l}, & R > R_t. \end{cases} \quad (6)$$

Similarly, satisfying the conditions of a reference target with reflectivity, λ is scanned at different incident angles θ and at a constant distance R_0 , deduced from Formula (3). The intensity after the incident angle correction $I_{\theta c}$ can be written as

$$I_{\theta c} = \frac{f_R(R_0) f_{\theta}(\cos \theta_0)}{f_R(R_0) f_{\theta}(\cos \theta)} \cdot I(\lambda, R, \theta) = \frac{f_{\theta}(\cos \theta_0)}{f_{\theta}(\cos \theta)} \cdot I(\lambda, R, \theta). \quad (7)$$

In the above incident angle-intensity data, $0 \leq \cos \theta \leq 1$. According to the Weierstrass theorem mentioned above, the intensity regression value $I(\lambda, R_0, \theta)$ under the reference reflectivity λ_0 and the reference incident angle θ_0 can also be obtained by polynomial regression fitting. A cosine polynomial is used for modeling $f_{\theta}(\cos \theta)$, as shown in the following formula:

$$f_{\theta}(\cos \theta) = I(\lambda, R_0, \theta) = \sum_{m=0}^M c_m (\cos \theta)^m, \quad (8)$$

where c_m is the coefficient of the incident angle cosine polynomial and M is the order of the incident angle cosine polynomial. By substituting the obtained $f_{\theta}(\cos \theta)$ into Formula (7), $I_{\theta c}$ is established as follows:

$$I_{\theta c} = \frac{f_{\theta}(\cos \theta_0)}{f_{\theta}(\cos \theta)} \cdot I(\lambda, R, \theta) = \frac{I(\lambda, R, \theta) \sum_{m=0}^M c_m (\cos \theta_0)^m}{\sum_{m=0}^M c_m (\cos \theta)^m}. \quad (9)$$

2.4. Parameter Estimation of the MLS Intensity Correction Model. In this paper, the elbow rule method and the least square method are used to determine the polynomial order and the fitting polynomial coefficients. For example, in the polynomial in the short-distance segment ($R \leq R_t$) of Formula (5), the values of K and a_k need to be determined in a way that enables the error sum of squares $S(a_0, a_1, \dots, a_K)$ between $I(\lambda, R, \theta_0)$ and $I(\lambda, R, \theta)$ to be minimized. $S(a_0, a_1, \dots, a_K)$ is the cost function. The root mean square error (RMSE) of the intensity sample is defined as

$$\begin{aligned} \text{RMSE} &= \sqrt{\frac{S(a_0, a_1, \dots, a_K)}{N}} \\ &= \sqrt{\frac{\sum_{n=1}^N (I(\lambda, R, \theta_0)_n - I(\lambda, R, \theta)_n)^2}{N}}, \end{aligned} \quad (10)$$

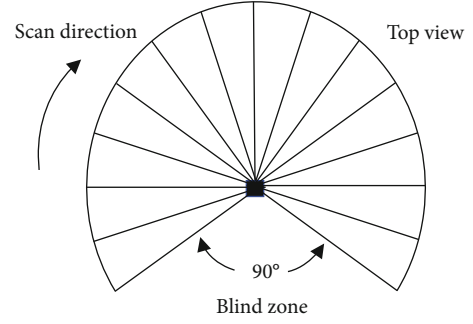


FIGURE 3: UTM-30LX scan area.



FIGURE 4: Standard diffuse reflector.

where N is the number of $I(\lambda, R, \theta)$ and $I(\lambda, R, \theta_0)$ and n is the serial number of the intensity sample. The smaller the RMSE value is, the lower the fitting error is. In general, the higher the polynomial order (the larger the K value), the smaller the RMSE value. However, it should be noted that if the fitting order K value is selected too high, it is easy to overfit when the distance correction model is applied to the actual scene for intensity correction and the correction effect is poor. If the K value is too low, the distance intensity correction cannot be carried out effectively. To avoid overfitting, according to the change curve of different polynomial order RMSE values, the elbow rule in machine learning is used to determine that the elbow position is the best polynomial order.

To obtain the polynomial coefficient a_k under $S(a_0, a_1, \dots, a_K)$ the minimum value, the polynomial in the short-distance segment ($R \leq R_t$) of Formula (5) is brought into Formula (10). Then, the partial derivative of each polynomial coefficient a_k is transformed into the problem of finding the extremum; thus,

$$\frac{\partial S(a_0, a_1, \dots, a_K)}{\partial a_k} = 2 \sum_{n=1}^N \left(\sum_{k=0}^K a_k R_n^k - I(\lambda, R, \theta)_n \right) R_n^k = 0. \quad (11)$$

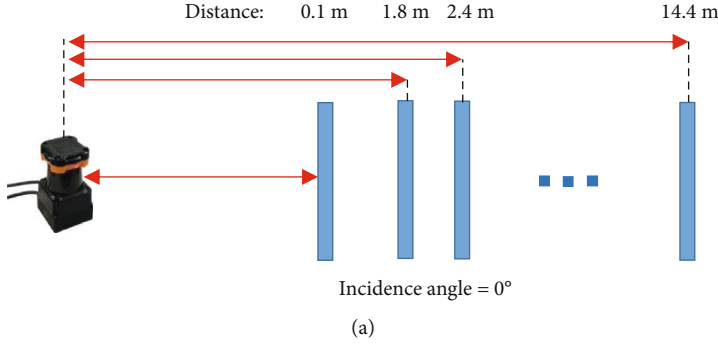


FIGURE 5: Intensity measurement at different sites in incident angle experiments: (a) intensity measurement at different sites of 0.1 m-14.4 m under the reference incident angle of 0° ; (b) the picture of the scanning scene.

It can be further obtained that

$$\sum_{k=0}^K a_k R_n^k = I(\lambda, R, \theta)_n. \quad (12)$$

The linear formulas can be obtained as follows:

$$\begin{bmatrix} R_1^0 & R_1^1 & \cdots & R_1^k \\ R_2^0 & R_2^1 & \cdots & R_2^k \\ \vdots & \vdots & \ddots & \vdots \\ R_n^0 & R_n^1 & \cdots & R_n^k \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_k \end{bmatrix} = \begin{bmatrix} I(\lambda, R, \theta)_1 \\ I(\lambda, R, \theta)_2 \\ \vdots \\ I(\lambda, R, \theta)_n \end{bmatrix}. \quad (13)$$

In this way, the polynomial coefficients $[a_0, a_1, \dots, a_k]$ of $f_R(R)$ can be solved by the Gaussian elimination method. Similarly, bringing the polynomial in the long-distance segment ($R > R_t$) of Formula (5) and the cosine polynomial of Formula (8) into Formula (10), based on the elbow method and the least square method, the corresponding polynomial coefficient parameters $[b_0, b_1, \dots, b_l]$ and $[c_0, c_1, \dots, c_m]$ are obtained.

2.5. Evaluation Indices of the MLS Intensity Correction Model. To evaluate the effect of distance and incident angle correction models on MLS intensity correction, the coefficient of variation CV is used to represent the degree of discretization before and after intensity correction. The coefficient of variation CV is defined by

$$CV = \frac{STD}{Mean} \times 100\%, \quad (14)$$

where STD is the standard deviation of intensity and Mean is the mean value of intensity. CV is positively correlated with the dispersion of intensity data. The smaller its value, the smaller the data dispersion. In addition, the evaluation index

variance-to-mean ratio ε of the intensity correction model is defined as follows:

$$\varepsilon = \frac{CV_{cor}}{CV_{ori}} = \frac{(STD/Mean)_{cor}}{(STD/Mean)_{ori}}. \quad (15)$$

CV_{ori} and CV_{cor} represent the variation coefficient of the original intensity and the variation coefficient of the corrected intensity, which can be calculated by Formula (14). When ε is less than 1, it means that the variability of the intensity value after correction is less than that before correction, and the intensity correction model is effective. The smaller the value of ε is, the better the intensity consistency of the model after correction.

3. Experiment and Data Acquisition

3.1. 2D LiDAR and Diffuse Reflector Plate. UTM-30LX 2D LiDAR [34] utilizes time-of-flight technology to measure the distance. The maximum detection distance of the 2D LiDAR is 30 m. The accuracy is ± 30 mm within the range of 0.1-10 m and ± 50 mm within the range of 10-30 m. The 2D LiDAR laser beam has a wavelength of 905 nm, an angular resolution of 0.25° , a scanning period of 25 ms, and a scanning range of 0° to 270° , as shown in Figure 3. A total of 1081 target distances and intensity values in different directions are included in one frame of scan data and stored in a file as 4-byte and 2-byte unsigned integer data.

The reference target is a standard diffuse reflectance plate with a size of 50 cm \times 50 cm and a reflectance of 50% (that is, the reflectance of the reference target is $\lambda_0 = 0.5$), as shown in Figure 4. The diffuse reflector plate is composed of highly diffuse reflector materials. When irradiated by a laser beam with a wavelength of 905 nm, the diffuse reflector plate can be regarded as a standard Lambert body.

The UTM-30LX 2D LiDAR was placed vertically against the diffuse reflectance plate in an indoor environment. Since

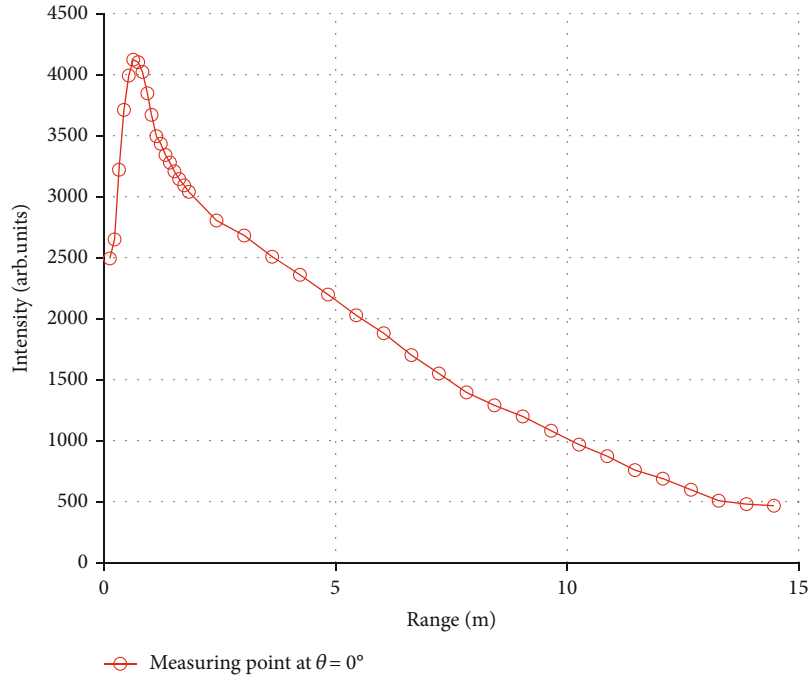


FIGURE 6: The relationship between the distance and intensity of different sites in the range of 0.1 m-14.4 m under the reference angle of 0°.

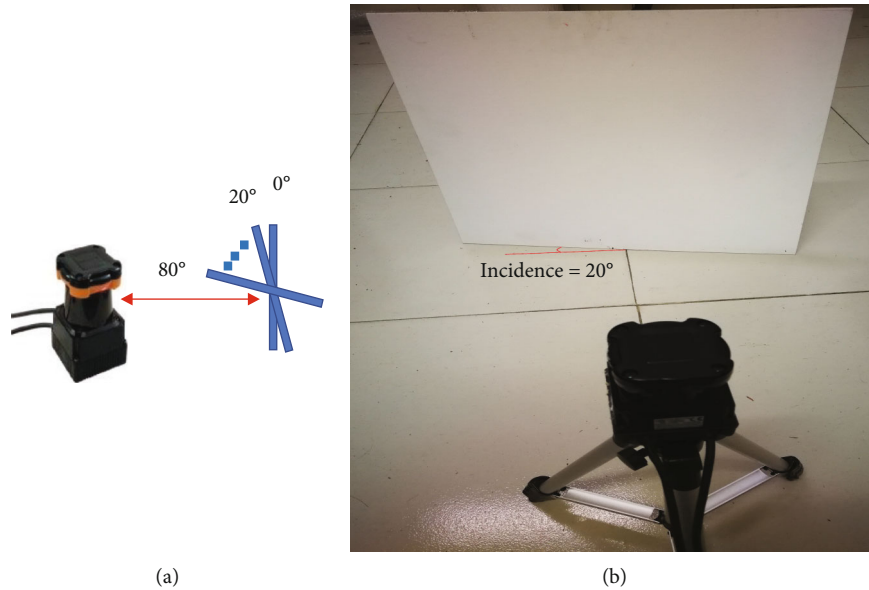


FIGURE 7: Intensity measurement at different sites in distance experiments: (a) intensity measurement at different sites of 0°-80° at a reference distance of 1.2 m; (b) the picture of the scanning scene.

the laser performance in different emission directions is the same, only the single laser beam in the middle was used to obtain the distance-intensity measurement value. The order and coefficient of polynomials $f_R(R)$ and $f_\theta(\cos \theta)$ were obtained by designing two groups of experiments of distance and incident angle correction model acquisition, respectively.

3.2. Distance Correction Model Acquisition Experiment. The reference incident angle was set to $\theta_0 = 0^\circ$, and the

distance-intensity data $(R_i, I(\lambda_0, R_i, \theta_0))$ were obtained by setting different distance sites to scan the diffuse reflectance plate. Since the intensity measurement error was large when the distance was long, the intensity data of the diffuse reflectance plate were collected within the range of 0.1 m-14.4 m by taking half of the maximum detection distance. Nonuniform sampling was adopted, and a total of 39 sites were set. Dense sampling with an interval of 0.1 m was carried out in the range of 0.1 m-1.8 m, and 18 sites were set; sparse

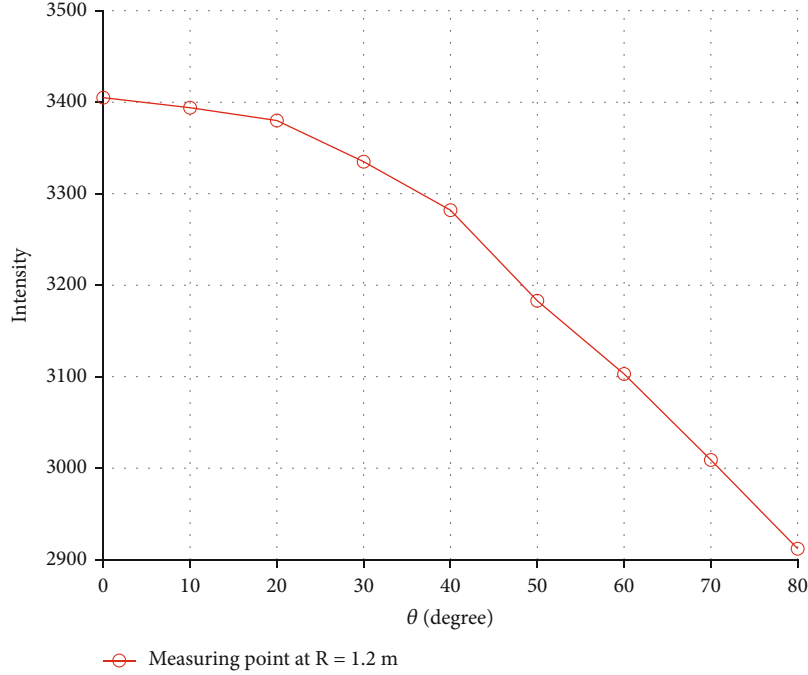


FIGURE 8: The relationship between the incident angle and intensity at different sites from 0° to 80° at a reference distance of 1.2 m.

sampling with an interval of 0.6 m was carried out in the range of 1.8 m-14.4 m, and 21 sites were set. The experimental scenario is shown in Figure 5.

The 2D LiDAR was fixed on the ground and placed horizontally opposite the diffuse reflector plate. To ensure that the distance-intensity data $(R_i, I(\lambda_0, R_i, \theta_0))$ were obtained at the reference angle $\theta_0 = 0^\circ$, the laser beam at the intermediate point of the LiDAR at each site (frame 541) was perpendicular to the diffuse reflector plate. To improve the measurement accuracy of distance-intensity data, 10 frames were collected continuously at each site, and then, the average intensity value was calculated. Setting the reference incident angle $\theta_0 = 0^\circ$, within the range of 0.1 m-14.4 m, the relationship between different distances and the corresponding intensity is shown in Figure 6.

In the range of 0.1-0.7 m, the intensity increases rapidly with an increasing distance value, and after 0.7 m, the intensity decreases slowly with an increasing distance value. Therefore, the polynomial function can be fitted by Formula (5) with 0.7 m as the distance cutoff point.

3.3. Incident Angle Correction Model Acquisition Experiment. The reference distance was set to $R_0 = 1.2$ m, and the intensities of different incident angles within the range of 0° - 80° were collected. To ensure that the incident angle-intensity data $(\theta_i, I(\lambda_0, R_0, \theta_i))$ were measured at $R_0 = 1.2$ m, the laser beam at frame 541 of LiDAR was positioned directly against the central axis of the diffuse reflector plate. The position of LiDAR was fixed, and the diffuse reflector plate central axis was taken to change the position within the range of 0° - 80° at intervals of 10° . The experiment is shown in Figure 7.

Ten frames were collected by continuous scanning at each angle site (frame 541), the incident angle-intensity data

TABLE 1: RMSE for K , L , and M values at different polynomial orders.

Order	RMSE		
	K	L	M
1	182.0	138.5	7.3
2	75.6	37.6	7.3
3	38.6	17.6	5.6
4	17.6	15.4	4.2
5	17.5	10.0	4.2
6	8.7	9.8	3.4

of the laser beam position were obtained, and the average intensity value was calculated. The experimental result is shown in Figure 8.

3.4. Determining the Order and Coefficient of the Polynomial Model. The least square method described in Section 2.4 is used to determine the order and coefficient of the polynomial model of distance and incident angle. To avoid overfitting the intensity data, according to the elbow rule in machine learning, the best order of the polynomial is determined to be the elbow position. To determine the ideal polynomial order, the distance correction model acquisition experiment in Section 3.2 and the incident angle correction model acquisition experiment in Section 3.3 were repeated 5 times to expand the dataset. Table 1 lists RMSE for different polynomial order K , L , and M values. The process of determining the optimal values of K , L , and M based on RMSE is shown in Figure 9. It is proven that $K = 4$ and $L = 3$ are the ideal distance piecewise polynomial orders in the short-distance and long-distance segments, respectively,

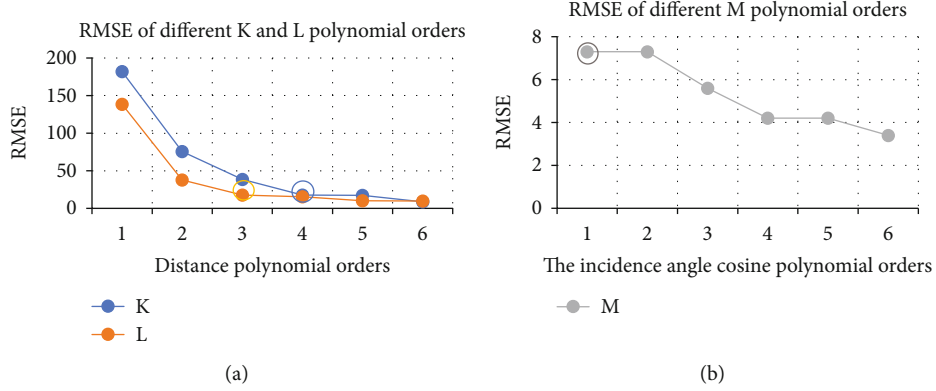


FIGURE 9: The process of determining the optimal values of K , L , and M based on RMSE. (a) The corresponding RMSE value change curve as the distance polynomial order K and L increase. (b) The corresponding RMSE value change curve as the incidence angle cosine polynomial order M increases.

TABLE 2: Specific expression of the distance and incident angle cosine polynomial correction function.

$$f_R(R) = 1.2328 \times 10^5 R^4 - 2.1138 \times 10^5 R^3 + 1.2268 \times 10^5 R^2 - 2.39 \times 10^4 R + 3.9332 \times 10^3, R \leq 0.7\text{m}.$$

$$f_R(R) = 6.0276 \times 10^3 (1/R)^3 - 1.5033 \times 10^4 (1/R)^2 + 1.2582 \times 10^4 (1/R) - 99.7915, R > 0.7\text{m}.$$

$$f_\theta(\cos \theta) = 607.177 \cos \theta + 2.8033 \times 10^3, 0^\circ \leq \theta \leq 80^\circ.$$

and $M=1$ is the ideal incident angle cosine polynomial order.

As shown in Figure 9(a), the RMSE decreases substantially with increasing K and L values in the distance correction experiment, and $K \geq 4$ and $L \geq 3$ tend to be flat. According to the principle of the elbow rule in machine learning, the order of the distance polynomial model is set to $K=4$ and $L=3$. The blue and orange circles in Figure 9(a) represent the optimal values of K and L , respectively. In Figure 9(b), compared with K and L , the different RMSE values and fluctuation ranges corresponding to the order M of the incident angle cosine polynomial are relatively small. To avoid overfitting, the order of the incident angle polynomial model is set to $M=1$. The gray circle in Figure 9(b) represents the optimal value of M .

In the short-distance segment ($R \leq R_t$), the distance-intensity data ($R_i, I(\lambda_0, R_i, \theta_0)$) are substituted into Formula (13) for calculation. Then, $a_0 = 3.9332 \times 10^3$, $a_1 = -2.39 \times 10^4$, $a_2 = 1.2268 \times 10^5$, $a_3 = -2.1138 \times 10^5$, and $a_4 = 1.2328 \times 10^5$. Similarly, in the long-distance segment ($R > R_t$), $b_0 = -99.7915$, $b_1 = 1.2582 \times 10^4$, $b_2 = -1.5033 \times 10^4$, and $b_3 = 6.0276 \times 10^3$. The incident angle polynomial coefficients $c_0 = 2.8033 \times 10^3$ and $c_1 = 607.177$ can also be obtained by using the incident angle-intensity data ($\theta_i, I(\lambda_0, R_0, \theta_i)$). The specific expressions of the distance and incident angle cosine polynomial correction functions are shown in Table 2.

The function expression in Table 2 is only applicable to the point cloud intensity data obtained by UTM-30LX 2D LiDAR scanning the same target reflectivity Lambert body and can only fit the intensity data within the range of

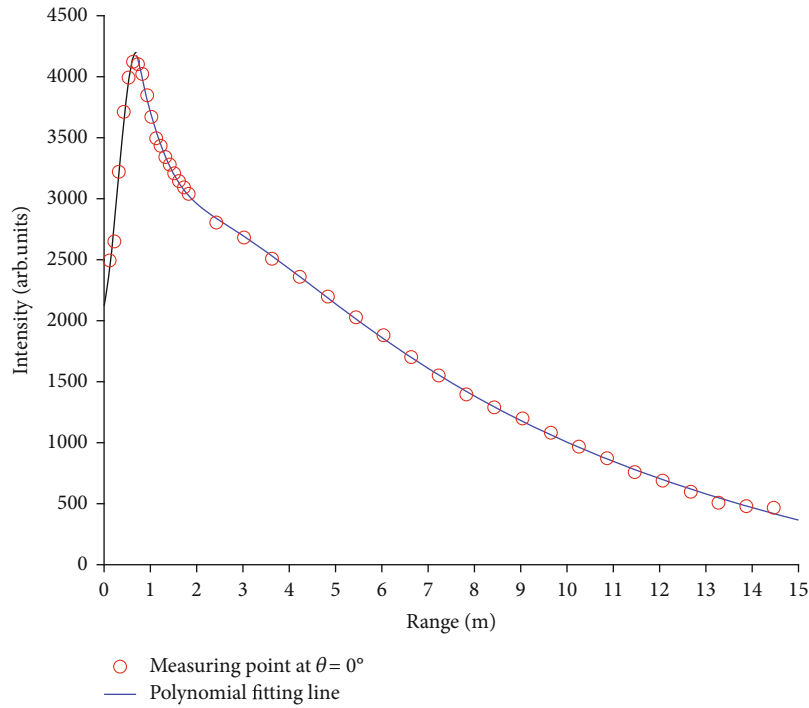
0.1 m-14.4 m and incident angle value of 0° - 80° . The fitting results of the piecewise distance and incident angle cosine polynomial are shown in Figure 10.

$R_0=1.2$ m and $\theta_0 = 0^\circ$ are taken and substituted by the distance-intensity data ($R_i, I(\lambda_0, R_i, \theta_0)$) and the incident angle-intensity data ($\theta_i, I(\lambda_0, R_0, \theta_i)$) mentioned above into Formulas (6) and (9) for intensity correction. The intensity after correction at different sites is shown in Figure 11. The intensity distribution of each site is approximately 3400, although there are some model errors.

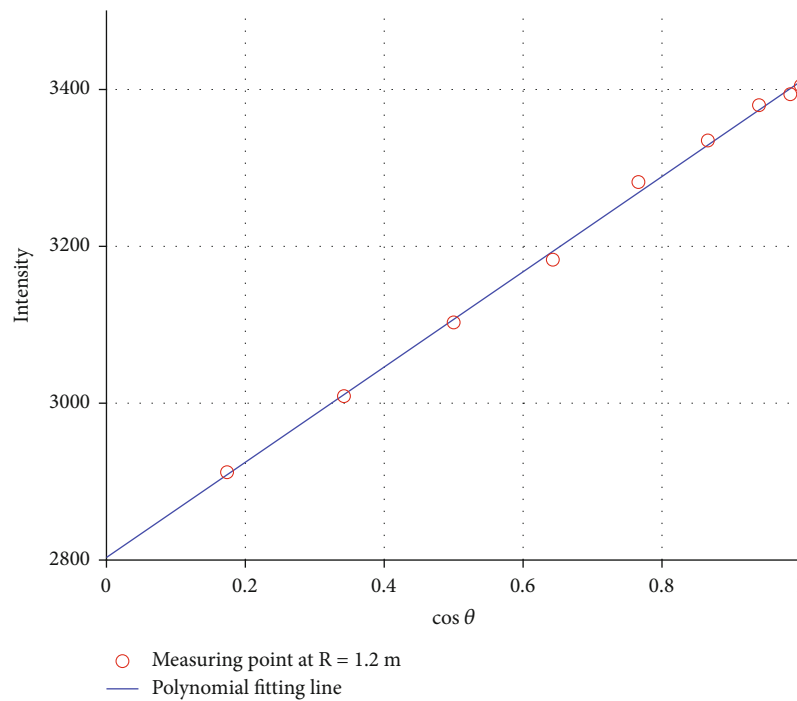
4. Results and Analysis

4.1. MLS 2D LiDAR Point Cloud Data Acquisition System and Coordinate System

4.1.1. MLS 2D LiDAR Point Cloud Data Acquisition System. To verify the validity of the intensity correction model for the distance and incident angle, a flat wall was selected as the experimental object, which was roughly considered a Lambert surface. To visually observe the intensity point cloud, the intensity value was converted into RGB in CloudCompare, a software point cloud development tool. The 3D point cloud RGB intensity map was also converted into a 2D pseudocolor map for display. The actual scene diagram of the experimental scanning wall is shown in Figure 12. The MLS 2D LiDAR measurement system emitted laser pulses in all directions through internal rotating optical components to form a 2D fan-shaped scanning surface. The moving platform carried the laser pulses along the direction perpendicular to the scanning surface to realize the 3D



(a) $f_R(R)$ ($K = 4, L = 3$)



(b) $f_\theta(\cos \theta)$ ($M = 1$)

FIGURE 10: Results of polynomial fitting: (a) the result of piecewise fitting of the distance polynomial; (b) the result of cosine polynomial fitting of the incident angle.

measurement of the target surface. A moving slide mounted UTM-30LX 2D LiDAR was used to move in a straight line at a constant speed of 0.01 m/s in a direction parallel to the wall. The fan-shaped scanning surface was placed directly against the wall. Inside the red rectangular box was the wall

study area, with a height of 1.5 m from top to bottom and a width of 1.2 m from left to right. The vertical distance from the mobile slide to the wall was 0.5 m, and the range of distance and incident angle of the wall study area was [0.52 m-1.55 m] and [0° - 74°], respectively.

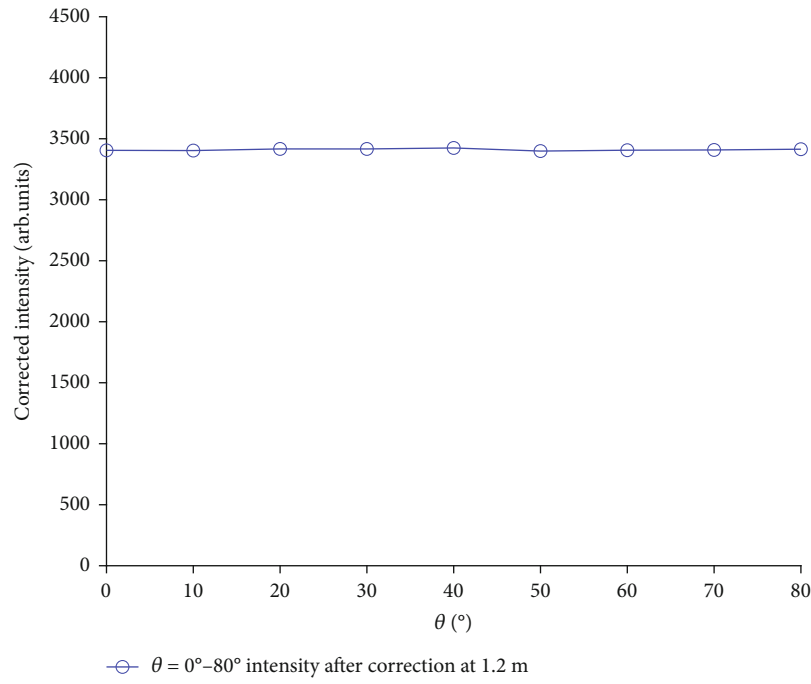
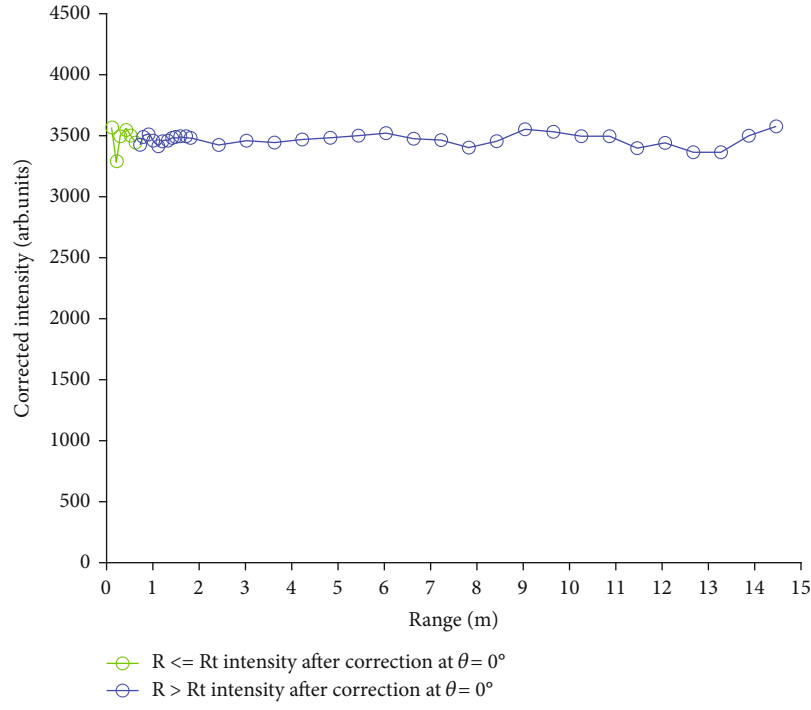


FIGURE 11: Corrected intensity at different sites in distance and incident angle experiments. (a) The intensity value of different distances from 0 to 15 m was corrected by the distance polynomial model under the reference incident angle of 0° . (b) The intensity value of different incident angles from 0° to 80° was corrected by the incident angle cosine polynomial model under the reference distance of 1.2 m.

4.1.2. *MLS 2D LiDAR Coordinate System.* The starting point position of LiDAR is taken as the origin, and the right-handed coordinate system of the point cloud of the MLS 2D LiDAR measurement system is established, as shown in Figure 13.

The x -axis is the moving direction of LiDAR, the y -axis is the scanning direction of LiDAR, the z -axis is vertical to the ground, $\alpha(i, j)$ is the scanning angle of the i^{th} measuring point in the j^{th} frame, and R is the distance between the LiDAR laser receiver and the measuring

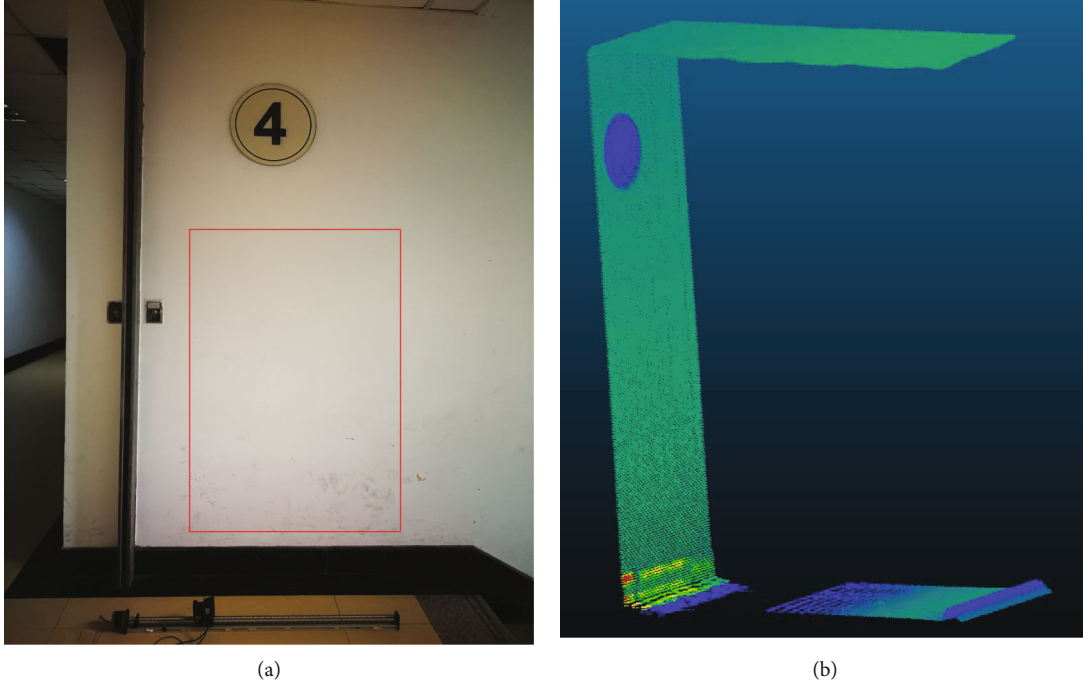


FIGURE 12: The scene of scanning the wall surface: (a) the picture of the scanning scene; (b) 2D pseudocolor intensity map of the scanning scene.

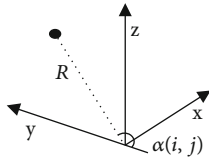


FIGURE 13: Coordinate system of point clouds acquired by MLS 2D LiDAR.

point. The calculation of the coordinate system is as follows:

$$\begin{cases} x(i, j) = v\Delta t \cdot j, \\ y(i, j) = -R \cdot \cos \alpha(i, j), \\ z(i, j) = R \cdot \sin \alpha(i, j), \end{cases} \quad (16)$$

where i is the in-frame number, j is the frame number of the measured points, and $x(i, j)$, $y(i, j)$, and $z(i, j)$ represent the X , Y , and Z coordinates of the i^{th} measuring point in the j^{th} frame. v is the speed of the moving slide platform in the direction of the motion, and Δt is the scan cycle of the 2D LiDAR.

The geometric relationship between the moving slide with 2D LiDAR and the wall is shown in Figure 14. The coordinate of LiDAR center point O at different scanning moments was set as $(x(i, j), 0, 0)$; the 3D coordinate of the i^{th} measurement point in frame j^{th} was $p(i, j) = (x(i, j), y(i, j), z(i, j))$, incident laser vector $\mathbf{l}(i, j) = (0, y(i, j), z(i, j))$, and the normal vector of scanning point $\mathbf{n}(i, j) = (n_1, n_2, n_3)$.

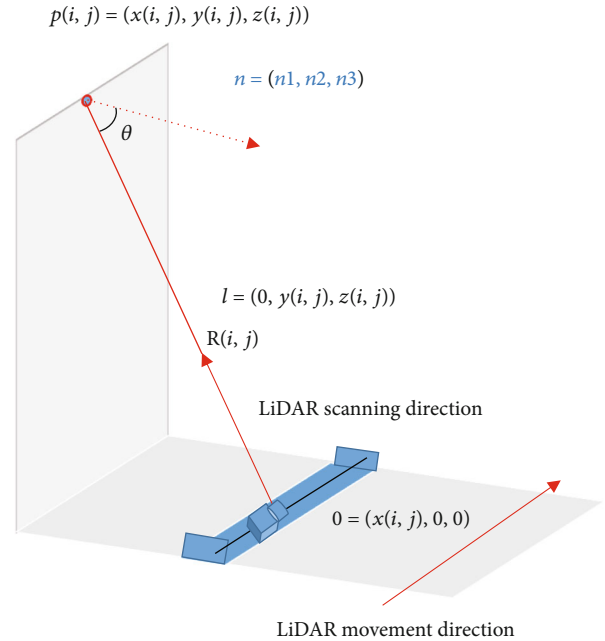
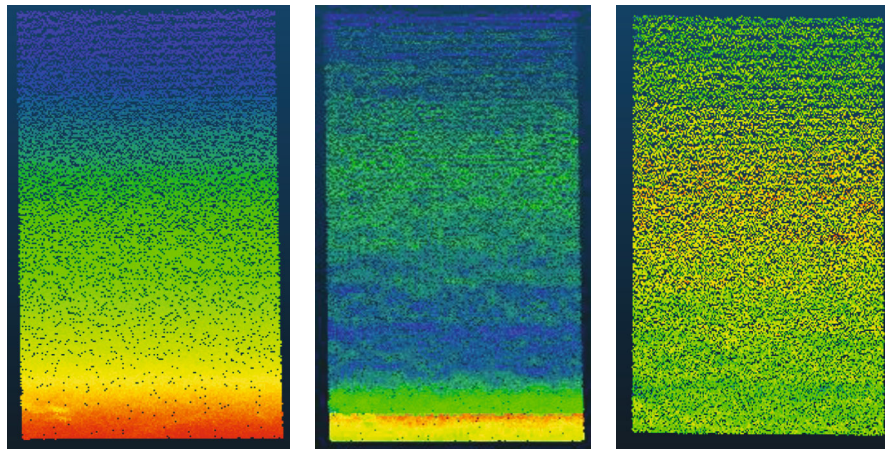
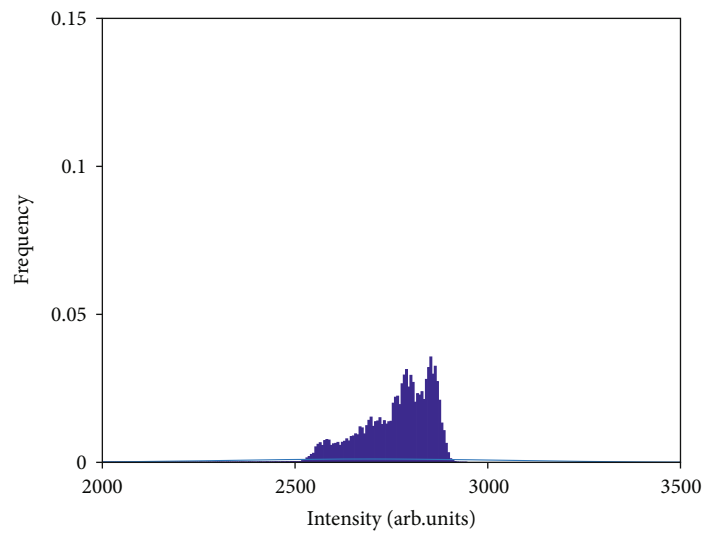


FIGURE 14: Moving slide and wall geometrical relationship diagram.

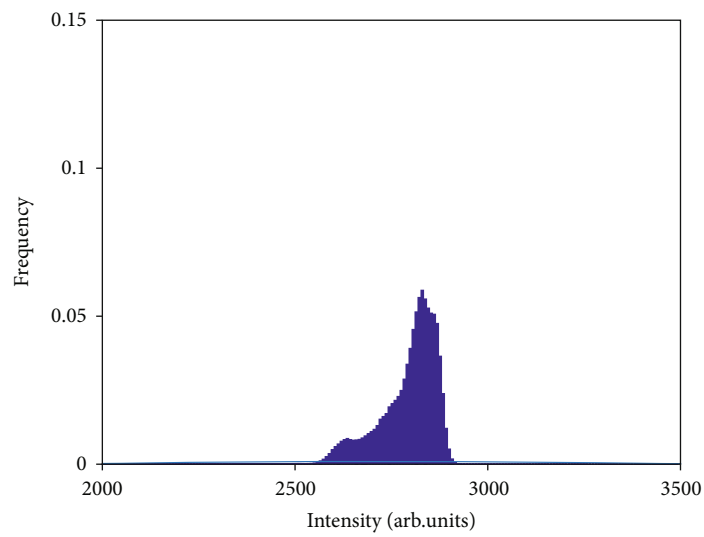
4.2. Obtaining R and $\cos \theta$. Generally, the density of point cloud data obtained by 3D LiDAR is relatively high. Usually, to reduce data redundancy, a support vector machine (SVM) is used to integrate different feature weights of the point cloud into the classifier for training [35–38]. In addition, to improve the efficiency of neighborhood searching, the k -D tree algorithm is usually used for downsampling 3D LiDAR



(a) (b) (c)



(d)



(e)

FIGURE 15: Continued.

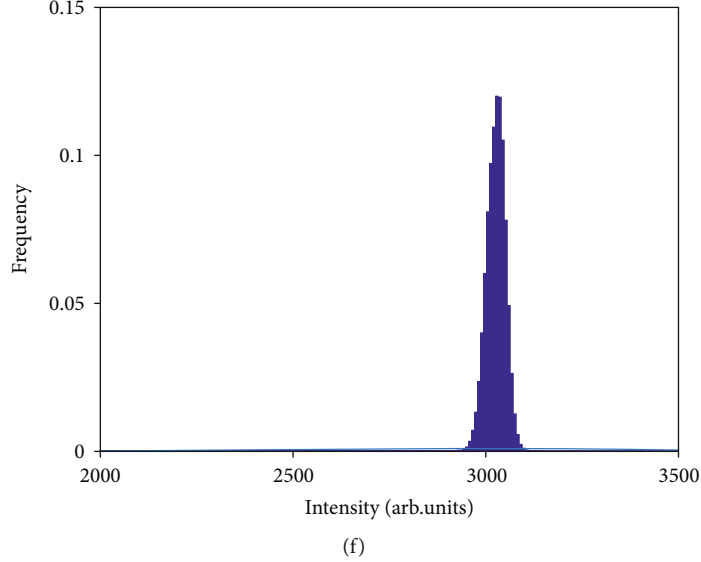


FIGURE 15: The intensity 2D pseudocolor map and intensity distribution histogram of the red rectangular wall area described in Figure 12 before and after intensity correction: (a) original intensity pseudocolor chart; (b) intensity pseudocolor chart after the incident angle correction; (c) intensity pseudocolor chart of distance correction after the incident angle correction; (d) histogram of original intensity distribution; (e) histogram of intensity distribution after incident angle correction; (f) intensity distribution histogram of distance correction after the incident angle correction.

point clouds [39]. Compared with 3D LiDAR point clouds, the data structure of 2D LiDAR point clouds is relatively simple. It only needs to be indexed by the grid in the established coordinate system, and the k -nearest neighbor algorithm [40] can be used to establish the point cloud neighborhood set. Referring to our previous research [41], a 3D spherical neighborhood $S(i, j)$ is defined as a set of the nearest adjacent points within a sphere centered at the measurement point $p(i, j)$ with a radius of δ . According to the idea of the least square method, the plane is fitted in the neighborhood set $S(i, j)$, and the normal vector $\mathbf{n}(i, j)$ corresponding to each measuring point $p(i, j)$ is obtained. Then, $\cos \theta(i, j)$ of the LiDAR laser receiver to each measuring point can be calculated according to Formula (17). Obviously, $R(i, j)$ can be obtained directly.

$$\cos \theta(i, j) = \frac{|\mathbf{l}(i, j) \cdot \mathbf{n}(i, j)|}{|\mathbf{l}(i, j)| |\mathbf{n}(i, j)|}. \quad (17)$$

Since $\mathbf{l}(i, j)$ does not change, the accuracy of $\cos \theta(i, j)$ depends on $\mathbf{n}(i, j)$. In general, the smaller the value of δ , the smaller the fitting plane, and the more accurate the value of $\mathbf{n}(i, j)$. However, if the value of δ is too small, the point cloud data in the two adjacent frames cannot be included in the neighborhood; then, the plane cannot be fitted either. In the MLS 2D LiDAR measurement system, the value of δ is closely related to distance $R(i, j)$ and the resolution of $\alpha(i, j)$. The resolution of the moving direction and the scanning direction of LiDAR is defined as Δx and Δs , respectively:

$$\begin{cases} \Delta x = v \Delta t, \\ \Delta s = R(i, j) \sin \Delta \alpha. \end{cases} \quad (18)$$

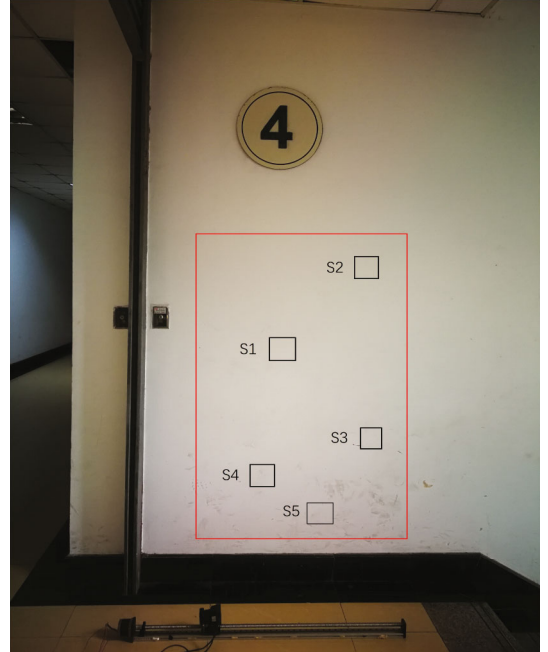


FIGURE 16: Different areas of S1-S5 randomly selected on the wall.

$\Delta \alpha$ is the 2D LiDAR angular resolution, and its value is constant. The larger the $R(i, j)$ is, the greater the value of Δs . Therefore, if the value of δ is less than Δs , in the LiDAR scanning direction, all of the in-frame data points adjacent to the measurement point $p(i, j)$ cannot be included in $S(i, j)$. Then, $\mathbf{n}(i, j)$ cannot be obtained by plane fitting. According to Formula (18), the necessary conditions for the measurement point $p(i, j)$ to have

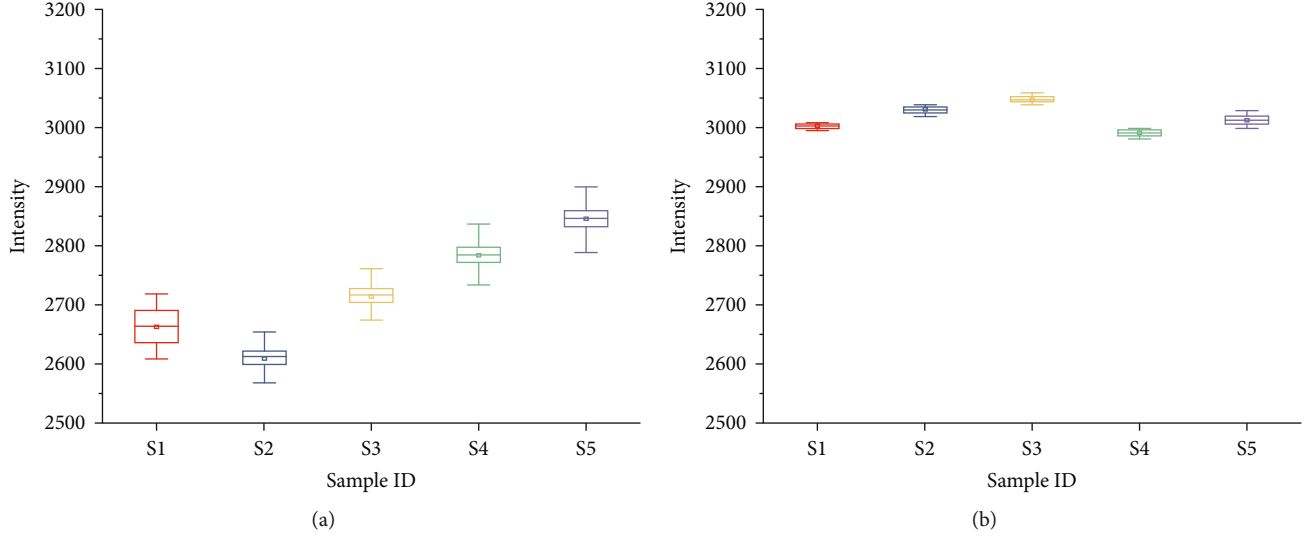


FIGURE 17: Intensity box diagram of different areas on the same wall before and after correction: (a) intensity box diagram of area S1-S5 before correction; (b) intensity box diagram of area S1-S5 after correction.

TABLE 3: The distribution of intensity MAX, MIN, Mean, STD, CV, and the evaluation index ε at S1-S5 different regions before and after correction.

	Intensity	Region S1	Region S2	Region S3	Region S4	Region S5
MAX	Original	2681	2668	2750	2832	2896
	Corrected	3042	3015	3061	3006	3025
MIN	Original	2598	2346	2657	2710	2775
	Corrected	3019	2990	3037	2981	3002
Mean	Original	2612	2557	2717	2785	2846
	Corrected	3031	3003	3049	2992	3013
STD	Original	16.7	16.9	15.9	18.1	20.44
	Corrected	5.8	4.3	5.6	5.3	8.3
CV	CV _{ori}	0.0064	0.0066	0.0058	0.0065	0.0072
	CV _{cor}	0.0019	0.0014	0.0018	0.0018	0.0028
ε		0.297	0.212	0.310	0.277	0.389

neighborhood points in the LiDAR scanning direction can be deduced:

$$\begin{aligned} \delta &> R(i, j)_{\max} \sin \Delta \alpha, \\ R(i, j) &\leq R(i, j)_{\max}. \end{aligned} \quad (19)$$

As for UTM-30LX, the value of $\Delta \alpha$ is 0.25° . In the wall intensity correction experiment, the maximum distance of LiDAR from the wall is not more than 5.5 m. According to Formula (19), $R(i, j)_{\max} \sin \Delta \alpha$ is 0.024. Considering that the range resolution of the LiDAR sensor within 10 m is ± 30 mm, the value of δ is set to 0.03 m.

4.3. Correction of the Wall Point Cloud Intensity Data. According to the calculation method of the incident angle

and distance, the incident angle θ and the distance value R of each measurement point in the wall point cloud intensity data were calculated. The study of Tan and Cheng [42] shows that since the impact of the incident angle and distance on intensity is independent, the sequence of the incident angle correction and distance correction does not affect the correction results. According to the established intensity correction model, the reference incident angle and reference distance are set to be $\theta_0 = 0^\circ$ and $R_0 = 1.2$ m, respectively. The incident angle and distance are corrected for the wall point cloud intensity data. Figure 15 shows the intensity 2D pseudocolor map and intensity distribution histogram of the red rectangular wall area described in Figure 12 before and after intensity correction.

As seen from the RGB diagram of the original intensity in Figure 15(a), the intensity value at different distances

TABLE 4: Comparison of ε values of the white wall intensity corrected by the Tan method and the method proposed in this paper.

Method	Region S1/A	Region S2/B	Region S3/C	ε	Region S4/D	Region S5/E	Region S6/F
The proposed method	0.297	0.212	0.310		0.277	0.389	—
Tan method	0.607	0.656	0.523		0.787	1.001 \uparrow	1.139 \uparrow

and incident angles is considerably different before correction under the same target reflectivity. The closer the distance or the smaller the incident angle, the higher the intensity value will be. In contrast, the farther the distance or the larger the incident angle, the lower the intensity value will be. The RGB of the intensity after the correction of the incident angle in Figure 15(b) shows that the uniformity of the wall point cloud intensity data has been improved. The RGB diagram of the distance correction intensity after the incident angle correction is shown in Figure 15(c). The influence of the distance and incident angle on intensity is eliminated, and the corrected intensity value is consistent and similar. The histogram of the intensity distribution in Figure 15(d) shows that the original intensity data of the wall point cloud under the same target reflectivity have no obvious distribution pattern. From Figure 15(f), it is clear that after the correction of the incident angle and distance, the intensity value of the wall is concentrated and presents a Gaussian distribution.

4.4. Using an Evaluation Index ε to Verify the Intensity Correction Model

4.4.1. Single-Site Multiregion Verification Experiment. In the wall point cloud intensity data obtained from the above experiment, 5 different areas of $20\text{ cm} \times 20\text{ cm}$ were randomly selected, as shown in Figure 16. According to the established intensity correction model, the intensity data in different areas of S1-S5 were also corrected with reference distance $R_0=1.2\text{ m}$ and incident angle $\theta_0=0^\circ$. In addition, the intensity value before and after correction was statistically analyzed. The intensity box chart is shown in Figure 17. The distribution of intensity maximum value (MAX), minimum value (MIN), mean, STD, CV, and the evaluation index ε before and after correction is shown in Table 3.

As shown in Figure 17(a), the intensity value of different regions is substantially different and fluctuates in a large range within each region before intensity correction. From Figure 17(b), it is evident that the difference in the intensity value between different regions decreases, and the fluctuation range of the intensity within the region also becomes considerably smaller after the correction of the incident angle and distance. It is clear that the consistency of intensity between different areas and within different areas has been substantially improved by comparing the box graphs of intensity data from different areas of walls S1-S5 before and after intensity correction under the same reflectivity of the target.

Table 3 also shows a conclusion similar to that in Figure 17. Due to the small area of the S1-S5 region, the dif-

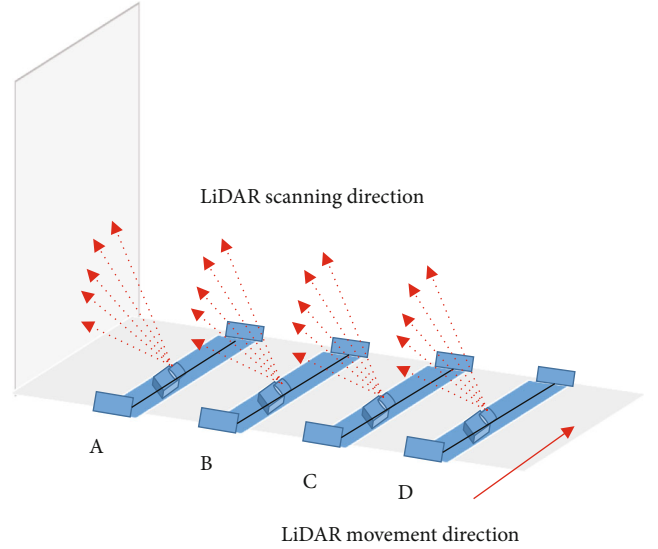


FIGURE 18: Mobile laser scan of 4 sites A, B, C, and D.

TABLE 5: Distance and incident angle range values of different sites A, B, C, and D.

		A	B	C	D
R (m)	MAX	3.20	3.74	4.53	5.34
	MIN	1.52	2.53	3.52	4.53
θ ($^\circ$)	MAX	62	48	39	32
	MIN	0	0	0	0
$\cos \theta$	MAX	1	1	1	1
	MIN	0.4695	0.6691	0.7771	0.8480

ference between the internal distance and incident angle of each region is not large. The STD values of the original intensity data are all below 21. In addition, due to the large difference in distance and incident angle value between different regions, their STD values are also relatively large, with a minimum value of 15.9 and a maximum value of 20.44. After intensity correction, the STD values of the intensity data in each region are all below 10, and the difference between them is small, indicating that the uniformity of the intensity distribution in each region has been substantially improved. The ε value fluctuates approximately 0.3 in S1-S5 different regions, indicating that the intensity consistency of the five same area regions with the range of distance and incident angle $[0.52\text{ m}-1.55\text{ m}, 0^\circ-74^\circ]$ has been massively improved by 70% after applying the proposed correction method.

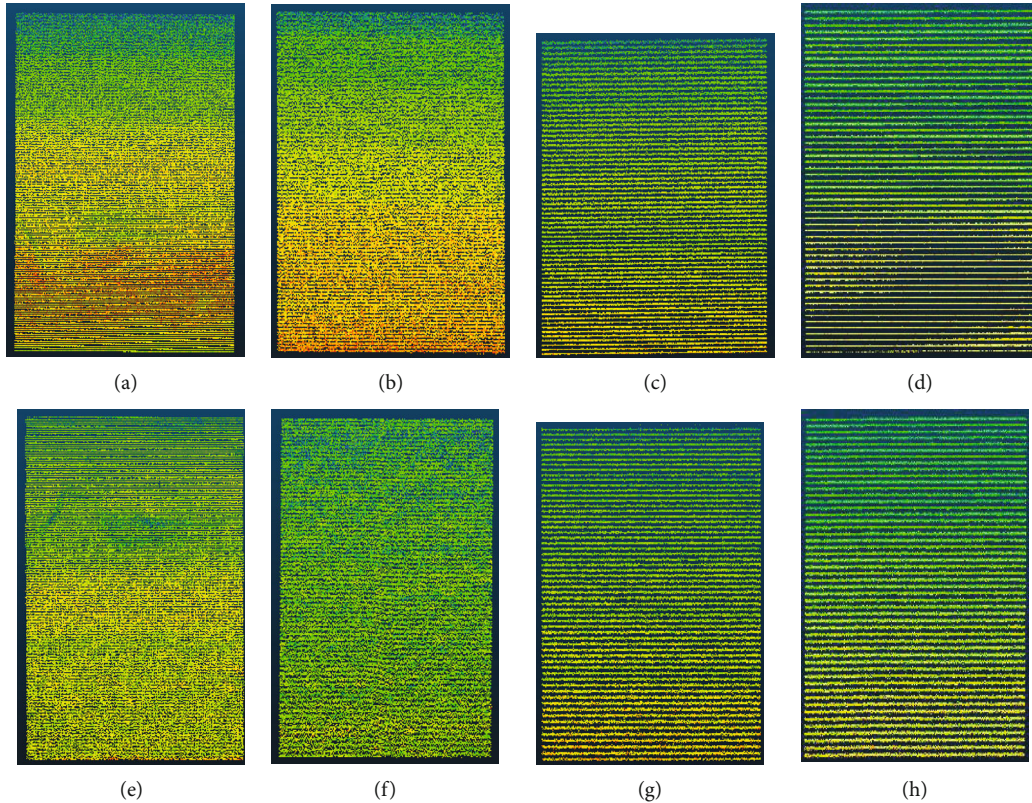


FIGURE 19: The intensity 2D pseudocolor map of the red rectangular area on the wall described in Figure 12 before and after intensity correction at A, B, C, and D 4 sites, respectively. (a–d) Are the original intensity 2D pseudocolor map of sites A, B, C, and D. (e–h) Are the corresponding corrected intensity 2D pseudocolor map of sites A, B, C, and D, respectively.

Similar to the single-site multiregion verification experiment in this paper, Tan and Cheng [29] utilized Faro Focus3D TLS to obtain intensity point cloud data of ordinary white wall surfaces. In the range of distance and incident angle [6.70 m–14.76 m, 0° – 80°], intensity correction experiments were carried out for the A–F region with the same area. The ε value corresponding to the intensity correction of the white wall obtained by the Tan method and the method proposed in this paper is shown in Table 4 below.

The ε value of the Tan method has a large fluctuation range, approximately 0.6, while the ε value of the method proposed in this paper has a small fluctuation range, approximately 0.3. In the Tan experiment, especially in regions E and F, the ε value is greater than 1, indicating that the consistency of the corrected intensity is not as good as the original intensity, and there is an overfitting phenomenon. Moreover, the Tan method only conducts intensity correction research at distances where the intensity value is relatively stable, ignoring the LiDAR short-distance effect, and does not correct the intensity at distances less than 1 m. Overall, we can say that the proposed method in this paper is better than the Tan method and provides a higher accuracy of intensity correction.

4.4.2. Multisite Scanning of the Whole Wall for the Verification Experiment. In the previous single-site experiment, the area of the S1–S5 region was small, and the inci-

dent angle and distance value of the point clouds in each region were not substantially different. The STD of the original intensity was also relatively small and could not effectively reflect the distribution characteristics of the original intensity data of the whole wall. A multisite experiment was devoted to the intensity correction of the same red rectangular area on the wall, as shown in Figure 12, under multiple sites. Four sites were set up: A, B, C, and D. The vertical distances between the sliding platform and the wall were 1.5 m, 2.5 m, 3.5 m, and 4.5 m, respectively. The rectangular area of the wall was scanned to obtain the point cloud intensity data, as shown in Figure 18. According to the established intensity correction model, the intensity data at different sites were corrected at reference distance $R_0=1.2$ m and incident angle $\theta_0=0^\circ$. Table 5 lists the distance and incident angle range values of different sites A, B, C, and D. Figure 19 shows the intensity pseudocolor map of the red rectangular area on the wall described in Figure 12 before and after intensity correction at sites A, B, C and D. The histogram of the intensity distribution before and after intensity correction is shown in Figure 20.

It is evident from Figures 19(a)–19(d) that as the distance increases, the density of point clouds on walls with the same area decreases, while the intensity value shows great differences within the whole wall of each site. It can be clearly seen from Figures 19(e)–19(h) that the variation of wall intensity of the four sites has been considerably

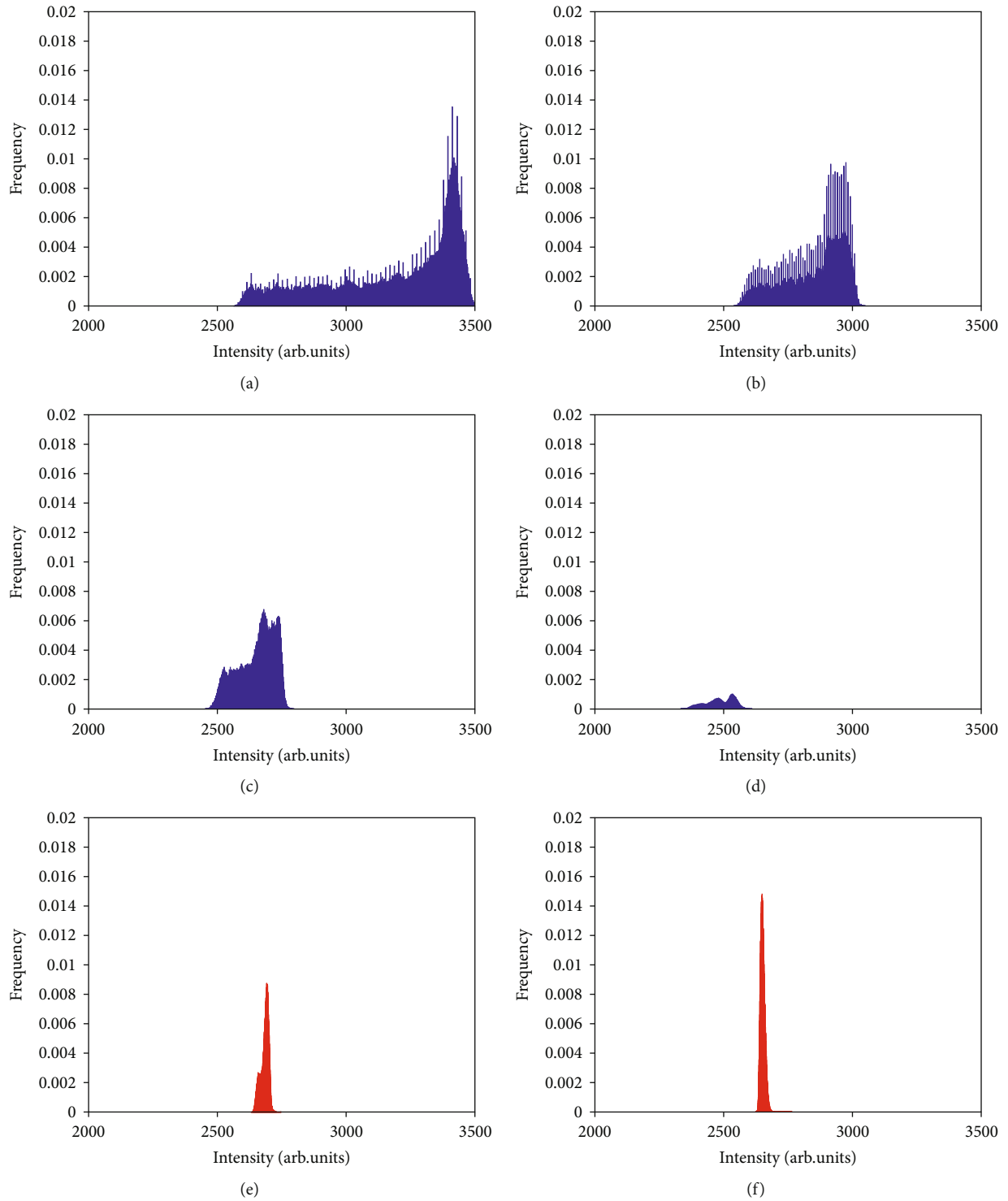


FIGURE 20: Continued.

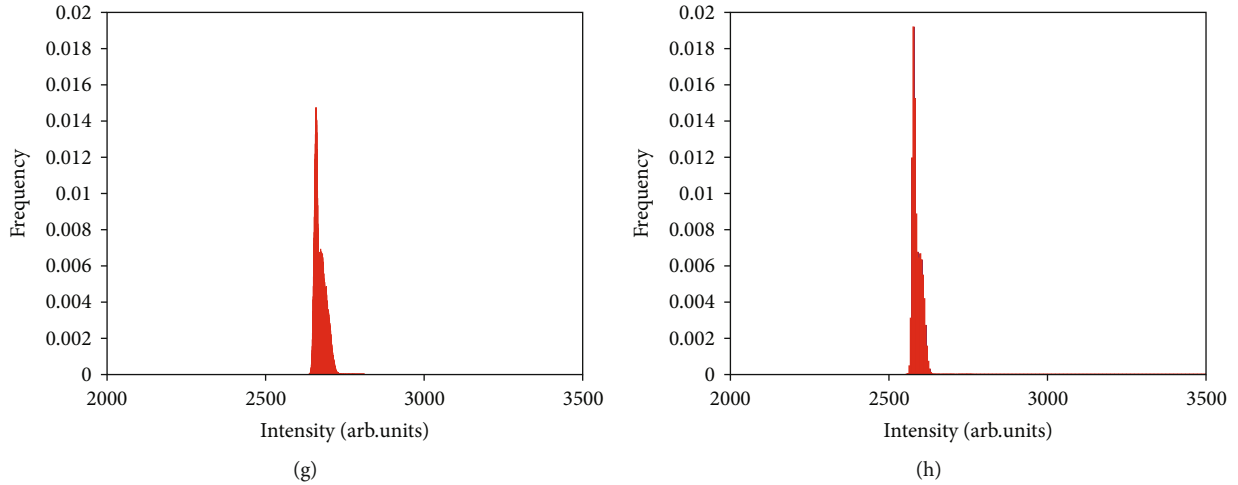


FIGURE 20: Histogram of intensity distribution before and after correction at A, B, C, and D different sites, respectively. (a–d) Are the original intensity distribution histogram of sites A, B, C, and D. (e–h) Are the corrected intensity distribution histograms of sites A, B, C, and D, respectively.

TABLE 6: The distribution of MAX, MIN, Mean, STD, CV of the intensity, and the evaluation index ε at different sites A, B, C, and D before and after intensity correction.

	Intensity	A	B	C	D
MAX	Original	3518	3052	2798	2610
	Corrected	2748	2765	2812	2762
MIN	Original	2563	2541	2454	2334
	Corrected	2632	2624	2638	2554
Mean	Original	3191	2854	2653	2486
	Corrected	2685	2652	2673	2601
STD	Original	254.40	119.50	71.70	53.43
	Corrected	15.50	8.77	16.82	15.86
CV	CV_{ori}	0.0797	0.0419	0.0270	0.0215
	CV_{cor}	0.0058	0.0033	0.0063	0.0061
ε		0.073	0.079	0.233	0.280

reduced after distance and angle correction. By comparing the original intensity distribution in Figures 20(a)–20(d), it was found that quantitatively, the original intensity of each site decreases with an increasing distance and that the degree of dispersion of the intensity distribution is obviously different. As seen from Figures 20(e)–20(h), after intensity correction, the degree of discretization of the intensity value at sites A, B, C, and D is massively reduced, and the intensity value distribution is centralized and approximates a Gaussian distribution. Table 6 lists the distribution of MAX, MIN, Mean, STD, CV of the intensity, and the evaluation index ε at different sites A, B, C, and D before and after intensity correction.

The original intensity distributions of sites A, B, C, and D are [2563, 3518], [2541, 3052], [2454, 2798], and [2334, 2610], respectively. The STD values are 254.40, 119.50, 71.70, and 53.43. The intensity distribution of each site is

successively [2632, 2748], [2624, 2765], [2638, 2812], and [2554, 2762] after correction. The STD value decreases to 15.50, 8.77, 16.82, and 15.86, which is better than the values before intensity correction.

Meanwhile, the values of CV_{ori} at A, B, C, and D are 0.0797, 0.0419, 0.0270, and 0.0215, respectively. The value of CV_{ori} decreases gradually with increasing distance between the sites and finally stabilizes at approximately 5 m. In combination with the distance and incident angle value in Table 5, it can be seen that the closer the site is to the wall, the larger the range of fluctuation of distance and incident angle in the same area will be, leading to the larger difference of original intensity value. Therefore, the values of CV_{ori} at sites A and B are slightly higher than those at sites C and D. After correction, the values of CV_{cor} at the four sites are 0.0058, 0.0033, 0.0063, and 0.0061, respectively. It has been shown that the intensity variability after correction is

smaller and the intensity correction effect is obvious. Among them, the CV_{cor} values of sites A, C, and D are approximately 0.006, while the CV_{cor} value of site B is only 0.0033, indicating that the corrected intensity consistency of site B is substantially better than that of other sites. The same conclusion can be obtained from Figure 19(f). The reason is that the original intensity distribution of site B is more concentrated than that of sites A, C, and D under the joint influence of distance and incident angle, as shown in Figure 20(b). Consequently, the intensity variability after correction at site B is lower than that at sites A, C, and D. Meanwhile, the ε values of sites A, B, C, and D are 0.073, 0.079, 0.233, and 0.280, respectively. This means that the intensity consistency of the four sites increased by 92.7%, 92.1%, 76.7%, and 72%, respectively. From Figure 19, we can intuitively see that the intensity correction effect of the short-distance sites A and B is better than that of the long-distance sites C and D.

Similar to the research topic in this paper, Tan and Cheng [30] adopted a linear interpolation method to fit the relationship between incident angle versus intensity and distance versus intensity. In the range of distances and incident angles [1 m-29 m, 0° - 80°], four reference targets with reflectance of 20%, 40%, 60%, and 80% were established for intensity correction models. Then, a total of 20 small regions with a size of approximately 15 cm \times 15 cm in the white lime wall surface were randomly sampled to verify the intensity correction effect over the whole range of distances and incident angles. After using the above four intensity correction models, the intensity variance-to-mean ratio ε was 0.26, 0.14, 0.19, and 0.21, meaning that the intensity consistency was improved by 74%, 86%, 81%, and 79%.

Compared with Tan's work, an intensity correction model based on a reference target with 50% reflectivity in the range of distance and incident angle [0.1-14.4 m, 0° - 80°] is proposed. The intensity correction effect in this paper is better than Tan's in eliminating the factors of the distance and incident angle. As mentioned above, the improvement rates of intensity consistency before and after correction for site A and site B are 92.7% and 92.1%, respectively. Obviously, regardless of what kind of reference target reflectance Tan's intensity correction model is based on, the intensity consistency of the corrected white wall at sites A and B in this paper is higher than that of his research work. The reasons are as follows: first, to eliminate the distance factor, distance-intensity data with a distance value less than 1 m are considered in this paper. Therefore, compared with Tan's method, the method presented in this paper performs better in short-distance intensity correction. Then, a new method of spherical neighborhood search fitting plane is proposed to accurately calculate the cosine of the incident angle. In particular, the relationship between neighborhood radius and distance is discussed in the process of plane fitting. The accuracy of incident angle measurement is improved by selecting an appropriate neighborhood radius. Finally, unlike Tan's interpolation method, this paper adopts a piecewise distance polynomial and an incident angle cosine polynomial to fit distance-intensity and incident angle-intensity data. In conclusion, compared with Tan's

research, the intensity correction model established in this paper has a higher reliability and accuracy.

5. Conclusion

In this paper, a new point cloud intensity correction method for 2D MLS based on theoretical derivation and empirical correction is proposed to solve the problem that intensity information cannot be directly used for target recognition. Based on the diffuse reflection Lambert body of the same target reflectance, the intensity correction model of the piecewise distance polynomial and incident angle cosine polynomial is adopted, and the model parameters are calculated by experiments. The effectiveness of the intensity correction method is verified by single-site and multisite experiments on a white wall using MLS 2D LiDAR. The experimental results show that the intensity consistency is substantially improved by 70% to 92.7% after correction within the range of the distance and incident angle [0.52 m-5.34 m, 0° - 74°]. Compared with the latest research, the intensity correction model proposed in this paper has a higher fitting accuracy and can effectively eliminate the MLS intensity deviation caused by distance and incident angle.

However, the intensity correction model is only suitable for the specific UTM-30LX 2D LiDAR and applicable to targets similar to the standard Lambert surface. The intensity deviation still exists after the correction, indicating that further research is needed, especially to reduce the model error. In addition, it is found in the incident angle correction experiment that the correction of the incident angle is not thorough in place with a large angle, which requires further study of a more rigorous correction model. Furthermore, considering the long running time of LiDAR, the influence of internal temperature rise on intensity can further improve the accuracy of model correction, which is also a research direction for the future.

Data Availability

No data were used to support this study.

Conflicts of Interest

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This research is supported by the National Natural Science Foundation of China (61473156), Project 948 of the State Forestry Administration of China (2015-4-56), and Basic Research Program of Jiangsu Province-Youth Foundation Project (BK20170930).

References

- [1] Y. Bisheng, L. Fuxun, and H. Ronggang, "Progress, challenges and perspectives of 3D LiDAR point cloud processing," *Acta Geodaetica et Cartographica Sinica*, vol. 46, no. 10, pp. 1509-1516, 2017.

- [2] T. Yun, K. Jiang, G. Li et al., "Individual tree crown segmentation from airborne LiDAR data using a novel Gaussian filter and energy function minimization-based approach," *Remote Sensing of Environment*, vol. 256, article 112307, 2021.
- [3] Z. Chuan, G. Haitao, L. Jun, Y. Donghang, and Z. Baoming, "Airborne LiDAR point cloud classification based on deep residual network," *Acta Geodaetica et Cartographica Sinica*, vol. 49, no. 2, pp. 202–213, 2020.
- [4] M. W. Lang, V. Kim, G. W. McCarty et al., "Improved detection of inundation below the forest canopy using normalized LiDAR intensity data," *Remote Sensing*, vol. 12, no. 4, p. 707, 2020.
- [5] S. Yan, G. Yang, Q. Li, and C. Wang, "Waveform centroid discrimination of pulsed LiDAR by combining EMD and intensity weighted method under low SNR conditions," *Infrared Physics and Technology*, vol. 109, article 103385, 2020.
- [6] J. Jin, L. de Sloover, J. Verbeurgt et al., "Measuring surface moisture on a sandy beach based on corrected intensity data of a mobile terrestrial LiDAR," *Remote Sensing*, vol. 12, no. 2, p. 209, 2020.
- [7] A. F. C. Errington and B. L. F. Daku, "Temperature compensation for radiometric correction of terrestrial LiDAR intensity data," *Remote Sensing*, vol. 9, no. 4, p. 356, 2017.
- [8] Y. Haotian, X. Yanqiu, P. Tao, and D. Jianhua, "Effects of different LiDAR intensity normalization methods on crotch pine forest leaf area index estimation," *Acta Geodaetica et Cartographica Sinica*, vol. 47, no. 2, pp. 170–179, 2018.
- [9] A. G. Kashani, M. J. Olsen, C. Parrish, and N. Wilson, "A review of LiDAR radiometric processing: from ad hoc intensity correction to rigorous radiometric calibration," *Sensors*, vol. 15, no. 11, pp. 28099–28128, 2015.
- [10] D. U. Song, L. I. Xiaohui, L. I. Zhaoyan et al., "Radiometric characteristics of the intensity data of laser scanner," *Journal of University of Chinese Academy of Sciences*, vol. 36, no. 3, pp. 392–400, 2019.
- [11] B. Dimitrios, "Terrestrial laser scanner intensity correction for the incidence angle effect on surfaces with different colours and sheens," *International Journal of Remote Sensing*, vol. 40, no. 18, pp. 7169–7189, 2019.
- [12] C. Xiaolong, C. Xiaojun, L. Quan, and X. Wenbing, "Laser intensity correction of terrestrial 3D laser scanning based on sectional polynomial model," *Laser & Optoelectronics Progress*, vol. 54, no. 11, pp. 427–435, 2017.
- [13] F. Wei, H. Xianfeng, Z. Fan, and L. Deren, "Mural image rectification based on correction of laser point cloud intensity," *Acta Geodaetica Et Cartographica Sinica*, vol. 34, no. 5, pp. 541–547, 2015.
- [14] L. Quan and C. Xiaojun, *Damage detection for historical architectures based on TLS intensity data*, ISPRS-International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 2018.
- [15] F. Coren and P. Sterzai, "Radiometric correction in laser scanning," *International Journal of Remote Sensing*, vol. 27, no. 15, pp. 3097–3104, 2006.
- [16] A. Vain, X. Yu, S. Kaasalainen, and J. Hyypä, "Correcting airborne laser scanning intensity data for automatic gain control effect," *IEEE Geoscience & Remote Sensing Letters*, vol. 7, no. 3, pp. 511–514, 2010.
- [17] I. Korpela, H. O. Ørka, J. Hyypä, V. Heikkinen, and T. Tokola, "Range and AGC normalization in airborne discrete-return LiDAR intensity data for forest canopies," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 65, no. 4, pp. 369–379, 2010.
- [18] I. Puente, H. González-Jorge, P. Arias, and J. Armesto, "Land-based mobile laser scanning systems: a review," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 3812, no. 1, pp. 163–168, 2011.
- [19] T. Yun, L. Cao, F. An et al., "Simulation of multi-platform LiDAR for assessing total leaf area in tree crowns," *Agricultural and Forest Meteorology*, vol. 276–277, article 107610, 2019.
- [20] E. Che, J. Jung, and M. J. Olsen, "Object recognition, segmentation, and classification of Mobile laser scanning point clouds: a state of the art review," *Sensors*, vol. 19, no. 4, p. 810, 2019.
- [21] S. Yang, Y. Li, K. Li, and J. Mao, "Tree extraction from vehicle-borne LiDAR data," *Surveying and Mapping Engineering*, vol. 23, no. 8, pp. 23–27, 2014.
- [22] Q. Li, J. Zheng, H. Zhou, R. Tao, and Y. Shu, "Point cloud recognition of street tree target based on variable-scale grid index and machine learning," *Transactions of the Chinese Society for Agricultural Machinery*, vol. 49, no. 6, pp. 32–37, 2018.
- [23] Q. Li, P. Yuan, X. Deng, and Y. Ru, "Calculation method of target leaf area based on mobile laser scanning," *Transactions of the Chinese Society for Agricultural Machinery*, vol. 51, no. 5, pp. 192–198, 2020.
- [24] H. Liu, N. Li, Y. Shen, and H. Xu, "Spray target detection based on laser scanning sensor and real-time correction of IMU attitude angle," *Transactions of the Chinese Society of Agricultural Engineering*, vol. 33, no. 15, pp. 88–97, 2017.
- [25] Q. Li, P. Yuan, Y. Lin, Y. Tong, and X. Liu, "Pointwise classification of mobile laser scanning point clouds of urban scenes using raw data," *Journal of Applied Remote Sensing*, vol. 15, no. 2, p. 024523, 2021.
- [26] B. Vallet, M. Brédif, A. Serna, B. Marcotegui, and N. Paparoditis, "TerraMobilita/iQmulus urban point cloud analysis benchmark," *Computers and Graphics*, vol. 49, pp. 126–133, 2015.
- [27] L. Xu, H. Zhang, H. Zhang et al., "Development and experiment of automatic target spray control system used in orchard sprayer," *Transactions of the Chinese Society of Agricultural Engineering*, vol. 30, no. 22, pp. 1–9, 2014.
- [28] Y. Nan, H. Zhang, J. Zheng et al., "Estimating leaf area density of *Osmanthus* trees using ultrasonic sensing," *Biosystems Engineering*, vol. 186, no. 186, pp. 60–70, 2019.
- [29] T. Kai and C. Xiaojun, "TLS laser intensity correction based on polynomial model," *Chinese Journal of Lasers*, vol. 42, no. 3, pp. 310–318, 2015.
- [30] K. Tan and X. Cheng, "Correction of Incidence angle and distance effects on TLS intensity data based on reference targets," *Remote Sensing*, vol. 8, no. 3, pp. 251–251, 2016.
- [31] T.-A. Teo and H.-L. Yu, "Empirical radiometric normalization of road points from terrestrial mobile LiDAR system," *Remote Sensing*, vol. 7, no. 5, pp. 6336–6357, 2015.
- [32] A. V. Jelalian, *Laser Radar Systems*, Artech House, Boston, 1992.
- [33] D. Perez and Y. Quintana, "A survey on the Weierstrass approximation theorem," *Divulgaciones Matemáticas*, vol. 16, no. 1, pp. 231–247, 2008.
- [34] P. Demski, M. Mikulski, and R. Koteras, "Characterization of Hokuyo UTM-30LX laser range finder for an autonomous

- mobile robot,” in *Advanced Technologies for Intelligent Systems of National Border Security*, Springer, 2013.
- [35] F. Zhu, J. Yang, S. Xu, C. Gao, N. Ye, and T. Yin, “Relative density degree induced boundary detection for one-class SVM,” *Soft Computing*, vol. 20, no. 11, pp. 4473–4485, 2016.
 - [36] F. Zhu, Y. Ning, X. Chen, Y. Zhao, and Y. Gang, “On removing potential redundant constraints for SVOR learning,” *Applied Soft Computing*, vol. 102, article 106941, 2021.
 - [37] W. Zheng, S. Chen, Z. Fu, F. Zhu, H. Yan, and J. Yang, “Feature selection boosted by unselected features,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 7, no. 99, pp. 1–13, 2021.
 - [38] F. Zhu, J. Gao, C. Xu, J. Yang, and D. Tao, “On selecting effective patterns for fast support vector regression training,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 8, pp. 3610–3622, 2018.
 - [39] T. Hackel, J. D. Wegner, and K. Schindler, “Fast semantic segmentation of 3d point clouds with strongly varying density,” *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. III-3, pp. 177–184, 2016.
 - [40] S. Arya, D. M. Mount, N. S. Netanyahu, R. Silverman, and A. Y. Wu, “An optimal algorithm for approximate nearest neighbor searching fixed dimensions,” *Journal of the ACM*, vol. 45, no. 6, pp. 891–923, 1998.
 - [41] Q. Li, P. Yuan, X. Liu, and H. Zhou, “Street tree segmentation from mobile laser scanning data,” *International Journal of Remote Sensing*, vol. 41, no. 18, pp. 7145–7162, 2020.
 - [42] T. Kai, C. Xiaojun, and Z. Jixing, “Correction for incident angle and distance effects on TLS intensity data,” *Geomatics and Information Science of Wuhan University*, vol. 42, no. 2, pp. 223–228, 2017.

Research Article

Efficient Semantic Enrichment Process for Spatiotemporal Trajectories

Bin Zhao ¹, Mingyu Liu ¹, Jingjing Han ², Genlin Ji ¹ and Xintao Liu ³

¹School of Computer and Electronic Information/School of Artificial Intelligence, Nanjing Normal University, Nanjing, China

²Quality Assurance Office, Jiangsu Open University, Nanjing, China

³Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, China

Correspondence should be addressed to Bin Zhao; zhaobin@njnu.edu.cn and Jingjing Han; hanjj@jsou.edu.cn

Received 5 August 2021; Accepted 1 October 2021; Published 12 November 2021

Academic Editor: Fa Zhu

Copyright © 2021 Bin Zhao et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The increasing availability of location-acquisition technologies has enabled collecting large-scale spatiotemporal trajectories, from which we can derive semantic information in urban environments, including location, time, direction, speed, and point of interest. Such semantic information can give us a semantic interpretation of movement behaviors of moving objects. However, existing semantic enrichment process approaches, which can produce semantic trajectories, are generally time-consuming. In this paper, we propose an efficient semantic enrichment process framework to annotate spatiotemporal trajectories by using geographic and application domain knowledge. The framework mainly includes preannotated semantic trajectory storage phase, spatiotemporal similarity measurement phase, and semantic information matching phase. Having observed the common trajectories in the same geospatial object scenes, we propose a semantic information matching algorithm to match semantic information in preannotated semantic trajectories to new spatiotemporal trajectories. In order to improve the efficiency of this approach, we build a spatial index to enhance the preannotated semantic trajectories. Finally, the experimental results based on a real dataset demonstrate the effectiveness and efficiency of our proposed approaches.

1. Introduction

Spatiotemporal trajectories record the spatiotemporal position sequences of moving objects. The increasing access to positioning device technologies, such as smartphones, GPS-enabled cameras and sensors, results in vast volumes of collected spatiotemporal trajectories. Analyzing and mining spatiotemporal trajectories can study in depth various fields such as traffic coordination and management (e.g., road flow monitoring), tourist route recommendation, and natural disaster early warning (e.g., typhoon prediction). However, many applications in the mobility domain require a semantic interpretation of movement information. This semantic interpretation is usually obtained by mining semantic trajectories, which is the fusion of spatiotemporal trajectories and semantic information. Location-based social networks (LBSN), such as Twitter and Weibo, produce multifaceted semantic information, which contains the moving

state of moving objects (e.g., speed and direction) and environment information (e.g., air temperature and spatial topological relationship) [1]. Combining semantic information, such as user's personalized characteristics, landmark names, user's interest, and occupation into the user's spatiotemporal trajectories, will contribute to the recommendation of nearby hot spots of interest for users [2, 3]. It can be seen that mining semantic trajectories [4] can better meet the needs of decision analysis applications.

Different from spatiotemporal trajectories obtained by position-aware devices, semantic trajectories must be generated through semantic trajectory modeling. Semantic trajectory modeling includes trajectory data preprocessing, trajectory segmentation, and semantic enrichment. Among them, the semantic enrichment process is the key stage, which annotates appropriate semantic information (e.g., behavior attributes, environment information, and domain knowledge) in spatiotemporal trajectories. With different

sources, complex types, and diverse forms of semantic information, there are different semantic enrichment process approaches.

The existing semantic enrichment process approaches can be divided into three categories: (1) Early approaches directly annotate velocity and direction in spatiotemporal trajectories. Due to lacking rich semantic information, the results of mining semantic trajectories annotated by early approaches have a low semantic interpretation. (2) Part approaches annotate domain knowledge in spatiotemporal trajectories through ontology. However, approaches based on ontology transform a semantic trajectory into RDF graph description [5], which causes the finding and reasoning semantic trajectories time-consuming. (3) Typical approaches annotate geographical object information, including areas of interest (ROIs), lines of interest (LOIs), and points of interest (POIs), through the *spatial join* [6] algorithm and *map matching* [7] algorithm. The execution time of [6, 7] is linearly correlated with the number of geospatial objects, which results in high time consumption. It can be seen that the existing semantic enrichment process approaches have the disadvantage of high time consumption.

On the other hand, given movement trajectories limited by topological relationship of urban road networks, there are common movement trajectories in the same geospatial object scenes. For example, commuters departing from the Tsinghua Park residential usually take Metro Line 4 to Beijing Zhongguancun SOHO Building. Due to traffic restrictions, it is easy to collect a large number of identical commuting trajectories. Obviously, new commuting trajectory information can be directly attached to historical commuting trajectories. Similarly, it is possible to directly annotate the semantic information in a preannotated semantic trajectory to new spatiotemporal trajectories. Using preannotated semantic trajectories for enrichment does not need a complicated computation and annotation process, which may avoid an inefficient semantic enrichment process.

In this paper, we propose a new semantic enrichment process approach named Efficient Semantic Enrichment Process for Spatiotemporal Trajectories based on Semantic Information Matching (SEPSIM), which firstly uses semantic information in preannotated semantic trajectories for annotating spatiotemporal trajectories. We first store preannotated semantic trajectories in the form of episodes. In this phase, we segment semantic trajectories into stop or move episodes. Then, we measure the spatiotemporal similarity between subtrajectories and episodes. The similarity of stop subtrajectories and move subtrajectories is measured, respectively. Finally, we propose a new algorithm named Semantic Information Matching Algorithm based on Similar Episodes (SESIM), which can match semantic information of episodes to a new trajectory. In order to put down the search cost of metrics and matching, we build a spatial index to store episodes of preannotated semantic trajectories.

In summary, this article makes the following contributions:

- (i) We propose an efficient semantic enrichment process framework (SEPSIM) for spatiotemporal trajectories based on semantic information matching. It includes three phases: preannotated semantic trajectory storage, spatiotemporal similarity measurement, and semantic information matching. In order to improve the efficiency of the SEPSIM approach, we establish a spatial index
- (ii) We propose a new standard to measure the effectiveness of semantic enrichment process approaches. Also, we compared different semantic enrichment process approaches in efficiency
- (iii) In order to verify the effectiveness and efficiency of the SEPSIM approach, experiments were performed by using the real trajectory dataset. The results prove the high effectiveness and efficiency of the SEPSIM approach

2. Related Work

There are different semantic enrichment process approaches with different sources, complex types, and various forms of semantic information. Early semantic enrichment process approaches directly annotate velocity and direction in spatiotemporal trajectories, which generate semantic trajectories as stop and move subtrajectory sequences. Ashbrook and Starner [8] calculated the moving speed (whether the speed is zero) to identify stop subtrajectories. Due to poor speed measurement and other reasons, semantic trajectory stop segments do not match actual situation. Krumm and Horvitz [9] calculated the speed and direction to identify stop subtrajectories; Palma et al. [10] set the subtrajectory below the average speed as a stop subtrajectory, generating the semantic trajectory consisting of stop and move subtrajectories. In addition to calculating the moving speed, Zheng et al. [11] also calculated the acceleration and speed change rate to discover move subtrajectories with different modes of transportation (e.g., bicycles, buses, and self-driving) to enrich the semantic trajectory. Although early semantic enrichment process approaches were fast in annotation, the semantic information was not rich enough.

Part semantic enrichment process approaches annotate domain knowledge as semantic information through ontology. Spaccapietra et al. [12] first proposed an ontological method for semantic trajectory modeling. Based on the concepts of “stop” and “move,” the ontology was used to define semantic trajectories, and the semantic information of trajectories was further enhanced using the reasoning ability of ontology. Baglioni et al. [13] extended the definition of Baglioni’s ontology and proposed the concept of core ontology, which formally describes the concepts of stop, move, time, place, and mode in human mobile behavior, further enriching the definition of semantic trajectories. In 2014, Vandecasteele et al. [14] combined semantic trajectories with semantic events. Nogueira et al. [15] proposed the QualiTraj ontology to describe the various motion characteristics of original trajectories, especially the derivative characteristics, such as speed, acceleration, and direction. Nogueira

and Martin [16] proposed a new ontology based on Quali-Traj ontology with stronger information description ability, namely, Semantic Trajectory Episodes (STEP) ontology. It can not only describe basic motion characteristics but also describe environmental characteristics of moving trajectories on a higher semantic level. In 2018, Nogueira et al. [17] proposed the FrameSTEP, a semantic trajectory labeling framework based on STEP ontology. This method can calculate various physical movements and spatial geometric features of trajectory segments and use external reliable resources (such as OSM and LinkedGeoData geographic knowledge base) to label the environmental features of trajectories. However, approaches based on ontology need to represent semantic trajectories as RDF graphical descriptions, which results in time consumption.

The main source of information on semantic enrichment is geospatial objects with geometric features in geographical objects, including regions of interest (ROI), lines of interest (LOI), and points of interest (POI) [18]. At present, the typical semantic enrichment processing method uses the spatial join algorithm [6] to find the regions of interest (ROI) that have a topological relationship with spatiotemporal trajectories and label the regions of interest associated with spatiotemporal trajectories and the corresponding topological relationship. This algorithm needs to combine the external environment information (e.g., OSM map and Baidu map) to select the regions of interest associated with spatiotemporal trajectories. The execution time of the algorithm is linearly related to the number of geospatial objects, resulting in high time complexity and low semantic enrichment performance in the spatial connection process. For points of interest (POI), Sun et al. used an implicit Markov model [19] to label the POI categories for staying segments of spatiotemporal trajectories, but in the regions with intensive POI, staying segments may be related to multiple interest points. Coupled with the low GPS sampling rate, it is difficult to identify effective POIs. On the other hand, the LOI labeling method often uses a global map matching algorithm [7] to determine the location of spatiotemporal trajectories. Parent et al. proposed a “point-segment distance” measurement method [7] to replace the original distance function in the global map matching algorithm, which is suitable for labeling lines of interest in geographical scenarios such as dense road networks, parallel roads, and intersections. The global matching algorithm needs to perform metric matching on trajectory segments where spatiotemporal trajectories are located, which easily results in high time complexity of algorithm execution and low semantic enrichment performance.

3. Preliminaries

In this section, we will present definitions of all necessary concepts used in this paper and formally state the problem.

3.1. Basic Concepts. The SEPSIM approach proposed in this paper is aimed at annotating semantic information of preannotated semantic trajectories in spatiotemporal trajectories. The input of this problem is a trajectory, short for a spatio-

temporal trajectory. Thus, we provide the definition of “trajectory” at first.

Definition 1 (trajectory). A trajectory T is a sequence of sampling points in the form $T = \{p_1, p_2, \dots, p_{|T|}\}$, $p_i = (\text{tid}, x_i, y_i, t_i)$, where tid is an object identifier and x, y and t are spatial coordinates and a time stamp, respectively. $|T|$ records the number of sampling points in trajectory T .

Definition 2 (subtrajectory). A subtrajectory is a substring of a trajectory, i.e., $T_s = \{p_{i+1}, p_{i+2}, p_{i+3}, \dots, p_{i+m}\}$, where $0 \leq i \leq |T| - m, m \geq 0$.

Definition 3 (stop subtrajectory and move subtrajectory). Given the distance threshold ε and the number of point threshold minpts, a DBSCAN cluster [20] analyzes the trajectory T . Each cluster is a stop subtrajectory of the trajectory. If each p_i in $T_s = \{p_{i+1}, p_{i+2}, p_{i+3}, \dots, p_{i+m}\}$ is an outlier, T_s is a stop subtrajectory (stop T_s). If point p_i is in the end of a stop subtrajectory and point p_{i+m+1} is in the beginning of another stop subtrajectory, $i + m < |T|$, T_s is a move subtrajectory (move T_s).

Then, we define “semantic trajectory” as the output of this problem. The main source of information on semantic enrichment is geospatial objects in geographical environment. For this reason, the semantic information matching in this paper refers to geospatial object information matching. First, we give the basic related to semantic information.

Definition 4 (geospatial object). According to geometric shapes, geographical objects are divided into three categories: region of interest (ROI), line of interest (LOI), and point of interest (POI). In this paper, we refer to ROIs, LOIs, and POIs collectively as geospatial objects. A geospatial object Go is defined as a uniquely identified specific space site (e.g., a park, a road, or a cinema). A Go is a quad (id, cat, loc, con), where id represents a geospatial object identifier and cat denotes the category of it (e.g., ROI, LOI, and POI), and loc denotes its corresponding location attribute in terms of longitude and latitude coordinates and con denotes its name.

Definition 5 (topological relation). For different types of geospatial objects, the topological relationship between subtrajectory T_s and the geospatial object Go is defined as the following seven types: T_s pass by Go (Go is a LOI), T_s pass by Go (Go is a POI), T_s pass by Go (Go is a ROI), T_s across Go (Go is a ROI), T_s enter Go (Go is a ROI), T_s leave Go (Go is a ROI), and T_s stop inside Go (Go is a ROI).

Definition 6 (episode). An episode [21] is a subtrajectory of semantically homogeneous sections of a trajectory, such as move episode and stop episode. We define an episode as a multilayered semantic sequence aligned in accordance with the time of a subtrajectory, i.e., episode = (T_s , sp, dir, geoinf), where T_s denotes the episode corresponding to trajectory segments, sp denotes the average speed of an episode, dir denotes the direction of an episode, and geoinf denotes the

episode corresponding geospatial information. The form of a specific episode is shown in Figure 1.

Definition 7 (semantic trajectory). A semantic trajectory ST is a sequence of episodes in a spatiotemporal order of a moving object, i.e., $ST = \{\text{episode}_1, \text{episode}_2, \dots, \text{episode}_{|ST|}\}$.

The list of major symbols and notations in this paper is summarized in Table 1.

3.2. Problem Statement. Given a trajectory T , a preannotated semantic trajectory dataset OST , two clustering thresholds ϵ and minpts , four radii r_1, r_2, r_3, r_4 , and a similarity threshold σ , our goal is to annotate semantic information of preannotated semantic trajectories in trajectory T , which can transform trajectory T to semantic trajectory ST .

4. Framework

In this section, we will present the SEPSIM framework including preannotated semantic trajectory storage phase, spatiotemporal similarity measurement phase, and semantic information matching phase. Figure 2 outlines this framework.

Preannotated Semantic Trajectory Storage. Given the preannotated semantic trajectory dataset OST , the first step is to store them. In order to prevent reducing the semantic information matching accuracy, preannotated semantic trajectories are stored in the form of episodes, which are representative and diverse. Semantic trajectories are segmented into episodes by the moving state (stop/move) of the moving object. The output of this phase is a set of stop episodes and move episodes, which can represent and describe a certain region.

Spatiotemporal Similarity Measurement. Given a trajectory T , the spatial-temporal similarity is measured between T and episodes obtained in the first phase. We first segment trajectory T into stop/move subtrajectories by DBSCAN clustering. Then, there are two subproblems that need to be solved: how to measure the similarity between the stop subtrajectory and stop episode and how to measure the similarity between the move subtrajectory and move episode. To solve the problem above, we propose the algorithms based on the Hausdorff distance [22] and based on the Longest Common Subsequence (LCS) [23], respectively. The output of this phase is stop and move episodes, which satisfy the specified similarity condition.

Semantic Information Matching. Semantic information of similar stop/move episodes is matched to trajectory T in this phase, through the proposed semantic information matching algorithm (SESIM). The algorithm consists of two subphases: candidate episode sorting and semantic information mapping. We aim to generate a semantic trajectory ST that contains the most semantic information. For part subtrajectories which have no matching information, we complete the semantic enrichment process of ST by using the typical approach.

4.1. Preannotated Semantic Trajectory Storage. After we get the preannotated semantic trajectory dataset OST , the first

task is to store them for the matching phase. Storing all preannotated semantic trajectories can reduce the workload of storage and search, but the effectiveness and efficiency of the matching phase between complete trajectories are poor. And storing complete preannotated semantic trajectories with corresponding episodes causes data redundancy. In order to ensure complete semantic information and avoid data redundancy, we choose to store preannotated semantic trajectories in forms of episodes. However, episodes can only be obtained through trajectory segmentation. There are two kinds of trajectory segmentation methods: segment according to geospatial objects and segment according to the moving state of the moving object. With complex and irregular distribution and a large number of geospatial objects, segmentation according to geospatial objects is easy to cause trajectory fragments and time consumption. Meanwhile, segmentation according to the moving state of moving objects has the advantages of high segmentation efficiency and clear segmentation rules. So, we choose to segment spatiotemporal trajectories by the moving state of moving objects. For the reason that the stop of the moving object produces trajectory point gathering, we segment preannotated semantic trajectories into stop/move episodes by DBSCAN clustering.

Given a preannotated semantic trajectory dataset OST and a new coming preannotated episode, there are three situations to compare with the episodes in the dataset OST . The first case is the newly episode not repeated in the dataset OST at all, the second case is partial repetition but not complete repetition compared with the dataset OST , and the third case is complete repetition. If all preannotated episodes were stored, it will cause querying multiple repeated episodes with the increasing dataset, which reduces the efficiency of the similarity measurement and matching phase. Therefore, there is a challenge: which preannotated episodes stored can guarantee to avoid redundancy and ensure the effectiveness and efficiency of matching.

To solve the challenge above, we choose to store representative and diverse preannotated episodes to build the dataset OST . The semantic information of the semantic episodes (spatial information and geospatial environment information) represents the geospatial environment characteristics of a certain region. Therefore, the representative semantic episode of a certain region is defined as the episode with the same or partial spatial information and incomplete semantic information compared with preannotated episodes in the set semantic trajectories OST . The diversity of episodes is reflected in the diversity of geospatial environment information, which can enrich the characteristics of a certain region. So, we define the diverse episode as an episode with new geospatial environment information compared with preannotated episodes in the dataset OST . In this paper, the representative and diverse episodes are obtained through trajectory classification in Figure 3. For a given preannotated episode dataset, we first classify it according to spatial information and then classify it according to geospatial information and topological relationship, and finally, the leaf nodes store fine-grained representative and diverse preannotated episodes for matching. The

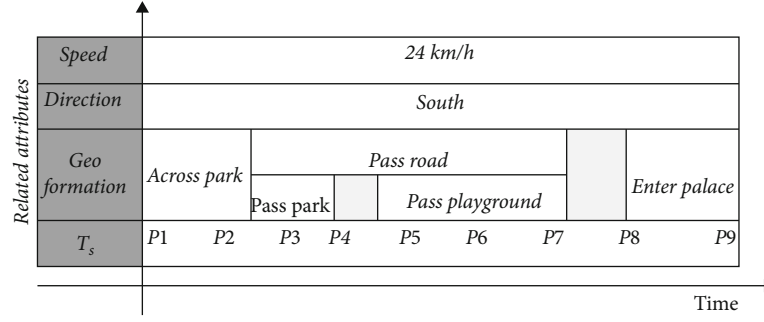


FIGURE 1: Example of an episode.

TABLE 1: Table of notations.

Notations	Definition
T	A spatio-temporal trajectory
T_s	A subtrajectory of T
stop T_s	A stop subtrajectory of the trajectory T
move T_s	A move subtrajectory of the trajectory T
Go	A geospatial object
episode	A subtrajectory of a semantic trajectory
stop episode	A stop subtrajectory of a semantic trajectory
move episode	A move subtrajectory of a semantic trajectory
ST	A semantic trajectory
OST $\{ST_1, ST_2, \dots\}$	The set of semantic trajectories
ε	The DBSCAN clustering distance threshold
minpts	The DBSCAN clustering point number threshold
r_1, r_2, r_3, r_4	Four similar region radii
σ	The similarity threshold
stop $T_{s\text{set}}$	The set of stop T_s of T
move $T_{s\text{set}}$	The set of move T_s of T
stop Episode _{Set}	The set of stop episodes of OST
move Episode _{Set}	The set of move episodes of OST

output of this phase is a set of representative and diverse stop/move episodes of the set semantic trajectory OST, which represent a certain region.

4.2. Spatiotemporal Similarity Measurement. For an incoming trajectory T , we compare it with episodes to find similar episodes. Once we find the similar episodes, we can match the semantic information of episodes to trajectory T . Giving the limitation of topological relationship of urban road networks, there are many similar or the same trajectory segments. So, we first segment trajectory T into stop/move T_s by DBSCAN clustering. Then, we solve the two problems: the similarity between stop subtrajectory and stop episode (stop trajectories) measurements and the similarity between move subtrajectory and move episode (move trajectories) measurements. Next, we will discuss the algorithm to solve these two problems, respectively, in the following algorithms.

The Algorithm to Determine the Similarity between Stop Trajectories. To our knowledge, there is no basic method for measuring the similarity of stop trajectories in the Euclidean space. In this paper, the stop T_s and stop episode are clusters of trajectory points obtained by DBSCAN clustering. The similarity measurement of the stop T_s and stop episode can be regarded as similarity measurement of point sets. Therefore, we view each stop trajectory, which is a stop T_s or a stop episode, as point sets. The algorithm proposed in this paper consists of two steps: (1) similar region determination and (2) similarity measurement based on the Hausdorff distance. Given the fact that the closer the space, the more similar the trajectories, we first narrow the metric range of stop episodes down and remain stop T_s with greater likelihood of similarity. Then, we calculate the Hausdorff distance between each stop T_s in stop $T_{s\text{set}}$ of T and stop episodes in stop Episode_{Set} of OST sequentially. Finally, stop episodes meeting similar conditions are remained.

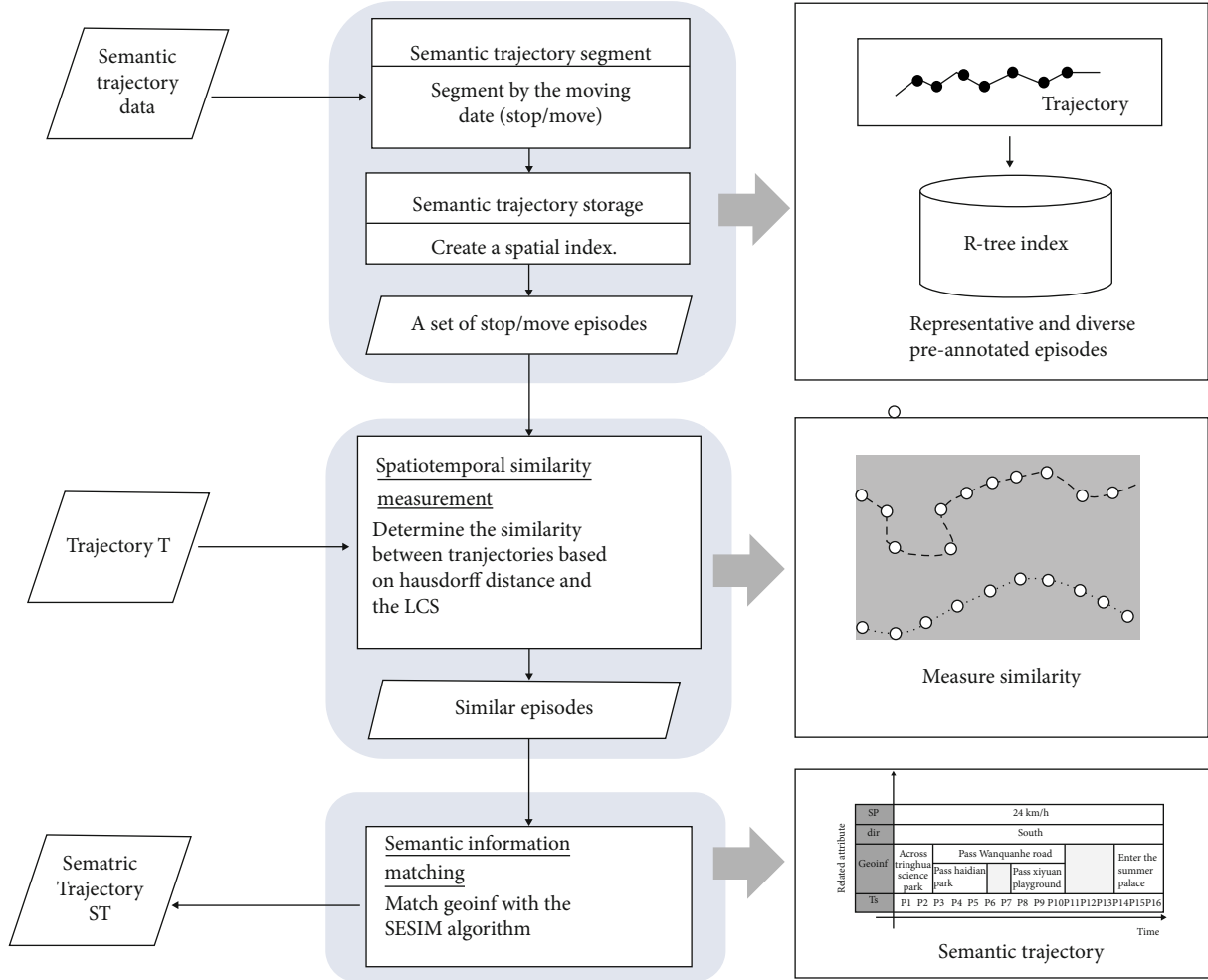


FIGURE 2: The framework of the SEPSIM process.

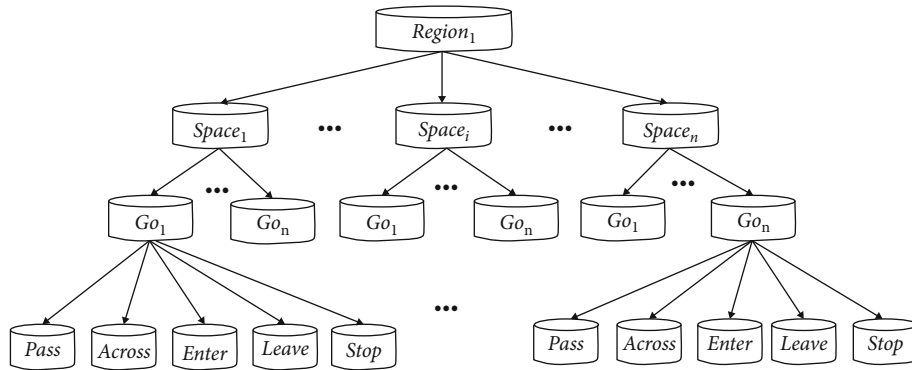


FIGURE 3: Principle of the trajectory classification.

In the first step, we narrow the number of stop episodes down and remain stop episodes with high similar probability to each stop T_s of T . Firstly, we convert each stop T_s to a point set P by assigning the latitude and longitude coordinates of each stop T_s to the x, y coordinates of the point set O (lines 1-5). According to the minimum circumscribed point o in point set P and the given radius r_1 , we draw a cir-

cular area $Circle_1$ as the similar region of the stop T_s (line 6). All the stop episodes that intersect with or are inside $Circle_1$ are extracted for similarity measurement. If there are no stop episodes in a similar region, there is no similar stop episode to the stop T_s . Otherwise, we convert stop episodes extracted in a similar region to point sets $E_{set}(E_1(\text{stop Episode}_1), \dots, E_n(\text{stop Episode}_n))$ in the second step

(lines 7-10). Figure 4 shows the similar region determination of each stop T_s .

Then, we calculate the Hausdorff distance between P and each point set E_i in a similar region (lines 11-13). Finally, the point set E_i , which has the minimum Hausdorff distance to point set P , was returned. The stop episode corresponding to the point set E_i is the most similar episode to the stop T_s (lines 14-16).

The Algorithm to Determine the Similarity between Move Trajectories. Generally, move episodes are not completely similar to the entire subtrajectory. In academia, this kind of similarity measurement is called the local matching of the trajectories. Existing local matching methods include the Frechet distance [24], Longest Common Subsequence (LCS) [23], and K Best Connected Trajectories [25]. The Frechet distance method is sensitive to a noise trajectory point; the K Best Connected Trajectory method can only query a few elements and is mainly used for recommending tourist routes. The Longest Common Subsequence (LCS) method is different from the previous similarity measurement methods. The previous methods focus on calculating the distance between point pairs of trajectories. The LCS method takes into account the movement of vehicles, which is restricted by the road network. If vehicles travel on the same road segment, the trajectories passing through the road segment may completely overlap, which is consistent with the thought of the SEPSIM approach. Therefore, the degree of overlap between trajectories can be used as a criterion for similarity.

The LCS method is only suitable for trajectory data generated on the road network, and the time complexity is $O(m * n)$. However, the LCS method has the advantage of not considering departure time and driving speed of trajectories and is robust to noise, which is consistent with the situation of the experimental data in this paper. Therefore, we propose the algorithm to determine the similarity between move trajectories based on the LCS. The detail of the LCS method can be found in [23].

This algorithm consists of three steps: (1) similar region determination, (2) measurement range determination, and (3) similarity measurement based on LCS. First, we filter move episodes that are likely similar in each move T_s similar region [26]. Then, the subtrajectory part of the move episode that is similar to T is determined. Finally, we calculate the similarity between move episodes and the corresponding similar subtrajectory of T based on the LCS method. The long common subsequence obtains the similarity and retains move episodes that meet the similarity threshold. The same operation is performed on each move T_s .

We use the same way to draw the similar region of each move T_s in move $T_{s_{set}}$ of T . In the first step, we draw a circular area $Circle_2$ with a given radius r_2 and a circle point o , which is the center of each move T_s , as the similar region of each move T_s (lines 1 and 2). Each move episode that intersects with or is inside the circle is extracted for measurement range determination, which is the candidate move episode set E_{set} (lines 3-5). For each move episode in a similar region, we draw two circular areas $Circle_3$ and $Circle_4$ with the given

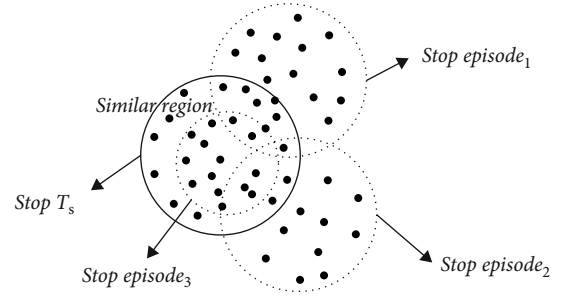


FIGURE 4: Similar region determination of each stop T_s of T .

radii r_3 and r_4 and two circle points, which are the beginning and end point of each move episode (lines 6-9). Given that the trajectories are partially similar, we then confirm the measurement ranges of trajectory T , where each move episode measures the similarity. The part of trajectory T , which is tangent to the two circles $Circle_3$ and $Circle_4$, is the measurement range corresponding to each move episode. Figure 5 shows similar region determination and measurement range of each stop T_s of T .

In the third step, we calculate the similarity $\text{simSeq}(\text{move episode, move } T_s)$ based on the Longest Common Subsequence (LCS) method (lines 9 and 10). If the simSeq is greater than or equal to the given similarity threshold σ , the move episode is similar to the part trajectory T . We remain the move episodes as $\text{simMoveEpisode}_{Set}$, which meet the similarity threshold (lines 11-13).

4.3. Semantic Information Matching. In this phase, we aim to match semantic information of episodes Episode_{Set} remained in the spatiotemporal similarity measurement phase to the trajectory T . The $\text{simStop Episode}_{Set}$ remained are the most similar ones corresponding to the part trajectory T , and all the similarities of $\text{simMove Episode}_{Set}$ are greater than or equal to 95%, which are identical to T in spatial information. Given a trajectory T and a set of similar episodes Episode_{Set} , the episodes corresponding to T have the following three matching ways shown in Figure 6. Obviously, there is a problem that needs to be solved: how to determine if the selected episodes are the best combination in the similar episode set for matching T to ST , which has the most semantic information.

To solve the problem, we propose a Semantic Information Matching Algorithm based on Similar Episodes (SESIM). This algorithm consists of two steps: (1) similar episode sorting and (2) semantic information matching. According to measurement range determination in the second phase, we first sort similar episodes meeting similar conditions by the spatial coordinate sequence of the trajectory T . Then, we model the problem as a knapsack problem to match semantic information.

Similar Episode Sorting. Given a trajectory T and a set of similar episodes $\text{simStop Episode}_{Set}$, we first measure the similar range of the trajectory T corresponding to similar episodes with the same solution in the step of measurement range determination. In this step, we convert the set of

```

Input : stop  $T_{s_{set}}$ , stop  $Episode_{set}$ ,  $r_1$ 
Output : simStop  $Episode_{set}$ 
1  for each stop  $T_s \in stop T_{s_{set}}$  do
2    for each  $P_i(x, y) \in stop T_s$  do
3       $O_i(x, y) \leftarrow P_i(x, y)$ ;
4       $i++$ 
5      Insert  $O_i$  into  $O_{set}(stop T_s)$ ;
6      Circle  $c = \text{minCircle}(O_{set}(stop T_s), r_1)$ ;
7    for each stop  $Episode_i \in stop Episode_{set}$  do
8      if stop  $Episode_i$  in or insert Circle  $c$  then
9         $E_i(stop Episode_i) \leftarrow \text{EpisodeTransferPoint}(stop Episode_i)$ 
10       Insert  $E_i$  into  $E_{set}(E_1(stop Episode_1), \dots, E_n(stop Episode_n))$ ;
11    for each  $E_i(stop Episode_i) \in E_{set}$  do
12      distance( $stop Episode_i$ )  $\leftarrow \text{Hausdorff}(stop Episode_i, stop T_s)$ 
13      insert distance( $stop Episode_i$ ) into  $distance_{set}$ ;
14      for each distance( $stop Episode_i$ )  $\in distance_{set}$  do
15        simStop  $Episode_i = \text{MinDistance}(distance(stop Episode_i))$ ;
16    return simStop  $Episode_{set}$ ;

```

ALGORITHM 1: Similarity measurement of the stop trajectory (SMST).

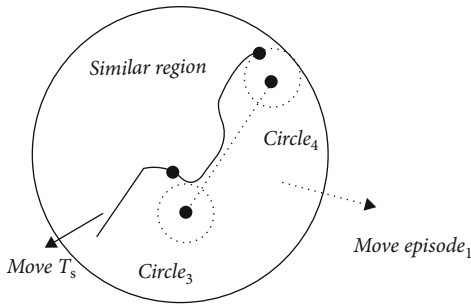


FIGURE 5: Similar region determination and measurement range of each stop T_s of T .

similar episodes to the candidate set $Episode_i(Episode_i, (P_{begin}, P_{end}), V(i), L(i))$, where P_{begin} and P_{end} are the beginning and end trajectory points of subtrajectory T_s , respectively, corresponding to $\text{simStop } Episode_{set}$, $L(i)$ is the number of sampling points in T_s , and $V(i)$ is the number of geospatial information in $\text{simStop } Episode_{set}$ (lines 1-5). Then, we sort the set E by the position of P_{begin} in trajectory T (lines 6-10).

Semantic Information Matching. In this step, we aim to select the best combination of episodes in set E for matching the semantic trajectory with most semantic information. We extend a knapsack algorithm, considering the number of sampling points of the trajectory T as the capacity of the backpack W and the number of geospatial information in E as the value of the episode. In start matching from the end sampling point P_{end} of the trajectory T , we aim to maximize the total value of the entire backpack. Given the candidate set $E\{Episode_i(Episode_i, (P_{begin}, P_{end}), V(i), L(i))\}$, we define the value of the trajectory T using the following formula: $\text{SemScore}(|T|) = \text{Max}(\text{SemScore}(|T| - 1), \text{SemScore}(|T| - L(i)) + V(i))$ (lines 11-18).

4.4. Space Index Establishment. To quickly get the preannotated semantic episodes similar to trajectory T , we use the space attribute of trajectory data to establish a space index for saving and querying episodes quickly, which will improve the efficiency of the SEPSIM approach.

The establishment of the space index is related to the query target. The index in this section is used to query episodes similar to the trajectory T_s . Therefore, the elements stored in the space index should be trajectory edge data. The common space index includes R -tree index [27], quad-tree index [28], and grid index [29]. The elements stored in the spatial index are episodes, which are essentially trajectory edge data. The quad-tree index is only adapted to query a trajectory point. The large number of unevenly distributed geospatial objects causes the grid index to be inefficient. Meanwhile, the R -tree index can be efficient in the unevenly distributed dataset in this paper by ensuring the balance of the tree. Therefore, we create and maintain an R -tree index for preannotated episodes. With this index, we can compare an incoming subtrajectory T_s with preannotated episodes in the index, which are inside or intersect with the subtrajectory T_s .

5. Experiments

In this section, we conduct extensive experiments on real trajectory datasets to compare the effectiveness and efficiency between the proposed approach SEPSIM in this paper and the typical approach based on the *spatial join* algorithm and *map matching* algorithm as the baseline approach.

5.1. Experimental Settings. We evaluate our approach on the GeoLife dataset. This trajectory dataset was collected in (Microsoft Research Asia) GeoLife project by 182 users in a period of over five years (from April 2007 to August 2012), which contains 17,621 trajectories with a total distance of 1,292,951 kilometers and a total duration of


```

Input:  $move T_{s_{set}}, move Episode_{set}, d, r_2, r_3, r_4, \sigma$ 
Output:  $simMove Episode_{set}$ 
1 for each  $move T_s \in move T_{s_{set}}$  do
2    $Circle C = minCircle(move T_s, r_2)$ ;
3 for each  $move Episode_i \in move Episode_{set}$  do
4   if  $move Episode_i$  in or insert  $Circle c$  then
5     insert  $move Episode_i$  into  $E_{set}(move Episode_1, \dots, move Episode_n)$ ;
6 for each  $move Episode_i \in E_{set}$  do
7    $Circle C_1 = minCircle(move Episode_i, P_{begin}, r_3)$ ;
8    $Circle C_2 = minCircle(move Episode_i, P_{end}, r_4)$ ;
9   if  $move T_s$  tangent  $Circle C_1$  and  $Circle C_2$  then
10     $SimSeq(move T_s, move Episode_i) = LCS(move T_s, move Episode_i)$ ;
11    if  $SimSeq(move T_s, move Episode_i) \geq \sigma$  then
12      insert  $move Episode_i$  into  $simMove Episode_{set}$ ;
13 return  $simMove Episode_{set}$ ;

```

ALGORITHM 2: Similarity measurement of the move trajectory (SMMT).

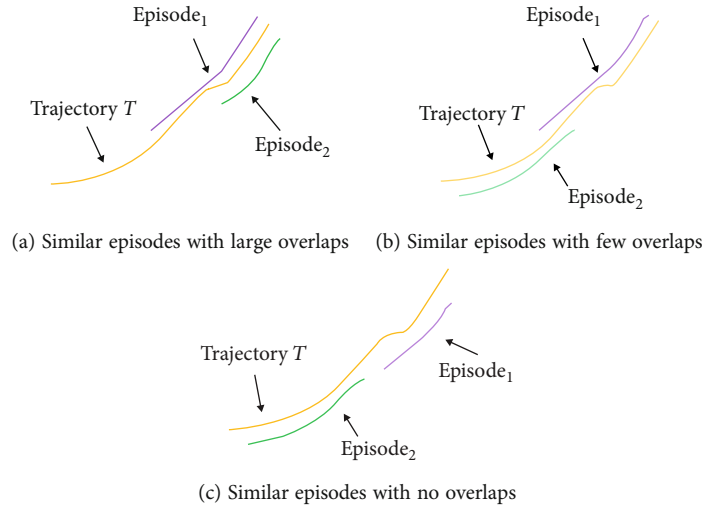


FIGURE 6: Tree matching types of similar episodes.

```

Input:  $T, Episode_{set} \{ simStop Episode_{set}, simMove Episode_{set} \}$ 
Output:  $ST$ 
1 for each  $Episode_i \in Episode_{set}$  do
2    $P_{begin}, P_{end} \leftarrow MeasurementRangesDetermination(Episode_i, T)$ ;
3    $V(i) \leftarrow GetGeoInfNumber(Episode_i)$ ;
4    $L(i) \leftarrow GetNumberTrajPoint(Episode_i, T)$ ;
5   insert  $P_{begin}, P_{end}, V(i), L(i)$  into  $Episode_i(P_{begin}, P_{end}, V(i), L(i))$ ;
6 for each  $Episode_i$  do
7   Insert  $Episode_i(P_{begin}, P_{end}, V(i), L(i))$  into set  $E$ ;
8   SortByTrajSpatial( $E$ );
9 for  $i = 0$  to  $|T|$  do
10   $SemScore(0) = 0$ ;
11  if  $P_{end}$  in  $Episode_i$  equal  $P_{end}$  in  $T$  do
12     $SemScore(|T|) = \text{Max}(SemScore(|T| - 1), SemScore(|T| - L(i) + V(i)))$ 
13  for each  $Episode_i$  in  $SemScore(|T|)$ 
14     $ST \leftarrow MatchSemanticInf(Episode_i)$ 
15 return  $ST$ ;

```

ALGORITHM 3: Semantic information matching algorithm based on similar episodes (SIM).

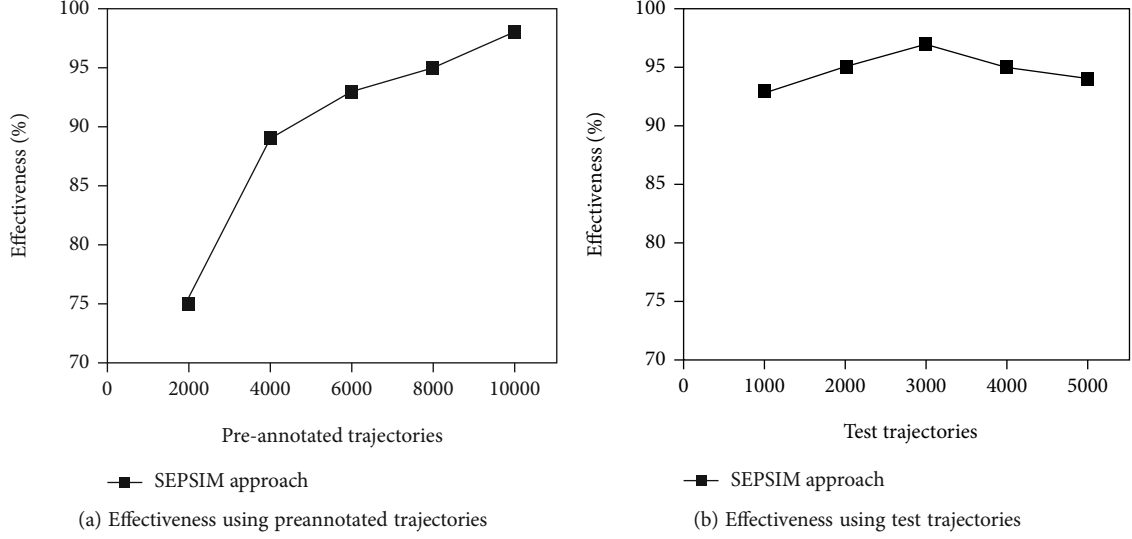


FIGURE 7: Effectiveness of the SEPSIM approach.

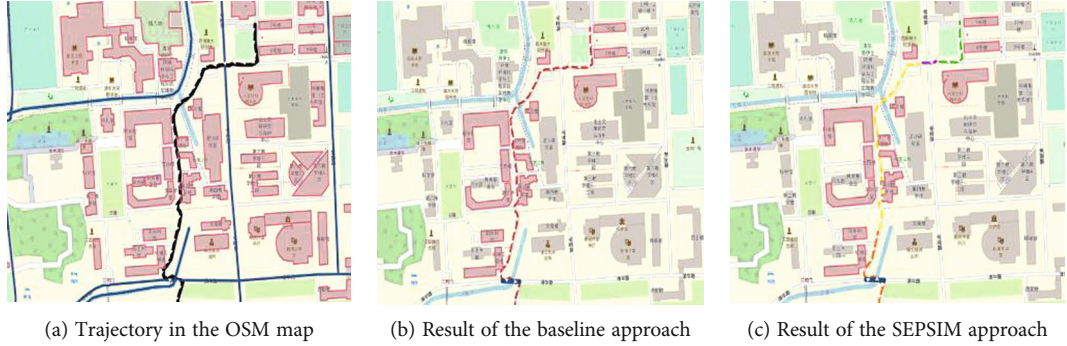


FIGURE 8: Trajectory enrichment results using different approaches.

50,176 hours. These trajectories were recorded by different GPS loggers and GPS phones and have a variety of sampling rates. The majority of the data was created in Beijing, China, and the data size is 1.87 GB. In this paper, all the preannotated semantic trajectories are generated by the typical approach. Both algorithms are implemented in Java and on computers with Intel(R) Xeon(R) CPU E5-2620 (2.10 GHz) and 32 GB memory.

5.2. Effectiveness. There is no clear and unified definition for the effectiveness of the semantic enrichment process. In this paper, we propose a new standard to measure the effectiveness of the algorithm proposed in this paper. For a trajectory T , we view the semantic trajectory ST_1 generated by the typical approach as the standard one and compare the semantic trajectory ST_2 generated by the SEPSIM approach with its difference. Firstly, we segment ST_1 and ST_2 by the move state. Then, we compared the accuracy of each pair of subtrajectories T_{s_1} and T_{s_2} between ST_1 and ST_2 . The effectiveness of ST_2 generated by the SEPSIM process approach is defined as the average accuracy of matched semantic information.

$$\text{Effectiveness}(ST_2) = \frac{\sum T_{s_2} \cdot \text{Accuracy} * T_{s_2} \cdot \text{Count}}{ST_2 \cdot \text{Count}}, \quad (1)$$

$$T_{s_2} \cdot \text{Accuracy} = \frac{\text{matchedGeoInf of } T_{s_2} \cdot \text{Count}}{\text{standardGeoInf of } T_{s_1} \cdot \text{Count}}, \quad (2)$$

where semantic trajectory ST_2 is generated by the SEPSIM approach of a given trajectory, $T_{s_2} \cdot \text{Accuracy}$ means the correct matched semantic information accuracy of the subtrajectory T_{s_2} compared to corresponding subtrajectory T_{s_1} in ST_1 , which is defined as the ratio of correct matched semantic information quantity in T_{s_2} (matchedGeoInf of $T_{s_2} \cdot \text{Count}$) to the standard semantic information quantity in T_{s_2} (standardGeoInf of $T_{s_1} \cdot \text{Count}$); $T_{s_2} \cdot \text{Count}$ and $ST_2 \cdot \text{Count}$ represent the number of sampling points contained in T_{s_2} and semantic trajectory ST_2 . Obviously, the higher the average accuracy of a matched subtrajectory, the more effective our proposed algorithm will be.

Figure 7(a) shows the change in effectiveness with the increasing preannotated trajectories. Obviously, after processing more and more preannotated trajectories, the

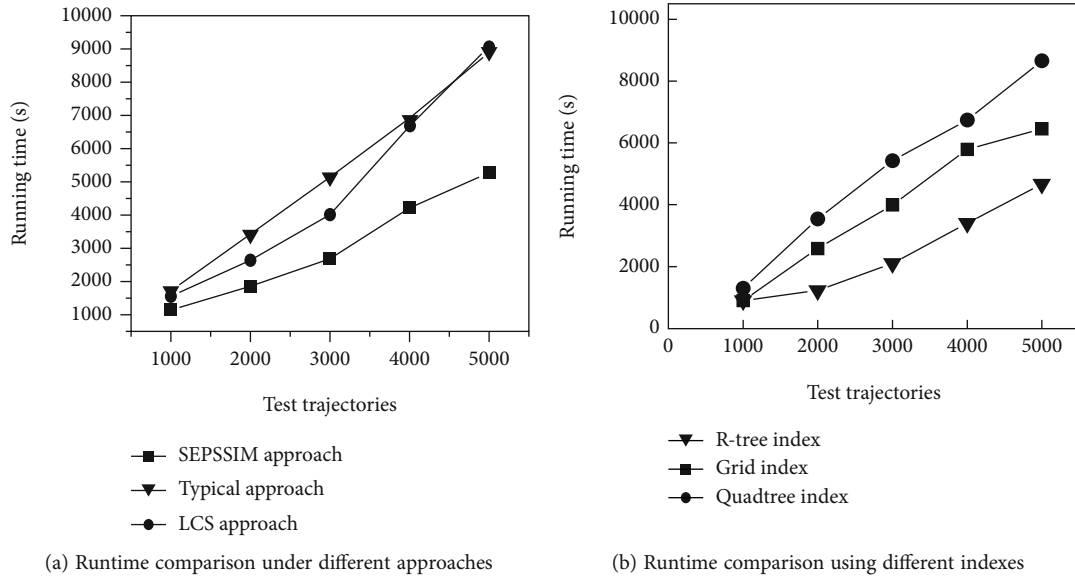


FIGURE 9: Performance comparisons among different approaches.

effectiveness of trajectories that need to be enriched is gradually increasing. When the number of preannotated trajectories reaches 4000, the effectiveness exceeds 90% and keeps increasing steadily. Figure 7(b) shows the change in effectiveness with the increasing test trajectories. It can be seen that the effectiveness of test trajectories keeps above 90%.

On the other hand, to evaluate the effectiveness of the SEPSIM algorithm, we compare the semantic trajectories generated by the baseline approach and by the SEPSIM approach in the form of visualization. Figure 8(a) shows the geographical object information represented by red boxes and corresponding topological relationships of a given trajectory in OSM map. Figure 8(b) shows the geographical object information and corresponding topological relationships enriched in the given trajectory by the baseline approach, which annotate all relevant and reasonable geographical object information. Figure 8(c) shows that trajectory matched with different episodes represented by different colors annotates the same geographical object information. It can be seen that the algorithm proposed in this paper can annotate reasonable semantic information for spatiotemporal trajectories in geospatial environment.

5.3. Efficiency. In this section, we study the efficiency of our proposed algorithms. We compare it with the baseline approach and the LCS approach, which can annotate the semantic information on the similar trajectories. For each trajectory in the GeoLife dataset, we generate the semantic trajectory by the SEPSIM approach, the baseline approach, and the LCS approach, respectively, to retrieve the running time. The results of comparison are shown in Figure 9(a). We can see that the baseline approach and the LCS approach take more time annotating the same number of test trajectories than the SEPSIM approach. With the increasing test trajectories, the time spent by the typical approach and the LCS

approach and the time spent by the SEPSIM approach gradually become more time-consuming.

Figure 9(b) shows the efficiency of the SEPSIM approach with different spatial indexes. Obviously, the time spent by the SEPSIM with the *R*-tree index is much less than that of the other two spatial indexes in the SEPSIM approach, which means the *R*-tree index is appropriate to the dataset in this paper. Meanwhile, the SEPSIM approaches with the three indexes are faster than the typical approach and the LCS approach, which represents the high efficiency of our proposed SEPSIM approach.

6. Conclusion

In this paper, we study the problem of the semantic enrichment process for spatiotemporal trajectories in geospatial environments. We first directly use semantic information in preannotated semantic trajectories for annotating spatiotemporal trajectories by the SEPSIM approach. It includes three phases: preannotated semantic trajectory storage, spatiotemporal similarity measurement, and semantic information matching. We propose an algorithm named Semantic Information Matching Algorithm based on Similar Episodes (SIM) for matching semantic information. In order to improve the performance of efficient enrichment processing, we establish an *R*-tree index to query preannotated semantic trajectories. Finally, we conduct extensive experiments over a real dataset. The experimental results verify the superiority of our proposed approach in terms of effectiveness and efficiency.

Data Availability

The trajectory dataset used to support the findings of this study can be made available at <https://www.microsoft.com/en-us/download/details.aspx?id=52367>.

Disclosure

This paper expands on the short paper “Efficient Semantic Enrichment Process for Spatiotemporal Trajectories,” which was published in 4th Asia-Pacific Web and Web-Age Information Management, Joint Conference on Web and Big Data, APWeb-WAIM 2020.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Funding

This study was supported by NSFC41971343 and NSFC61702271.

Acknowledgments

This study was supported by the NSF of Jiangsu Province (BK20200725) and the Postgraduate Research Innovation Program of Jiangsu Province (KYCX201258).

References

- [1] D. Daowd and S. Mallappa, “Semantic analysis techniques using twitter datasets on big data: comparative analysis study,” *Computer Systems Science and Engineering*, vol. 35, no. 6, pp. 495–512, 2020.
- [2] L. H. Qi, R. G. Chen, and X. Wen, “Research on the LBS matching based on stay point of the semantic trajectory,” *Journal of Geo-Information Science*, vol. 16, no. 5, pp. 720–726, 2014.
- [3] A. Hussain, B. N. Keshavamurthy, and R. Prasad, “Accurate location prediction of social-users using mHMM,” *Intelligent Automation & Soft Computing*, vol. 25, no. 3, pp. 473–486, 2019.
- [4] F. Zhu, J. Gao, and C. Xu, “On selecting effective patterns for fast support vector regression training,” *IEEE transactions on neural networks and learning systems*, vol. 29, no. 8, pp. 3610–3622, 2018.
- [5] T. Bry and T. Fureche, “Web and semantic web query languages: a survey,” *Reasoning Web, Msida, Malta: Computer Science*, vol. 3564, pp. 35–133, 2005.
- [6] Z. X. Yan, D. Chakraborty, and C. Parent, “Semantic trajectories,” *ACM TIST*, vol. 4, no. 3, pp. 1–38, 2013.
- [7] C. Parent, S. Spaccapietra, C. Renso et al., “Semantic trajectories modeling and analysis,” *ACM Computing Surveys*, vol. 45, no. 4, pp. 1–32, 2013.
- [8] A. Daniel and S. Thad, “Using GPS to learn significant locations and predict movement across multiple users,” *Personal and Ubiquitous Computing*, vol. 7, no. 5, pp. 275–286, 2003.
- [9] K. John and H. Eric, “Predestination: inferring destinations from partial trajectories,” in *International Conference on Ubiquitous Computing*, pp. 243–260, Orange County, CA, USA, 2006.
- [10] A. T. Palma, V. Bogorny, B. Kuijpers, and L. O. Alvares, “A clustering-based approach for discovering interesting places in trajectories,” in *Proceedings of the 2008 ACM symposium on Applied computing - SAC '08*, pp. 863–868, Fortaleza, Ceara, Brazil, 2008.
- [11] Y. Zheng, L. Zhang, Z. Ma, X. Xie, and W.-Y. Ma, “Recommending friends and locations based on individual location history,” *ACM Transactions on the Web*, vol. 5, no. 1, pp. 1–44, 2011.
- [12] S. Spaccapietra, C. Parent, M. L. Damiani, J. A. de Macedo, F. Porto, and C. Vangenot, “A conceptual view on trajectories,” *Data & Knowledge Engineering*, vol. 65, no. 1, pp. 126–146, 2008.
- [13] M. Baglioni, J. Macêdo, and C. Renso, “Towards semantic interpretation of movement behavior,” in *12th AGILE Conference Advances in GIS*, pp. 271–288, Hannover, 2009.
- [14] A. Vandecasteele, R. Devillers, and A. Napoli, “From movement data to objects behavior using semantic trajectory and semantic events,” *Marine Geodesy*, vol. 37, no. 2, pp. 126–144, 2014.
- [15] P. T. Nogueira, R. B. Braga, and H. Martin, “An ontology-based approach to represent trajectory characteristics,” in *The 5th International Conference on Computing for Geospatial Research and Application*, pp. 102–107, USA, 2014.
- [16] T. P. Nogueira and H. Martin, “Qualitative representation of dynamic attributes of trajectories,” in *17th AGILE Conference on Geographic Information Science*, Castellón, Spain, 2014.
- [17] T. P. Nogueira, R. B. Braga, and C. T. Oliveira, “FrameSTEP: a framework for annotating semantic trajectories based on episodes,” *Expert Systems with Applications*, vol. 92, pp. 533–545, 2018.
- [18] L. G. Xiang, T. Wu, and J. Y. Gong, “A geo-spatial information oriented trajectory model and spatio-temporal pattern querying,” *Acta Geodactica et Catographica Sinica*, vol. 43, no. 9, pp. 982–988, 2014.
- [19] T. Sun, Z. Huang, H. Zhu, Y. Huang, and P. Zheng, “Congestion pattern prediction for a busy traffic zone based on the hidden Markov model,” *IEEE Access*, vol. 9, pp. 2390–2400, 2021.
- [20] E. Martin, K. P. Hans, and S. Jorg, “A density-based algorithm for discovering clusters in large spatial databases with noise,” KDD, Portland, Oregon, USA, 1996.
- [21] D. Mountain and J. Raper, “Modelling human spatio-temporal behaviour: a challenge for location-based services,” in *Proceedings of 6th International Conference on Geocomputation*, The University of Queensland, Brisbane, Australia, 2001.
- [22] M. P. Dubuisson and A. K. Jain, “A modified Hausdorff distance for object matching,” in *Proceedings of 12th international conference on pattern recognition*, pp. 566–568, Jerusalem, Israel, 1994.
- [23] J. Kima and S. Mahmassanibhan, “Spatial and temporal characterization of travel patterns in a traffic network using vehicle trajectories,” *Symposium on Transportation and Traffic Theory*, vol. 9, pp. 164–184, 2015.
- [24] M. M. Fréchet, “Sur quelques points du calcul fonctionnel,” *Rendiconti del Circolo Matematico di Palermo (1884-1940)*, vol. 22, no. 1, pp. 1–72, 1906.
- [25] Z. Chen and H. T. Shen, “Searching trajectories by locations: an efficiency study,” in *Proceedings of the 2010 ACM SIGMOD International Conference on Management of data*, pp. 255–266, Indianapolis, Indiana, USA, 2010.
- [26] F. Zhu, J. Yang, J. Gao, C. Xu, S. Xu, and C. Gao, “Finding the samples near the decision plane for support vector learning,” *Information Sciences*, vol. 382–383, pp. 292–307, 2017.
- [27] A. Guttman, “R-trees: a dynamic index structure for spatial searching,” in *Proceedings of the 1984 ACM SIGMOD*

international conference on Management of data, pp. 47–57, Boston, Massachusetts, USA, 1984.

- [28] R. Kanth, S. Ravada, and D. Abugov, “Quadtree and R-tree indexes in oracle spatial: a comparison using GIS data,” in *Proceedings of the 2002 ACM SIGMOD international conference on Management of data*, pp. 546–557, Madison, Wisconsin, USA, 2002.
- [29] X. F. Xu, L. Xiong, and V. S. Sunderam, “D-Grid: An in-memory dual space grid index for moving object databases,” in *2016 17th IEEE International Conference on Mobile Data Management (MDM)*, pp. 252–261, Porto, Portugal, 2016.

Research Article

A Deep Learning-Based Inventory Management and Demand Prediction Optimization Method for Anomaly Detection

Chuning Deng  and Yongji Liu 

School of Business Administration, Liaoning Technical University, 125105 Liaoning, China

Correspondence should be addressed to Yongji Liu; liuyongji@lntu.edu.cn

Received 26 July 2021; Accepted 21 September 2021; Published 11 October 2021

Academic Editor: Fa Zhu

Copyright © 2021 Chuning Deng and Yongji Liu. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The rapid development of emerging technologies such as machine learning and data mining promotes a lot of smart applications, e.g., Internet of things (IoT). The supply chain management and communication are a key research direction in the IoT environment, while the inventory management (IM) has increasingly become a core part of the whole life cycle management process of the supply chain. However, the current situations of a long supply chain life cycle, complex supply chain management, and frequently changing user demands all lead to a sharp rise in logistics and communication cost. Hence, as the core part of the supply chain, effective and predictable IM becomes particularly important. In this way, this work intends to reduce the cost during the life cycle of the supply chain by optimizing the IM process. Specifically, the IM process is firstly formulated as a mathematical model, in which the objective is to jointly minimize the logistic cost and maximize the profit. On this basis, a deep inventory management (DIM) method is proposed to address this model by using the long short-term memory (LSTM) theory of deep learning (DL). In particular, DIM transforms the time series problem into a supervised learning one and it is trained using the back propagation pattern, such that the training process can be finished efficiently. The experimental results show that the average inventory demand prediction accuracy of DIM exceeds about 80%, which can reduce the inventory cost by about 25% compared with the other state-of-the-art methods and detect the anomaly inventory actions quickly.

1. Introduction

The emergence of artificial intelligence, big data, data mining, and other technologies, as well as the rapid development of computer hardware performance, has exerted a profound influence on the performance of the supply chain technology [1]. For example, the efficiency of the whole supply chain life cycle (e.g., technical support and product delivery) can be optimized based on big data analysis technology [2], while potential customers of products can be dug out to improve profits based on data mining technology [3]. The traditional supply chain process usually only involves procurement, inventory, production, and distribution. Nowadays, with the rapid development of the information technology, the customers who were originally excluded from the supply chain management process have now become particularly important. In this way, the whole life cycle management

process of the supply chain becomes longer and longer, which involves the procedures of manufacturing, supply, storage, transportation, distribution, and retail [4]. During each procedure, the activities of operation, control, and optimization are all required. Besides, the collaboration among these procedures is also required. More importantly, since the customers are included in the supply chain management process, their frequently changing demands would increase the uncertainty of key product provision in supply chain management [5], which not only aggravate the complexity of supply chain management but also greatly increase the overall cost.

Inventory management (IM) [6], as the key part of supply chain management, plays a very important role in reducing the overall cost of supply chain management. Generally, too much or too little inventory can have a bad result. For example, excessive inventory can result in the oversupply

situation, since the amount of stored products has exceeded the market demands greatly. In this case, the corresponding inventory cost would also be high, because most products will have to be stored for a long time, which will then lead to the situations of slow or insufficient resource turnover of enterprises [7]. In addition, too little inventory may cause the insufficient coverage situation of customer demands, which will then gradually lead to the shortage of products and the reduction of customer trust and even the profit [8]. Therefore, the inventory management is becoming more and more important for supply chain management, since effective IM will both reduce the cost and increase the profit. Based on this consideration, more and more attention is paid to the research area of inventory management and optimization.

Generally, the performance and function of IM are largely affected by the prediction accuracy of the future customer demands [9], since most of the IM decisions are made according to the predicted results. The bad prediction results can reduce the sales volume of products, while the good prediction results can naturally improve the number of products sold. In this way, most researches would like to improve the customer demand prediction accuracy to achieve a better performance of IM. In fact, the customer demand prediction accuracy can be improved by deeply analyzing the customers' demand for products. However, how to accurately obtain the customers' demand for product becomes a hot topic in the supply chain and inventory management fields.

Most traditional enterprises analyze the potential demand of customers by studying their historical order data on the basis of some traditional statistical analysis and data mining technologies; for example, reference [10] adopted the data mining technology to discover the relationship between customer needs and the market trends. This can indeed obtain some efficient information, but it is also usually unable to adapt to the rapidly changing customer demands, thus leading to a low demand prediction accuracy [11]. To solve this awkward situation, some other related studies tried to adapt to the actual supply and demand ratio by using the strategy of single-point inventory and bulk order. For instance, the single-point inventory and bulk order strategy was used by the reference [12] to achieve a reliable inventory management. Despite this, this single-point inventory means a centralized processing method, which always suffers from the performance bottleneck. Hence, some researchers also begin to study the distributed and dynamic scheduling strategy to optimize the inventory management process which includes the inventory replenishment and distribution [13]. Nevertheless, the diversity of products and the dynamic nature of customer requirements both increase the uncertainty of inventory management and provision, making the existing methods no longer applicable.

The rapid development of artificial intelligence (AI) technology and the computer hardware capabilities allows us to make many decision-making parts of the supply chain management process intelligent, which includes the inventory management. Intelligent inventory decision-making

can adapt to the changes of environment and customer demands, so as to cope with the continuous and long-lasting customer demands [14]. Based on this consideration, this paper proposes a deep inventory management (DIM) method using the long short-term memory (LSTM) theory of deep learning (DL) [15]. DIM intends to predict customers' demands, according to which the intelligent decisions for inventory management can be made. Usually, the key to the application of LSTM lies in the comprehensibility of the learning model and the accuracy of the prediction. Although most research show that the LSTM-based neural network can offer a high prediction accuracy, the incomprehensibility of its prediction behavior hinders its application in solving the inventory management problem efficiently [16]. In this regard, the proposed DIM method firstly introduces the state unit before the hidden layer, thus to save more long-term information and gradient information, so as to alleviate the problem of gradient disappearance to some extent. After that, DIM converts the prediction results from the LSTM training model to a corresponding product popularity rating indicator, which will then be used to guide and optimize the inventory management.

The main contributions of this paper are summarized as follows:

- (i) This work formulates the supply chain and inventory management problem into a novel multiobjective optimization model which comprehensively considers multiple factors of inventory management. In particular, the objective mainly includes cost minimization and profit maximization
- (ii) Based on the formulated model, this work proposes the deep inventory management method DIM to address the challenges faced by inventory management. Particularly, by using the LSTM theory, DIM offers intelligent decision-making ability for the inventory management
- (iii) The experimental results indicate that the proposed DIM method can effectively predict the customer demand trends with the prediction accuracy exceeding about 80% and reduce the overall cost by about 25%

The rest of this work is organized as follows: Section 2 mainly discusses the relevant research work in recent years to show the main stream of the research direction of IM. Section 3 presents the constructed mathematical model of inventory management in this work. Section 4 explains the corresponding inventory management and optimization algorithm proposed in this work. Section 5 discusses the experimental results with deep and detailed analysis given. Section 6 makes a summary.

2. Related Work

The related work of supply chain and inventory management is separated into two categories which are the

traditional inventory management methods and the intelligent inventory management methods, as follows.

2.1. Traditional Inventory Management. The main purpose of inventory management is to store a certain amount of physical resources for a company or enterprise, which can then be transformed into profits via effective product sale or other operations [17]. As explained, the inventory management is a core part of the supply chain management process, which is now becoming a vital focus of many enterprises and companies.

Generally, the performance of the inventory management can be affected by a lot of factors. Since it is one key part of the supply chain management, a great deal of researches have been proposed to optimize the efficiency of the inventory management process to finally promote the performance of the supply chain. For example, reference [18] would like to optimize the process of inventory management by using the performance management technology to promote the activities of the whole supply chain management. Specifically, this work classified irregular demands into three kinds which are erratic, slow moving, and lumpy ones. Then, three corresponding periodic review policies were proposed to maintain the lowest holding inventory. The results indicated that this work was very effective especially for dealing with the erratic inventory demands. Despite this, this work did not take the dynamic changing environment into consideration, which may not be applicable.

In order to show a clear direction of the inventory management, reference [19] discussed the challenges of IM by dividing the corresponding challenges into five categories, that is, the technology, the organization, the finance, the management, and the information involved in the process of inventory management. In particular, such classification was made based on the decision variable, the demand type, the quality deterioration function, and the method of settlement used, such that the classification could not only provide the detailed description about IM but also show a gap to be developed. Similarly, reference [20] also explored the potential challenges of IM by dividing the whole process into multiple different aspects which included the safety inventory, the procurement efficiency, the demand prediction, and the training and interaction. Such two kinds of classification both promote the development of inventory management, but the difference is that [20] studied each aspect deeply and provided optimization directions for each of them.

Compared with the above research work that intended to review inventory management, many other work were more inclined to study the technical aspects of inventory management. For example, reference [12] focused more on addressing the unnecessary out-of-stock and the oversupply issues that happened during the process of inventory management. In order to address these issues, this work proposed to optimize both the transport and inventory, such that the strategies of the single inventory and bulk order were jointly taken into consideration. In this way, the market supply-demand ratio could be dynamically adapted. But the drawback still

existed, that is, it would take a long time distributing the products from the warehouse to the retailer, when there was a shortage of products at the retail level. As for reference [13], it tried to optimize the entire inventory management process via introducing a novel inventory replenishment and distribution model. Specifically, this work first analyzed the characteristics of the existing distributed inventory model, and then, it combined the advantages of cloud computing and the distributed inventory model to finally build the hierarchy-control distributed inventory model. By simulation and calculation, the results indicated that the proposed distributed inventory model was correct and effective. However, it should be noted that both references [12, 13] did not consider the uncertainty of the market, which may cause great loss especially when the oversupply situations happen.

In order to prevent such incidents from happening, references [21, 22] tried to build a safety inventory and start the research from the perspective of reducing the loss caused by the uncertainty of customer demands. For [21], it focused on optimizing the deficiency of traditional inventory management methods and forecasting the demands for all kinds of emergency supplies using the Euclidean algorithm. For [22], it developed an automated inventory system based on the passive radio frequency ID. Compared with the manual system, the product delivery time was reduced from 15.45 minutes to 2.92 minutes on average. However, the two references were different, which was mainly reflected by the fact that the safety inventory in [21] mainly considered the number of customers, customer satisfaction, delivery reliability, and supply reliability, while the safety inventory in [22] mainly considered the product usage frequency, service quality level, and sales situation. Therefore, the former realized a safety inventory from the view of customers, while the latter realized a safety inventory from the perspective of products. Nevertheless, none of them are totally intelligent, which means that the human intervention will be required more or less, and then, the inventory error would occur with a high probability when the amount of product becomes extremely large.

2.2. Intelligent Inventory Management. The turning point of the development from traditional inventory to intelligent inventory is about the prediction technology of customer demands. Generally speaking, there are many prediction methods and they can be divided into two categories. The first one mainly relies on static mathematical statistical analysis methods, while the second one mainly relies on the machine learning methods [23].

The traditional static mathematical analysis-based methods (e.g., statistics and data mining) rely heavily on the quality of history data (e.g., product order). Hence, high-quality data would lead to a more accurate prediction accuracy than low-quality data. Usually, the mathematical analysis is executed and applied on these history data for the purpose of digging out the potential pattern and trends about the market needs. Such needs are actually proportional to the customer demands, based on which we can optimize the inventory management process. For example,

reference [10] tried to predict the customer demands mainly by studying their historical order data. Specifically, this work first searched the corresponding product order information on the web. Then, the data mining technology was applied to dig the potential customer demands in the future and finally to establish one simple but concise inventory policy. The results indicated that by using the history sales data, this work could reduce the total cost of inventory more efficiently. As explained, the quality of the data is of vital importance. However, the collected data from the web usually has very low quality and needs a lot of extra procedures before putting them into use.

Similarly, another work, that is, reference [24], also used the data mining technology to address the inventory management problem. In particular, this work focused more on studying the correlation among the historical data. Based on the intercorrelation discovered, a more complete analysis was carried out to improve the prediction accuracy and this work claimed to reduce both the cost and energy consumption for enterprises. Despite this, mainly relying on static mathematical statistics leads to an awkward situation that the prediction accuracy achieved by these methods is not very high. Another factor greatly influencing the achieved prediction accuracy is the data quality. However, it is generally known that the open-source data quality cannot be guaranteed and their format may not satisfy the corresponding requirements, such that the robustness and scalability of inventory management cannot be guaranteed neither.

Compared to the above work using the statistical methods, the second kind of research work on intelligent inventory has higher intelligence, since they mainly rely on the well-known machine learning models and methods (e.g., deep learning and reinforcement learning methods). The general workflow for most work using the machine learning method is that they first establish and train a learning model. It is noted that the learning model should be trained by a large number of historical dataset to generate a common knowledge system. After that, the customer demand prediction can be carried out based on this trained model. The more mature this model is, the higher the accuracy of the prediction results is.

For instance, reference [25] adopted the artificial neural network (ANN) to deal with the process of inventory management. The intermediate process of inventory was modeled as the ANN's hidden layer. After that, the learning model was continuously trained to approximate a solution with the optimal prediction accuracy. Despite the case that this work claimed to have achieved better results, there are also some limitations for this work, for example, it assumed that the demand changes regularly, while such changes were uncertain in the real world. Another limitation was that this work did not take enough consideration on the impact factors of the safety stock, which resulted in the situation that the actual safety stock was inadequate.

Reference [26] also adopted ANN as the technology to establish a learning model for inventory management. The difference between [25, 26] was that the latter constructed an additional set of knowledge discovery system to convert the results obtained from the learning model into more

accurate knowledge, so as to guide the process of inventory management. Specifically, we can discover that the inventory management of [26] was actually handled by a cloud-based customer relationship management framework. This work claimed that the proposed framework could help the enterprises about future plans of their inventory based on the past history of paid invoice data. Meanwhile, the JSON script language was used to conduct the experiments, which indirectly increased the burden, since it needed to be parsed before putting into use.

In addition, reference [27] relied on using the back propagation neural network (BPNN) technology to construct an inventory management and learning model. Then, a simple and practical inventory strategy was calculated based on the training model. Different from the above linear prediction, this work presented the nonlinear prediction due to the uncertainty and diversity of the market needs. To fulfill such objective, the step size of BPNN was set to be variable, based on which the prediction accuracy would be more precise and the investment risk would be reduced. The drawback of back propagation is also very obvious, that is, it cannot evolve automatically. In this regard, reference [28] adopted the technology of reinforcement learning together with a heuristic strategy to finally address the multilayer inventory management and optimization problem. On one hand, the reinforcement learning was used to build the inventory management model under a global view. On the other hand, under the guidance of the RL model, the efficient and rapid inventory decision process could be realized via using a local heuristic strategy. The experimental results indicated that this work could improve the performance of profitability, adaptability, and solution time.

2.3. Discuss. The rapid development of information technology, the dynamic nature of customer demand, and the complexity of business have now far exceeded the application scope of traditional inventory management methods. Therefore, the demand prediction-driven methods are born, which mainly depend on the mathematical statistics and machine learning technologies. However, the mathematical statistics relies heavily on the quantity and quality of historical data, which leads to the situation that they often fail to keep pace with the rapid change of today's customer demands. In this way, the prediction accuracy is reduced. On the other hand, the machine learning-based methods can improve the prediction accuracy for inventory management greatly based on continuous learning. However, the corresponding research is still in the early stage and there is still much room to improve the accuracy of demand prediction for inventory management. Under these conditions, the inventory management method DIM is proposed in this work to adapt the changes of the environment and customer demands by dynamically adjusting the prediction scope and accuracy, so as to reduce the overall inventory management cost and increasing the profit.

3. Problem Model and Objective

This section mainly focuses on building the inventory management and optimization model, which includes the

inventory management problem, the objective, and the constraints.

3.1. Supply Chain Inventory Model. Firstly, this work assumes that the whole supply chain logistics and warehousing system is composed of n inventory nodes and m external supplier nodes. Then, given any warehousing node $p \in [1, n]$ (p is an integer discrete variable), the inventory amount of this node at time t is denoted by $I_p(t)$, while the market demanding amount is denoted by $d_p(t)$. Apparently, when the demanding amount exceeds the inventory amount, product supplement will be required. Now, using $u_p(t)$ to indicate the supplement amount of products at time t for the warehousing node p , then, $u_p(t)$ may be supplied by multiple provision nodes at different times (e.g., t'). Hence, it is expanded as follows:

$$u_p(t) = \sum_{q=1, q \neq p}^{m+n} \lambda_{qp} u_p(t') X_p^q, \quad \forall t' < t, \quad (1)$$

where $X_p^q \in \{0, 1\}$ means whether node p demands the product from node q ($q \in [1, m+n]$ and $q \neq p$) and λ_{qp} means the ratio between the amount of products demanded by p and the overall requirement of node p .

Based on the above definitions, we can build the relationship among the three notations (i.e., $I_p(t)$, $d_p(t)$, and $u_p(t)$), as follows:

$$I_p(t+1) = I_p(t) - d_p(t) + u_p(t) = I_p(t) - d_p(t) + \sum_{q=1, q \neq p}^{m+n} \lambda_{qp} u_p(t') X_p^q. \quad (2)$$

According to the calculation of equation (2), it is easy to observe that the values of $d_p(t)$ and $u_p(t)$ have a great impact on the amount of current inventory. Since $d_p(t)$ indicates the market demanding amount of products and $u_p(t)$ indicates the supplement amount of products for warehouse node p , we can now abstract them as the input and output the node p , respectively. Specifically, $u_p(t)$ is the input of p and $d_p(t)$ is the output of p . Since the input is already formulated in equation (1), we now formulate the output price for node p . Assuming that the output price is denoted by $\text{price}(d_p(t))$, then, we have

$$\text{price}(d_p(t)) = \sum_i w_i x_i, \quad \forall i > 0, \quad (3)$$

where i means the category of output products, w_i means the value of product i , and x_i means the number of product i .

3.2. Objective and Constraints. For the inventory management, the more products we sell (i.e., $d_p(t)$), the more profits we gain (i.e., $\text{price}(d_p(t))$). However, we should also guarantee the product update speed in the warehouse node, since the faster we update, the lower the inventory cost per unit of products. Jointly taking the two factors into consideration,

we establish the following two-objective optimization model on the basis of equations (1)–(3):

$$\begin{aligned} \text{Maximize :} & \quad \{f_1(t), f_2(t)\} \\ \text{s.t.} & \quad f_1(t) = u_p(t) \\ & \quad f_2(t) = \text{price}(d_p(t)), \end{aligned} \quad (4)$$

where $f_1(t)$ indicates the input condition of warehousing node p . The more products imported per time unit, the faster the updating speed of warehousing node p . $f_2(t)$ indicates the price condition of exported products. The higher the price, the higher the profit.

In particular, as explained, $u_p(t)$ is the input and $d_p(t)$ is the output amount of products of warehouse node p . Apparently, the larger value of $d_p(t)$, the higher the prices, that is,

$$f_1(t) \propto d_p(t). \quad (5)$$

On the other hand, the higher value of $d_p(t)$ also means that more products are sold out, such that more products should be supplemented as well, which is equivalent to the value of $u_p(t)$. From such deduction, we also have

$$f_2(t) \propto d_p(t). \quad (6)$$

Combining equations (5) and (6), we can conclude that the two objective functions $f_1(t)$ and $f_2(t)$ are not on the opposite, such that we do not need to make a tradeoff between them. In addition, the implementation of equation (4) must also satisfy the following constraints:

$$\begin{aligned} \sum_q \lambda_{qp} &= 1, \quad \forall 1 \neq p, \\ \lambda_{qp} &\neq \lambda_{pq}, \\ \sum_q X_p^q &< m+n, \quad \forall q \neq p, \\ d_p(t) &\leq I_p(t) + u_p(t), \quad \forall t > 0, \\ I_p(t) &\geq 0, \quad \forall t > 0, \end{aligned} \quad (7)$$

where the first constraint means that the overall amount of products demanded by node p from all the other nodes cannot exceed the original requirement of p ; the second constraint means that the supply and demand ratio between any two nodes may not be constant; the third constraint means that the product supply nodes are within the range of the already known m product providers and the other $n-1$ warehousing nodes; the fourth constraint means that the output amount of products on node p cannot exceed the sum amount of input and current inventory; the last constraint means that the inventory amount of products on node p at time t cannot be negative at any time.

4. Deep Inventory Management Algorithm Design

As known, the LSTM is a kind of the deep learning technology, which makes predictions toward different metrics according to the time series data. Nowadays, the customer demands usually change greatly at different stages of time with a high frequency, which leads to the situation that the product inventory should be managed accordingly to satisfy such changes. Based on this kind of characteristic, we can regard the corresponding generated data as some kind of time series data. Hence, in this work, we propose the deep learning-based inventory management method DIM which greatly leverages the characteristics of LSTM (i.e., time series and back propagation-based prediction) to optimize the inventory management process. The system framework of DIM is shown in Figure 1, which consists of several modules including the data collection module, the data preprocessing module, the training module, and the prediction module. The main functions of these modules are explained as follows:

- (i) Data collection module: it is used to collect the historical order data which are raw, disordered, and massive. Despite this, this module is the foundation of the other modules
- (ii) Data preprocessing module: it is used to handle the raw and disordered data, for example, cleaning the useless data and extracting the effective data features for the following training
- (iii) LSTM module: it is used to create a learning model based on the input data and to finally output a value for future demand prediction
- (iv) Prediction module: it is used to calculate the demanding level of any product in the market based on both the learning model and the product popularity

In this work, the prediction accuracy of DIM is mainly evaluated by the popularity of products, which is first defined as follows:

Definition 1. Product popularity means the popular trend and importance of the product in the current market. In this work, it is calculated based on the product demand frequency and value per time unit. Given the type of product and denoting it by i , then, the corresponding product popularity can be calculated as follows:

$$\text{Pop}_i = f_i \times \frac{w_i - w_{i,\min}}{w_{i,\max} - w_i}, \quad (8)$$

where $w_{i,\min}$ and $w_{i,\max}$ indicate the lower and upper value bounds of the i th type of products, respectively; $(w_i - w_{i,\min})/(w_{i,\max} - w_i)$ is used as the standard operation to reduce the influence on product popularity calculation, which is caused by the price gap between different products.

Generally, the higher the value of Pop_i , the more popular the i th kind of product, which means that this product is frequently demanded by customers, such that more customers may demand such kind of product in the future. Based on Definition 1, we next elaborate the main procedures of DIM.

4.1. Training Data Collection. The training data are very important for product popularity prediction. Hence, we need to collect the data as much as possible and as high quality as possible. On one hand, we can obtain a lot of historical order data from the open-source websites. On the other hand, we can also regularly collect the product order information from the online shops. These information can be directly used to reflect the distribution of the customer needs in a certain extent. However, the deeper relationship between these information and the more customer demands should be explored. From the perspective of inventory management, the product order has a lot of attributes, among which nine of them are mainly used in this work for product popularity prediction. Specifically, the nine attributes are the order date, the current popularity of the product, the name of the product, the type of the product, the weight of the product, the number of the product, the price of the product, the brand of the product, and the origin of the product. Despite this, we should be aware that there may be a lot of useless data. In this way, a simple criterion is defined to filter out these useless data, that is, if we cannot extract the required nine attributes from the data, then, we discard the data. Hence, after such operation, all the data left behind can meet our requirements. Now, for each piece of the collected data, the first thing that we should do is to extract the values of the above nine attributes. After that, we can easily describe the current distribution of the customer demands according to these obtained information. More importantly, we need to predict the future distribution of customer demands based on these obtained information. In order to fulfill this target, we define the product popularity into 8 levels which are shown in Figure 2. The higher the level, the more popular the product is. Hence, for the proposed model and mechanism, given any input (i.e., the product), the output is the service level (i.e., the popularity) that this product would have in the future. Reviewing the details in Figure 2, we should also note that, the corresponding learning model associated with the proposed mechanism should be trained well before being used to predict the future customer demands.

4.2. Data Preprocessing. According to the actual situation, the value of the same attribute in terms of different products may differ quite a lot, which then may lead to the unfair comparison between the achieved prediction result and the actual situation. In this regard, it is necessary to carry out further data preprocessing before training the model and predicting the results. On this basis, it becomes convenient to compare different indicators. On the other hand, the prediction of customer demands will be more accurate. Starting from this point, this work selects the min-max operation [29] to standardize and normalize these attributes of the product. Then, the value of these attributes will be mapped

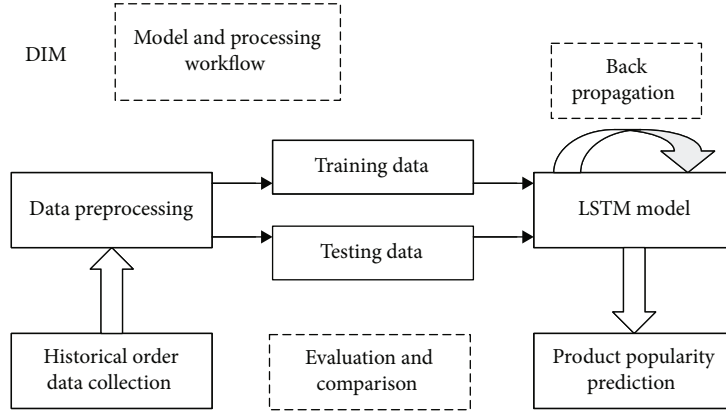


FIGURE 1: The system framework and workflow of DIM.

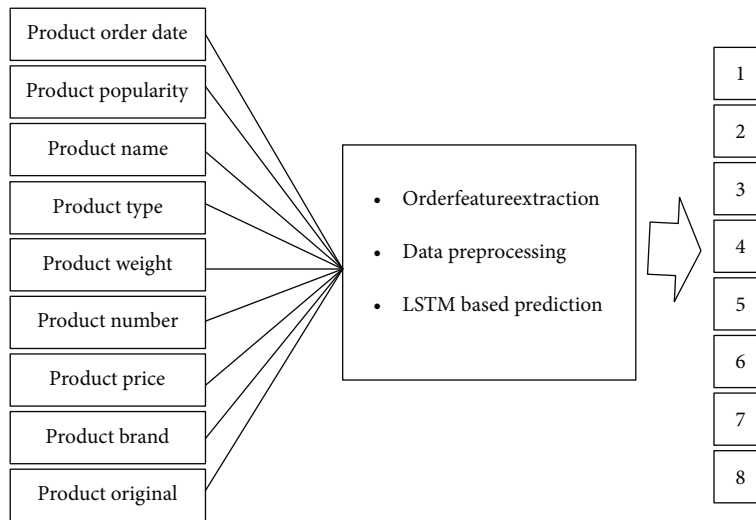


FIGURE 2: Illustration diagram of product order attributes and prediction results.

into the same scope of $[0,1]$. The specific preprocessing equation is shown in (9):

$$v = \begin{cases} \frac{|w_i - w_{i,\min}|}{|w_i - w_{i,\max}|}, & v_{\min} < v < v_{\max}, \\ 1, & v = v_{\max}, \\ 0, & v = v_{\min}, \end{cases} \quad (9)$$

where $v_{\min} = \min \{v_1, v_2, v_3, \dots\}$ means the minimum value among all the data in terms of the same attribute, $v_{\max} = \max \{v_1, v_2, v_3, \dots\}$ means the maximum value among all the data in terms of the same attribute, and v is the data after mapping.

4.3. Prediction Model Establishment and Training. The proposed DIM method adopts LSTM to build the prediction model. Compared with the RNN model, DIM introduces a state unit on the basis of the hidden layer for the purpose of retaining more long-term information. Accordingly, such operation will also retain more gradient information. In this way, LSTM-based DIM can alleviate the problem of gradient

disappearance to a certain extent compared to RNN. The prediction structure of DIM is shown in Figure 3, where (1) the input parameters will be trained using the back propagation method, (2) the time series problem of LSTM will be transformed into a supervised learning problem via the hidden layer, and (3) the output of the previous layer will be used as the input of the next layer, so as to iteratively complete the training process.

Observing Figure 3, we can see that the proposed prediction model is composed of three inputs which are the vector of product order $V = \{v_1, v_2, v_3, \dots\}$, the state unit vector $C = \{c_1, c_2, c_3, \dots\}$, and the hidden layer vector $H = \{h_1, h_2, h_3, \dots\}$. Hence, we need to firstly initialize all the vectors and the corresponding weight matrixes. On this basis, we train the prediction model of DIM. As shown in Figure 4, we mainly focus on the calculation of the forgetting gate F_t , the input gate I_t , the output gate O_t , and final output prediction value.

First of all, the forgetting gate F_t is mainly used to control the number of states of c_t which remained in c_{t-1} . The input gate I_t is mainly used to control the number of states of c_t that should be maintained by the input v_t at time t .

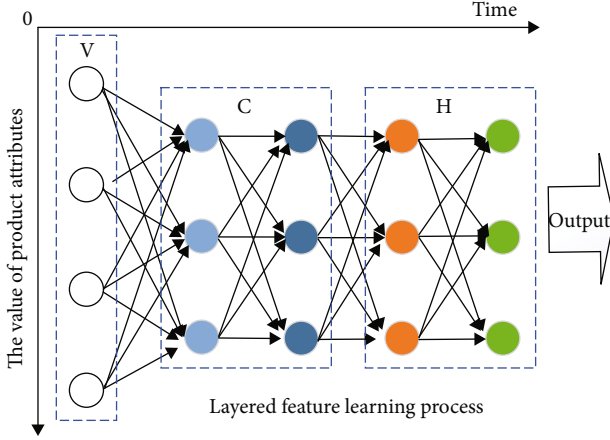


FIGURE 3: Prediction structure of DIM.

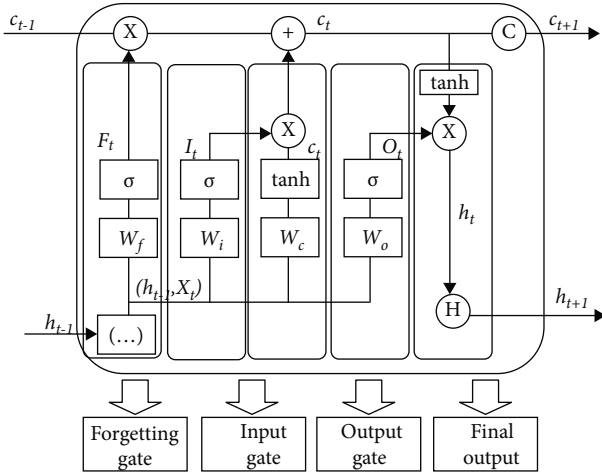


FIGURE 4: Illustration of training model.

The output gate is mainly used to control the number of output h_t with the state of c_t . Based on these, the forgetting gate is calculated as follows:

$$F_t = \sigma(W_f \cdot [h_{t-1}, v_t] + b_f), \quad (10)$$

where W_f is the weight matrix of the forgetting gate, $[h_{t-1}, v_t]$ indicates the connection between h_{t-1} and v_t , b_f indicates the offset item of the forgetting gate, and σ is a function with the value in the scope of $[0,1]$.

Typically, the weight matrix for prediction model training is of vital importance. For the calculation of the input gate I_t , the weight matrix should also be used. Denoting the weight matrix by W_i , then, the calculation of I_t is as follows:

$$I_t = \sigma(W_i \cdot [h_{t-1}, v_t] + b_i). \quad (11)$$

Similarly, W_o and b_o indicate the corresponding weight matrix and offset item of O_t , respectively, as shown in (11). Since the structure form of the output gate is the same with

those of the forgetting gate and the input gate, now given the weight matrix W_o and the offset item b_o for O_t , it follows that

$$O_t = \sigma(W_o \cdot [h_{t-1}, v_t] + b_o). \quad (12)$$

The final output result of DIM is jointly determined by the output gate and the unit state, such that we have

$$H_t = O_t \circ \tanh h(C_t), \quad (13)$$

where the notation \circ indicates the operator of matrix multiplication.

4.4. Prediction-Based Inventory Management. As explained and shown in Figure 2, the final output of the proposed DIM should be mapped to the 8 kinds of product popularity levels which are then leveraged to further predict the product demands in the future. According to equation (13), the final output of DIM based on LSTM is denoted by H_t and we expand it as follows:

$$H_t = \{h_t^1, h_t^2, h_t^3, \dots\}, \quad (14)$$

where h_t^i is the output function of product i and it is calculated as follows:

$$h_t^i = \left\lceil \frac{h_t^i - \min \{H_t\}}{\max \{H_t\} - \min \{H_t\}} \times 8 \right\rceil. \quad (15)$$

Based on equation (15), given any product, the output value of this model is constrained in the set $\{1, 2, 3, 4, 5, 6, 7, 8\}$. As explained, the higher this value, the more popular the product. Hence, we can arrange the inventory allocation and warehousing in advance according to the popularity value of different products. Then, assuming that there are k groups of data in total, the whole objective function can be formulated as follows:

$$\begin{aligned} \text{Maximize :} & \quad \sum_{i=1}^K w_i \cdot x_i \\ \text{s.t.} & \quad (4), (5), (6), (7). \end{aligned} \quad (16)$$

Now, with the formulated objective, the overall working process of the proposed DIM can be described in several steps, that is, (1) we need to calculate the current product popularity and collect the history data, which are then regarded as the input of the predicting model; (2) with the predicted product popularity, we should first check if the amount of the product with the highest popularity is enough or not in the warehouse node. If it is enough, then, repeatedly check the product with the second highest popularity until all the popular products are guaranteed; and 3) if the popular products are not enough, then, we should try to store enough products in the nearest warehouse. It is noted that the nearest warehouse node may not have enough room to store the popular products; in this case, another

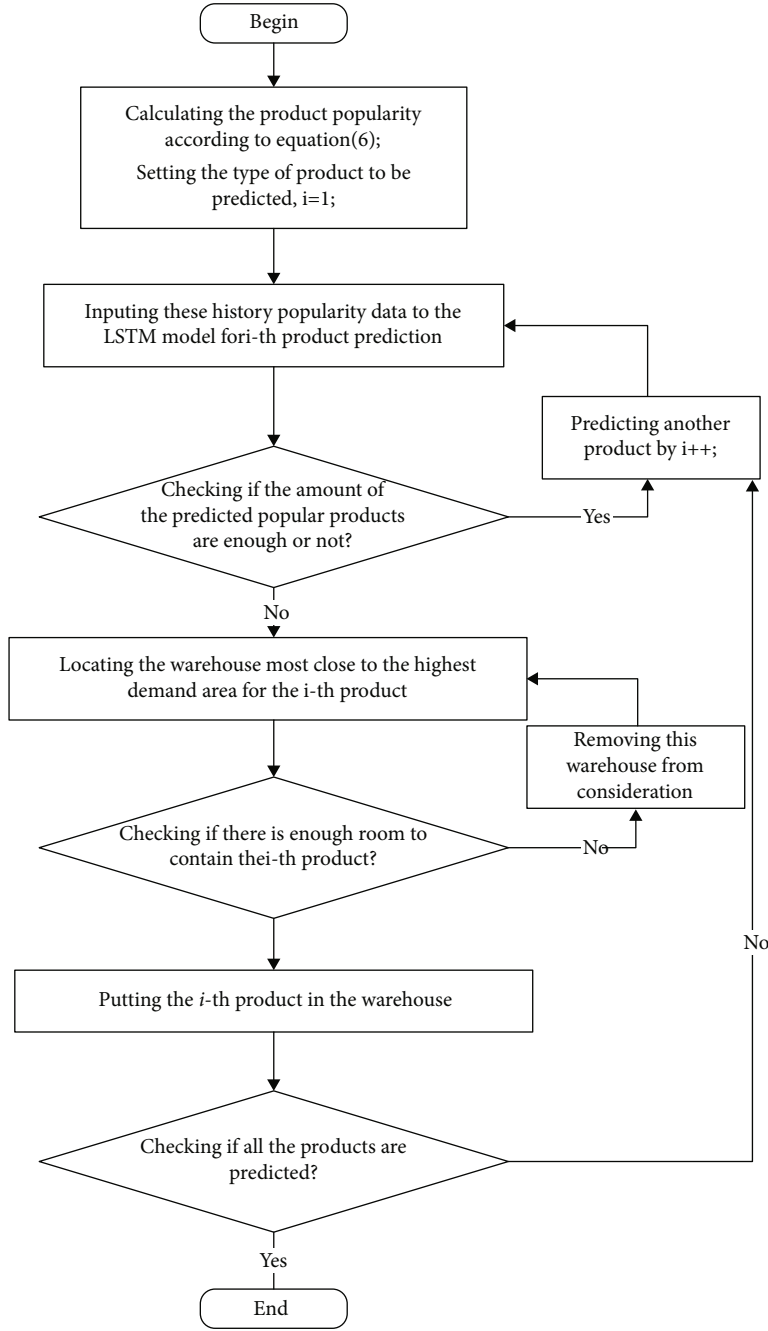


FIGURE 5: Flowcharts of DIM.

warehouse with available room should be used instead. The corresponding flowchart of the above working process of DIM is presented in Figure 5.

5. Performance Evaluation

In this section, the proposed algorithm are tested and evaluated following the steps of the parameter setup, the benchmark algorithms and metrics, and the experimental results.

5.1. Setup. In the experiment, we consider four inventory nodes and two external supplier nodes, that is, $n = 4$ and

$m = 2$. The distribution structure is shown in Figure 6, where external supplier node 1 provides products for inventory nodes 1 and 3, while external supplier node 2 provides products for inventory nodes 2 and 4. In addition, the adjacent nodes can also supply products for each other, such that the pairs include $\{(1, 2), (1, 3), (2, 4), (3, 4)\}$. Each arrow in Figure 6 has two attributes denoted by $\langle \lambda_{i,j}, w_{i,j} \rangle$, where $\lambda_{i,j}$ is the ratio between the amount of products demanded by node j from node i and the amount of all demanding products, while $w_{i,j}$ is the corresponding price.

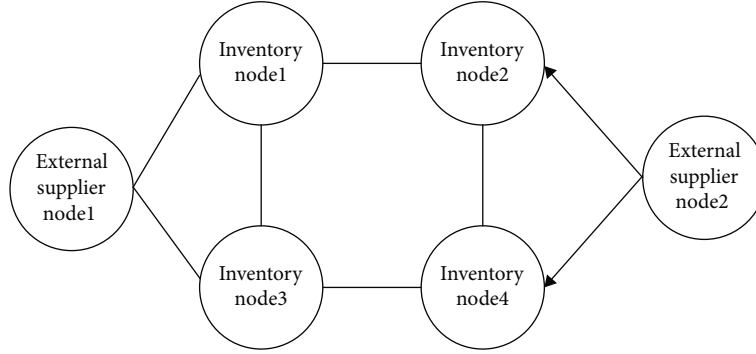


FIGURE 6: Connection structure of inventory nodes and supplier nodes.

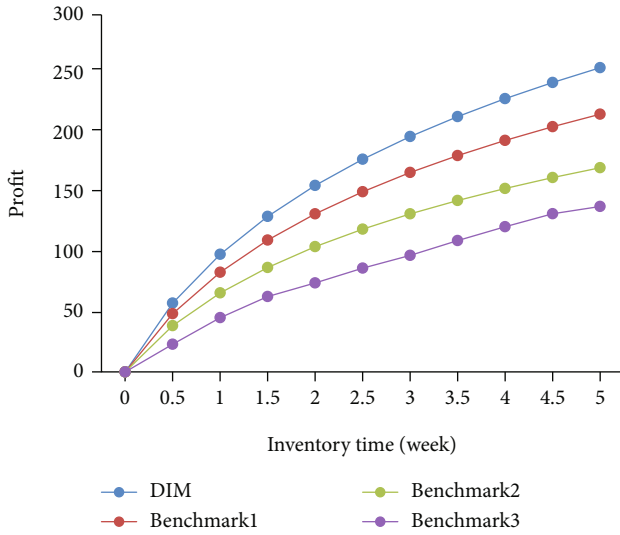


FIGURE 7: Results of the inventory sale profit.

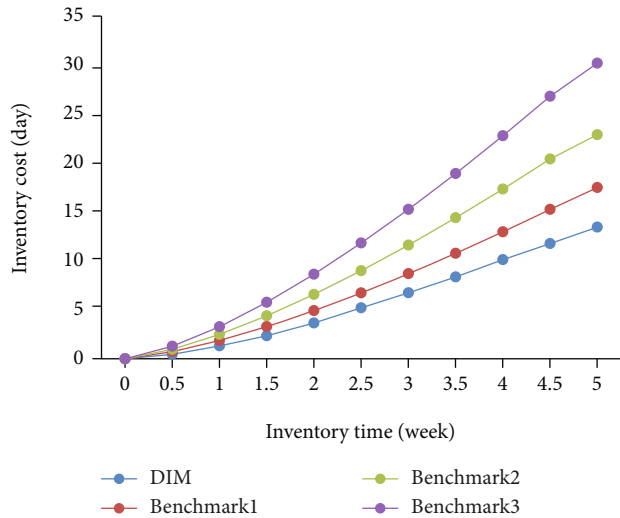


FIGURE 8: Results of the cost on inventory time.

In order to train the LSTM model and verify the effectiveness of the proposed mechanism, we use the crawler software to obtain the product order data of 25 weeks from the

Taobao websites, which exceeds 100 thousand records. However, due to the chaos of the raw data, we select 10 thousand high-quality records as the dataset, where 8 thousand records are uniformly distributed in the first 20 weeks and 2 thousand records are uniformly distributed in the last 5 weeks. In particular, among the 10 thousand records, there are around 100 kinds of products. As explained, each record has nine product attributes, that is, date, popularity, name, type, weight, number, price, brand, and original, and we should note that the attribute of popularity is actually calculated in this work and attached to the obtained order data as one attribute. Moreover, the first 8 thousand records of data in the first 20 weeks are used to train the product popularity prediction model and the remaining 2 thousand records of data in the last 5 weeks are used as the test data for simulation and performance evaluation.

The experimental environment and the proposed algorithm are implemented by using the JAVA language on the basis of the Microsoft Windows 10 (64 bits) OS, Intel(R) Core(TM) i5-7400 CPU @3.00 GHz, 16 GB.

5.2. *Benchmarks and Metrics.* There are three kinds of evaluation metrics according to the objective shown in equation (4), as follows:

- (i) The profit is evaluated by the value earned after the inventory sales and calculated according to equation (16)
- (ii) The cost is mainly evaluated by the update cycle of the inventory. For any item, the longer the inventory time, the higher the cost. The corresponding calculation is expressed as $(1/K) \sum_{i=1}^K |t_i^{\text{end}} - t_i^{\text{start}}|$, where t_i^{start} means the time that product i enters the inventory and t_i^{end} means the time that product i leaves the inventory
- (iii) The prediction accuracy is expressed by the average absolute error, that is, the average of the absolute values between the predicted values and the observed values. The corresponding calculation is expressed as $(1/K) \sum_{i=1}^K |h_i - \hat{h}_i|$, where h_i is the prediction value and \hat{h}_i is the actual observation value

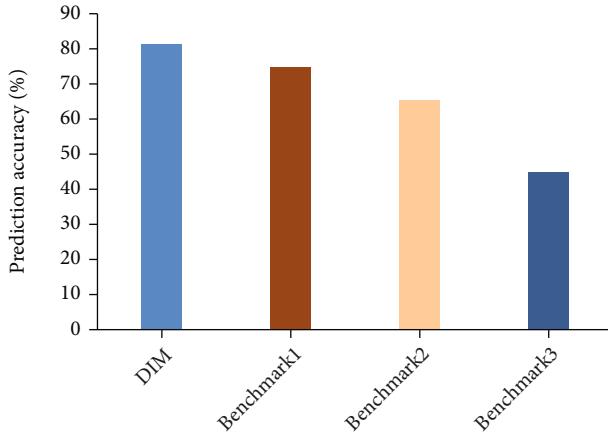


FIGURE 9: Results of prediction accuracy.

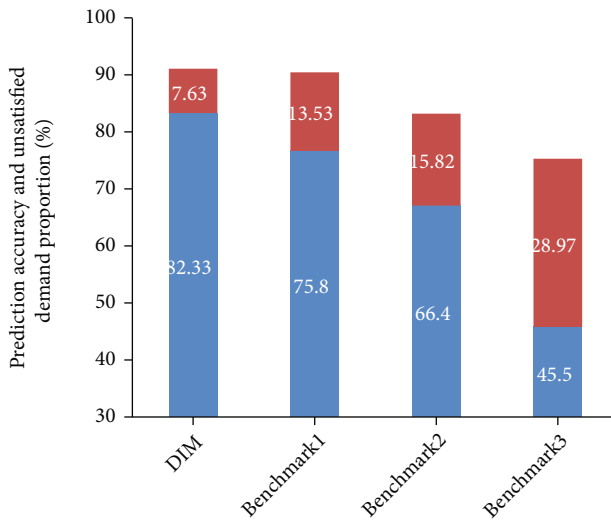


FIGURE 10: Results of the percentage between the prediction accuracy and the unsatisfied demands.

The following three benchmarks are used for comparison in order to evaluate the performance of DIM.

- (i) benchmark1 [10]: it mainly uses the data mining technology to implement the customer demand prediction for inventory management
- (ii) benchmark2 [26]: it mainly uses the neural network to realize the customer demand prediction for inventory management
- (iii) benchmark3 [30]: it mainly uses the random strategy to fulfill the inventory management

5.3. Results. As explained, 10 thousand records of data uniformly distributed in 25 weeks are used for the experiment. In particular, the data in the first 20 weeks are used for training, while the data in the last 5 weeks are used for evaluating the performance of profit, cost, prediction accuracy, and

unsatisfied demand proportion. The specific results are shown in Figures 7–11.

Firstly, the results of profit achieved by the four algorithms on inventory sales are shown in Figure 7. Obviously, the sale profit increases over time for the four algorithms in Figure 7, because the longer the time, the more products will be sold in general and the profit naturally increases. In addition, by comparing the four algorithms, we can observe that the DIM method achieves the highest profit. The second higher one is benchmark1 and the third higher one is benchmark2, while the last one is benchmark3. There are several reasons behind this phenomenon: (1) the benchmark3 randomly satisfies the inventory management, that is, when the warehouse node is running out, the manager will replenish products for this warehouse randomly. In this way, the replenished products may not be those required in the near future, which indirectly hinders the sale of products and then leads to the reduce of profit; (2) benchmark2 mainly adopts the data statistical analysis and mining methods to address the inventory management, which relies heavily on data quantity and quality, such that the self-adaptability of benchmark2 is lost. Despite this, compared with benchmark3, benchmark2 still has benefits, since it executes simple principles when choosing which products to store; (3) compared with benchmark2, DIM and benchmark1 both use the machine learning method to predict the customer demands, which can not only dynamically adapt the actual environment but also offer better inventory management decisions. Hence, we can see that the performance of DIM and benchmark1 exceed that of benchmark2 and benchmark3; and (4) as for DIM and benchmark1, the latter only has the hidden layer as the intermediate state, while the former introduces a state layer on the basis of the hidden layer, which can keep more long-term information. Hence, DIM achieves higher prediction accuracy, which is the main cause leading to higher profit. Nevertheless, we should be aware that the profit gained by the four algorithms will gradually tend to stable condition. That is because the popular products are usually sold quickly and the unpopular products may not be sold. Then, the longer the inventory time, the more unpopular products will be left. Despite the fact that the total profit is still increasing, we can see that the increasing trend of profit is going to be stable.

Secondly, the cost results of inventory management are shown in Figure 8, where we can easily see that DIM has the lowest cost and benchmark3 has the highest cost. The cost performance of benchmark1 and benchmark2 is in the middle, where benchmark1 outperforms benchmark2. As explained, the inventory cost is mainly measured by the time a product is being stored in the warehouse node. Because the longer time spent in the warehouse, the higher the inventory cost of this product. By optimizing the inventory management process, the products stored in the warehouse can be sold at a fast speed, which in turn decreases the inventory cost. At the same time, the gained profit is also increased.

For the four algorithms, they all have the ability to reduce the average inventory time of products in a certain extent. Hence, their performance will increase against the inventory time. However, DIM still achieves the best

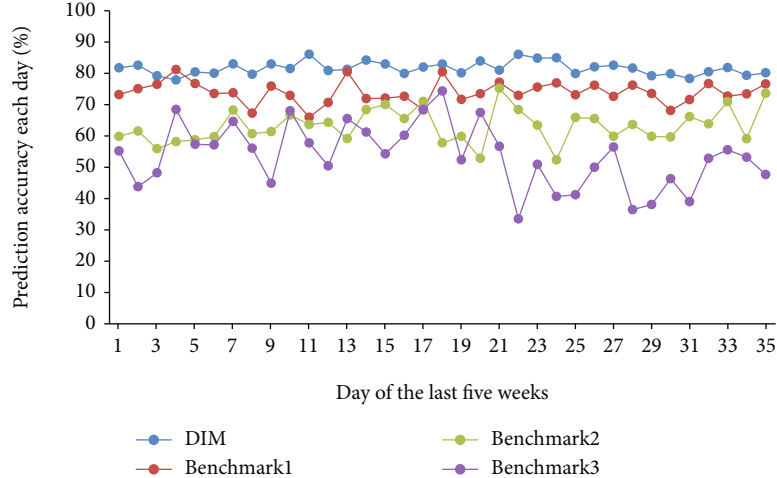


FIGURE 11: Details of the prediction accuracy each day in the remaining five weeks.

TABLE 1: Distribution of prediction accuracy using the testing data in the remaining five weeks.

Time (week)	DIM	benchmark1	benchmark2	benchmark3
1	81.97	76.21876	59.63312	53.02354
2	82.25	74.35157	61.43788	57.20488
3	82.36	74.9271	65.14423	60.70123
4	81.9	74.25434	62.96547	46.05241
5	81.25	73.85469	63.54434	49.24474

performance, since its inventory cost increases the slowest. The main reasons are because the prediction accuracy of customer demands of DIM is the highest as indicated in Figure 9, where DIM has the highest prediction accuracy of customer demands; benchmark1 is in the second place; benchmark2 is in the third place, while benchmark3 has the lowest prediction accuracy. Due to such high prediction accuracy, DIM can reduce the inventory cost by (1) about 16.44% compared to the performance of benchmark1, (2) about 33.57% compared to the performance of benchmark2, and (3) about 48.9% compared to the performance of benchmark3. Thus, DIM reduces approximately 25% of the inventory cost on average. Despite this, it is noted that the inventory cost will increase continuously. According to the definition of the inventory cost, it is proportional to the gap between the time that products enter the inventory and the time that products leave the inventory. Hence, the longer the inventory time, the longer the gap between the time that one product enters and leaves the inventory.

Thirdly, DIM is claimed to have the highest prediction accuracy in the above results and we now present the prediction results in Figure 9, where the specific prediction accuracies of DIM, benchmark1, benchmark2, and benchmark3 are 82.33%, 75.8%, 66.4%, and 45.5%, respectively. At the same time, we also calculate the prediction accuracy of each day in the last five weeks and the average prediction accuracy of each week is calculated and summarized in Table 1. Via comparison, we can discover that the average result in Table 1 accords with the results in Figure 9 with the devia-

tion smaller than 1%, which reflects the correctness of the proposed algorithm to a certain extent. Besides, according to the results in Table 1, we further calculate the standard deviations of the prediction accuracy of different algorithms to analyze the dispersion degree of these results. Specifically, the dispersion degree of DIM is about 0.433; benchmark1 is about 0.92; benchmark2 is about 2.099, and benchmark3 is about 5.89. Such phenomenon indicates that the prediction accuracy of DIM is the most stable in different time periods, while the prediction accuracy of benchmark3 shows the largest fluctuation. The more stable the prediction accuracy is, the better the robustness and adaptability of DIM. Hence, DIM outperforms the other three methods in this regard.

Finally, on the basis of the prediction accuracy results, we also calculate the proportion of customer demands that cannot be satisfied. The corresponding results are shown in Figure 10. Since the prediction accuracy results are already explained in Figure 9, we now focus on the opposite aspect, that is, the proportion of unsatisfied customer requirements. As can be seen in Figure 10, although DIM has a good prediction accuracy (i.e., 82.33%), it cannot fully meet the needs of all customers especially when the supply shortage happens. On the other hand, for either algorithm, it is noted that the sum of demand prediction accuracy and the proportion of unsatisfied customer demand cannot be equivalent to 100%, because these two indicators are actually calculated in completely different ways. Nevertheless, the proportion of unsatisfied customer demands of DIM is only about 7.63% which is far less than the other three algorithms. Therefore, DIM still has some advantages.

Reviewing the results in Table 1, we present the average prediction accuracy using one week as the unit. However, the statistical results of each day in terms of the prediction accuracy should be discussed, since it is a key metric in this work. In this way, the details of this metric are calculated and shown in Figure 11, where we can see the results against 35 days (5 weeks * 7 days/each week). Apparently, the proposed DIM remains very stable, while the performance of benchmark3 fluctuates greatly. The mean values of the four methods are 81, 74, 63, 53, respectively. Such phenomenon

means that the proposed DIM can adapt to different situations, while benchmark3 cannot. The performance of benchmark2 is slightly better than that of benchmark3, but far less than that of the proposed approach. As for benchmark1, we can see that it also has very stable performance. Despite this, its prediction accuracy is lower than that of DIM in almost each day. One exception is the fourth day, where benchmark1 has higher prediction accuracy than DIM. Nevertheless, it is noted that the gap is not large. Overall, the proposed DIM is superior to the other three benchmarks.

6. Conclusion

The current situations of long supply chain life cycle, complex inventory management process, and frequently changing customer demands all lead to the rapid rise of logistics cost. In this regard, this work firstly formulates the inventory management process as a mathematical model with the goals of minimizing the cost and maximizing the profit. On this basis, DIM is proposed, which offers effective inventory management by using the LSTM theory. In particular, the time series and back propagation pattern are jointly leveraged by DIM to achieve high prediction accuracy which then will be used to optimize the inventory management process. The experimental results show that the average prediction accuracy of DIM is more than 80% and the overall cost can be reduced by about 25%. Future research directions include large-scale logistics, warehousing, and distribution problems in inventory management.

Data Availability

The experimental data and code used to support the findings of this study are available from the first author and the corresponding author upon request.

Disclosure

This research did not receive specific funding but was performed as part of the employment of the author Yongji Liu.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Acknowledgments

We would like to thank all the experts for their comments on this work.

References

- [1] S. Lei, W. Haiying, L. Haiyue, and T. Weiyu, "Research of innovative business classification in bulk commodity digital supply chain finance," in *2020 International Conference on Computer Engineering and Application (ICCEA)*, pp. 170–173, Guangzhou, China, 2020.
- [2] A. Villalva-Cataño, E. Ramos-Palomino, K. Provost, and E. Casal, "A model in agri-food supply chain costing using ABC costing: an empirical research for Peruvian coffee supply chain," in *2019 7th International Engineering, Sciences and Technology Conference (IESTEC)*, pp. 1–6, Panama, Panama, 2019.
- [3] Q. Zhang, M. Zhou, C. Li, X. Zheng, K. Wang, and S. Yang, "Case analysis on value creation and sustainable development path of supply chain integrators," in *2018 15th International Conference on Service Systems and Service Management (ICSSSM)*, pp. 1–6, Hangzhou, 2018.
- [4] F. Zhu, Y. Ning, X. Chen, Y. Zhao, and Y. Gang, "On removing potential redundant constraints for SVOR learning," *Applied Soft Computing*, vol. 102, p. 106941, 2021.
- [5] A. Alzamendi-Ramirez, J. Yoshida-Chiney, E. Ramos-Palomino, and R. Mesia, "Supply chain agility in manufacturing companies: a literature review," in *2019 7th International Engineering, Sciences and Technology Conference (IESTEC)*, pp. 467–472, Panama, Panama, 2019.
- [6] R. Zare, P. Chavez, C. Raymundo, and J. Rojas, "Collaborative culture management model to improve the performance in the inventory management of a supply chain," in *2018 Congreso Internacional de Innovación y Tendencias en Ingeniería (CON-IITI)*, pp. 1–4, Bogota, 2018.
- [7] N. Nemtajela and C. Mbohwa, "Inventory management models and their effects on uncertain demand," in *2016 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)*, pp. 1046–1049, Bali, 2016.
- [8] S. Guo, T. Choi, B. Shen, and S. Jung, "Inventory management in mass customization operations: a review," *IEEE Transactions on Engineering Management*, vol. 66, no. 3, pp. 412–428, 2019.
- [9] F. Zhu, J. Yang, C. Gao, S. Xu, N. Ye, and T. Yin, "A weighted one-class support vector machine," *Neurocomputing*, vol. 189, pp. 1–10, 2016.
- [10] X. Guo, C. Liu, W. Xu, H. Yuan, and M. Wang, "A prediction-based inventory optimization using data mining models," in *2014 Seventh International Joint Conference on Computational Sciences and Optimization*, pp. 611–615, Beijing, 2014.
- [11] R. Gustriansyah, D. I. Sensuse, and A. Ramadhan, "Decision support system for inventory management in pharmacy using fuzzy analytic hierarchy process and sequential pattern analysis approach," in *2015 3rd International Conference on New Media (CONMEDIA)*, pp. 1–6, Tangerang, 2015.
- [12] Y. Sutanto and R. Sarno, "Inventory management optimization model with database synchronization through internet network," in *2015 International Conference on Electrical Engineering and Informatics (ICEEI)*, pp. 115–120, Denpasar, 2015.
- [13] W. Li, Z. Lin, X. Zhang, and J. Zhou, "Research on distributed logistics inventory model based on cloud computing," in *2016 12th International Conference on Computational Intelligence and Security (CIS)*, pp. 73–77, Wuxi, 2016.
- [14] N. El Haoud and Z. Bachiri, "Stochastic artificial intelligence benefits and supply chain management inventory prediction," in *2019 International Colloquium on Logistics and Supply Chain Management (LOGISTIQUA)*, pp. 1–5, Paris, France, 2019.
- [15] F. Zhu, N. Ye, W. Yu, S. Xu, and G. Li, "Boundary detection and sample reduction for one-class support vector machines," *Neurocomputing*, vol. 123, pp. 166–173, 2014.
- [16] N. Xue, I. Triguero, G. P. Figueredo, and D. Landa-Silva, "Evolving deep CNN-LSTMs for inventory time series

- prediction,” in *2019 IEEE Congress on Evolutionary Computation (CEC)*, pp. 1517–1524, Wellington, New Zealand, 2019.
- [17] Y. Xue, “A classical inventory model amendment based on management accounting,” in *2013 6th International Conference on Information Management, Innovation Management and Industrial Engineering*, pp. 101–103, Xi’an, 2013.
- [18] F. Wang and L. X. Xiao Xia, “Evaluation and selection of periodic inventory review policy for irregular demand: a case study,” in *2015 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)*, pp. 275–279, Singapore, 2015.
- [19] D. Satiti, A. Rusdiansyah, and R. S. Dewi, “Review of refrigerated inventory control system for perishable products,” in *2018 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)*, pp. 36–40, Bangkok, 2018.
- [20] P. Yao, D. Jiang, and T. Zhu, “Analysis of the risk pooling to inventory management for a three-stage supply chain,” in *2010 International Conference on Optoelectronics and Image Processing*, pp. 263–266, Haikou, 2010.
- [21] H. Li, Y. Ru, and X. Xu, “Emergency logistics based on inventory management method,” in *2010 International Conference on Machine Learning and Cybernetics*, pp. 1338–1341, Qingdao, 2010.
- [22] E. Raguindin and D. J. Ronquillo, “Development of an automated laboratory assets inventory control with security system,” in *2019 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE)*, pp. 501–504, Dubai, United Arab Emirates, 2019.
- [23] T. Inprasit and S. Tanachutiwat, “Reordering point determination using machine learning technique for inventory management,” in *2018 International Conference on Engineering, Applied Sciences, and Technology (ICEAST)*, pp. 1–4, Phuket, 2018.
- [24] L. Lin, W. Xuejun, H. Xiu, W. Guangchao, and S. Yong, “Enterprise lean catering material management information system based on sequence pattern data mining,” in *2018 IEEE 4th International Conference on Computer and Communications (ICCC)*, pp. 1757–1761, Chengdu, China, 2018.
- [25] P. Zhao and J. Liu, “The product safety stock prediction method based on artificial neural network,” in *2010 International Conference of Information Science and Management Engineering*, pp. 299–302, Xi’an, 2010.
- [26] J. Rehman, J. Uddin, A. Khan, and A. Zeb, “A cloud based CRM architecture for neural network inventory control,” in *2019 International Conference on Electrical, Communication, and Computer Engineering (ICECCE)*, pp. 1–5, Swat, Pakistan, 2019.
- [27] H. Lican, Z. Yuhong, X. Xin, and F. Fan, “Prediction of investment on inventory clearance based on improved BP neural network,” *First International Conference on Networking and Distributed Computing*, 2010, pp. 73–75, Hangzhou, 2010.
- [28] J. Zhou and X. Zhou, “Multi-echelon inventory optimizations for divergent networks by combining deep reinforcement learning and heuristics improvement,” in *2019 12th International Symposium on Computational Intelligence and Design (ISCID)*, pp. 69–73, Hangzhou, China, 2019.
- [29] E. Alhroob and N. A. Ghani, “Fuzzy min-max classifier based on new membership function for pattern classification: a conceptual solution,” *8th IEEE International Conference on Control System, Computing and Engineering (ICCSCE)*, 2018, pp. 131–135, Penang, Malaysia, 2018.
- [30] X. Li and P. Li, “Simulation optimization under random conditions TG business model of spare parts inventory,” in *2019 4th International Conference on Mechanical, Control and Computer Engineering (ICMCCE)*, pp. 1025–10253, Hohhot, China, 2019.

Research Article

Comparison Analysis of Different Time-Scale Heart Rate Variability Signals for Mental Workload Assessment in Human-Robot Interaction

Shiliang Shao^{1,2}, Ting Wang^{1,2}, Yawei Li^{1,2,3}, Chunhe Song^{1,2}, Yihan Jiang^{1,2,4} and Chen Yao^{1,2}

¹State Key Laboratory of Robotics, Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang 110016, China

²Institutes for Robotics and Intelligent Manufacturing, Chinese Academy of Sciences, Shenyang 110169, China

³University of Chinese Academy of Sciences, Beijing 100049, China

⁴Shenyang Ligong University, School of Automation and Electrical Engineering, Shenyang 110159, China

Correspondence should be addressed to Shiliang Shao; shaoshiliang@sia.cn and Ting Wang; wangting@sia.cn

Received 21 June 2021; Revised 19 August 2021; Accepted 31 August 2021; Published 6 October 2021

Academic Editor: Fa Zhu

Copyright © 2021 Shiliang Shao et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Excessive mental workload affects human health and may lead to accidents. This study is motivated by the need to assess mental workload in the process of human-robot interaction, in particular, when the robot performs a dangerous task. In this study, the use of heart rate variability (HRV) signals with different time scales in mental workload assessment was analyzed. A humanoid dual-arm robot that can perform dangerous work was used as a human-robot interaction object. Electrocardiogram (ECG) signals of six subjects were collected in two states: during the task and in a relaxed state. Multiple time-scale (1, 3, and 5 min) HRV signals were extracted from ECG signals. Then, we extracted the same linear and nonlinear features from the HRV signals at different time scales. The performance of machine learning algorithms using the different time-scale HRV signals obtained during the human-robot interaction was evaluated. The results show that for the per-subject case with a 3 min HRV signal length, the *K*-nearest neighbor classifier achieved the best mental workload classification performance. For the cross-subject case with a 5 min time-scale signal length, the gentle boost classifier achieved the best mental workload classification accuracy. This study provides a novel research idea for using HRV signals to measure mental workload during human-robot interaction.

1. Introduction

Nowadays, robots, instead of humans, work in unstructured environments, expanding the scope of human work. Humans interact with robots through visual, tactile, and other feedback [1–4]. The robot can be operated remotely to complete a dangerous task; this operation can be challenging for humans. At present, research in the field of robotics primarily focuses on how robots perform human control instructions, how they perceive environmental information, and how autonomous operation can be achieved [5, 6]. However, this research neglects the robot's assessment of the human's psychological activity and the emotions of humans interacting with the robot. Therefore, it is of great

significance to accurately measure the mental workload of the operator during their interaction with the robot [7, 8].

Mental workload can be measured continuously and objectively using physiological signals. In particular, heart rate variability (HRV) signals have been widely studied because they are easy to collect. In [9], the relationships between mental workload and time-domain, frequency-domain, and Poincare plot features of 5 min signals were analyzed. In [10], 5 min HRV signal segments were used to detect the mental workload of a worker. Several linear features (time and frequency domains) were utilized. Then, the combination of principal component analysis and support vector machine (SVM) achieved 84.4% accuracy. In fact, the physiological system of the human body can be

regarded as a nonlinear system. However, the nonlinear nature of HRV signals cannot be reflected by linear analysis methods [11, 12]. In [13], the mental workload of performing MATA-II tasks was measured using 5 min scale HRV signals. This study extracted the multiscale entropy features of the HRV. Using those, it obtained a higher accuracy for mental workload recognition than using traditional time- and frequency-domain features. In [14], 5 min length HRV signal segments were utilized to evaluate the mental workload of hospital staff. A variety of conventional and multiscale HRV features were extracted, and SVM was used as the classifier. The results showed that the multiscale features obtain a better mental workload recognition effect. In [15], the respiratory and HRV signals were extracted using 5 min scale electrocardiogram (ECG) signals. This study introduced a novel method that fused respiratory and HRV signals to assess subtle variations in sympathovagal balance using ECG recordings during the MATA-II mission. Standard short-term HRV analysis is usually performed on 5 min recordings [16], and shorter recordings of HRV signals are being researched, aiming at a faster detection of mental workload. In [17], human HRV signals were collected during human-robot interaction through different types of wearable devices. Using signals of 3 min length, the linear features of HRV signals collected by different wearable devices were extracted, compared, and analyzed under different mental workload levels. In [18], 3 min HRV signals were used, and linear and nonlinear features were utilized. Several machine learning algorithms have been utilized for assessing the mental workload of humans while operating a dual-arm robot. In [19], 2.5 min HRV signals were detected by a consumer smart watch. Subsequently, the mental workload of human interaction with multiple robots was studied. However, analysis of mental workload recognition with HRV signals at different time scales is not sufficiently researched. In [20], a nonparametric statistical test method was utilized to analyze the significant differences between rest and stress phases with time scales of 30 s and 1, 2, 3, and 5 min. However, HRV signals were obtained from healthy subjects during an examination and in a resting condition, not during human-robot interaction.

Humans use visual, haptic, and other feedback information to remotely perceive the environment information during human-robot interaction, and the robot is remotely operated to complete the task. The entire human-robot interaction process requires the joint perception and decision-making of human hands, eyes, ears, brain, and other limbs and organs, which may be very challenging for the operator. At present, there is a lack of mental workload measurement analysis during human-robot interactions using HRV with different time scales. Therefore, in this study, the differences among HRV signals of multiple time scales in measuring mental workload were analyzed; six traditional machine learning methods were used to evaluate the performances of HRV signals with different time scales. Traditional machine learning methods were used because they are more suitable for small sample sizes. Although deep learning methods have been widely studied, many training samples are required.

The contribution of this study can be summarized as follows:

- (i) During human-robot interaction, HRV signals were collected based on a single physiological signal. In addition, linear and nonlinear features of different time-scale HRV signals were extracted, and statistical differences between the mental workloads in the two states were analyzed
- (ii) A variety of representative machine learning algorithms were applied. Differences in the performances of the machine learning algorithms with statistically different linear and nonlinear features extracted from HRV signals of different time scales in evaluating mental workload were analyzed
- (iii) Finally, the different performances of the algorithms with HRV signals of various time scales in evaluating mental workload are discussed

The remainder of this paper is organized as follows. Section 2 introduces the data collection and preprocessing algorithms. The mental workload assessment results of algorithms using different time scales of HRV signals and a discussion of the results are presented in Section 3. The concluding remarks are presented in Section 4.

2. Data and Method

The research block diagram is shown in Figure 1. It can be seen that the ECG signals were obtained from volunteers while they operated the dual-arm robot and in the rest state. HRV signals were then extracted from the ECG signals. Using a sliding window of different time scales (1, 3, and 5 min), the HRV signals were divided to obtain a collection of sample data of different time scales. Then, linear and nonlinear features of different time scales were extracted. In addition, an SVM, K -nearest neighbor (KNN) classifier, gentle boost (GB), linear discriminant analysis (LDA), naive Bayes (NB), and decision tree (DT) were utilized to identify the task-performing and rest states. The performance differences of the classifiers in the mental workload evaluation with HRV signals at different time scales were compared and analyzed.

2.1. Data. In this subsection, the subjects and data acquisition processes are described. Then, a preprocessing algorithm is introduced to obtain the HRV signals from the collected ECG signals. In addition, multiple time-scale HRV signal segments are obtained using sliding windows of different time scales.

2.1.1. Participants. The ECG signals used for mental workload assessments were obtained from six male participants. A description of the six subjects is provided in Table 1. They were recruited from the Shenyang Institute of Automation, Chinese Academy of Sciences. Their average age was 25.16 (± 2.93). They had normal or corrected vision and were all healthy, with no nervous system diseases. Before starting the experimental data collection, all participants were

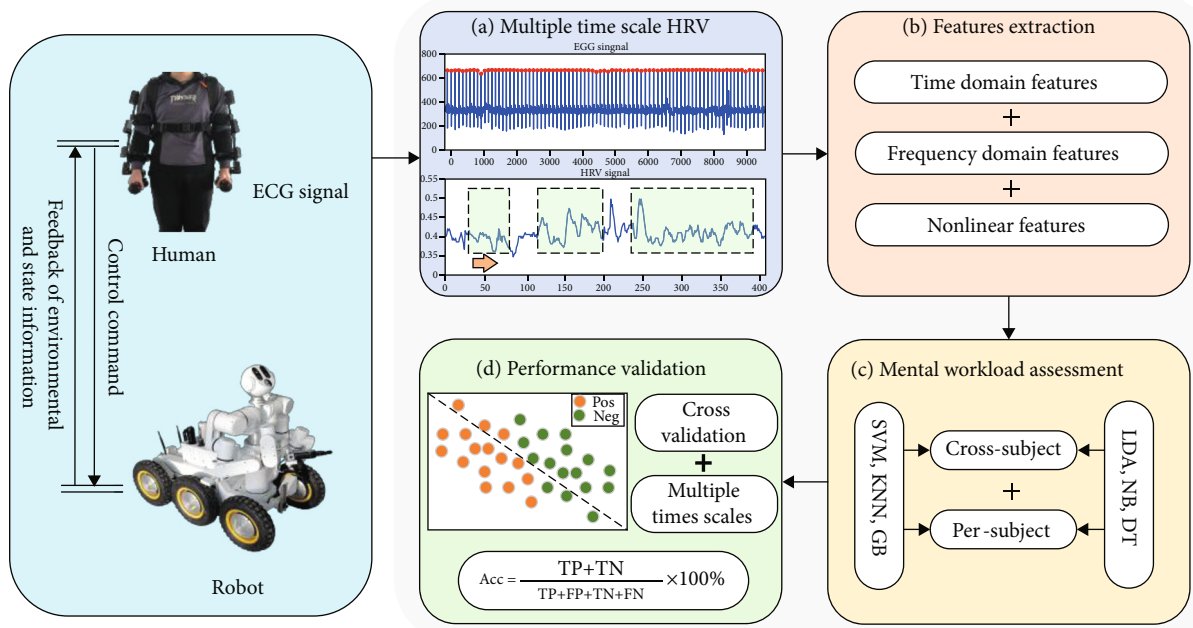


FIGURE 1: Framework of multiple time-scale HRV analysis for mental workload assessment. HRV: heart rate variability; ECG: electrocardiogram; SVM: support vector machine; KNN: K -nearest neighbors; GB: gentle boost; LDA: linear discriminant analysis; NB: naive Bayes; DT: decision tree; TP: true positive; TN: true negative; FP: false positive; FN: false negative.

TABLE 1: Participant characteristics.

	Subject 1	Subject 2	Subject 3	Subject 4	Subject 5	Subject 6
Height (cm)	180	175	173	180	175	178
Weight (kg)	67.5	78.5	58	55	75	72.5
Age (years)	24	24	31	23	24	25
Body mass index (kg/m^2)	20.8	25.6	19.4	17.0	24.5	22.9

informed of the entire data collection process and precautions.

2.1.2. Data Acquisition. In this study, the operating object was a dual-arm robot shown in Figure 2. It can be seen that the robot consists of six wheels and two arms. Moreover, each wheel is independent, and each arm has seven degrees of freedom to access all positions in space. In addition, the top of the robot is equipped with a binocular camera for environmental observations. The robot controller is an exoskeleton device that can be worn by an operator (Figure 3). The exoskeleton controller also has two arms, and each arm has seven degrees of freedom, similar to the dual-arm robot. The ECG signal collection process is shown in Figure 4. A portable sensor was placed on the chest of the operator for the acquisition of ECG signals. The captured ECG signals were sent to a computer via Bluetooth for processing. ECG signals were collected in two states of the operator: during the operation of the dual-arm robot and during rest.

2.1.3. Signal Preprocessing. The HRV signals refer to a time series consisting of intervals between each pair of heartbeats.

Therefore, to obtain the HRV signals, it is necessary to detect the peak and trough values of the ECG signals. Therefore, the Q, R, and S waves of the ECG signal were detected using a QRS wave group detection method [21]. However, there may be an abnormal point in the output RR interval sequence. Therefore, a classical median-filtering algorithm was applied to the output RR interval sequence [22]. The RR interval sequence was regarded as an HRV signal. As shown in Figures 5–7, sliding windows at different time scales (1, 3, and 5 min) were used with an overlap of 30 s. HRV signals were then divided into six groups: M-1, R-1, M-3, R-3, M-5, and R-5 groups. The M group signals represent the operator in the task-performing state, and the R group signals represent the operator in the rest state.

The proposed mental workload assessment preprocessing algorithm is described in Algorithm 1, where $x_i(t)$ is the ECG data recorded from the i th participant, and I is the number of participants. The purpose of Steps 1 to 6 is to obtain the HRV signals $y_i(t)$ from $x_i(t)$ signals. The HRV signals $y_i(t)$ are segmented into different time-scale (1, 3, and 5 min) segments $y_i^1(t)$, $y_i^3(t)$, and $y_i^5(t)$ in Steps 7 to 10.

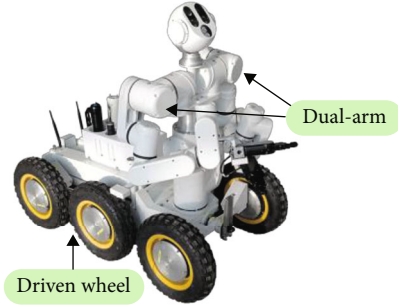


FIGURE 2: Dual-arm robot experiment platform.



FIGURE 3: Exoskeleton robot controller.

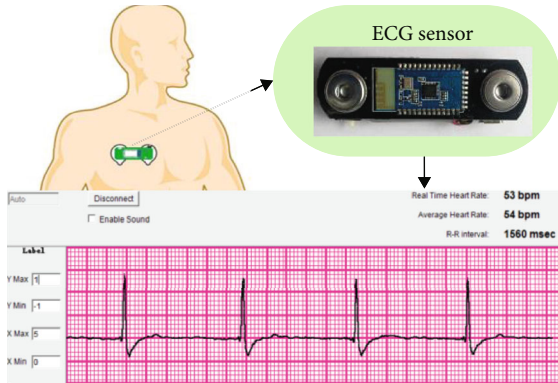


FIGURE 4: Process of ECG signal collection. ECG: electrocardiogram.

2.2. Method. Linear and nonlinear analysis methods are the most commonly used HRV signal analysis methods. Therefore, in this subsection, the linear and nonlinear features used in this study are described. The collection of physiological signals during human-robot interaction requires considerable manpower and energy; thus, it is difficult to collect large-scale sample data. However, machine learning algorithms do not require large-scale sample data for efficient feature recognition [23, 24]. Therefore, in this study, several different types of machine learning algorithms (SVM, KNN, GB, LDA, NB, and DT) were used to compare the effects of feature recognition.

2.2.1. Feature Extraction. First, the linear and nonlinear features used in this study are presented. In human-robot inter-

action, the fluctuation of the operator's mental workload is related to the fluctuation of the human autonomic nervous system (ANS). The ANS consists of the sympathetic and parasympathetic nervous systems. The time- and frequency-domain features of HRV signals can reflect fluctuations in the sympathetic and the parasympathetic nervous systems. In addition, nonlinear features can reflect the nonlinear dynamic characteristics of the HRV signal [25, 26].

The linear features include time- and frequency-domain features. First, we introduce time features.

SDNN denotes the standard deviation of all RR intervals:

$$SDNN = \sqrt{\frac{1}{N} \sum_{i=1}^N \left(RR_{s_i} - \frac{1}{N} \sum_{i=1}^N RR_{s_i} \right)^2}. \quad (1)$$

RMSSD denotes the root mean square of the adjacent RR interval difference:

$$RMSSD = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N-1} (RR_{s_{i+1}} - RR_{s_i})^2}. \quad (2)$$

pNN50 denotes the ratio of the number of pairs of adjacent RR intervals with a difference of more than 50 ms:

$$pNN50 = \frac{\text{num}[(RR_{s_{i+1}} - RR_{s_i}) > 50 \text{ ms}]}{N-1}. \quad (3)$$

In addition, all RR intervals were integrated and divided by the maximum density distribution parameter, and the mean and median of the HRV signals were also extracted as time-domain features.

In this study, all frequency-domain features were obtained based on the power spectral density [27]. Furthermore, the basic frequency-domain features are defined as the sum of the power spectra at different frequency ranges: aTotal = 0 – 0.4 Hz; aVLF = 0.003 – 0.04 Hz; aLF = 0.04 – 0.15 Hz; and aHF = 0.15 – 0.4 Hz. The ratio of aLF and aHF is defined as

$$\frac{LF}{HF} = \frac{aLF}{aHF}. \quad (4)$$

The percentage of aVLF, aLF, and aHF are defined as

$$\begin{aligned} pVLF &= \frac{aVLF}{aTotal}, \\ pLF &= \frac{aLF}{aTotal}, \\ pHF &= \frac{aHF}{aTotal}. \end{aligned} \quad (5)$$

The respective ratios of aLF and aHF to aLF + aHF are

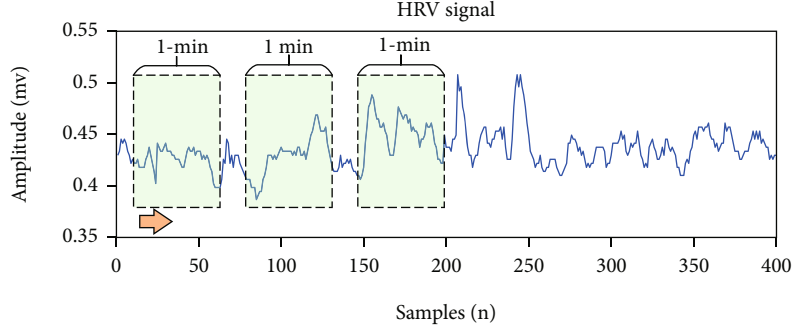


FIGURE 5: Heart rate variability segmented by 1 min time scale.

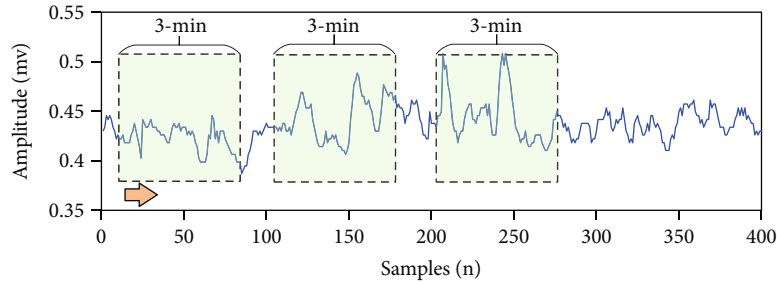


FIGURE 6: Heart rate variability segmented by 3 min time scale.

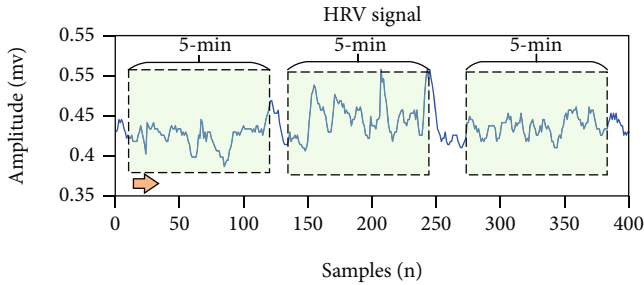


FIGURE 7: Heart rate variability segmented by 5 min time scale.

defined as

$$\begin{aligned} \text{nLF} &= \frac{\text{aLF}}{\text{aLF} + \text{aHF}}, \\ \text{nHF} &= \frac{\text{aHF}}{\text{aLF} + \text{aHF}}. \end{aligned} \quad (6)$$

Finally, two typical nonlinear analysis methods applied in this study are presented. These are sample entropy (SaEn) and detrended fluctuation analysis (DFA). On the one hand, SaEn is a method for investigating the dynamics of HRV signals. It has the advantages of strong antinoise and antijamming abilities. In addition, it can be used to analyze shorter HRV signals. In the case of large differences in the parameter value range, good consistency is still achieved. On the other hand, DFA is suitable for the analysis of non-stationary time series, and HRV signals have this character-

istic. In addition, the DFA method can filter out the trend components in the HRV signal. Therefore, it can effectively avoid the disturbance of false correlations owing to noise and signal instability.

2.2.2. Mental Workload Recognition. In this subsection, the abstracted feature vector of HRV signals at different time scales is used to evaluate the mental workload. The different time-scale HRV features were analyzed using the t -test to obtain the statistical significance of the difference between task-performing and relaxed states; $p < 0.05$ was considered statistically significant [28]. Then, linear and nonlinear features with statistical differences were used to construct feature vectors as inputs to machine learning algorithms. Six different machine learning methods, SVM, KNN, LDA, GB, NB, and DT, were used in this study to exclude the effects of performance differences in machine learning algorithms.

After the initial HRV signal preprocessing, 1, 3, and 5 min time-scale HRV signals for mental workload can be assessed using Algorithm 2. The linear and nonlinear features of the i th subject were extracted in Steps 2 to 5. $\vec{\mathbf{F}}_i^s$ is defined as the feature vector in the human task-performing state, and $\vec{\mathbf{F}}_i^r$ is defined as the feature vector in the human relaxed state. Steps 6 to 11 define the process of per-subject mental workload assessment. $\vec{\mathbf{F}}_{i_Train}^s$ and $\vec{\mathbf{F}}_{i_Test}^s$ are the training and testing sets of the i th subject, respectively. Steps 12 to 18 define the process of cross-subject mental workload assessment. The extracted HRV features $\vec{\mathbf{F}}_i^s$ and $\vec{\mathbf{F}}_i^r$ of all subjects in task-performing and relaxation states are merged

Input: ECG signals $x_i(t)$ for each i subject.
Output: Multiple time-scale HRV signals for all subjects.
1: For each i such that $1 \leq i \leq I$ do
2: Q wave, R wave, and S wave of ECG signals $x_i(t)$ are detected.
3: RR internal sequence is obtained.
4: Abnormal points in the output RR internal sequence are removed by median filtering.
5: HRV signals $y_i(t)$ are obtained.
6: End for
7: For each i such that $1 \leq i \leq I$ do
8: Segment the $y_i(t)$ signals into 1 min, 3 min, and 5 min time-scale segments defined as
9: $y_i^1(t)$, $y_i^3(t)$, and $y_i^5(t)$, respectively.
10: End for

ALGORITHM 1: Mental workload assessment preparation.

Input: Multiple time-scale HRV segments for all subjects.
Output: Per-subject and cross-subject probability of mental workload.
1: For each time scale s , $s = 1, 3, 5$.
2: For each i such that $1 \leq i \leq I$ do
3: Extract linear and nonlinear features $\tilde{\mathbf{F}}_i^s$ for each $y_i^s(t)$ signal at task-performing state.
4: Extract linear and nonlinear features \mathbf{F}_i^s for each $y_i^s(t)$ signal at relaxation state.
5: End for
6: If per-subject mental workload assessment
7: For each i such that $1 < i < I$ do
8: Train classifiers (SVM, KNN, LDA, GB, NB, and DT) based on the training set $\mathbf{F}_{i_Train}^s$ randomly selected from \mathbf{F}_i^s .
9: Obtain the probability of mental workload based on the testing set $\mathbf{F}_{i_Test}^s$, which is defined as $\mathbf{F}_i^s - \mathbf{F}_{i_Train}^s$.
10: End for
11: End if
12: If cross-subject mental workload assessment
13: Merge matrices $\tilde{\mathbf{F}}_1^s, \tilde{\mathbf{F}}_2^s, \dots, \tilde{\mathbf{F}}_I^s$ into one matrix $\tilde{\mathbf{F}}^s$.
14: Merge matrices $\mathbf{F}_1^s, \mathbf{F}_2^s, \dots, \mathbf{F}_I^s$ into one matrix \mathbf{F}^s .
15: Train machine learning method (SVM, KNN, LDA, GB, NB, and DT) based on the training set \mathbf{F}_{Train}^s and $\tilde{\mathbf{F}}_{Train}^s$ randomly selected from \mathbf{F}^s and $\tilde{\mathbf{F}}^s$, respectively.
16: Obtain probability of mental workload based on the testing set \mathbf{F}_{Test}^s and $\tilde{\mathbf{F}}_{Test}^s$, which are defined as $\mathbf{F}^s - \mathbf{F}_{Train}^s$ and $\tilde{\mathbf{F}}^s - \tilde{\mathbf{F}}_{Train}^s$, respectively.
17: End if
18: End for

ALGORITHM 2: Mental workload assessment after preprocessing.

into matrices $\tilde{\mathbf{F}}^s$ and \mathbf{F}^s in Steps 13 and 14, respectively. Then, in Steps 15 to 17, the merged matrices $\tilde{\mathbf{F}}^s$ and \mathbf{F}^s are prepared for model construction and mental workload assessment.

3. Experimental Results

In this section, the mental workload recognition performance of classifiers with HRV signals of different time scales is presented. The statistical differences of the linear and nonlinear features extracted in this study among different mental workload levels were analyzed via a t -test, and the feature vectors were composed of per-subject and cross-subject mental workload assessments.

To evaluate the performance of mental workload classification with different time scales, accuracy was used, which is defined as follows:

$$\text{Accuracy : Acc} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} \times 100\%, \quad (7)$$

where TP is true positive, FP is false positive, FN is false negative, and TN is true negative.

3.1. Per-Subject Mental Workload Evaluation. The results of per-subject mental workload evaluation at different time scales (1, 3, and 5 min) are presented. The samples of each subject were randomly divided into two sets. One was used for training the machine learning model, and the other was

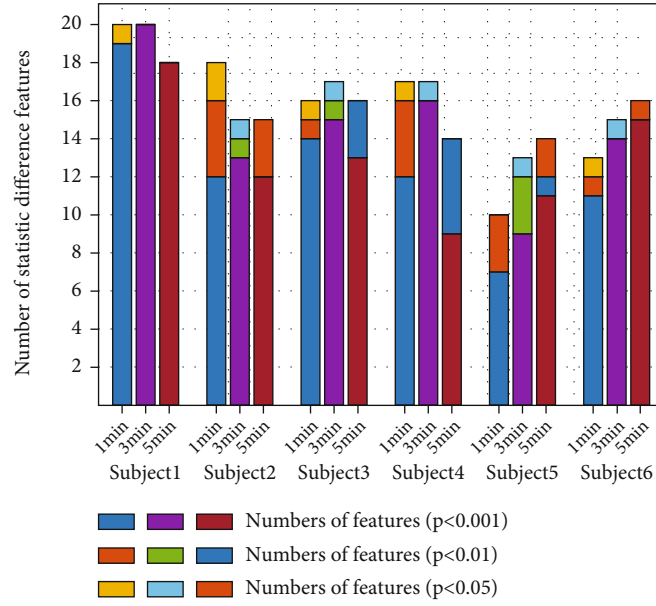


FIGURE 8: Results with statistically significant features of per-subject analysis at different time scales.

used for testing the model. In addition, to increase the reliability of the results, the average of the results repeated 500 times was regarded as the final classification result.

3.1.1. Results of Statistically Significant Features of 1, 3, and 5 min Length. Figure 8 shows the statistics of the significantly different ($p < 0.001$, $p < 0.01$, and $p < 0.05$) features at different time scales of each subject. It can be seen that Subject 1 has more significantly different ($p < 0.001$) features at the 3 min time scale, followed by the 1 min and 5 min time scales. Subject 2 showed more significantly different features at the 3 min time scale and at the 1 min time scale; the sum of the most significantly different ($p < 0.001$) features and the significantly different ($p < 0.01$ and $p < 0.05$) features was the largest. Subject 3 and Subject 4 have the most significantly different ($p < 0.001$) features at the 3 min time scale. Subject 5 and Subject 6 have the most significantly different ($p < 0.001$) features at the 5 min time scale.

3.1.2. Classification Accuracy of Different Classifiers with Different Time Scales. Figure 9 shows the classification accuracy of the mental workload using different classifiers with different time scales. Figure 9(a) shows the classification accuracy using SVM. It can be seen that the time scale with which the SVM achieved the highest average recognition accuracy was 3 min. In addition, the average classification accuracies of Subject 1 to Subject 6 with the 1, 3, and 5 min time scales were 95.30%, 97.54%, and 95.11%, respectively. Figure 9(b) shows the classification accuracy using KNN. It can be seen that the time scale with which the KNN obtained the highest average recognition accuracy was 3 min. In addition, the average classification accuracies of Subject 1 to Subject 6 with the 1, 3, and 5 min time scales were 96.09%, 98.77%, and 96.21%, respectively. Figure 9(c) shows the classification accuracy using GB; it achieved the highest average recognition accuracy with the 3 min time

scale. In addition, the average classification accuracies of Subject 1 to Subject 6 with the 1, 3, and 5 min time scales were 93.17%, 95.90%, and 90.61%, respectively.

Figure 9(d) shows the classification accuracy using LDA; it did not achieve good classification performance with any of the three types of time scales. The average classification accuracies of Subject 1 to Subject 6 with the 1, 3, and 5 min time scales were 52.02%, 52.27%, and 52.28%, respectively. Figure 9(e) shows the classification accuracy using NB. It can be seen that NB achieved the highest average recognition accuracy with the time scale of 3 min. The average classification accuracies of Subject 1 to Subject 6 with the 1, 3, and 5 min time scales were 80.52%, 84.99%, and 80.07%, respectively. Finally, Figure 9(f) shows the classification accuracy using DT. The average classification accuracies of Subject 1 to Subject 6 with the 1, 3, and 5 min time scales were 80.52%, 84.99%, and 80.07%, respectively.

3.2. Cross-Subject Mental Workload Evaluation. The results of cross-subject mental workload evaluation at different time scales (1, 3, and 5 min) are presented in this subsection. The sample data of five of the six subjects were selected to train the machine learning model. At the same time, the sample data of the remaining subject were selected to test the machine learning model.

3.2.1. Statistically Significant Analysis of Features. Table 2 shows the statistical differences between the two groups at the time scales of 1, 3, and 5 min. From Table 2, we can see that there were 17 features in the most significantly different category ($p < 0.001$) and 2 features with significant differences ($p < 0.01$) between groups M-1 and R-1. There were eighteen features in the most significantly different category ($p < 0.001$) and two features of the significantly different category ($p < 0.01$) between groups M-3 and R-3. There were 17 features in the most significantly different category ($p < 0.001$) between groups M-5 and R-5.

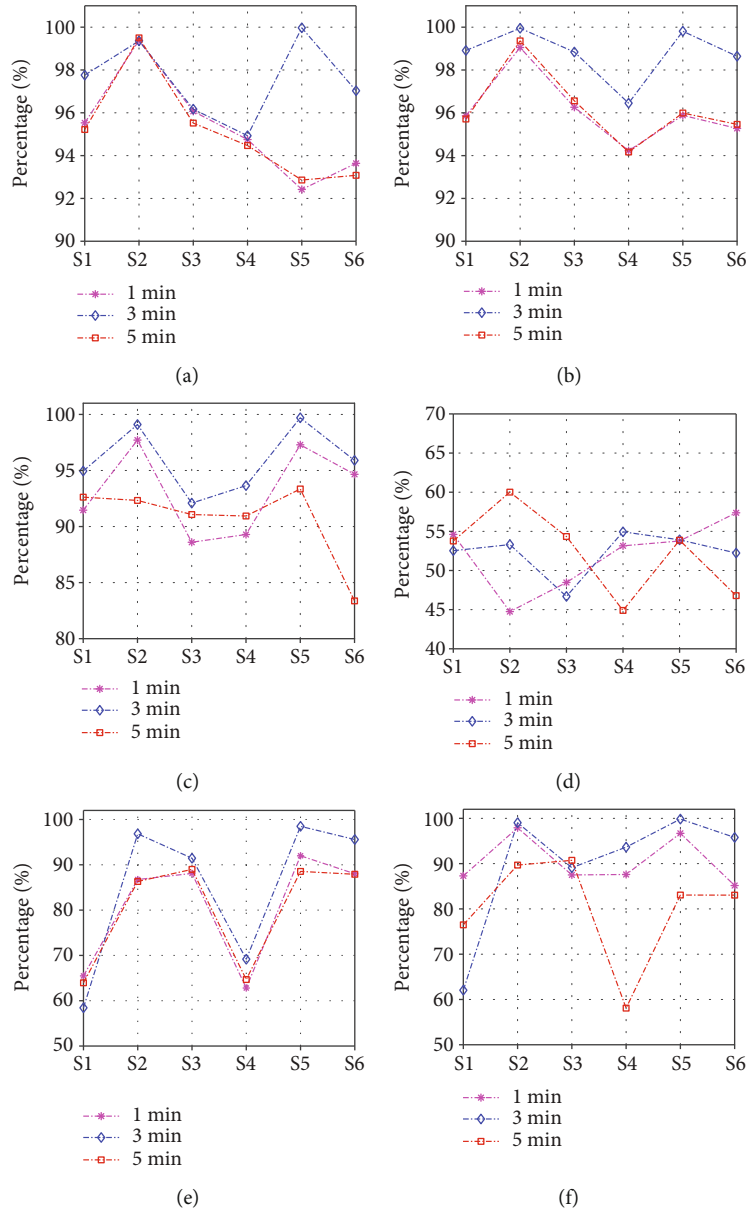


FIGURE 9: Classification accuracies of per-subject mental workload by different classifiers at different time scales: (a) support vector machine, (b) K -nearest neighbors, (c) gentle boost, (d) linear discriminant analysis, (e) naïve Bayes, and (f) decision tree.

3.2.2. Classification Accuracy of Different Classifiers with Different Time Scales. Figure 10 shows the classification accuracy of the mental workload using different classifiers at different time scales. Figure 10(a) shows the classification accuracy using SVM. It can be seen that when Subject 3 was used as the test subject, the classifier achieved the worst classification accuracy. The average classification accuracies of the classifier across all subjects with the 1, 3, and 5 min time scales were 77.59%, 75.06%, and 78.51%, respectively. Figure 10(b) shows the classification accuracy using KNN. Again, when Subject 3 was the test subject, the worst classification accuracy was achieved. The average classification accuracies of the classifier across all subjects with the 1, 3, and 5 min time scales were 69.24%, 70.40%, and 73.53%, respectively. Figure 10(c) shows the classification accuracy using GB. It can be seen that GB showed the worst classifica-

tion accuracy with the time scale of 1 min and the best accuracy with the time scale of 5 min, both when Subject 2 was the test subject. The average classification accuracies of Subject 1 to Subject 6 with the 1, 3, and 5 min time scales were 63.53%, 71.55%, and 80.56%, respectively. Figure 10(d) shows the classification accuracy using LDA. It can be seen that the classifier showed the worst classification accuracy with the time scale of 3 min and the best accuracy with the time scale of 5 min, both when the data of Subject 3 were used as the test set. The average classification accuracies with the 1, 3, and 5 min time scales were 44.44%, 35.92%, and 53.92%, respectively. Figure 10(e) shows the classification accuracy using NB. It achieved the worst classification accuracy with Subject 3 as the test subject and the time scale of 5 min. It obtained the best accuracy with Subject 2 and the time scale of 5 min. The average classification accuracies

TABLE 2: Statistical analysis results of features under multiple time scales.

		M-1 and R-1	M-3 and R-3	M-5 and R-5
Time domain	HRVTi	0***	0***	0***
	Mean	0***	0***	0***
	SDNN	0***	0***	0***
	Median	0***	0***	0***
	pNN50	0***	0***	0***
	RMSSD	0***	0***	0***
Frequency domain	aHF	0***	0***	0***
	aLF	0***	0***	0***
	aTotal	0***	0***	0***
	aVLF	0***	0***	0***
	LF/HF	0***	0.054	0***
	nHF	0***	0***	0.001**
	nLF	0***	0***	0.002**
	pHF	0.80	0***	0***
	pLF	0.003**	0.60	0.194
	pVLF	0.006**	0***	0***
Nonlinear	SaEn	0***	0***	0***
	Alpha	0***	0***	0***
	Alpha1	0***	0***	0***
	Alpha2	0***	0***	0***

*, **, and *** represent $p < 0.05$, $p < 0.01$, and $p < 0.001$, respectively. M-1: 1 min signals of the task-performing state; R-1: 1 min signals of the rest state; M-3: 3 min signals of the task-performing state; R-3: 3 min signals of the rest state; M-5: 3 min signals of the task-performing state; R-5: 5 min signals of the rest state.

with the 1, 3, and 5 min time scales were 64.53%, 66.48%, and 66.50%, respectively. Figure 10(f) shows the classification accuracy using DT. It can be seen that DT showed the worst classification accuracy with Subject 1 and the time scale of 5 min and the best accuracy with Subject 4 and the time scale of 5 min. The average classification accuracies with the 1, 3, and 5 min time scales were 65.03%, 67.91%, and 59.48%, respectively.

3.3. Discussion. Studies have shown that HRV can be used to measure and evaluate the mental workload of operators during human-robot interaction. Different time scales of HRV signals for mental workload measurement analysis have been widely studied. However, they were not based on a dataset of human-robot interaction. In addition, for the same dataset, the mental workload measurement analysis of human-robot interaction using HRV signals of different time scales was not reported, and there is no relevant public dataset. Hence, in this study, ECG signals were collected from six volunteers during task performance and rest. The fluctuation in the mental workload is closely related to the fluctuation state of the ANS, and HRV signals can react to the fluctuating state of the ANS. HRV signals of different

lengths show levels of nervous activity information about the mental workload. This study presented a detailed comparative analysis.

First, the HRV signals at different time scales (1, 3, and 5 min) of the same individual were analyzed. Using a t -test, the statistical differences between the task-performing and rest states were analyzed. The results are shown in Figure 8. These are the p values of 1, 3, and 5 min time-scale HRV signals and the results with statistically significant features per subject at different time scales. It can be seen from Figure 8 that Subject 1 to Subject 4 show the most significantly different features at the 3 min time scale, whereas Subject 5 and Subject 6 have slightly less than the 5 min time scale. Moreover, there were a total of 75, 87, and 78 features with the most significant differences ($p < 0.001$) for the 1 min, 3 min, and 5 min time-scale HRV signals of the six subjects, respectively. It is shown that at the time scale of 3 min, there are more significantly different features than at the other time scales. The classification analysis of mental workload was performed using the features with statistical differences ($p < 0.05$) and six types of classifiers. The results are shown in Figure 9. It can be seen that the average accuracy across the six subjects with the 3 min time scale was the highest, i.e., 98.77% with the KNN classifier. The average accuracy across the six subjects at 1 min and 5 min were 96.09% (KNN) and 96.21% (KNN), respectively. This difference may be because the 1 min time-scale signal contains a limited amount of information. Although the 5 min time-scale signal contains a sufficient amount of information, the number of samples split from the collected signal is relatively small, which affects the training accuracy of the classification model. The signal length of 3 min contains sufficient time- and frequency-domain information, and more samples can be divided from the collected signals. Therefore, at a time scale of 3 min, the HRV signal analysis of the same individual obtained a high average classification accuracy. In addition, using 1, 3, and 5 min signals achieved high overall recognition accuracy and further verified that HRV signals can reflect the operator's mental workload changes during human-robot interaction.

HRV signals between different individuals were then analyzed. Using a t -test, the statistical differences between the task-performing and rest states were analyzed. The results are presented in Table 2. Table 2 shows that 17, 18, and 17 features were the most significantly different ($p < 0.001$) for 1 min, 3 min, and 5 min time-scale HRV signals of the six subjects, respectively. The classification analysis of mental workload was performed using the features with statistical differences ($p < 0.05$) and six types of classifiers. The sample data of five of the six individuals were used as the training set, and the sample data of one individual were left as the test set. The results are shown in Figure 10. It can be seen that the average accuracy of cross-subject identification is highest at 80.56% (GB) with the 5 min time scale, and the accuracies with 1 and 3 min time scales were 77.59% (SVM) and 75.06% (SVM), respectively. We found that the accuracy of cross-subject mental workload recognition was much lower than the per-subject mental workload recognition. This is because there are strong individual

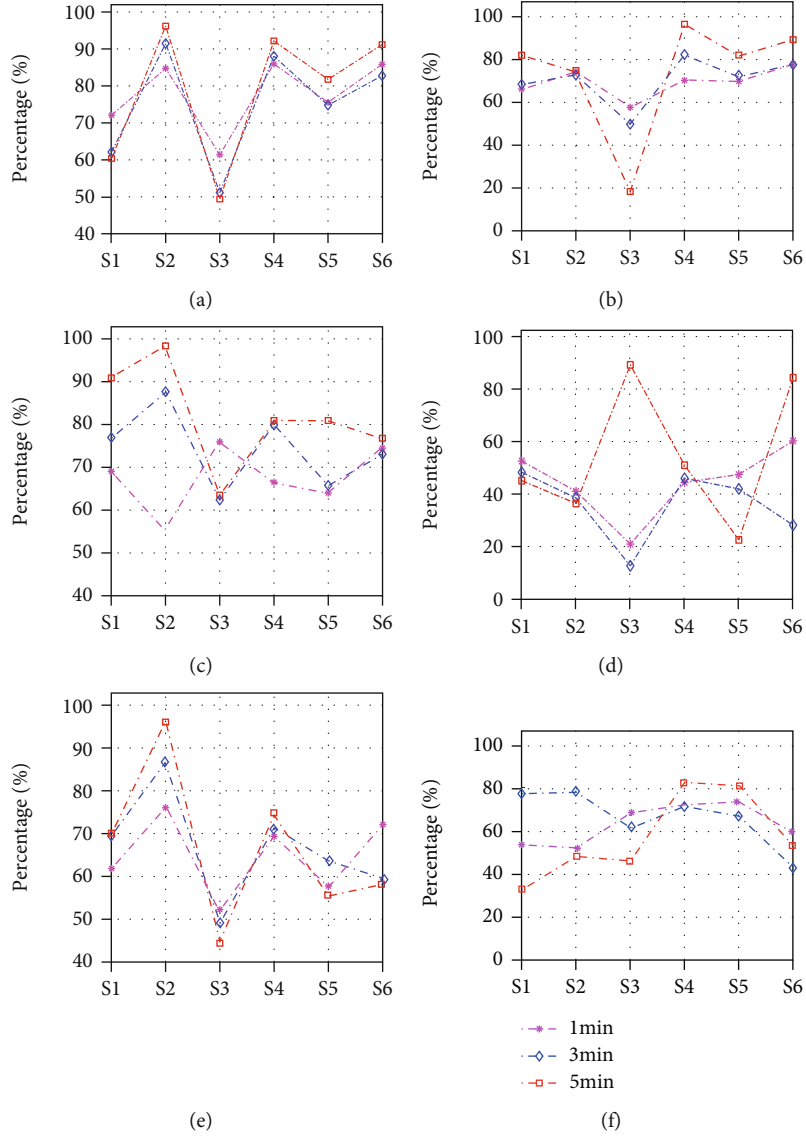


FIGURE 10: The classification accuracy of cross-subject mental workload by different classifiers at different time scales: (a) support vector machine, (b) K -nearest neighbors, (c) gentle boost, (d) linear discriminant analysis, (e) naïve Bayes, and (f) decision tree.

differences in HRV signals. Although HRV signals can reflect the fluctuating state of the ANS, there are differences in the psychological and physical qualities of different individuals. Therefore, to study cross-subject mental workload recognition, we need to further investigate the HRV signal to reflect the common characteristics of different individuals and to establish a universal mental workload recognition model.

4. Conclusion

In this paper, the differences in the recognition of the mental workload during human-robot interaction using multiple time-scale HRV signals were analyzed. First, ECG signals were obtained from six subjects while they were performing a task and while staying relaxed. Then, HRV signals were extracted based on the ECG signals. Furthermore, the HRV signals were divided into different groups using sliding windows of 1, 3, and 5 min. Then, several linear and nonlinear features of

HRV signals were extracted for these different groups. Finally, six different machine learning algorithms were used to assess the mental workload performance. For the per-subject evaluation of mental workload with different time scales, the HRV signals of each individual were used for training, and then this individual's mental workload was assessed by the trained model. In the case of a 3 min signal length, the KNN method obtained an average accuracy of 98.77%. For the cross-subject mental workload evaluation, the HRV signals of five of six individuals were used to train the model. Then, the trained model identified the mental workload of the remaining individual. The highest average classification accuracy was obtained by the GB algorithm using the 5 min time scale, and its average accuracy was 80.56%. This study explores the problems of the operator's mental workload recognition during human-robot interaction using different time-scale HRV signals. However, the sample size in this study was limited; in the future, more data will be collected for analysis to provide

generalizable experimental results. In addition, online identification of human-robot interaction mental workload will be studied. Furthermore, different machine learning algorithms will be combined to choose the best recognition result of mental workload by voting.

Data Availability

Because the physiological signal of the human body involves personal privacy, so the experimental data will not be made public temporarily.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Acknowledgments

This research was funded by the National Natural Science Foundation of China (Grant number U20A20201), the Liaoning Province Doctoral Scientific Research Foundation (Grant number 2020-BS-025), the Liaoning Revitalization Talents Program (Grant number XLYC1807018), and the National Key Research and Development Program of China (Grant number 2016YFE0206200).

References

- [1] A. Costes, F. Danieau, F. Argelaguet, P. Guillotel, and A. Lecuyer, "Towards haptic images: a survey on touchscreen-based surface haptics," *IEEE Transactions on Haptics*, vol. 13, no. 3, pp. 530–541, 2020.
- [2] S. Shao, T. Wang, Y. Su, C. Yao, C. Song, and Z. Ju, "Multi-IMF sample entropy features with machine learning for surface texture recognition based on robot tactile perception," *International Journal of Humanoid Robotics*, vol. 18, no. 2, p. 2150005, 2021.
- [3] W. Zheng, H. Liu, and F. Sun, "Lifelong visual-tactile cross-modal learning for robotic material perception," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 3, pp. 1192–1203, 2021.
- [4] P. Falco, S. Lu, C. Natale, S. Pirozzi, and D. Lee, "A transfer learning approach to cross-modal object recognition: from visual observation to robotic haptic exploration," *IEEE Transactions on Robotics*, vol. 35, no. 4, pp. 987–998, 2019.
- [5] Jongdae Jung, Seung-Mok Lee, and Hyun Myung, "Indoor mobile robot localization and mapping based on ambient magnetic fields and aiding radio sources," *IEEE Transactions on Instrumentation and Measurement*, vol. 64, no. 7, pp. 1922–1934, 2015.
- [6] W. Yuan, Z. Li, and C.-Y. Su, "Multisensor-based navigation and control of a mobile service robot," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 51, no. 4, pp. 2624–2634, 2021.
- [7] E. Debie, R. Fernandez Rojas, J. Fidock et al., "Multimodal fusion for objective assessment of cognitive workload: a review," *IEEE Transactions on Cybernetics*, vol. 51, no. 3, pp. 1542–1555, 2021.
- [8] Z. Pei, H. Wang, A. Bezerianos, and J. Li, "EEG-based multi-class workload identification using feature fusion and selection," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, article 4001108, 2021.
- [9] S. Delliaux, A. Delaforge, J.-C. Deharo, and G. Chaumet, "Mental workload alters heart rate variability, lowering non-linear dynamics," *Frontiers in physiology*, vol. 10, article 565, 2019.
- [10] K. Tsunoda, A. Chiba, K. Yoshida, T. Watanabe, and O. Mizuno, "Predicting changes in cognitive performance using heart rate variability," *IEICE Transactions on Information and Systems*, vol. E100.D, no. 10, pp. 2411–2419, 2017.
- [11] S. Shao, T. Wang, C. Song, X. Chen, E. Cui, and H. Zhao, "Obstructive sleep apnea recognition based on multi-bands spectral entropy analysis of short-time heart rate variability," *Entropy*, vol. 21, no. 8, p. 812, 2019.
- [12] S.-L. Shao, T. Wang, C.-H. Song, E. N. Cui, H. Zhao, and C. Yao, "A novel method of heart rate variability measurement," *Acta Physica Sinica*, vol. 68, no. 17, article 178701, 2019.
- [13] A. Tiwari, I. Albuquerque, M. Parent et al., "Multi-scale heart beat entropy measures for mental workload assessment of ambulant users," *Entropy*, vol. 21, no. 8, p. 783, 2019.
- [14] A. Tiwari, S. Narayanan, and T. H. Falk, "Stress and anxiety measurement "In-the-Wild" using quality-aware multi-scale HRV features," in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 7056–7059, Montreal, Canada, 2019.
- [15] P. Gilfriche, L. M. Arzac, Y. Daviaux et al., "Highly sensitive index of cardiac autonomic control based on time-varying respiration derived from ECG," *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, vol. 315, no. 3, pp. R469–R478, 2018.
- [16] M. Malik, J. T. Bigger, A. J. Camm et al., "Heart rate variability: Standards of measurement, physiological interpretation, and clinical use," *European Heart Journal*, vol. 17, no. 3, pp. 354–381, 1996.
- [17] V. Villani, M. Righi, L. Sabattini, and C. Secchi, "Wearable devices for the assessment of cognitive effort for human-robot interaction," *IEEE Sensors Journal*, vol. 20, no. 21, pp. 13047–13056, 2020.
- [18] S. Shao, T. Wang, Y. Wang, Y. Su, C. Song, and C. Yao, "Research of HRV as a measure of mental workload in human and dual-arm robot interaction," *Electronics*, vol. 9, no. 12, article 2174, 2020.
- [19] V. Villani, B. Capelli, C. Secchi, C. Fantuzzi, and L. Sabattini, "Humans interacting with multi-robot systems: a natural affect-based approach," *Autonomous Robots*, vol. 44, pp. 601–616, 2020.
- [20] R. Castaldo, L. Montesinos, P. Melillo, C. James, and L. Pecchia, "Ultra-short term HRV features as surrogates of short term HRV: a case study on mental stress detection in real life," *BMC Medical Informatics and Decision Making*, vol. 19, no. 1, article 12, 2019.
- [21] J. Pan and W. J. Tompkins, "A real-time QRS detection algorithm," *IEEE Transactions on Biomedical Engineering*, vol. -BME-32, no. 3, pp. 230–236, 1985.
- [22] L. Chen, X. Zhang, and C. Song, "An automatic screening approach for obstructive sleep apnea diagnosis based on single-lead electrocardiogram," *IEEE Transactions on Automation Science and Engineering*, vol. 12, no. 1, pp. 106–115, 2015.
- [23] F. Zhu, J. Yang, S. Xu, C. Gao, N. Ye, and T. Yin, "Incorporating neighbors' distribution knowledge into support vector

- machines,” *Soft Computing*, vol. 21, no. 21, pp. 6407–6420, 2017.
- [24] F. Zhu, Y. Ning, X. C. Chen, Y. Zhao, and Y. Gang, “On removing potential redundant constraints for SVOR learning,” *Applied Soft Computing*, vol. 102, no. 4, article 106941, 2021.
- [25] Y. Li, W. Pan, K. Li, Q. Jiang, and G. Liu, “Sliding trend fuzzy approximate entropy as a novel descriptor of heart rate variability in obstructive sleep apnea,” *IEEE Journal of Biomedical and Health Informatics*, vol. 23, no. 1, pp. 175–183, 2019.
- [26] K. Machetanz, L. Berelidze, R. Guggenberger, and A. Gharabaghi, “Brain-heart interaction during transcutaneous auricular vagus nerve stimulation,” *Frontiers in Neuroscience*, vol. 15, article 632697, 2021.
- [27] G. D. Clifford and L. Tarassenko, “Quantifying errors in spectral estimates of HRV due to beat replacement and resampling,” *IEEE Transactions on Biomedical Engineering*, vol. 52, no. 4, pp. 630–638, 2005.
- [28] W. Zheng, S. Chen, Z. Fu, F. Zhu, H. Yan, and J. Yang, “Feature selection boosted by unselected features,” *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–13, 2021.

Research Article

Unsupervised Anomaly Detection for Glaucoma Diagnosis

Wei Zhou ¹, Yuan Gao,² Jianhang Ji,¹ Shicheng Li ³, and Yugen Yi ³

¹School of Computer, Shenyang Aerospace University, Shenyang 110136, China

²Coast Guard Academy, Naval Aviation University, Yantai 264000, China

³School of Software, Jiangxi Normal University, Nanchang 330022, China

Correspondence should be addressed to Yugen Yi; yiyg510@jxnu.edu.cn

Received 22 July 2021; Revised 27 August 2021; Accepted 15 September 2021; Published 1 October 2021

Academic Editor: Fa Zhu

Copyright © 2021 Wei Zhou et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the rapid development of high tech, Internet of Things (IoT) and artificial intelligence (AI) achieve a series of achievements in the healthcare industry. Among them, automatic glaucoma diagnosis is one of them. Glaucoma is second leading cause of blindness in the world. Although many automatic glaucoma diagnosis approaches have been proposed, they still face the following two challenges. First, the data acquisition of diseased images is extremely expensive, especially for disease with low occurrence, leading to the class imbalance. Second, large-scale labeled data are hard to obtain in medical image domain. The aforementioned challenges limit the practical application of these approaches in glaucoma diagnosis. To address these disadvantages, this paper proposes an unsupervised anomaly detection framework based on sparse principal component analysis (SPCA) for glaucoma diagnosis. In the proposed approach, we just employ the one-class normal (nonglaucoma) images for training, so the class imbalance problem can be avoided. Then, to distinguish the glaucoma (abnormal) images from the normal images, a feature set consisting of segmentation-based features and image-based features is extracted, which can capture the shape and textural changes. Next, SPCA is adopted to select the effective features from the feature set. Finally, with the usage of the extracted effective features, glaucoma diagnosis can be automatically accomplished via introducing the T^2 statistic and the control limit, overcoming the issue of insufficient labeled samples. Extensive experiments are carried out on the two public databases, and the experimental results verify the effectiveness of the proposed approach.

1. Introduction

Internet of Things (IoT) techniques are emerging, which are known to be among the most critical sources of data streams that produce massive amounts of data continuously from numerous applications, such as transportation systems, security systems, intrusion systems, and fault detection in industry [1–10]. Among them, anomaly detection has drawn tremendous interest in the past couple of years. Since science and technology play a vital role in the medical department, how to establish an anomaly detection big data analysis model supported by medicine becomes a hot topic.

With the rapid growth of the worldwide population, the count of eye diseases is also increased. Among them, glaucoma is one of the serious eye diseases that can lead to irreversible vision loss [11]. According to the recent report [12], the number of glaucoma patients is predicted to increase to

80 million by 2020 and to 111.8 million by 2040. Since glaucoma can cause blindness, early detection and timely treatment are good ways to slow down the progress, preventing further vision loss [13]. In clinical, ophthalmologists always employ the color fundus images to assess the optic nerve head (ONH) for diagnosing glaucoma [14] due to its low cost. Figure 1 depicts the structure of ONH, in which the optic disc (OD) appears as a bright yellowish elliptical region consisting of the central bright region as optic cup (OC) and a peripheral region as the neuroretinal rim. Since ONH assessment requires the qualified ophthalmologists to delineate the OC and OD, it is subjective, labor-intensive, time-consuming, and not suitable for population screening [15].

For large-scale glaucoma screening, it is necessary to use automatic ONH assessment approaches. Since the glaucoma caused by the enlargement OC, several quantitative indicators are presented to identify the symptoms of glaucoma

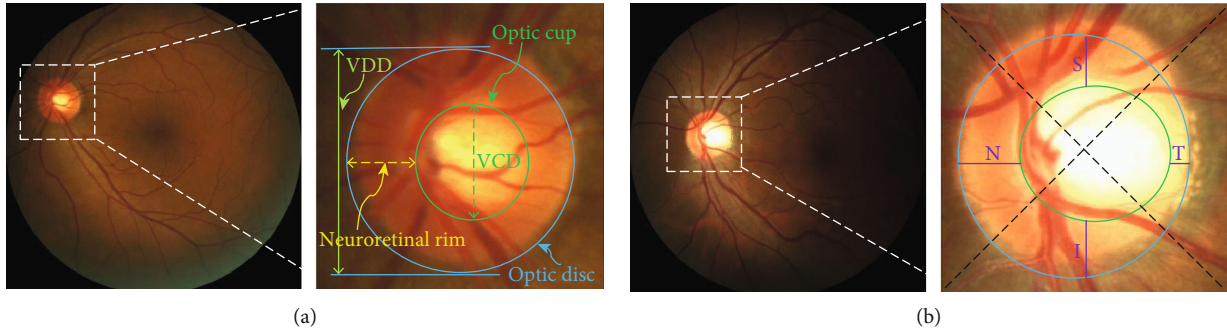


FIGURE 1: The structures of the optic nerve head. (a) Normal, (b) glaucoma.

from the fundus images, such as vertical cup to disc ratio (CDR), disc diameter, rim area, and the ISNT (inferior (I), superior (S), nasal (N), and temporal (T)) rule [16, 17] (as shown in Figure 1). Among these indicators, CDR is well accepted and widely employed indicator by clinicians. The CDR value can be calculated by the ratio of vertical cup diameter (VCD) to vertical disc diameter (VDD). Normally, a larger CDR value means the higher risk of glaucoma and vice versa. Besides, ISNT rule includes inferior (I), superior (S), nasal (N), and temporal (T) rules, which is used for the detection of glaucoma. For a normal subject, I has the highest value of neuro retinal rim configuration followed by S, N, and T.

With the rapid development of pattern recognition and machine learning, a series of automatic glaucoma diagnosis approaches have been proposed recently [18–25]. Although these approaches can achieve automatic glaucoma diagnosis, there are still two challenges in the practical applications. On one hand, the data acquisition of abnormal images is extremely expensive in medical image domain due to the large variations in appearance of OD. On the other hand, it is a time consuming and tedious task to labeling the medical images by the experienced clinicians. Therefore, the cost of obtaining large-scale labeled medical images is often expensive in clinical. In contrast, collecting a large number of normal data is relatively easy.

As we all know, human are good at distinguishing the abnormal images from the normal images [26]. Inspired by this, this paper designs an automatic approach for glaucoma diagnosis, which just uses the normal images for training model. The main advantages of the proposed approach are given as below:

- (1) In order to improve the detection accuracy and reduce the computational cost, the region of interest (ROI) in retinal image is extracted by exploring the OD location with unsupervised boundary extraction and Hough transform
- (2) Segmentation-based features and image-based features are extracted from the obtained ROIs to capture the shape and textural changes of the fundus images
- (3) To reduce the effects caused by noises and redundancy features, an effective feature set can be selected by combining the elastic net penalty and PCA together

- (4) In order to solve the issue of class imbalance, the T^2 statistic and the control limit are designed on the normal images, which can be used to distinguish the abnormal images from the normal images, achieving automatic glaucoma diagnosis
- (5) Extensive experiments are carried out on the two public fundus databases, and the experimental results indicate that the proposed approach is effective

The remainder of this paper is arranged as below. Section 2 gives some related works, and then, the proposed approach is introduced in details in Section 3. Next, the experimental results and analyses are presented in Section 4. Finally, the whole work is concluded in brief in Section 5.

2. Related Work

The existing automatic glaucoma diagnosis approaches can be divided into two categories: machine learning- (ML-) based approaches and deep learning- (DL-) based approaches.

2.1. Machine Learning- (ML-) Based Approaches. ML-based glaucoma diagnosis approaches follow a fixed procedure [27], e.g., (1) input an image; (2) preprocessing; (3) feature extraction; and (4) classification (diagnosis). In preprocessing stage, Contrast Limited Adaptive Histogram Equalization (CLAHE) [28] is commonly used to reduce the negative effects caused by noise and artifact, improving the quality of the fundus images. In feature extraction stage, a series of important and distinctive hand-crafted features will be extracted to explore the concealed pixel variations in the retinal images. The extracted hand-crafted features are classified into wavelet decomposition-based features [29], morphological-based features [30], nonlinear-based features [31], textural-based features [32], and image descriptor-based features [33]. After feature extraction, the last process is classification. Usually, with the usage of the extracted features, a classification model can be trained which identifies the normal class versus abnormal class. Many classifiers have been employed to distinguish the two classes based on the extracted features, for instance, artificial neural network (ANN) [34], K -nearest neighbor (KNN) [32], support vector machine (SVM) [35], least square support vector machine

(LS-SVM) [29], and extreme learning machine (ELM) [36]. A major challenge in the machine learning-based approaches is that the hand-crafted appropriate features should be set beforehand. Seen from these extracted features, most of them belong to the image-based features. Nevertheless, segmentation-based features are ignored, which can be regarded as one of the most important clinical indicators for glaucoma diagnosis. Besides, usually, the number of normal images is much larger than the abnormal images, leading to the class imbalance. Under this circumstance, the classifier of machine learning approaches can hardly train well, which will affect the final diagnosis performance.

2.2. Deep Learning- (DL-) Based Approaches. DL-based approaches follow the same sequence as ML approaches for glaucoma diagnosis. However, the major difference between them is the deep learning network can self-learn during the training of the network, without extracting a series of hand-crafted features for classification. For example, convolutional neural network (CNN) is the most commonly employed form of deep learning. Generally, a structure of CNN contains (1) convolutional layers, (2) pooling layers, and (3) fully connected layers. The convolutional layer is used to extract the nonlinear features. The pooling layer is utilized to reduce the space dimensionality of sample and keep the important information unchanged. The fully connected layer is to connect the every neuron in the previous layer with the neurons in the current layer of the CNN model. Furthermore, some variations of DL, e.g., adversarial learning [37], FCNs [38], and modified U-net [24], are used to glaucoma diagnosis, improving the diagnosis performance.

Although DL-based approaches have achieved many breakthroughs in medical image analysis, these approaches mainly rely on large-scale labeled data. However, in medical image domain, the samples are always small-scale and unlabeled, so DL-based approaches can hardly perform well [39, 40].

3. Our Approach

Our approach consists of the following three stages, including ROI extraction, features extraction (segmentation-based features and image-based features), and glaucoma diagnosis. The flowchart of the proposed approach is shown in Figure 2.

3.1. ROI Extraction. Generally, the resolution of original fundus image is relatively large. Meanwhile, the region of interest for glaucoma diagnosis is just a small region according to the clinical prior knowledge. In order to reduce the influence of the unnecessary background information and the computation cost, improving the diagnosis accuracy, the extraction of ROI around the OD is necessary (as shown in Figure 3).

The process of ROI extraction consists of the following three stages: first, this paper employs our previous proposed approach named as an adaptive rough OD boundary curve extraction based on unsupervised learning [16] to extract the OD region (as shown in Figure 3(b)). After that, in order to accurately locate the center of OD, the Circular Hough Transform (CHT) is employed to the extracted OD region

according to the fact that the prior knowledge of OD is circle in shape, depicted in Figure 3(c). At last, with the usage of the extracted center of OD as the ROI center, according to the clinical experience [16], this paper cuts the original images into small images with the resolution of 400×400 on the REFUGE and RIM-ONE r2 databases, which are called ROIs around the OD. An example from the REFUGE database is shown in Figure 3(d).

Successful optic disc center localization is considered to be that the distance between the estimated optic disc center and the manually selected center is less than the optic disc radius [41, 42]. Compared with the state-of-the-art OD detection approaches, our approach is more robust to the image quality, contrast, brightness, and different lesions. Since human vision is more sensitive to green color, only green channel of ROI is extracted from the RGB image for further processing. Meanwhile, CLAHE is introduced into the extracted green channel of ROI [28] to enhance the contrast (illustrated in Figure 3(e)).

3.2. Feature Extraction. OD with varying intrinsic principal features, e.g., distinct vessel structures, sizes, and brightness, gives more challenges for glaucoma diagnosis. To fully explore the difference between the normal images and abnormal images, improving the effectiveness and accuracy of the glaucoma diagnosis, this paper describes the OD from the following two sides. For one side, inspired by the clinical glaucoma diagnosis, a series of segmentation-based features are extracted to accurately capture the shape deformations to characterize glaucoma. For another, some image-based features are extracted to capture the shape and textual changes in OD. We believe that taking the segmentation-based and image-based features into a united framework can provide complementary and independent information for OD descriptions. The detailed descriptions of the extracted features are given as below.

3.2.1. Extraction of Segmentation-Based Features. Based on the extracted ROI around the OD, this paper adopts [11] to segment the OD and OC. In clinical, for a normal fundus image, the rim is the thickest in the inferior (I) sector and thinnest in the temporal (T) sector (as shown in Figure 1). Usually, experienced ophthalmologists observe inferior (I), superior (S), nasal (N), and temporal (T) quadrants for glaucoma diagnosis. Therefore, a series of clinical features are extracted from the segmented OD and OC, denoted as F1 to F5 (as shown in Table 1). F1-F3 are based on rim-disc ratio, and F4-F5 are calculated by rim profile and ISNT, respectively. Among them, the calculation of F5 is based on a disk-shaped region named as the NRR, obtained by removing the OC from the OD.

The calculation process of F5 is given as below: Figure 4(a) is the original image, and Figure 4(f) is the NRR. The masks employed in this paper are used to measure the NRR in each quadrant which are depicted in Figures 4(b)–4(e). The NRR area in superior, nasal, inferior, and temporal are, respectively, shown in Figures 4(g)–4(j), which can be used to calculate F5.

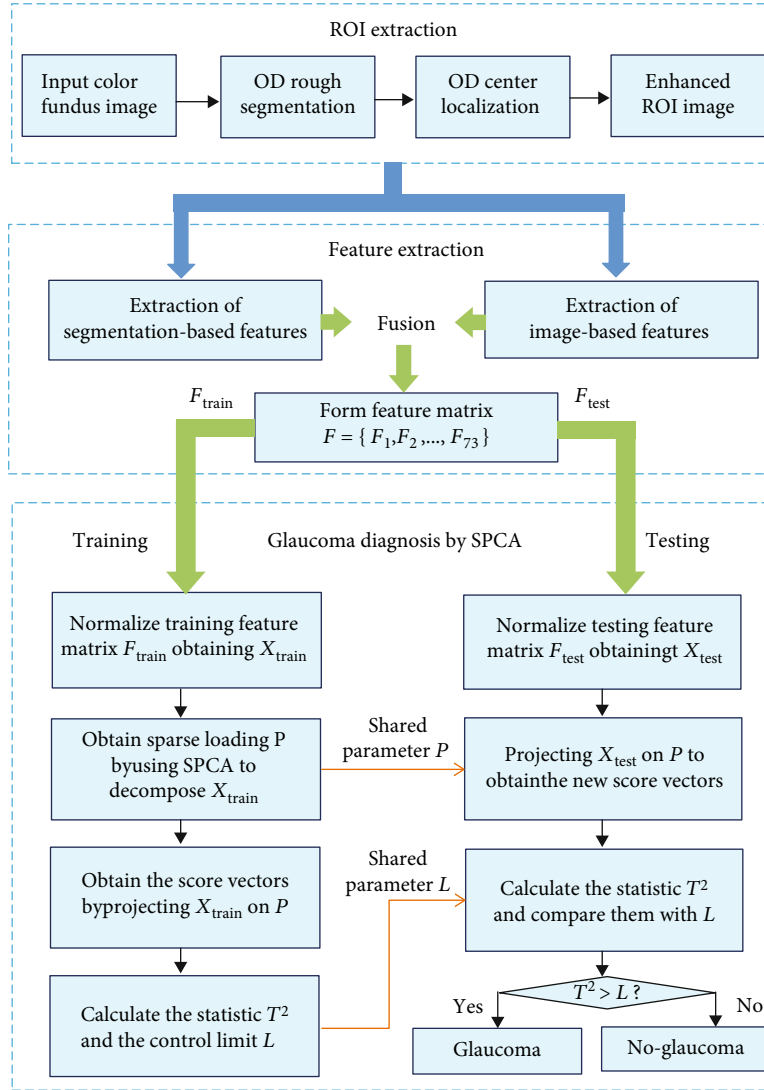


FIGURE 2: The flowchart of the proposed approach.

3.2.2. *Extraction of Image-Based Features.* Since the gray level of the image on the edge is discontinuous, it has singularity. Recent studies have shown that the multiresolution features of wavelet coefficients and the localization characteristics of wavelet analysis can be used to obtain the domain features at different scales. In addition, the high frequency information decomposed by wavelet can also be used for multiscale edge detection [43], which is very fit for medical image processing and analysis. Inspired by the advantages of wavelet features, this paper employs an adaptive signal decomposing approach, named as two-dimensional (2D) EWT with Littlewood-Paley as an empirical wavelet [44] to decompose the image into various frequency bands. Based on the decomposed components of (2D) EWT, a series of image-based features consisting of chip histogram, gray level cooccurrence matrix (GLCM), and moment invariance are extracted. Here, the enhanced green channel image is utilized as the input of (2D) EWT. For each input image, four frequency bands can be decomposed. The detailed image-based features descriptions are given as below:

(a) Chip histogram features

Chip feature belongs to statistical textural feature, which can be extracted from the second-order histogram. There are six features in it including mean, variance, skewness, kurtosis, energy, and entropy. For a given gray image, $t(i)$ is the probability density function of the intensity level i , which is denoted as $t(i) = \text{Histogram}(i)/T$, $i = 1, 2, 3, \dots, N$. T is the total pixels in the gray image, and N is the total number of gray levels, in which Q is the gray level vector represented as $[1, 2, 3, \dots, N]$. The expressions of the chip histogram features are depicted in Table 2.

(b) GLCM features

GLCM is utilized to extract the texture in an image by doing the transition of gray level between two pixels, which is one of the earliest approaches for texture feature extraction [45]. For each GLCM, four directions can be computed, in which each direction has four different characteristics, i.e.,

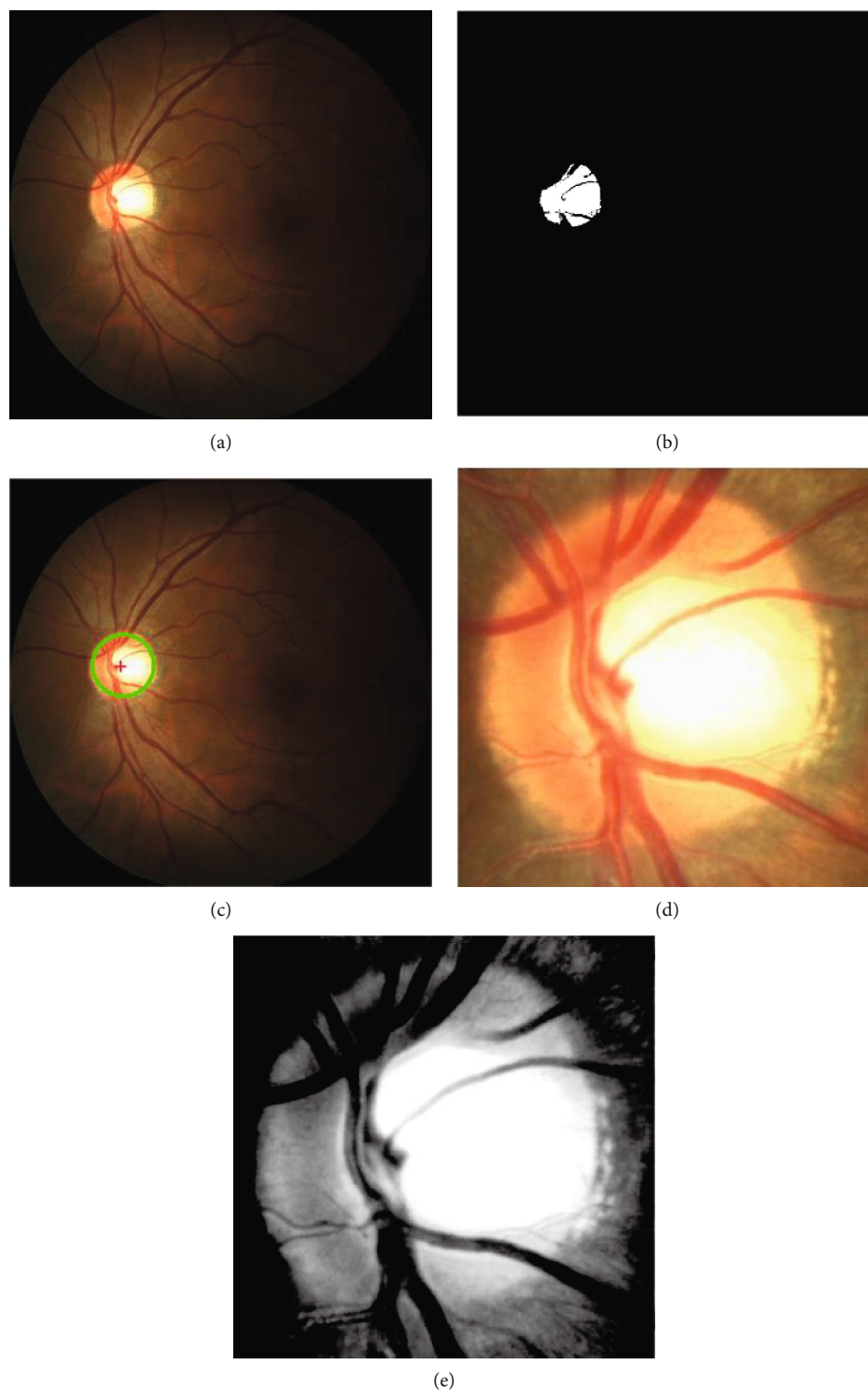


FIGURE 3: (a) Original image; (b) coarse segmentation image; (c) OD localization-based CHT; (d) extracted ROI; (e) enhanced green channel image by CLAHE.

correlation, contrast, energy, and homogeneity. For each characteristic, we extract the mean value of four different directions. Therefore, there are four GLCM features for each image. Suppose that $G(u, v)$ denotes (u, v) entries in normalized GLCM, and the mean and standard deviation along

u and v coordinates are expressed as $(\mu_x$ and $\mu_y)$ and $(\sigma_x$ and $\sigma_y)$, respectively. Table 3 gives the expressions of the GLCM features.

(c) Moment Invariance features

TABLE 1: Extraction of segmentation-based features and the corresponding expressions.

Segmentation-based features	Expressions
F1: vertical CDR	vCDR = VCD/VDD
F2: the area ratio of rim to disc	Area_ratio = Rim_Area/OD_Area
F3: the major axis length of OD	Length_Major_axis_OD = L_OD
F4: the area of rim	OD_Area - OC_Area
F5: neuro-retinal rim (NRR)	NRR in I quadrant area + NRR in S quadrant area/NRR in N quadrant area + NRR in T quadrant area

In order to overcome the changes of object shape, position, and orientation [46], this paper employs the moment invariance features for describing the OD and OC. There are seven features in moment invariant features, and the mathematical expressions of these features are shown in Table 4.

The invariants λ_{mn} can be calculated by $\lambda_{mn} = \mu_{mn} / \mu_{00}^{1+(m+n)/2}$, in which μ_{mn} and μ_{00} are, respectively, denoted as the center moment and the zeroth central moment.

After feature extraction, 73 features in which 5 segmentation-based features and $68 \times (6 + 4 + 7)$ image-based features are computed for each input image, forming a 73-dimension feature set $F = \{F_1, F_2, \dots, F_{73}\}$. Each feature F_i is normalized to zero mean and unit variance by using $F_i' = (F_i - \mu_i) / \sigma_i$ in which μ_i is the mean of the i th feature, and σ_i is the corresponding standard deviation.

3.3. Anomaly Detection for Glaucoma with SPCA. In this subsection, we firstly review the (principal component analysis, PCA) and sparse PCA. Then, T^2 statistic of glaucoma and the corresponding control limit are given for glaucoma diagnosis.

3.3.1. PCA. PCA is one of the most popular dimensionality reduction approaches, which is aimed at maximizing the variance of projections on the new directions. For PCA, a series of load vectors can be constructed by the orthogonal vectors. Supposing that $X \in R^{N \times D}$ is a given training sample set, N and D denote the number of the samples and variables (features), respectively. The objective function of PCA is given as

$$\max_{u \neq 0} \frac{u^T X^T X u}{u^T u}, \quad (1)$$

where $u \in R^D$. The solution of Equation (1) can be computed by singular value decomposition (SVD) as

$$\frac{1}{\sqrt{N-1}} X = W \Sigma U^T, \quad (2)$$

where $W \in R^{N \times N}$ and $U \in R^{D \times D}$ are unitary matrices, in which the orthogonal column vectors in matrix U are called the load vectors. Projecting X on the i th column has the variance as σ_i^2 . $\Sigma \in R^{N \times D}$ is a diagonal matrix where the elements in main diagonal are the nonnegative singular

values in descending order ($\sigma_1 \geq \sigma_2, \dots, \geq \sigma_{\min(N,D)} \geq 0$), and the others are zeros.

A new load matrix $P \in R^{D \times A}$ can be obtained by selecting the first A columns in U , named as the principal component subspace (PCS), which is represented as $T = XP$. For a given new sample x , it can be projected on the PCS as

$$t = P^T x \in \text{PCS}, \quad (3)$$

In Equation (3), A is a parameter, which is calculated by cumulative percent variance (CPV), and this paper adopts reference [47] to resolve it.

3.3.2. Sparse PCA. Seen from the PCA, its major work is to acquire the maximum variance on the certain loading vectors. However, in practice, some principal components are linear relevant with each other and some of them have large noises, which will weaken the precision of detection. Inspired by the advantages of sparsity, sparse PCA has been proposed by performing the maximization of objective function under the L1 norm constraint as

$$\sum_{j=1}^D |u_j| \leq s, \quad (4)$$

where s is the number of nonzero elements in a load vector. $u_j (j \in \{1, 2, \dots, D\})$ denotes the j th element of the load vector. By introducing the aforementioned sparse constraint, each principle component has lesser original variables. Therefore, it not only enhances the interpretability of the principal components but also reduces the storage space.

Many approaches have been designed to solve Equation (4). Among them, Jolliffe and Uddin [48] proposed the most effective approach to resolve it. The main processes are depicted as below: first, PCA can be recast exactly by a regression problem, such as ridge regression. Then, the regression problem is changed to an elastic-net regression by introducing the L1 norm constraint. For more details, please refer to reference [49].

3.3.3. Index and Control Limit. A series of statistics have been applied to multivariable statistical analysis. Among them, one of the most preventative ways, namely, Hotelling's T^2 statistic, is utilized to represent the variability in the PCS, which is defined as

$$T^2 = x^T P \Lambda^{-1} P^T x, \quad (5)$$

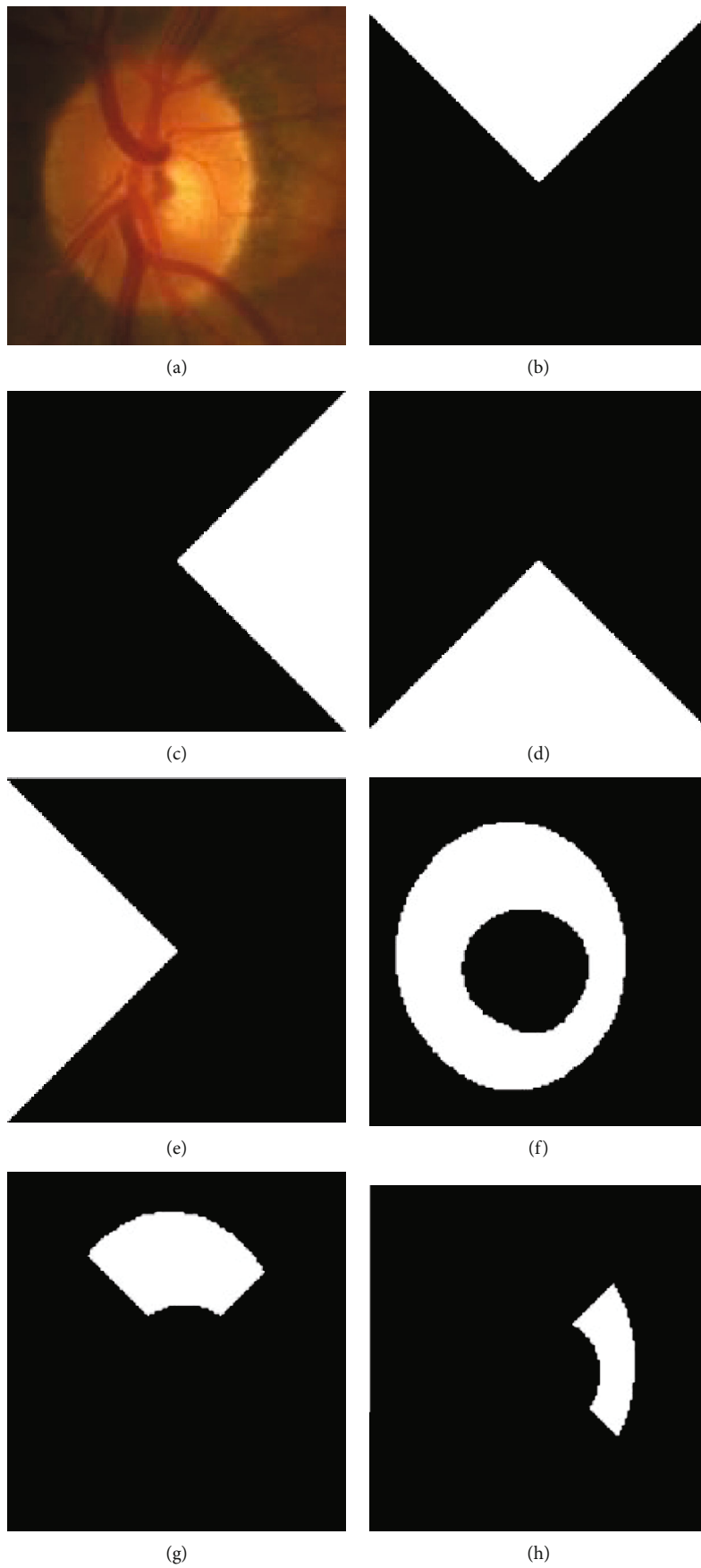


FIGURE 4: Continued.

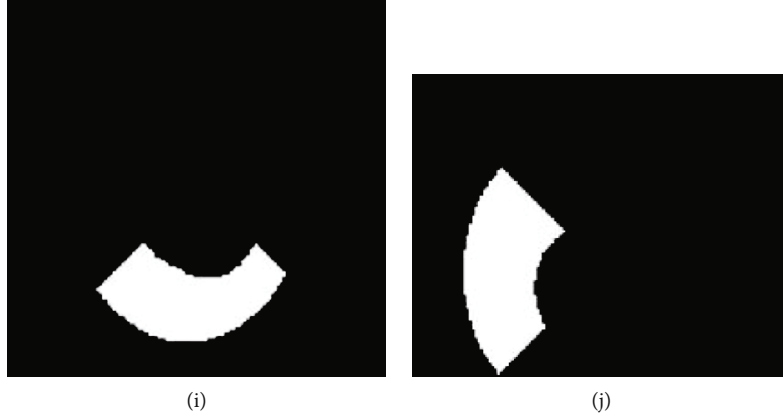


FIGURE 4: (a) Original color fundus image; (b–e) the masks in superior, nasal, inferior, and temporal, respectively. (f) NRR; (g–j) NRR in superior, nasal, inferior, and temporal, respectively.

TABLE 2: Expressions of chip histogram features.

Chip histogram features	Expressions
Mean	$\sum_{i=1}^N (t(i) \times Q(i))$
Variance	$\sum_{i=1}^N (t(i) \times Q(i)^2)$
Skewness	$\left[\sum_{i=1}^N (t(i) \times Q(i) - \text{Mean})^3 \right] \times \left[(\sqrt{\text{Variance}})^{-3} \right]$
Kurtosis	$\left[\sum_{i=1}^N (t(i) \times Q(i) - \text{Mean})^4 \right] \times \left[(\sqrt{\text{Variance}})^{-4} \right]$
Energy	$\sum_{i=1}^N t(i) \times t(i)$
Entropy	$-\sum_{i=1}^N t(i) \times \log t(i)$

TABLE 3: Expressions of GLCM feature.

GLCM features	Expressions
Contrast	$\sum_{u=1} \sum_{v=1} (u - v)^2 \times G(u, v)$
Correlation	$\sum_{u=1} \sum_{v=1} uv \times G(u, v) - \mu_x \mu_y / \sigma_x \sigma_y$
Energy	$\sum_{u=1} \sum_{v=1} G(u, v)^2$
Homogeneity	$\sum_{u=1} \sum_{v=1} G(u, v) / 1 + (u - v)^2$

where $\Lambda = \Sigma^T \Sigma$. The sample vector x follows a multivariate normal distribution.

$$\frac{N(N-A)}{A(N^2-1)} T^2 \sim F_{A, N-A}, \quad (6)$$

where $F_{A, N-A}$ is an F distribution with A and $N - A$ degrees of freedom. For a given significance level α , the detected image is regarded as glaucoma if

$$T^2 \leq T_\alpha^2 \equiv \frac{A(N^2-1)}{N(N-A)} F_{A, N-A; \alpha}. \quad (7)$$

The whole process of the proposed glaucoma diagnosis approach can be divided into the following four stages. In the first stage, an unsupervised boundary curve extraction model and Circular Hough Transform (CHT) are used to extract ROI. In the second stage, a series of features containing segmentation-based and image-based features are extracted from the ROI, forming the feature matrix. The third stage is the offline training. In this part, the glaucoma model is constructed based on sparse PCA, and the control limit L is also given. The last stage is the online testing. For a new fundus image, first, repeat the process of stages one and two, and then, compute the corresponding statistic; at last, compare them with the control L for distinguishing glaucoma from normal.

4. Experiments and Results

4.1. Databases. In experiment, we employ two public databases to verify the effectiveness of the proposed approach. The detailed descriptions of the databases are given as below:

Retinal Fundus Glaucoma Challenge (REFUGE) database [50] consists of 1200 color retinal fundus images stored in JPEG format, in which 120 are glaucomatous and 1080 are nonglaucoma images. REFUGE database is divided into three parts: 400 training images, 400 validation images, and 400 testing images, in which 400 training images are acquired with a Zeiss Visucam 500 fundus camera with the resolution of 2124×2056 pixels, and the 400 validation images and 400 testing images are acquired with Canon CR-2 device with the resolution of 1634×1634 pixels. In REFUGE database, only 400 training images and 400 validation images which are labeled as ‘‘ground truth’’ with the segmentation results of OD and OC and the rest 400 testing images without labeling the ground truth. Therefore, in this experiment, we adopt 400 training images (40 glaucoma images and 360 nonglaucoma images) and 400 validation images (40 glaucoma images and 360 nonglaucoma images)

TABLE 4: Expressions of moment invariance features.

Moment invariance features	Expressions
MV1	$\lambda_{20} + \lambda_{02}$
MV2	$(\lambda_{20} - \lambda_{02})^2 + 4\lambda_{11}^2$
MV3	$(\lambda_{30} - 3\lambda_{12})^2 + (3\lambda_{21} - \lambda_{03})^2$
MV4	$(\lambda_{30} + \lambda_{12})^2 + (\lambda_{21} + \lambda_{03})^2$
MV5	$(\lambda_{30} - 3\lambda_{12})(\lambda_{30} + \lambda_{12})((\lambda_{30} + \lambda_{12})^2 - 3(\lambda_{21} + \lambda_{03})^2)$ $+ (3\lambda_{21} - \lambda_{03})(\lambda_{21} + \lambda_{03})(3(\lambda_{30} + \lambda_{12})^2 - (\lambda_{21} + \lambda_{03})^2)$
MV6	$(\lambda_{20} - \lambda_{02})[(\lambda_{30} + \lambda_{12})^2 - (\lambda_{21} + \lambda_{03})^2] + 4\lambda_{11}(\lambda_{30} + \lambda_{12})(\lambda_{21} + \lambda_{03})$
MV7	$(3\lambda_{21} - \lambda_{03})(\lambda_{30} + \lambda_{12})[(\lambda_{30} + \lambda_{12})^2 - 3(\lambda_{21} + \lambda_{03})^2] - (\lambda_{30} - 3\lambda_{12})$ $(\lambda_{21} + \lambda_{03})[3(\lambda_{30} + \lambda_{12})^2 - (\lambda_{21} + \lambda_{03})^2]$

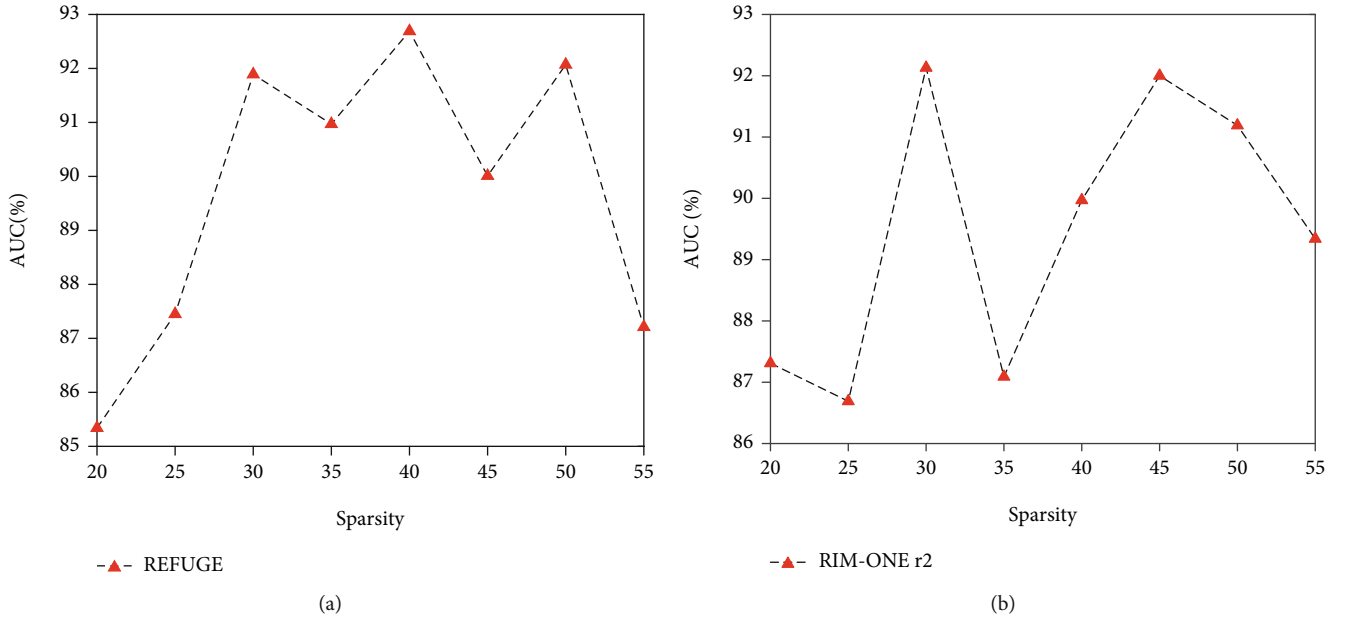


FIGURE 5: The classification performance (AUC) with different values of sparsity on the REFUGE and RIM-ONE r2 databases. (a) REFUGE database, (b) RIM-ONE r2 database.

to train and verify the effectiveness of the proposed approach. In total, there are 80 glaucoma images and 720 nonglaucoma images. In experiment, 500 nonglaucoma images are randomly selected constructing the training set. 20 glaucoma images and 160 nonglaucoma images are regarded as the validation set (20 + 160). And the rest 60 nonglaucoma images and 60 glaucoma images are regarded as the testing set (60 + 60). The random sample selection process is repeated 10 times, and the average result is regarded as the final result.

RIM-ONE r2 database [51] comprises of 455 retinal fundus images, in which 255 are normal images and 200 are glaucoma images. In experiment, first, all of the images in this database are resized in the same dimensionality. And then, the ROI extraction depicted in Section 2 is used to these images. At last, a series of features are extracted from the obtained ROIs, which are regarded as the inputs for our approach.

In experiment, 200 nonglaucoma images are randomly selected constructing the training set. 25 glaucoma images and 25 nonglaucoma images are randomly selected as the validation set (25 + 25). The rest 30 nonglaucoma images and 175 glaucoma images are regarded as the testing set (30 + 175). The random sample selection process is repeated 10 times, and the average result is regarded as the final result.

The proposed approach is trained and tested by using a desktop computer having 16 GB random access memory (RAM), Intel[®] Core™ i7 CPU950@3.7 GHz. We develop our approach using MATLAB 2021a for training and testing.

4.2. Evaluation Criterion. In order to evaluate the effectiveness of the proposed approach, three evaluation criteria, namely, accuracy, sensitivity, and specificity, are employed

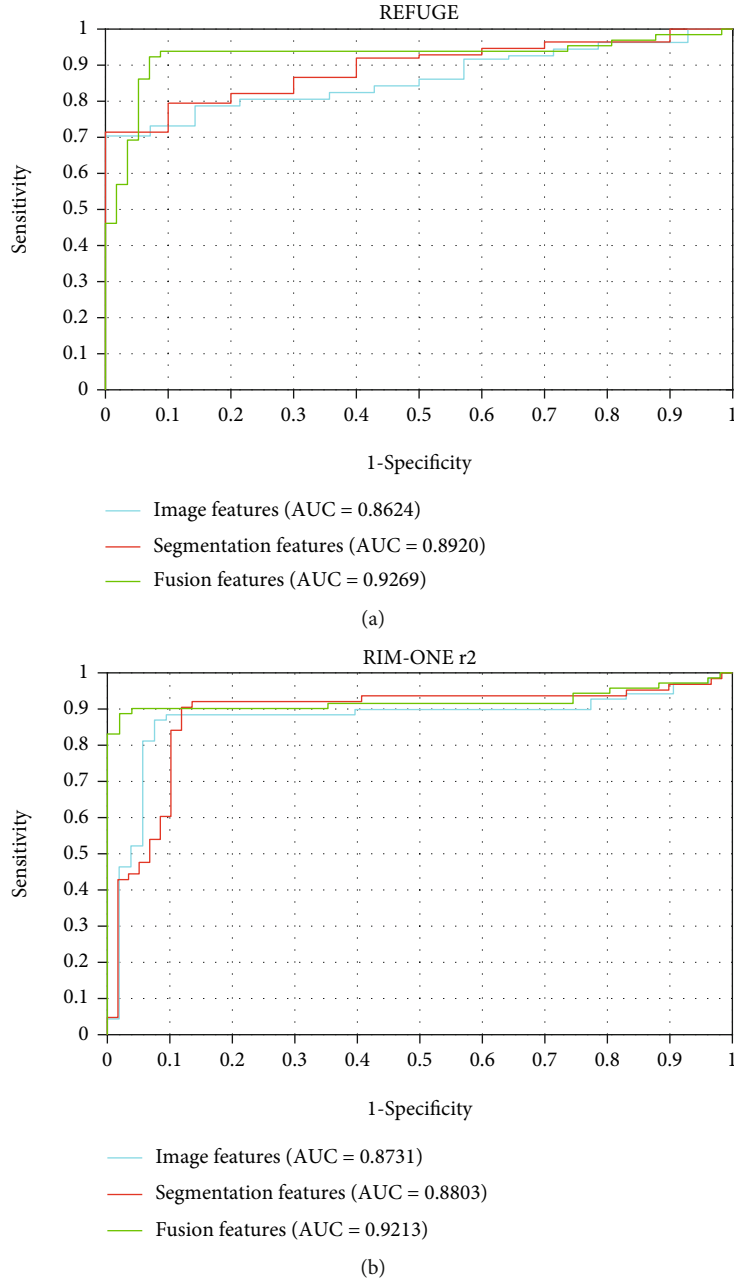


FIGURE 6: The ROC curves of the proposed glaucoma diagnosis approach with three different kinds of feature extraction ways on the REFUGE and RIM-ONE r2 databases.

in this experiment. The corresponding mathematical expressions of these evaluation criteria are depicted as below:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \times 100\%, \quad (8)$$

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \times 100\%, \quad (9)$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \times 100\%, \quad (10)$$

where true positive (TP) is the number of glaucoma images that are correctly identified; false negative (FN) is the number of incorrectly found as nonglaucoma images; false positive (FP) is the number of incorrectly found as glaucoma images. true negative (TN) is the number of nonglaucoma images that are correctly identified.

In order to evaluate the effectiveness of the proposed approach, receiver-operating characteristics (ROC) curve is utilized in this paper. In ROC curve, the vertical axis is the sensitivity and the horizontal axis is (1-specificity). The area under the ROC curve denotes as the AUC value

TABLE 5: Comparison of the proposed approach against the state-of-the-art approaches on the testing set of REFUGE database and RIM-ONE r2 database.

Methods	Accuracy (%) REFUGE	Accuracy (%) RIM-ONE r2
Wavelet features [52]	58.70	59.81
Superpixel segmentation [53]	63.53	78.64
Semisupervised clustering [54]	81.95	82.76
Deep learning [55]	80.12	84.46
AWLCSC [31]	85.54	85.56
Ours	92.44	93.65

for measuring and describing the algorithm performance. A higher AUC indicates that the performance of the approach is better.

4.3. Parameter Settings and Analysis. There are two hyper-parameters, i.e., the sparseness criterion for the SPCA stage and the sigma-like threshold on the actual classification stage. Here, sparsity will make principal component coefficients (the coefficients in front of each variable when forming principal components) be sparse. In other words, most of the coefficients will become zero values. In this way, we can highlight the major parts of principal components, so that the principal components will become easier to explain. In order to choose suitable sparsity, this paper employs the commonly used ROC curve. In our experiment, the validation set is employed for parameter selection and validation. For REFUGE database, 160 nonglaucoma images and 20 glaucoma images construct the validation set. For RIM-ONE r2 database, the validation set consists of 20 nonglaucoma images and 20 glaucoma images.

In our experiment, we tune the values of sparsity parameter by searching the grid {20, 25, 30, 35, 40, 45, 50, 55} for REFUGE and RIM-ONE r2 databases. We test the diagnosis performance (AUC) of the proposed approach under different values of sparsity parameter on the REFUGE and RIM-ONE r2 databases (as depicted in Figures 5(a) and 5(b)). As seen from these figures, we can learn that the proposed glaucoma diagnosis approach can achieve maximum AUC values on the REFUGE and RIM-ONE r2 databases when the values of sparsity are set as 40 and 30. After the sparsity is determined, the threshold for T^2 can be computed by Equation (7).

4.4. Experimental Results and Analysis. In this subsection, we will carry out two experiments to verify the effectiveness of the proposed approach. For one side, the proposed approach with different features including segmentation-based features, image-based features, and their fusion features will be tested, respectively, and the obtained diagnosis performances are shown in Figure 6.

Seen from the results depicted in Figure 6, we can learn that when the proposed approach employs the fusion features, it can achieve the best performance on the testing set. The main reason is that the segmentation-based and image-based features can capture the shape and textural

changes, respectively. Nevertheless, the fusion of these features provides complementary information for glaucoma diagnosis, which can improve diagnosis performance.

For another, the performance of the proposed approach is compared against the state-of-the-art approaches, i.e., wavelet features [52], superpixel segmentation [53], semisupervised clustering [54], deep learning [55], and AWLCSC [31]. Table 5 shows the diagnosis results obtained by different algorithms on the REFUGE database and RIM-ONE r2 database, respectively. Among them, superpixel segmentation, wavelet features, and deep learning are supervised learning approaches; semisupervised clustering is the semisupervised learning approach; AWLCSC and the proposed approach belong to unsupervised learning approaches. For supervised learning approaches, they can achieve good performance when a large number of labeled samples are used to train model. However, a large number of labeled data are hard to obtain, especially for medical domain. Under this situation, semisupervised learning and unsupervised learning have gained tremendous attention. Although these approaches can overcome the problem of insufficient samples, the class imbalance problem still lies in these approaches reducing their classification performance. Different from the existing glaucoma diagnosis approaches, this paper just employs one-class normal data for constructing model and designs the control limit for detecting the abnormal data. Therefore, the aforementioned two limitations can be avoided. According to the comparison results depicted in Table 5, we can learn that the proposed approach is more reliable than other tested approaches in terms of diagnosis accuracy, indicating the effectiveness of the proposed approach.

5. Conclusions

In this paper, we propose an unsupervised anomaly detection approach via sparse PCA for glaucoma diagnosis. Since the ODs vary in sizes, shapes, and appearances for glaucoma images, it can hardly construct an effective model to diagnose glaucoma. Instead of using the glaucoma images, this paper just constructs the diagnosis model based on the nonglaucoma images. Therefore, the proposed approach overcomes the class imbalance issue, which is a hard problem in classification. Furthermore, a series of features including segmentation-based features and image-based features are designed, which can capture the shape and textural changes, respectively, improving the discrimination of the OD and OC. Experimental results indicate that the proposed approach can achieve good glaucoma diagnosis performance.

The main disadvantage of the proposed approach is that the fundus images were obtained by different equipment and institutions, so that the trained model parameters by our proposed approach cannot perform stable detection results. Therefore, it is necessary to improve the generalization of the model from different datasets, which is one of our future research directions. Besides, more large-size databases will be introduced to further verify the effectiveness of the proposed approach.

Data Availability

The data are derived from public domain resources.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work is supported in part by grants from the National Natural Science Foundation of China (Nos. 62062040, 61772091, 61962006, and 62102270), the China Postdoctoral Science Foundation (No. 2019M661117), the Scientific Research Fund Project of Liaoning Provincial Department of Education (Nos. JYT19040 and JYT19053), the Scientific Research Funds of Shenyang Aerospace University under grant (Nos. 18YB01 and 19YB01), the Natural Science Foundation of Liaoning Province Science and Technology Department (No. 2019-ZD-0234), National Natural Science Foundation of Liaoning Province (No. 2020-MS-239), Key Scientific Research Projects of Liaoning Provincial Department of Education (No. LZD202002), and the Science and Technology Innovation Program of Hunan Province (No. 2020SK50106).

References

- [1] F. Zhu, J. Yang, C. Gao, S. Xu, N. Ye, and T. Yin, "A weighted one-class support vector machine," *Neurocomputing*, vol. 189, pp. 1–10, 2016.
- [2] Y. He, G. Duan, G. Luo, and X. Liu, "Robust visual relationship detection towards sparse images in Internet-of-Things," *Wireless Communications and Mobile Computing*, vol. 2021, 10 pages, 2021.
- [3] F. Zhu, Y. Ning, X. Chen, Y. Zhao, and Y. Gang, "On removing potential redundant constraints for SVOR learning," *Applied Soft Computing*, vol. 102, p. 106941, 2021.
- [4] Y. Yi, W. Zhou, Y. Shi, and J. Dai, "Speedup two-class supervised outlier detection," *IEEE Access*, vol. 6, pp. 63923–63933, 2018.
- [5] Z. Zhang, H. Chen, X. Yin, and J. Deng, "EAWNet: an edge attention-wise objector for real-time visual Internet of Things," *Wireless Communications and Mobile Computing*, vol. 2021, 15 pages, 2021.
- [6] F. Zhu, N. Ye, W. Yu, S. Xu, and G. Li, "Boundary detection and sample reduction for one-class support vector machines," *Neurocomputing*, vol. 123, pp. 166–173, 2014.
- [7] W. Zhou, Z. Gong, W. Guo, N. Han, and S. Qiao, "Robust graph structure learning for multimedia data analysis," *Wireless Communications and Mobile Computing*, vol. 2021, 12 pages, 2021.
- [8] Y. Yi, Y. Shi, W. Wang, G. Lei, J. Dai, and H. Zheng, "Combining boundary detector and SND-SVM for fast learning," *International Journal of Machine Learning and Cybernetics*, vol. 12, no. 3, pp. 689–698, 2021.
- [9] S. Li, Q. Liu, J. Dai, W. Wang, X. Gui, and Y. Yi, "Adaptive-weighted multiview deep basis matrix factorization for multimedia data analysis," *Wireless Communications and Mobile Computing*, vol. 2021, 12 pages, 2021.
- [10] R. Al-amri, R. K. Murugesan, M. Man, A. F. Abdulateef, M. A. Al-Sharafi, and A. A. Alkahtani, "A review of machine learning and deep learning techniques for anomaly detection in IoT data," *Applied Sciences*, vol. 11, no. 12, p. 5320, 2021.
- [11] W. Zhou, Y. Yi, Y. Gao, and J. Dai, "Optic disc and cup segmentation in retinal images for glaucoma diagnosis by locally statistical active contour model with structure prior," *Computational and Mathematical Methods in Medicine*, vol. 2019, 16 pages, 2019.
- [12] Y.-C. Tham, X. Li, T. Y. Wong, H. A. Quigley, T. Aung, and C.-Y. Cheng, "Global prevalence of glaucoma and projections of glaucoma burden through 2040: a systematic review and meta-analysis," *Ophthalmology*, vol. 121, no. 11, pp. 2081–2090, 2014.
- [13] L. Li, M. Xu, H. Liu et al., "A large-scale database and a CNN model for attention-based glaucoma detection," *IEEE Transactions on Medical Imaging*, vol. 39, no. 2, pp. 413–424, 2020.
- [14] H. Fu, J. Cheng, Y. Xu, D. W. K. Wong, J. Liu, and X. Cao, "Joint optic disc and cup segmentation based on multi-label deep network and polar transformation," *IEEE Transactions on Medical Imaging*, vol. 37, no. 7, pp. 1597–1605, 2018.
- [15] R. Zhao, X. Chen, and X. Liu, "Direct cup-to-disc ratio estimation for glaucoma screening via semi-supervised learning," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 4, pp. 1104–1113, 2020.
- [16] Y. GAO, W. U. Chengdong, Y. U. Xiaosheng, W. ZHOU, and W. U. Jiahui, "Full-automatic optic disc boundary extraction based on active contour model with multiple energies," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. E101.A, no. 3, pp. 658–661, 2018.
- [17] J. Guo, G. Azzopardi, C. Shi, N. M. Jansonius, and N. Petkov, "Automatic determination of vertical cup-to-disc ratio in retinal fundus images for glaucoma screening," *IEEE Access*, vol. 7, pp. 8527–8541, 2019.
- [18] A. Aquino, M. E. Gegúndez-Arias, and D. Marín, "Detecting the optic disc boundary in digital fundus images using morphological, edge detection, and feature extraction techniques," *IEEE Transactions on Medical Imaging*, vol. 29, no. 11, pp. 1860–1869, 2010.
- [19] B. Dashtbozorg, A. M. Mendonça, and A. Campilho, "Optic disc segmentation using the sliding band filter," *Computers in Biology and Medicine*, vol. 56, pp. 1–12, 2015.
- [20] A. Chakravarty and J. Sivaswamy, "Joint optic disc and cup boundary extraction from monocular fundus images," *Computer Methods and Programs in Biomedicine*, vol. 147, pp. 51–61, 2017.
- [21] Y. Zheng, D. Stambolian, and J. O'Brien, *Optic Disc and Cup Segmentation from Color Fundus Photograph Using Graph Cut with Priors*, in *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, Berlin, Heidelberg, 2013.
- [22] J. Zilly, J. M. Buhmann, and D. Mahapatra, "Glaucoma detection using entropy sampling and ensemble learning for automatic optic cup and disc segmentation," *Computerized Medical Imaging and Graphics*, vol. 55, pp. 28–41, 2017.
- [23] K. K. Maninis, J. Pont-Tuset, and P. Arbeláez, "Deep retinal image understanding," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 140–148, Springer, Cham, 2016.
- [24] A. Sevastopolsky, "Optic disc and cup segmentation methods for glaucoma detection with modification of U-Net

- convolutional neural network,” *Pattern Recognition and Image Analysis*, vol. 27, no. 3, pp. 618–624, 2017.
- [25] S. M. Shankaranarayana, K. Ram, and K. Mitra, “Joint optic disc and cup segmentation using fully convolutional and adversarial networks,” in *Fetal, Infant and Ophthalmic Medical Image Analysis*, pp. 168–176, Springer, Cham, 2017.
- [26] W. Zhou, C. Wu, D. Chen, Y. Yi, and W. Du, “Automatic microaneurysm detection using the sparse principal component analysis-based unsupervised classification method,” *IEEE Access*, vol. 5, pp. 2563–2572, 2017.
- [27] A. Sarhan, J. Rokne, and R. Alhaji, “Glaucoma detection using image processing techniques: a literature review,” *Computerized Medical Imaging and Graphics*, vol. 78, p. 101657, 2019.
- [28] S. M. Pizer, E. P. Amburn, J. D. Austin et al., “Adaptive histogram equalization and its variations,” *Computer vision, graphics, and image processing*, vol. 39, no. 3, pp. 355–368, 1987.
- [29] S. Maheshwari, R. B. Pachori, and U. R. Acharya, “Automated diagnosis of glaucoma using empirical wavelet transform and correntropy features extracted from fundus images,” *IEEE Journal of Biomedical and Health Informatics*, vol. 21, no. 3, pp. 803–813, 2017.
- [30] T. R. Kausu, V. P. Gopi, K. A. Wahid, W. Doma, and S. I. Niwas, “Combination of clinical and multiresolution features for glaucoma detection and its classification using fundus images,” *Biocybernetics and Biomedical Engineering*, vol. 38, no. 2, pp. 329–341, 2018.
- [31] W. Zhou, Y. Yi, J. Bao, and W. Wang, “Adaptive weighted locality-constrained sparse coding for glaucoma diagnosis,” *Medical & Biological Engineering & Computing*, vol. 57, no. 9, pp. 2055–2067, 2019.
- [32] U. R. Acharya, S. Bhat, J. E. W. Koh, S. V. Bhandary, and H. Adeli, “A novel algorithm to detect glaucoma risk using texture and local configuration pattern features extracted from fundus images,” *Computers in Biology and Medicine*, vol. 88, pp. 72–83, 2017.
- [33] R. Bock, J. Meier, and G. Michelson, *Classifying Glaucoma with Image-Based Features from Fundus Photographs*, in *Proceedings of the Joint Pattern Recognition Symposium*, Springer, Berlin, Heidelberg, 2007.
- [34] J. Nayak, U. Rajendra Acharya, P. S. Bhat, N. Shetty, and T.-C. Lim, “Automated diagnosis of glaucoma using digital fundus images,” *Journal of Medical Systems*, vol. 33, no. 5, pp. 337–346, 2009.
- [35] W. Zhou, H. Wu, C. Wu, X. Yu, and Y. Yi, “Automatic optic disc detection in color retinal images by local feature spectrum analysis,” *Computational and Mathematical Methods in Medicine*, vol. 2018, 12 pages, 2018.
- [36] W. Zhou and S. Qiao, “Automatic optic disc detection using low-rank representation based semi-supervised extreme learning machine,” *Automatic optic disc detection using low-rank representation based semi-supervised extreme learning machine*, in *Proceedings of the International Journal of Machine Learning and Cybernetics*, vol. 11, no. 1, pp. 55–69, 2020.
- [37] K. Zhou, S. Gao, and J. Cheng, “Sparse-Gan: sparsity-constrained generative adversarial network for anomaly detection in retinal oct image,” in *Proceedings of the 17th International Symposium on Biomedical Imaging (ISBI)*, pp. 1227–1231, IEEE, 2020.
- [38] V. G. Edupuganti, A. Chawla, and A. Kale, “Automatic optic disk and cup segmentation of fundus images using deep learning,” in *Proceedings of the 25th IEEE International Conference on Image Processing (ICIP)*, pp. 2227–2231, IEEE, 2018.
- [39] S. Yu, S. F. Di Xiao, and Y. Kanagasigam, “Robust optic disc and cup segmentation with deep learning for glaucoma detection,” *Computerized Medical Imaging and Graphics*, vol. 74, pp. 61–71, 2019.
- [40] S. Wang, L. Yu, X. Yang, C.-W. Fu, and P.-A. Heng, “Patch-based output space adversarial learning for joint optic disc and cup segmentation,” *IEEE Transactions on Medical Imaging*, vol. 38, no. 11, pp. 2485–2495, 2019.
- [41] H.-K. Hsiao, C.-C. Liu, C.-Y. Yu, S.-W. Kuo, and S.-S. Yu, “A novel optic disc detection scheme on retinal images,” *Expert Systems with Applications*, vol. 39, no. 12, pp. 10600–10606, 2012.
- [42] E. F. Kao, P. C. Lin, and M. C. Chou, “Automated detection of fovea in fundus images based on vessel-free zone and adaptive Gaussian template,” *Computer Methods and Programs in Biomedicine*, vol. 117, no. 2, pp. 92–103, 2014.
- [43] P. K. Chaudhary and R. B. Pachori, “Automatic diagnosis of glaucoma using two-dimensional Fourier-Bessel series expansion based empirical wavelet transform,” *Biomedical Signal Processing and Control*, vol. 64, p. 102237, 2021.
- [44] J. Gilles, G. Tran, and S. Osher, “2D empirical transforms. Wavelets, ridgelets, and curvelets revisited,” *SIAM Journal on Imaging Sciences*, vol. 7, no. 1, pp. 157–186, 2014.
- [45] R. M. Haralick, K. Shanmugam, and I. H. Dinstein, “Textural features for image classification,” *IEEE Transactions on systems, man, and cybernetics*, vol. SMC-3, no. 6, pp. 610–621, 1973.
- [46] M. Mercimek and K. Gulez, “Real object recognition using moment invariants,” *sadhana*, vol. 30, no. 6, pp. 765–775, 2005.
- [47] S. A. Valle, W. Li, and S. J. Qin, “Selection of the number of principal components: the variance of the reconstruction error criterion with a comparison to other methods,” *Industrial & Engineering Chemistry Research*, vol. 38, no. 11, pp. 4389–4401, 1999.
- [48] I. T. Jolliffe and T. M. Uddin, “A modified principal component technique based on the LASSO,” *Journal of Computational and Graphical Statistics*, vol. 12, no. 3, pp. 531–547, 2003.
- [49] B. Efron, T. Hastie, and J. R. Tibshirani, “Least angle regression,” *Annals of Statistics*, vol. 32, no. 2, pp. 407–451, 2004.
- [50] J. I. Orlando, H. Fu, J. B. Breda et al., “Refuge challenge: a unified framework for evaluating automated methods for glaucoma assessment from fundus photographs,” *Medical Image Analysis*, vol. 59, p. 101570, 2020.
- [51] F. Fumero, S. Alayon, J. L. Sanchez, J. Sigut, and M. Gonzalez-Hernandez, “RIM-ONE: an open retinal image database for optic nerve evaluation,” in *Proceedings of the 24th International Symposium on Computer-Based Medical Systems (CBMS)*, pp. 1–6, IEEE, Bristol, UK, 2011.
- [52] S. Dua, U. R. Acharya, and P. Chowriappa, “Wavelet-based energy features for glaucomatous image classification,” *IEEE Transactions on Information Technology in Biomedicine*, vol. 16, no. 1, pp. 80–87, 2012.
- [53] J. Cheng, J. Liu, Y. Xu et al., “Superpixel classification based optic disc and optic cup segmentation for glaucoma screening,” *IEEE Transactions on Medical Imaging*, vol. 32, no. 6, pp. 1019–1032, 2013.

- [54] E. Santos, R. Veras, and M. Frazao, "A semiautomatic super-pixel based approach to cup-to-disc ratio measurement," in *in: Proceedings of the 2018 IEEE Symposium on Computers and Communications (ISCC)*, pp. 00621–00626, IEEE, 2018.
- [55] B. Liu, D. Pan, and H. Song, "Joint optic disc and cup segmentation based on densely connected depthwise separable convolution deep network," *BMC Medical Imaging*, vol. 21, no. 1, pp. 1–12, 2021.

Research Article

EAWNNet: An Edge Attention-Wise Objector for Real-Time Visual Internet of Things

Zhichao Zhang ¹, Hui Chen,² Xiaoqing Yin ³, and Jinsheng Deng³

¹College of Computer, National University of Defense Technology, Changsha 410000, China

²Science and Technology on Integrated Logistics Support Laboratory, National University of Defense Technology, Changsha 410000, China

³College of Advanced Interdisciplinary Studies, National University of Defense Technology, Changsha 410000, China

Correspondence should be addressed to Xiaoqing Yin; yinxiaoqing89@163.com

Received 22 April 2021; Accepted 9 June 2021; Published 12 July 2021

Academic Editor: Fa Zhu

Copyright © 2021 Zhichao Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the upgrading of the high-performance image processing platform and visual internet of things sensors, VIOT is widely used in intelligent transportation, autopilot, military reconnaissance, public safety, and other fields. However, the outdoor visual internet of things system is very sensitive to the weather and unbalanced scale of latent object. The performance of supervised learning is often limited by the disturbance of abnormal data. It is difficult to collect all classes from limited historical instances. Therefore, in terms of the anomaly detection images, fast and accurate artificial intelligence-based object detection technology has become a research hot spot in the field of intelligent vision internet of things. To this end, we propose an efficient and accurate deep learning framework for real-time and dense object detection in VIOT named the Edge Attention-wise Convolutional Neural Network (EAWNNet) with three main features. First, it can identify remote aerial and daily scenery objects fast and accurately in terms of an unbalanced category. Second, edge prior and rotated anchor are adopted to enhance the efficiency of detection in edge computing internet. Third, our EAWNNet network uses an edge sensing object structure, makes full use of an attention mechanism to dynamically screen different kinds of objects, and performs target recognition on multiple scales. The edge recovery effect and target detection performance for long-distance aerial objects were significantly improved. We explore the efficiency of various architectures and fine tune the training process using various backbone and data enhancement strategies to increase the variety of the training data and overcome the size limitation of input images. Extensive experiments and comprehensive evaluation on COCO and large-scale DOTA datasets proved the effectiveness of this framework that achieved the most advanced performance in real-time VIOT object detection.

1. Introduction

Intelligent vision internet of things (VIOT) uses all kinds of image sensors, including surveillance cameras, mobile phones, and digital cameras, to obtain people, cars, objects, images, or video data; extract visual tags; and use intelligent analysis technology to process information, so as to provide support for follow-up applications as shown in Figure 1. The intelligent visual internet of things can directly, vividly, and efficiently reflect the monitoring data of the observed object and output the results of intelligent analysis. Therefore, VIOT is widely used in important places such as social

public safety, intelligent vehicles, parking lots, community monitoring, land and sea traffic monitoring, urban security, and military reconnaissance. However, the performances of supervised learning are often limited by the disturbance of abnormal data and unbalanced scale of latent objects, which impair the automatic inference speed and recognition accuracy.

The intelligent visual internet of things can provide information assistance for public security departments, such as real-time monitoring, suspect tracking, and crime early warning. At the same time, it can also provide a large number of real-time traffic information for traffic management

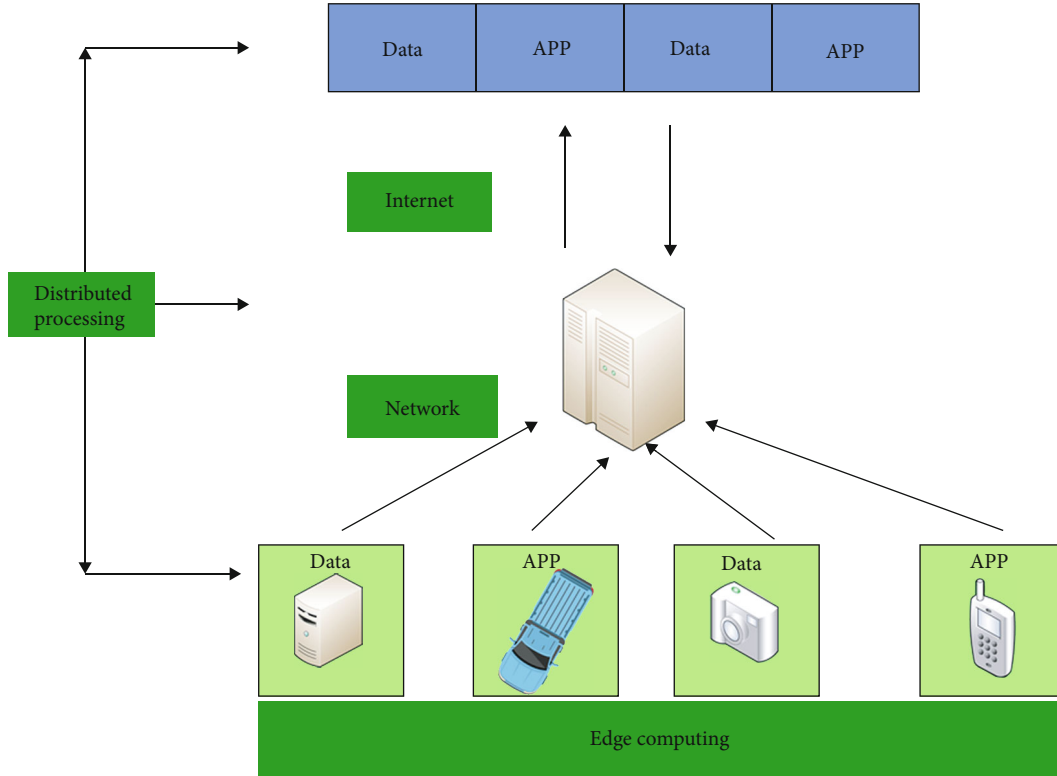


FIGURE 1: The whole object detection process of intelligent vision internet of things (VIOT).

departments to facilitate their traffic supervision. There is no doubt that the emergence of the internet of things in intelligent vision has brought great convenience to people's lives. The intelligent visual internet of things system needs to extract accurate image features in the application.

However, the object detection task for an outdoor internet of things vision system is very sensitive to weather conditions, especially in frequent and widely distributed blurry scenes. In addition, when vehicles and people move quickly, the captured images may be blurry. Out-of-focus cameras can also lead to a decrease in detection accuracy. Finding specific key information from traffic surveillance, astronomical remote sensing, public security investigation, and other applications therefore remains a significant challenge. Because of the lack of information, these low-quality images seriously affect the effectiveness of the intelligent visual internet of things system.

In order to conquer these challenges, we need more advanced object detectors. Advanced object detection methods have greatly improved over the past few years, and several methods have been introduced to optimize the network structure, which can be divided into single-stage and double-stage; however, the use of an attention module to improve the efficiency of searching is not well investigated. They are divided into two mainstreams: two-stage detector and single-stage detector. Two-stage detector: the R-CNN [1] directly performs the selective search [2] and classifies objects using a CNN. Compared with the traditional methods, the use of the R-CNN significantly improved the accuracy of classification, marking the beginning of the era

of target detection using deep learning. In its variants (for example, fast R-CNN [3]), the two-phase framework was updated, which helps them achieve even better performance. In addition, to further improve the accuracy, some highly effective extensions were proposed, such as R-FCN [2] and mask R-CNN [4]. Single-stage detector: The most representative single-stage detectors are YOLO [5, 6] and SSD [7]. They use feature maps to predict the confidence and location of a multitarget receptive field block (RFB) network to achieve accurate and fast object detection. Both these detectors use lightweight backbones for acceleration; however, their accuracy clearly lags behind the top two-stage approach. Recently, more advanced single-stage detectors (such as DSSD [8] and RetinaNet [9]) have updated their original lightweight backbone using a deeper ResNet-101 and by applying some techniques, such as deconvolution [8] or caustics [9]. Their results are comparable to or even better than most advanced two-stage methods. However, these detectors could not achieve the balance of speed and accuracy simultaneously.

In view of the abovementioned limitations, we propose to design a generative edge detection method to perceive the object structure, make full use of attention mechanism to dynamically screen different kinds of objects, and perform object detection on multiple scales. By doing so, not only does the edge recovery effect become better, but also the object detection performance for long-distance aerial objects is significantly improved.

We design a lightweight single-stage attention rotation intelligent object detection network for wireless internet of

things. Inspired by the previously proposed methods, our framework, using an edge attention-wise conventional neural network, EAWNet, and based on attention mechanism, has the following advantages:

- (i) Edge prior depiction greatly reduces the amount of attention-aware computation. We propose an edge attention-wise network beneficial for extracting features effectively and reducing the ground truth position shrinkage. The framework performs well in terms of accuracy (state-of-the-art stability for multiclass) and speed (real-time video recognition). Edge prior reconstruction and attention-wise modules are embedded into the EAWNet, which helps in performing efficient latent search and localization
- (ii) With the combination of intelligent connection and residual link, rotating bounding box, and synthesis loss function, the visual loss of intensive detection is reduced to a minimum. Pass-wise connection follows a straight way to pass the initial patch information to different last stage fusion layers to restore the recognition fusion. It propagates semantically strong features and enhances all features with a reasonable classification capability. In addition, residual connections for local convolutional layers and pass-wise connections for global feature dataflow are designed to modify the architecture for faster and lighter inference
- (iii) A dual parallel attention module is used to improve the efficiency of multiscale object detection. Attention-wise modules including context attention-wise module (CAW) and position refinement-wise module (PRW) are designed to reduce computing cost and improve effectiveness. These modules can match the right object and position instead of searching the entire background. A rotating bounding box, designed for aerial image object detection, proved to be beneficial for the recognition of dense and tiny objects

2. Related Work

2.1. Object Detection. In the wave of artificial intelligence sweeping the world, the intelligent visual internet of things is expected to achieve a significant social and economic promotion of the internet of things. It has become a typical successful representative of the application of the internet of things. Object detection methods are mainly based on CNNs; one-stage object detectors play a remarkable role in object detection. Most existing VIOT object detectors are classified according to whether they have suggested steps for regions of interest (two-stage [3, 4, 10]) or not (one-stage [6, 7, 11]). Although two-stage detectors are more flexible and accurate, single-stage detectors are generally considered faster and more efficient by using pretrained anchors [12]. Single-stage detectors have attracted wide attention because of their high efficiency and simplicity. Here, we mainly follow the design of single-stage detectors and prove that higher effi-

ciency and higher accuracy can be achieved by optimizing the network structure.

A recent single-stage detector [7, 13] was designed to match the accuracy of more complex two-stage detection methods. Although these detectors show impressive results on large- and medium-sized objects, their performance on small objects is lower than expected [14]. (The size of an object is related to the pixels it occupies in the picture.) When using the most advanced single-stage RetinaNet [13], it achieves unbalanced results with a COCO AP-large of 47 but only 14 for AP-small objects (as defined in [15]). Small object detection is a challenging problem, which requires not only low intermediate information to accurately describe it but also high-level semantics to distinguish the target from the others or background.

There are five types of YOLO from YOLOv1 to YOLOv5 [16, 17]. Based on YOLOv1 to YOLOv3 [18], researchers propose an efficient and powerful object detection model called YOLOv4 [19]. A variety of modules are mixed in the YOLOv4, such as Weighted Residual Connections (WRC), Cross Stage Partial (CSP) connections, Cross Minibatch Normalization (CmBN), Self-Adversarial Training (SAT), Mish-activation, Mosaic data augmentation, CmBN, DropBlock regularization, and CIOU loss. The introduction of these modules increased the calculation time yet greatly improved the accuracy. While YOLOv5 shows the fastest speed among the series of YOLO algorithms and is comparable to YOLOv4 in terms of accuracy, YOLOv5 is remarkably lightweight.

The following methods are the most classic deep learning object detection methods from different schools in the past two years. RFBNet [20] simply employs dilated convolutions to enlarge the receptive field and achieves good vision of the extracted features; although it could be learnable for image recognition, its performance is not satisfactory. LRF [15] is a lightweight scratch network (LSN) that is trained from scratch taking a downsampled image as input and passing it through a few convolutional layers to efficiently construct low- and middle-level features. However, learning from both scratch and pretrained sacrifices too much time efficiency, and the network is so complex that the inference speed is slower than in one-stage methods. CenterMask [21] combines instance segmentation and object detection into one task, and it goes even further by using the instance segmentation to achieve the recognition of class and position that object detection demands. EfficientDet [22] balances the network depth for speed and feature flow connection strategy for accuracy, ignoring the attention mechanism to enhance the performance without occupying large resources.

Considering the mentioned advantages and weaknesses, we propose an attention-wise YOLO, which handles the feature extraction flow in a reasonable way by designing a multi-path refinement framework. In addition, pass-wise connection meets the demands in terms of balancing time efficiency and prediction accuracy. To learn semantic information wisely, attention-wise modules are introduced between the backbone and the neck.

2.2. Attention Module. The main focus of attention models in computer vision is to focus on interesting things and ignore

irrelevant information. Recently, attention models have been classified into three groups: hard attention [23] and soft attention [16], global attention [16, 17] and local attention [24, 25], and self-attention [16]. Hard attention models have been widely used for a long model without preprocessing. The computational cost of local attention is lower than that of global attention because it does not need to consider the hidden layer state of all encoders. The self-attention mechanism improves the attention model, which reduces the dependence on external information and is capable of capturing internal correlation of data features. As self-attention shows good performance, it is widely used on computer vision tasks.

The attention model mechanism is important in deep learning methods. The first to propose the self-attention mechanism were Vaswani et al. [25]. It relies on global dependencies between inputs and outputs and was applied in machine translation. In computer vision area, attention modules have been also adopted. Zhang et al. [26] created an image generator that leverages adjacent regions to object shapes rather than local regions of fixed shape for image generation by the self-attention mechanism. An adapted attention module for object detection that uses the relative geometry between objects was proposed [27]. There is a successful application in space-time dimension for videos with nonlocal operation [28]. Fu et al. designed DANet [29] based on the newly Fully Convolutional Networks (FCNs) [30] with position attention mechanism (PAM) and channel attention mechanism (CAM). DANet settles the problems of object detection in some confusing categories and objects with different appearance. In addition, Fu et al. [31] proposed a DRANet which makes an improvement in self-attention modules based on DANet. DRANet adopts the compact PAM (CPAM) and the compact CAM (CCAM), reducing computational complexity.

2.3. Detection Bounding Boxes. Current object detection algorithms may not perform good results on detecting oriented targets [32]. State-of-the-art object detection methods rely on rectangular-shaped, horizontal/vertical bounding boxes drawn over an object to accurately localize its position. Such orthogonal bounding boxes ignore object pose, resulting in reduced object localization, and limiting downstream tasks such as object understanding and tracking. Rotated faster R-CNN [33] based on the faster R-CNN [34] adds a regression branch to predict the oriented bounding boxes for aerial images. It could improve the performance on tiny things in high resolution by introducing balanced FPN. R4Det [35] is an end-to-end detector which could address the problems of images with large aspect ratios, dense distributions, and extremely imbalanced categories. Moreover, from experimental results, we could see that the detector shows strong robustness against adversarial attacks.

3. Implementation

3.1. Network Architecture. The whole network architecture is shown in Figure 2. We propose a heavily reconstructed Edge Attention-wise Convolutional Neural Network (EAWNNet). It

employs the multipath refinement flow network (MRFNet) [36] as the backbone, which makes it easier to extract multi-level scale features from patches. We consider not only the efficiency between the backbone and neck through MRFNet but also the fusion effect of extracting features aided by a pass-wise connection (PWC) strategy. Furthermore, this framework is learnable for multiclass and multiscale objects because we design various attention-wise modules to make the training and validation more reasonable and extract features in a global perspective. Also, we modify the above model by adding rotating bounding boxes and design a synthesis loss function to constrain the training process and to boost the training convergence using the multipath refinement flow network (MRFNet).

The MRFNet architecture is shown in Figure 2(b). It combines each convolution layer and dataflow branches. Specifically, there are two path channels $a_0 = [a_0', a_0'']$. Every stage has a downsampled fusion layer, $[a_0', a_1, \dots, a_k]$, which will be downsampled to lower dimensions and larger output numbers. Then, the output of this refinement transfer results in a_τ , which will be concatenated with a_0' and undergo another transition layer to finally generate the output a_U . Equations (1) and (2) are the feed-forward pass and weight update of MRFNet, respectively. w_k , w_τ , and w_U are the weights of ground truth patches g_0' , g_k , and g_0'' , respectively. f_k , f_T , and f_U are the transformation function of downsampled layers, transfer results, and transition layer outputs, respectively.

$$\begin{aligned} a_k &= w_k * [a_0', a_1, \dots, a_{k-1}], \\ a_\tau &= w_\tau * [a_0'', a_1, \dots, a_k], \\ a_U &= w_U * [a_0', a_\tau], \end{aligned} \quad (1)$$

$$\begin{aligned} w_k' &= f_k(w_k, \{g_0'', g_1, \dots, g_{k-1}\}), \\ w_T' &= f_T(w_T, \{g_0'', g_1, \dots, g_k\}), \\ w_U' &= f_U(w_U, \{g_0', g_T\}). \end{aligned} \quad (2)$$

We compared our architecture with mainstream CNN architectures (ResNeXt, ResNet, and DenseNet). The results are usually a linear and nonlinear combination of the outputs of intermediate layers. Thus, the output of a k -layer CNN is

$$y = F(a_0) = a_k = H_k(a_{k-1}, H_{k-1}(a_{k-2}), H_{k-2}(a_{k-3}), \dots, H_1(a_0), a_0), \quad (3)$$

where the whole model of the convolutional neural network is denoted by F , the mapping function from a_0 to y , and H_k is the k -th layer of CNN. Generally, a set of convolutional layers and a nonlinear activate function consist of the H_k . As for ResNet and DenseNet, we can also formulate their models into the following:

$$a_k = R_k(a_{k-1}) + a_{k-1} = R_k(a_{k-1}) + R_{k-1}(a_{k-2}) + \dots + R_1(a_0) + a_0, \quad (4)$$

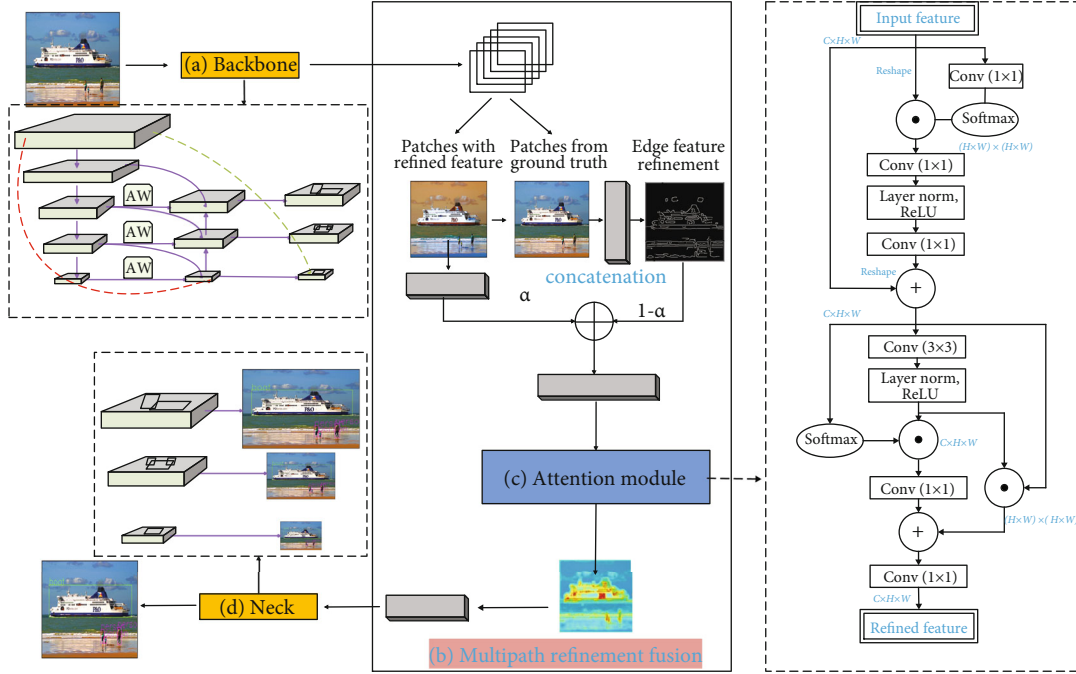


FIGURE 2: Overall network architecture of EAWNet comprising four parts: (a) backbone architecture, (b) multipath refinement fusion unit, (c) attention modules, and (d) neck consisting of the network. In addition, inside the dotted lines is the explicit implementation of the counterparts. (a) The backbone is responsible for the multiscale feature extraction and is optimized by pass-wise connection to avoid gradient disappearance. (b) The multipath refinement fusion (MRF) unit is responsible for making fusion from the edge prior which is extracted from the ground truth and refined patches. (c) The attention modules learn information of category and structure wisely quickly aided by the edge prior. The position-wise and channel-wise attention modules (in the dotted lines) consist of the attention modules in a parallel manner. (d) The neck is the decoder for object detection which is modified into rotated bounding boxes for better visual effects. All four parts are illustrated in the following sections.

$$a_k = [C_k(a_{k-1}), a_{k-1}] = [C_k(a_{k-1}), C_{k-1}(a_{k-2}), \dots, C_1(a_0), a_0], \quad (5)$$

where R and C represent the operational computation of the residual layers and convolutional layers, respectively; both are reproduced by two or three convolutional blocks.

From the above equations, it follows that the inputs of convolutional layers originated from the previous convolutional outputs. Under this circumstance, the gradient flow could be propagated more efficiently due to the minimum path length of the gradient. However, this design would result in the reverse propagation into all layers from $k-1$ to 1, which is redundant for a repeated training process. Figure 3 illustrates the EAWNet reusing the initial features and simultaneously preventing iteratively propagating gradient information by cutting down the gradient flow. The insightful vision of the design is to separate gradient flow and refinement features and fuse the last convolutional layers, which enhances feature extraction efficiency.

The specific multipath refinement flow network exhibits the advantage of multiscale feature extraction as RefineNet [36] and CSPDarkNet-53 [19], deeply modified by introducing lightweight and residual strategy, pass-wise connection, and attention modules to enhance speed and accuracy.

3.2. Pass-Wise Connection. In this section, we design two extra paths to pass features to the next extraction stage:

pass-wise and residual connections. The main purpose of pass-wise and residual connections is to learn robust features and train deeper networks. They can address gradient vanishing problems and enhance the capabilities of locating positions and propagating strong responses of low-level patterns.

It is based on the fact that high-level features responding to edges or instance parts are a strong indicator to accurately localize instances. To this end, regardless of the complicated multipath refinement dataflow, we additionally add two direct connections to pass feature maps. As depicted in Figure 3, one line in red directly passes the first layer patches of the backbone to the last layer of the neck. Another line in green directly passes the first convolutional results to the last layer of data augmentation. The data augmentation layer and the neck layer extract features in a parallel way, sacrificing memory usage to enhance the accuracy and benefit feature extraction.

3.3. Attention-Wise Module. At present, of the multimodel method can be used for semantic segmentation and object detection [37], but the multimodel method leads to too much training cost.

We keenly find that when the edge generation method is used to assist the attention module to perceive the object structure, the object category is guided and searched by dynamically adjusting the receptive field of the recognition

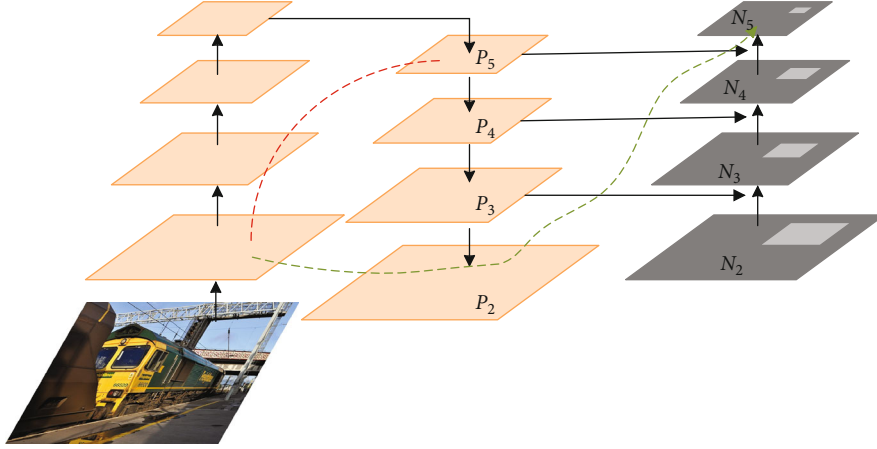


FIGURE 3: Pass-wise connection benefits the feature extraction and fusion.

frame, and then, the target is detected. This not only improves the performance of object detection but also overcomes the performance loss of semantic segmentation guiding the attention module. Therefore, we propose the EAWNet network structure, which perceives the object structure in a lightweight way through edge description and uses the attention module to improve the search efficiency. Finally, with the help of the multiscale network structure for object detection, it achieves the purpose of improving the training convergence detection performance.

We use a multipath refinement fusion (MRF) unit to fuse the information from the edge prior that is extracted from the ground truth and refined patches. Then, attention modules learn category and structure information and are quickly aided by the edge prior. The position-wise and channel-wise attention modules (depicted in the dotted lines of Figure 2(c)) consist of the attention modules in a parallel configuration.

3.4. Edge Prior and Attention Mechanism. Attention plays an important role in human visual recognition [30, 38, 39]. An important feature of the human visual system is that people do not try to deal with the whole scene at once. Instead, to better capture the visual structure, humans use a series of local glimpses and selectively focus on significant parts [40]. We propose a residual attention network using encoding and decoding attention modules. By improving the feature mapping, the network not only has good performance but also has strong robustness to noise input. Instead of calculating attention scores directly, we decompose the process into a learning channel and position attention information. The individual attention score generation process of the feature map is less than that of [37]; thus, it can be regarded as a plug-and-play module for the existing basic convolutional neural networks. Hu et al. [41] introduced a compact module to take advantage of the relationship between channels. In their squash and trigger modules, they used the global average set feature to calculate channel attention. We find that these are suboptimal features, and we recommend using the maximum set feature. They also missed out on spatial attention, which plays an important role in determining the “location” of focus, as shown in [42]. Here, we utilize both channel

and position attention based on an effective architecture and verify that using these two kinds of attention is better than using channel attention only [41]. Experiments show that the model is effective in detection tasks (MS-COCO and DOTA). We only need to place our module on the existing single detector [33] in the DOTA test set to achieve the most advanced performance.

Among the edge detection methods, Canny usually has the best edge restoration effect on small local objects, holistically nested edge detection (HED) [16] has the best edge restoration effect on the whole contour, and the edge of generative edge restoration training is often slightly intermittent, but the overall complex structure is the best, so we choose generative edge restoration for joint training.

On the one hand, one-stage detectors aim to handle images in a lightweight manner, making instance recognition fast and easy. On the other hand, one vital property of a visual CNN system is that it does not attempt to handle the whole scene at once. We propose to resolve this contradiction by adding the attention module between the backbone and the neck. The whole network MRF belongs in the one-stage method, while also using the attention-wise (AW) modules to preprocess the feature patches and assessing the contextual and position information several times in a multiscale manner. The design of contextual attention-wise unit and position refinement-wise unit is depicted in Figure 2(c).

Local features generated by traditional convolutional layers would result in misrecognition of specific things. It could also lead to a high computation cost searching for specific objects from the background. To model rich contextual relationships over local features, we continue to analyze the refinement features from the context attention-wise unit and pass the patches to the position refinement-wise unit to figure out the coherency transformation to the latent position. The position refinement-wise unit enhances their representation capability by encoding a wider range of channel and spatial information into local features,

A target associated with background scenarios could be grasped by the contextual attention module. In the meantime, object positions are located by a position refinement-wise module. Specifically, the attention results could be seen

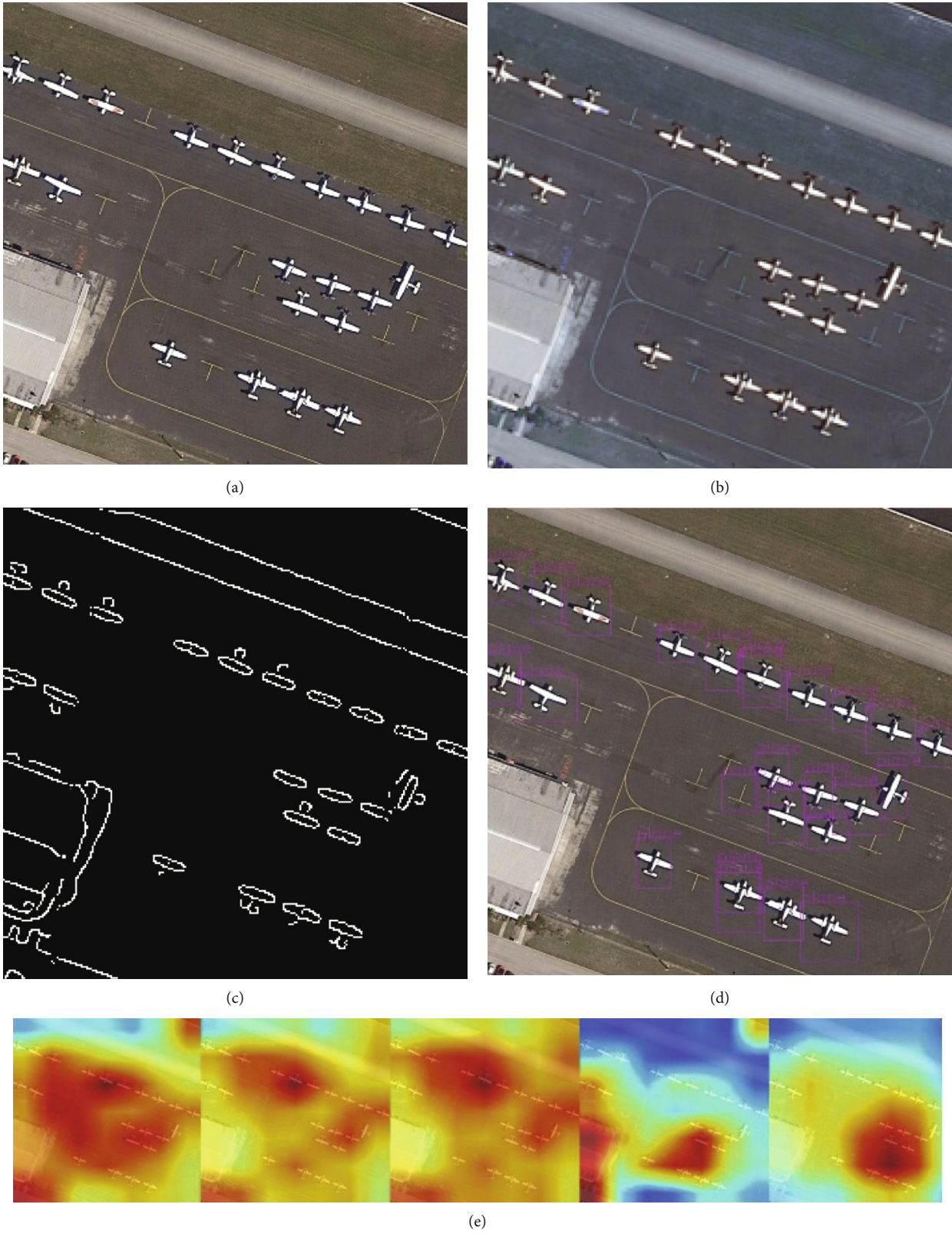
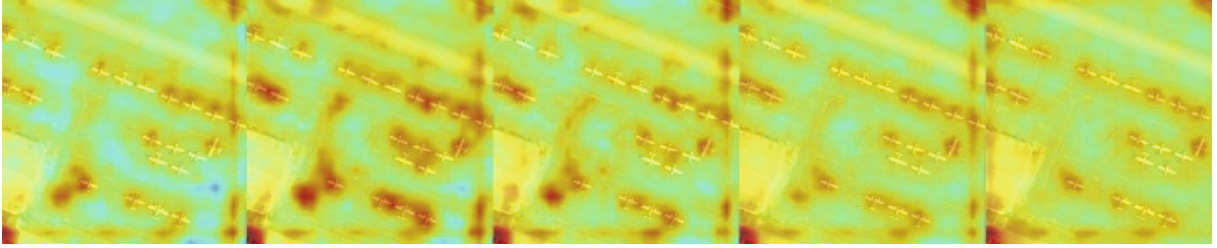
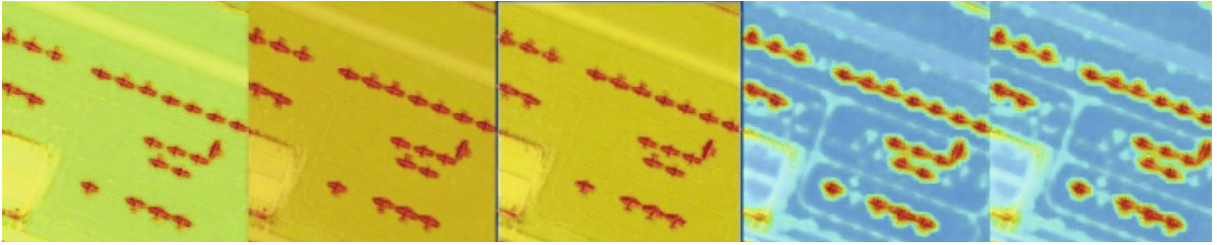


FIGURE 4: Continued.



(f)



(g)

FIGURE 4: Multifeature extraction for edge and sharpness.

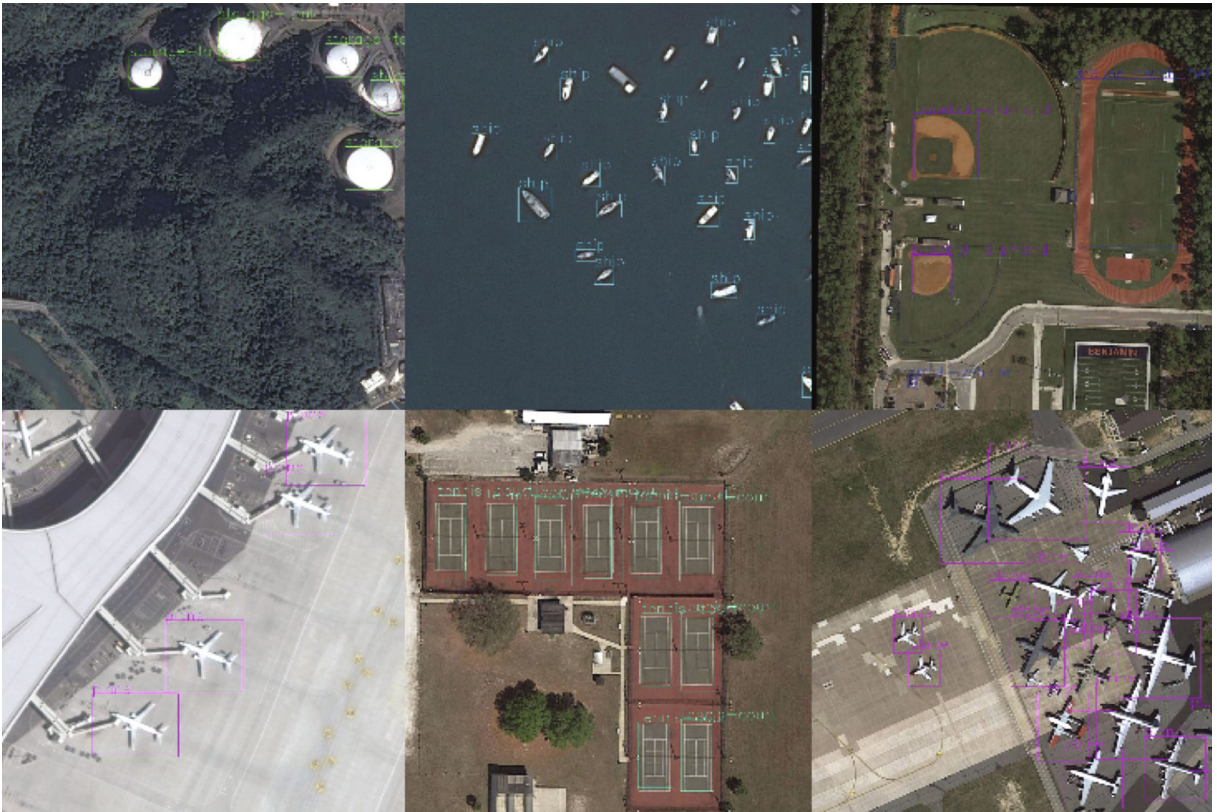


FIGURE 5: Our validation visual results on DOTA using EAWNet backbones.

in Figures 4(e), 4(f), and 4(g). First, the multipath refinement patches flow into the context attention-wise module, which is aimed at calculating the coherency between the bounding box and the background. This unit simplifies the channel, height, and width ($C \times H \times W$) patches into a softmax function and then combines the copies with matrix multiplica-

tion. We apply a softmax layer to obtain the context attention map m_{ij} :

$$m_{ij} = \frac{\exp(P_i \cdot P_j)}{\sum_{i=1}^C \exp(P_i \cdot P_j)}. \quad (6)$$

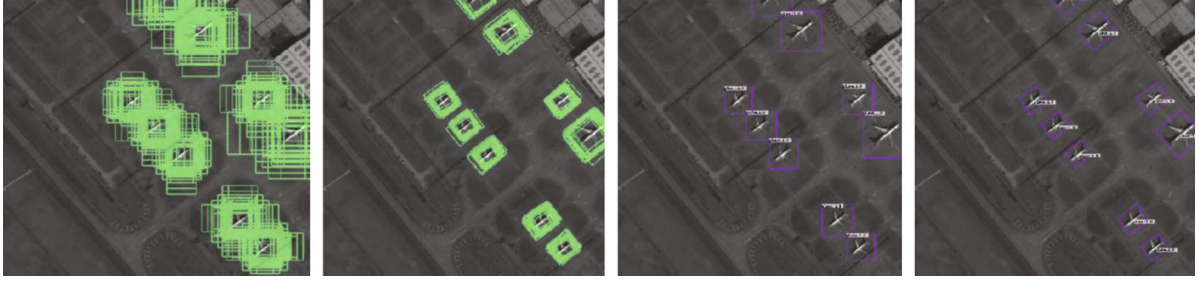


FIGURE 6: Visual comparison of general and rotated training processes.

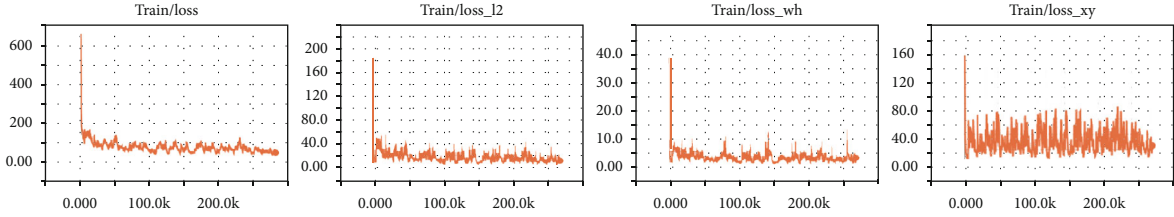


FIGURE 7: Different training loss function strategies are adopted in the training process which is beneficial for convergence speed.

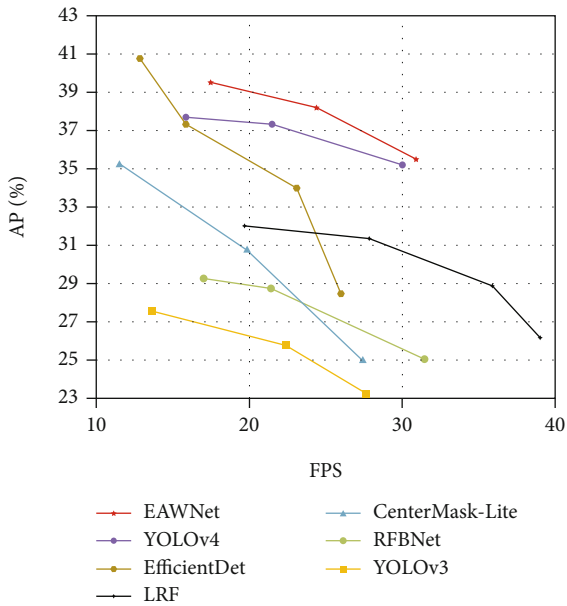


FIGURE 8: Our model EAWNet shows excellent performance in balancing average precision accuracy and frames per second speed compared to the SOTA methods.

The original patches P_i and reshaped branches P_j are aggregated into an average. In addition, we multiply the result by a scalable item α and add m_{ij} to obtain the output result R_{ij} :

$$R_{ij} = \alpha \sum_{i=1}^C (x_{ij} P_i) + m_{ij}. \quad (7)$$

The output result R is separated into the width, abscissa, and ordinate: W_i, X_j, Y_i ; then, W_i and x_{ij} are reshaped and

sent to the softmax function and combined as R_{ij} :

$$R_{ij} = \frac{\exp(W_i \cdot X_j)}{\sum_{i=1}^N \exp(W_i \cdot X_j)}. \quad (8)$$

Meanwhile, the Y_i is also combined with the reshaped S_{ij} using the sum function. We multiply it by a scalar item α and do an element-wise sum operation with the features Y_i to obtain the result $S_{ij} \in R^{C \times H \times W}$ as follows, and A_j is a fixed constant:

$$S_{ij} = \alpha \sum_{i=1}^N (R_{ij} Y_i) + A_j. \quad (9)$$

According to recent studies of the single object detectors, there are three ways to obtain the features that concern us. Firstly, the network uses a softmax function to weight the importance of the latent meaningful objects obtained from the background.

Then, the algorithm uses the location and class coherency to get the attention scores. This is not a promising method because the weak supervised methods cannot achieve high enough detection precision. Secondly, the object detection and the instance segmentation are combined; however, the abundant feature extraction and training computational cost are too high. Thirdly, we combine these two branches and follow a trade-off strategy: we adopt the lightweight spatial and class feature extraction channels to recognize the latent object classes, and then, the attention features are refined by the edge information to reinforce the boundary features for further inference. In this way, we use the edge information instead of instance segmentation to avoid the iterative training process and its computational cost. Thus, the

TABLE 1: Our model is attention-wise. The LRF also uses learnable strategies and acts more lightweight and fast; however, the accuracy is much lower than that of EAWNet. CenterMask learns in an efficient way by searching from the object center and achieves balanced performance. However, EAWNet showed a more significant improvement in learning strategy (adding attention module and rotated tight bounding boxes makes a significant progress on reconstructed deep learning models) and does achieve comparable results to similar algorithms such as LRF, RFBNet, CenterMask, EfficientDet, and YOLOv3. We can conclude that EAWNet outperforms most existing methods in terms of both accuracy and speed. The percentage of average precision on category on DOTA shows our model performs well on unbalanced and anomaly data categories.

Method	PL	BD	BR	GTF	SV	LV	SH	TC	BC	ST	SBF	RA	HA
RFBNet [20]	40.57	10.21	1.68	14.12	1.32	1.43	2.19	17.22	28.57	10.34	28.26	10.11	4.12
LRF [15]	40.59	21.29	37.74	24.20	9.93	2.19	5.86	45.44	39.45	35.72	17.22	38.73	48.34
CenterMask [21]	90.60	81.97	6.57	67.08	71.12	79.66	79.16	91.81	86.26	85.42	62.91	64.77	69.12
EfficientDet [22]	90.02	82.31	47.11	72.86	72.96	78.34	80.54	91.96	85.14	85.62	57.69	62.13	65.25
YOLOv4 [19]	91.13	82.13	50.28	72.64	72.78	80.43	80.47	91.89	85.76	85.73	60.12	62.64	68.09
EAWNet	90.08	86.56	54.01	74.94	76.75	82.52	81.32	91.83	87.96	86.34	65.14	61.85	70.17

TABLE 2: Ablation studies of network architecture (size 512×512).

Model	AP (%)	AP50 (%)	AP75 (%)
MRFNet [36]	37.1	58.2	38.2
MRFNet+PWC	37.3	58.1	39.8
MRFNet+RC	37.6	58.6	41.5
MRFNet+PWC+RC	36.9	59.1	44.7
EAWNet	37.9	59.7	45.2

attention-wise modules also attain the location and coherence information with lightweight and refinement.

As shown in Figure 2(b), we can consider the training process in the feature extraction view instead of the network dataflow. Patches with convolutional refined features are combined with the edge feature patches which are extracted from the ground truth counterparts. We control the proportion by α in Equation (7) of the edge information and the background recognition feature extraction.

Then, the fusion middle results are sent to the attention-wise modules and finally make the inferences for object detection average precision.

Figures 4(b) and 4(c) show the input images and the edge feature maps and the attention heat maps as middle outputs which could benefit the EAWNet for efficient recognition. When we change the perceptive field radius for different object scales, heat map visualization shows the dense small objects. Owing to the smart design of network feature extraction and efficient searching latent attention strategy, this rotated attention-wise network EAWNet achieves fast real-time recognition speed and significant precision enhancement among the state-of-the-art methods.

3.5. Rotated Bounding Box and Loss Design. Horizontal and vertical bounding boxes are drawn over an object for accurate localization. However, for dense VIOT object detection, the anchors are close and boundaries are overlapped. Therefore, we designed rotating bounding boxes to obtain tighter and more precise detections.

We use five parameters x, y, w, h, θ to represent the location of the rotating bounding boxes. If (x, y) are the coordi-

TABLE 3: Average precision for ablation experiments of attention modules (size 512×512).

Model (with optimal setting)	AP (%)	AP50 (%)	AP75 (%)
MRFNet [36]	37.6	59.8	41.3
MRFNet+CAW	37.5	59.6	41.0
MRFNet+PRW	37.5	59.3	41.2
MRFNet+CAW+PRW	37.6	60.2	41.5

nates of the center of the latent object, wh are its width and height, and θ is the angle of rotation in polar coordinate, then t is the angle of each coordinate:

$$t_x = \frac{x - x_a}{\omega_a}, t_y = \frac{y - y_a}{h_a},$$

$$t_w = \log\left(\frac{\omega}{\omega_a}\right), t_h = \log\left(\frac{h}{h_a}\right), t_\theta = \theta - \theta_a, \quad (10)$$

$$t'_x = \frac{x' - x_a}{\omega_a}, t'_y = \frac{y' - y_a}{h_a}, t'_w = \log\left(\frac{\omega'}{\omega_a}\right), t'_h = \log\left(\frac{h'}{h_a}\right), t'_\theta = \theta' - \theta_a. \quad (11)$$

x is the anchor boxes, and x' is a prediction of bounding boxes. Thus, the loss function is expressed as

$$L = \frac{\lambda_1}{N} L_{\text{attention}} + \frac{\lambda_2}{N} L_2 + \frac{\lambda_3}{N} L_{\text{AW}} + \frac{\lambda_4}{N} L_{\text{CLS}} + \frac{\lambda_5}{N} L_{\text{Obj}} + \frac{\lambda_6}{N} L_{X,Y,W,H}, \quad (12)$$

where N denotes the number of anchors and the hyper-parameters λ_k control the trade-off setting to one by default [1, 43]. The classification loss L_{CLS} is implemented by focal loss and smooth L_2 loss. L_{Obj} is the object detection loss. L_{AW} is the attention-wise loss. We also add the xy loss and wh loss $L_{X,Y,W,H}$ for the bounding box position precision and the object loss to analyze how many objects are missing. Figure 5 shows that the model is trained to detect dense and small objects in real-time VIOT fast and accurately. Thus, the rotated bounding boxes and hybrid loss function design are

TABLE 4: Item FPS and AP of different object detectors.

Method	Backbone	Size	FPS	AP (%)	AP50 (%)	AP75 (%)	APs (%)	APm (%)	API (%)
<i>YOLOv4: optimal speed and accuracy of object detection</i> [19]									
YOLOv4	CSPDarknet-53	416	30	35.6	57.8	38.2	17.3	39.2	52.1
YOLOv4	CSPDarknet-53	512	22	37.9	60.0	41.9	19.8	41.5	49.8
YOLOv4	CSPDarknet-53	608	16	38.2	60.9	42.5	21.7	42.1	47.4
<i>Learning rich features at high speed for single-shot object detection</i> [15]									
LRF	VGG-16	300	39.0	26.8	46.7	29.4	8.3	30.3	42.6
LRF	ResNet-101	300	36.2	29.2	50.0	32.2	8.6	33.2	45.9
LRF	VGG-16	512	27.9	31.9	51.7	33.8	14.7	35.4	44.3
LRF	ResNet-101	512	19.7	32.6	53.2	35.1	15.2	38.0	45.4
<i>Receptive field block net for accurate and fast object detection</i> [20]									
RFBNet	VGG-16	300	32.0	25.3	44.5	27.1	6.9	27.2	41.3
RFBNet	VGG-16	512	22.5	29.2	50.1	31.2	11.7	32.4	42.5
RFBNet-E	VGG-16	512	17.3	29.6	50.7	31.5	12.9	31.8	42.7
<i>YOLOv3: an incremental improvement</i> [18]									
YOLOv3	Darknet-53	320	27	23.3	46.8	25.1	7.2	25.8	38.4
YOLOv3	Darknet-53	416	23	26.4	50.6	27.8	10.3	28.2	38.2
YOLOv3	Darknet-53	608	14	28.0	53.0	29.6	13.7	30.7	36.9
YOLOv3-SPP	Darknet-53	608	16	31.2	56.2	33.5	15.6	33.4	41.4
<i>CenterMask: real-time anchor-free instance segmentation</i> [21]									
CenterMask-Lite	MobileNetV2-FPN	600x	27	25.5	—	—	9.2	27.1	36.3
CenterMask-Lite	VoVNetV-19-FPN	600x	20	31.2	—	—	14.9	33.0	41.2
CenterMask-Lite	VoVNetV-39-FPN	600x	12	35.7	—	—	17.8	38.6	48.5
<i>EfficientDet: scalable and efficient object detection</i> [22]									
EfficientDet-D0	Efficient-B0	512	26	29.0	47.2	31.2	7.3	33.3	46.2
EfficientDet-D1	Efficient-B1	640	23	34.6	53.8	37.5	13.1	39.8	51.0
EfficientDet-D2	Efficient-B2	768	16	38.2	57.3	41.2	17.9	42.4	53.6
EfficientDet-D3	Efficient-B3	896	13	41.3	60.6	44.6	21.6	45.1	55.1
<i>EAWNet: an Edge Attention-wise Convolutional Neural Network for real-time object detection</i>									
EAWNet	EAWNet	416	31	36.1	59.3	39.1	18.5	40.0	53.3
EAWNet	EAWNet	512	24	38.8	60.8	42.7	20.8	42.4	50.7
EAWNet	EAWNet	608	17	39.7	62.2	43.3	23.1	42.8	48.6

FPS, AP, AP50, AP75, APs, APm, and API represent the frame per second, average precision, reach to 50% average precision, reach to 75%, average precision, and average precision for small-, medium-, and large-scale objects, respectively.

beneficial for better visual effect and training process convergence as displayed in Figures 6 and 7.

4. Experiments

We conducted comparative experiments between EAWNet, RFBNet [20], LRF [15], YOLOv3 [18], CenterMask [21], and EfficientDet [22] in FPS, AP, and visual effects.

Frames per second (FPS) are raised by 6.25%, and AP average precision (AP) is increased by 1.51%. The results obtained with other state-of-the-art object detectors are displayed in Figure 8. Our EAWNet on the red line is on the Pareto optimality curve and is the fastest and most accurate detector.

4.1. Experimental Setup. We implemented our model with PyTorch. The model was trained with Adam ($\beta_1 = 0.9$, $\beta_2 = 0.999$). A batch size of 16 was used for training in four NVIDIA RTX2080Ti GPUs with 11 GB RAM. At the beginning of each epoch, the learning rate was initialized as 10^{-4} and subsequently diminished by half every 10 epochs. We trained 100 epochs on COCO and 150 epochs on DOTA.

4.2. Dataset and Augmentation. We assessed our method on two well-known benchmarks in VIOT for city and aerial scenarios: COCO [44] and DOTA [45]. The comparative experiments were performed under equivalent conditions (training on same GPU and dataset). The image size from DOTA was 1024×1024 , while that from COCO was 256×256 . The COCO benchmark is a large-scale object detection,



FIGURE 9: Some visual result tests on EAWNet show that the attention modules benefit prediction stability and training precision.

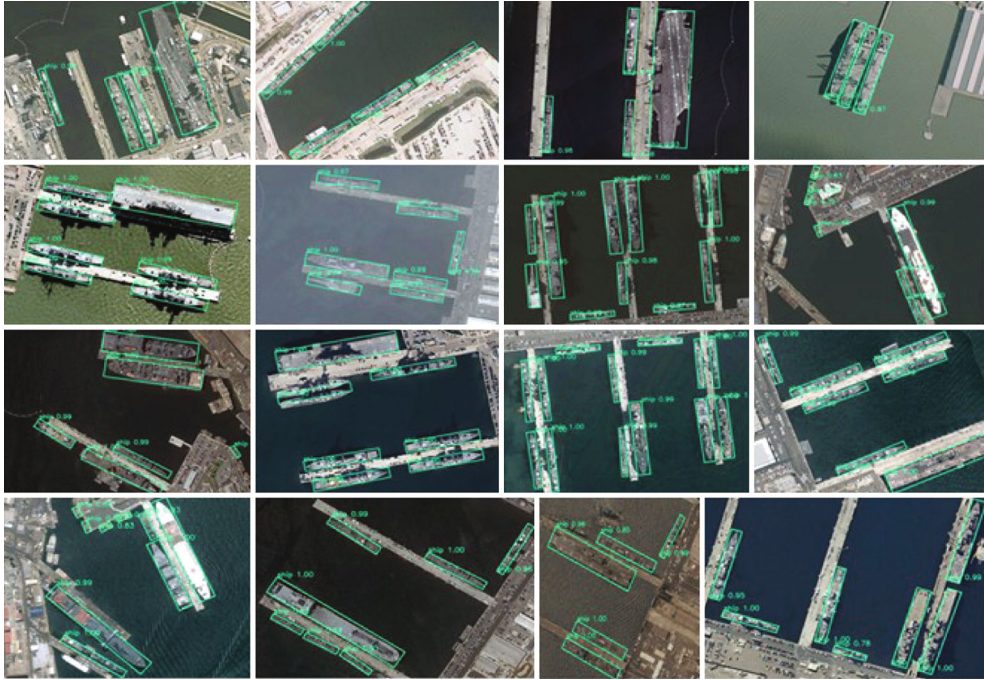


FIGURE 10: Visual results of rotated bounding boxes.

segmentation, and captioning dataset. It has 330k images, more than 200k labelled images, and 80 object categories, which is beneficial for object detection training.

The DOTA benchmark is the largest and most challenging dataset with oriented bounding box annotations for aerial image object detection. These images have been annotated by experts using 16 common object categories, and Table 1 shows that our approach has an excellent performance in terms of category balance and accuracy. The object categories include helicopter (HC), large vehicle (LV), small vehicle (SV), tennis court (TC), ground track field (GTF), basketball court (BC), soccer field (SBF), baseball diamond (BD), storage tank (ST), swimming pool (SP), and roundabout (RA).

In terms of data augmentation, images are flipped horizontally and vertically and rotated at random angles. For

color, RGB channels are replaced randomly. For image color degradation, saturation in the HSV color space is multiplied by a random number in $[0, 5]$.

We also conducted ablation experiments on COCO and DOTA by adopting different attention modules as shown in Table 2. Here, MRFNet is considered as the benchmark. PWC represents the pass-wise connection. RC represents the benchmark which adopts the residual connection techniques. EAWNet adopts the above strategies and modules to enhance the percentage of average precision (AP) and the inference speed of frames per second (FPS).

4.3. Experiment Analysis. We adopted different feature extraction methods and network structures as presented in Table 3. CAW represents the contextual attention-wise

modules and PRW represents the position refinement-wise modules. EAWNet adopted all the above approaches. Tables 1 and 4 illustrate the details of Figure 9 as well as the details of the experiments conducted. The experiments on COCO and DOTA validate the visual effect in dense and real-time object detection; the average precision is better than all other approaches in the comparative experiments. In addition, the results of the ablation experiments also shed light on different aspects of the connection strategies and enhancements on applying attention modules. The results further highlight the necessity of streamlining and optimizing the network structure and the effectiveness of using the attention mechanism to improve the efficiency of visual perception.

EAWNet outperforms YOLOv4 in terms of accuracy as shown in Table 4. With the same training process and dataset, we simply use the advanced MRFNet backbone as the benchmark and then add an attention-wise module.

The speed is slightly higher than YOLOv4 except for the additional module. Considering the significant accuracy improvement and fast training convergence speed, it is worthwhile to modify the model's name into an attention-wise multipath refinement flow counterpart. The average precision is shown in Figure 10. Tighter and specific detection bounding boxes benefit the training process.

As for FPS, the speed is higher than in comparative experiments. Compared to the LRF, YOLOv3, and YOLOv4, our method is slower but shows a significant improvement in terms of accuracy. Our method outperforms RFBNet, CenterMask, and EfficientDet with regard to both speed and accuracy. Therefore, our method presents a trade-off between accuracy and cost compared with YOLOv4 and outperforms most of the recent state-of-the-art methods.

5. Conclusions

Object detection has been widely used in the field of VIOT. Therefore, it is an important issue for reconstructing a smart city. However, very large images, complex image backgrounds, uneven size, and quantity distribution of training samples make detection tasks challenging, especially for small and dense objects. To solve these problems, an object detector Edge Attention-wise Convolutional Neural Network (EAWNet) is proposed in this paper. Firstly, a better training method with multiflow fusion network is designed to improve the detection accuracy. Secondly, self-attention modules are adopted to underline the meaningful information of feature maps while disregarding useless information. Finally, pass-wise connection makes key semantic features propagate effectively. Comparative experiments are conducted on the benchmark dataset COCO with state-of-the-art methods. The results indicate that our proposed object detection methods outperform the existing models. Extensive experiments and comprehensive evaluations on large-scale DOTA and daily COCO datasets demonstrate the effectiveness of the proposed framework on real-time and dense object detection inference.

In this work, we proposed a framework called EAWNet with edge attention-wise modules for real-time visual inter-

net of things. The patches flow in the multipath refinement flow network, and features are extracted by a pass-wise connection that contributes to a considerable training efficiency. The model was evaluated on two public datasets and compared to state-of-the-art approaches. It performed quite satisfactorily in terms of both accuracy and speed under the same conditions.

In the future, we will redesign the attention modules for lower computation cost. Then, continuous improvements on object detection detectors could be conducted by applying different data augmentation skills and various feature extraction methods and network enhancement approaches as well. We are also interested in establishing whether rotated boundaries could be replaced by the instance segmentation to achieve better results on specific tasks such as an adversarial training process.

Data Availability

The DOTA dataset used to support the findings of this study have been deposited in the DOTA repository (<https://captain-whu.github.io/DOTA/dataset.html>). The COCO dataset used to support the findings of this study have been deposited in the COCO repository (<https://cocodataset.org/>).

Conflicts of Interest

There are no conflicts of interest with any affiliation and person.

Acknowledgments

We thank the National University of Defense Technology for providing the GPU clusters. This research is supported by the College of Advanced Interdisciplinary Studies. This work is supported by the National Natural Science Foundation of China (grant number: 62001493), It is also funded by the Postgraduate Scientific Research Innovation Project of Hunan Province (grant number: CX20200043).

References

- [1] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *2014 IEEE conference on computer vision and pattern recognition*, pp. 580–587, Columbus, OH, 2014.
- [2] J. Dai, Y. Li, K. He, and J. Sun, "R-fcn: object detection via region-based fully convolutional networks," *NIPS*, 2016.
- [3] R. Girshick, "Fast R-CNN," in *IEEE International Conference on Computer Vision (ICCV)*, Las Condes Araucano Park, Chile, 2015.
- [4] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," *ICCV*, 2017.
- [5] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: unified, real-time object detection," in *2016 IEEE conference on computer vision and pattern recognition (CVPR)*, pp. 779–788, Las Vegas, NV, 2016.
- [6] J. Redmon and A. Farhadi, "YOLO9000: better, faster, stronger," in *2017 IEEE conference on computer vision and pattern recognition (CVPR)*, pp. 6517–6525, Honolulu, HI, 2017.

- [7] W. Liu, D. Anguelov, D. Erhan et al., "SSD: single shot multi-box detector," in *European conference on computer vision*, pp. 21–37, Cham, 2016.
- [8] C.-Y. Fu, W. Liu, A. Ranga, A. Tyagi, and A. C. Berg, "DSSD: deconvolutional single shot detector," 2017, <https://arxiv.org/abs/1701.06659>.
- [9] T. Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proceedings of the IEEE international conference on computer vision*, Venice, Italy, 2017.
- [10] Z. Cai and N. Vasconcelos, "Cascader-cnn: delving into high quality object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 6154–6162, Salt Lake City, Utah, U.S., 2018.
- [11] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. Le Cun, "Overfeat: integrated recognition, localization and detection using convolutional networks," 2014, <https://arxiv.org/abs/1312.6229>.
- [12] J. Huang, V. Rathod, C. Sun et al., "Speed/accuracy trade-offs for modern convolutional object detectors," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, vol. 2, Honolulu, Hawaii, USA, 2017.
- [13] H. Law and J. Deng, "Cornersnet: detecting objects as paired keypoints," in *Proceedings of the European conference on computer vision (ECCV)*, vol. 1, pp. 734–750, Munich, Germany, 2018.
- [14] Y. Bai, Y. Zhang, M. Ding, and B. Ghanem, "Sod-mtgan: small object detection via multi-task generative adversarial network," in *Proceedings of the European Conference on Computer Vision (ECCV)*, vol. 2, Munich, Germany, 2018.
- [15] T. Wang, R. M. Anwer, H. Cholakkal, F. S. Khan, Y. Pang, and L. Shao, "Learning rich features at high-speed for single-shot object detection," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1971–1980, Seoul, Korea, 2019.
- [16] S. Zheng, Z. Zhu, J. Cheng, Y. Guo, and Y. Zhao, "Edge heuristic GAN for non-uniform blind deblurring," *IEEE Signal Processing Letters*, vol. 26, no. 10, pp. 1546–1550, 2019.
- [17] Y. Yang and H. Deng, "Gc-yolov3: you only look once with global context block," *Electronics*, vol. 9, no. 8, pp. 1–14, 2020.
- [18] J. Redmon and A. Farhadi, "YOLOv3: an incremental improvement," in *2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Salt Lake City, Utah, U.S., 2018.
- [19] A. Bochkovskiy, C. Wang, and H. Mark Liao, "YOLOv4: optimal speed and accuracy of object detection," in *2020 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, WA, USA, 2020.
- [20] S. Liu, D. Huang, and Y. Wang, "Receptive field block net for accurate and fast object detection," in *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 11215 LNCS, pp. 404–419, 2018.
- [21] Y. Lee and J. Park, "CenterMask: real-time anchor-free instance segmentation," in *2020 IEEE/CVF conference on computer vision and pattern recognition (CVPR)*, pp. 13903–13912, Seattle, WA, USA, 2020.
- [22] M. Tan, R. Pang, and Q. V. Le, "EfficientDet: scalable and efficient object detection," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 10778–10787, Seattle, WA, USA, 2020.
- [23] V. Mnih, N. Heess, A. Graves, and K. Kavukcuoglu, "Recurrent models of visual attention," *arXiv preprint arXiv:1406.6247*, 2014.
- [24] T. Xiao, Y. Xu, K. Yang, J. Zhang, Y. Peng, and Z. Zhang, "The application of two-level attention models in deep convolutional neural network for fine-grained image classification," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 842–850, Boston, MA, USA, 2015.
- [25] A. Vaswani, N. Shazeer, N. Parmar et al., "Attention is all you need," *Advances in Neural Information Processing Systems*, vol. 30, pp. 6000–6010, 2017.
- [26] H. Zhang, I. Goodfellow, D. Metaxas, and A. Odena, "Self-attention generative adversarial networks," 2018, <http://arxiv.org/abs/1805.08318>.
- [27] H. Hu, J. Gu, Z. Zhang, J. Dai, and Y. Wei, "Relation networks for object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3588–3597, Salt Lake City, Utah, U.S., 2018.
- [28] X. Wang, R. Girshick, A. Gupta, and K. He, "Non-local neural networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 7794–7803, Salt Lake City, Utah, U.S., 2018.
- [29] J. Fu, J. Liu, H. Tian et al., "Dual attention network for scene segmentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 3146–3154, Long Beach, California, USA, 2019.
- [30] Z. Tian, C. Shen, H. Chen, and T. He, "FCOS: fully convolutional one-stage object detection," in *2019 IEEE/CVF international conference on computer vision (ICCV)*, pp. 9626–9635, Seoul, Korea (South), 2019.
- [31] J. Fu, J. Liu, J. Jiang, Y. Li, Y. Bao, and H. Lu, "Scene segmentation with dual relation-aware attention network," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 6, pp. 2146–2154, 2019.
- [32] M. Tohidian, S. A. R. Ahmadi-Mehr, and R. B. Staszewski, "A tiny quadrature oscillator using low-Q series LC tanks," *IEEE Microwave and Wireless Components Letters*, vol. 25, no. 8, pp. 520–522, 2015.
- [33] S. Yang, Z. Pei, F. Zhou, and G. Wang, "Rotated faster R-CNN for oriented object detection," in *Proceedings of the 2020 3rd International Conference on Robot Systems and Applications*, pp. 35–39, Chengdu, China, 2020.
- [34] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: towards real-time object detection with region proposal networks," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, 2017.
- [35] P. Sun, Y. Zheng, Z. Zhou, W. Xu, and Q. Ren, "R⁴ Det: refined single-stage detector with feature recursion and refinement for rotating object detection in aerial images," *Image and Vision Computing*, vol. 103, p. 104036, 2020.
- [36] G. Lin, A. Milan, and C. Shen, "RefineNet: multi-path refinement networks for high-resolution semantic segmentation," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, Hawaii, USA, 2017.
- [37] M. Zhen, J. Wang, L. Zhou et al., "Joint semantic segmentation and boundary detection using iterative pyramid contexts," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, WA, USA, 2020.
- [38] E. Yang, C. Deng, C. Li, W. Liu, J. Li, and D. Tao, "Shared predictive cross-modal deep quantization," *IEEE transactions on neural networks and learning systems*, vol. 29, no. 11, pp. 5292–5303, 2018.

- [39] C. Deng, E. Yang, T. Liu, and D. Tao, "Two-stream deep hashing with class-specific centers for supervised image search," *IEEE transactions on neural networks and learning systems*, vol. 31, no. 6, pp. 2189–2201, 2019.
- [40] H. Larochelle and G. E. Hinton, "Learning to combine foveal glimpses with a third-order Boltzmann machine," *Advances in neural information processing systems*, vol. 23, pp. 1243–1251, 2010.
- [41] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," 2017, <https://arxiv.org/abs/1709.01507>.
- [42] L. Chen, H. Zhang, J. Xiao et al., "Sca-cnn: spatial and channel-wise attention in convolutional networks for image captioning," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, Honolulu, Hawaii, USA, 2017.
- [43] F. Zhu, J. Yang, C. Gao, S. Xu, N. Ye, and T. Yin, "A weighted one-class support vector machine," *Neurocomputing*, vol. 189, pp. 1–10, 2016.
- [44] T. Y. Lin, M. Maire, S. Belongie et al., "Microsoft COCO: common objects in context," in *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 740–755, Cham, 2014.
- [45] G. S. Xia, X. Bai, J. Ding et al., "DOTA: a large-scale dataset for object detection in aerial images," in *2018 IEEE/CVF 8 conference on computer vision and pattern recognition*, pp. 3974–3983, Salt Lake City, UT, 2018.

Research Article

An Optimized Fingerprinting-Based Indoor Positioning with Kalman Filter and Universal Kriging for 5G Internet of Things

Shuai Huang ^{1,2}, Kun Zhao ^{1,2}, Zhengqi Zheng^{1,2}, Wenqing Ji^{1,2}, Tianyi Li³, and Xiaofei Liao⁴

¹Engineering Center of SHMEC for Space Information and GNSS, East China Normal University, Shanghai 200241, China

²Shanghai Key Laboratory of Multidimensional Information Processing, East China Normal University, Shanghai 200241, China

³Shanghai R&D Center, Ericsson, Shanghai 310000, China

⁴School of Information Science and Technology, Donghua University, Shanghai 201620, China

Correspondence should be addressed to Kun Zhao; kzhao@ce.ecnu.edu.cn

Received 4 April 2021; Revised 25 May 2021; Accepted 4 June 2021; Published 19 June 2021

Academic Editor: Fa Zhu

Copyright © 2021 Shuai Huang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Fingerprinting technique for indoor positioning based on 5G system has attracted attention. Kalman filter (KF) is used as preprocessing of raw data to reduce the disturbance of Received Signal Strength (RSS) values. After preprocessing, Universal Kriging (UK) algorithm is adopted to reduce the efforts of establishing a fingerprinting database by Spatial Interpolation. A machine learning algorithm named *K*-Nearest Neighbour (KNN) is used to calculate user equipment's position. Real experiments are setup with 5G signals over the air. Two indoor scenarios are considered depending whether the base station is located in the same room with user equipment or not. In test room A, the proposed KF and UK algorithms achieve 53% positioning accuracy improvement. In test room B, 43% performance improvement is obtained by the proposed algorithm. 1.44-meter positioning error is observed as the best case for 80% test samples.

1. Introduction

Global Navigation Satellite System (GNSS) has provided enough accuracy for outdoor positioning but not good indoor. 5G Internet of Things (IoT) is a popular research topic including various application scenarios such as indoor positioning, smart transportation, smart manufacturing, and smart security [1–4]. A variety of indoor positioning systems have emerged, including Ultra-Wide Band (UWB), Wi-Fi, Bluetooth, and Long-Term Evolution (LTE) [5–7]. The Base Stations (BSs) of LTE are widely distributed, which has shown advantages for IoT, Machine Learning (ML), and edge intelligence. 5G New Radio (NR) continues to evolve to further enhance LTE performance [8–12]. The number of connected devices in 5G is increasing rapidly and continues to grow exponentially.

Reference Signal Receiving Power (RSRP), Received Signal Strength (RSS), Sounding Reference Signal (SRS) and

other signals are used for positioning [13–15]. RSS-based positioning system includes a radio propagation distance loss model and fingerprinting method [16, 17]. The radio propagation distance loss model requires multiple BSs to perform trilateral positioning and applies in simple environments, while it is not easy to observe multiple NR BSs in a room in the early deployment phase. Hence, we choose the fingerprinting technique in this paper. Fingerprinting technique includes offline and online stages. In the offline stage, NR, RSS, and coordinates of each reference point are extracted to form fingerprints and input into a fingerprinting database. In the online stage, the RSS of the test point is measured in real time and compared with the offline fingerprints to calculate positions. It is important to build a reliable fingerprinting database

Varying multipath, Non-Line-of-Sight (NLOS) always makes RSS biased and reduces the reliability of fingerprints. To solve the problem, preprocessing methods are introduced

to mitigate multipath effects. Reference [18] proposes a method that reduces the effect of signal multipath fading in RSS-distance estimation using Kalman filter. Zhang et al. proposed an indoor positioning method combining MEMS sensors and wireless fingerprints. They used Kalman filter to constrain WIFI fingerprints, which can improve positioning accuracy and computational efficiency [19]. Besides, constructing offline fingerprints requires a lot of manpower and resources. Spatial interpolation methods are considered to improve the spatial resolution of fingerprints with less manual efforts. In [20], Zuo et al. proposed a time-variant multiphase fingerprint map indoor localization method based on Kriging interpolation. Reference [21] introduces a variant of inverse distance weight (IDW) interpolation which is a Modified Shepard method. Son et al. proposed Universal Kriging interpolation based on drift function [22]. This method showed a better performance than linear interpolation, inverse distance weighing, and Ordinary Kriging. Intelligent fingerprinting techniques widely use machine learning as the algorithm to calculate the positions of things [23]. A novel multimodal complete tracking system based on statistic and DL techniques is presented by reference [24]. The authors used a multiphase statistical fingerprint and deep learning to estimate target indoor position. In [25], KNN method was used to achieve the position based on RSS data received by the module to be located. And the authors used KF to optimize the positioning information.

In this paper, Kalman filter (KF) is used as preprocessing optimization method. Specifically, it consists of two stages. In the offline stage, the raw RSS is filtered to obtain reliable data. In the online stage, the RSS collected in real time can be filtered to eliminate the influence of varying multipath. We use spatial interpolation to interpolate the fingerprinting database and compare a variety of interpolation methods including Universal Kriging (UK) to improve the resolution of fingerprints. K -Nearest Neighbour (KNN) algorithm is taken as the positioning algorithm.

2. System Model

In the positioning system, we collect RSS signal and use the signal to calculate location of the mobile phone. As shown in Figure 1, our positioning system consists of two stages, offline and online. In the offline stage, we collect the RSS signal of the reference point and build the raw fingerprint database. After RSS preprocessing, we can build a RSS preprocessed fingerprinting database. By performing spatial interpolation on the database, we can build a database that is reliable and accurate. In the online stage, we capture the RSS signal of the test point and preprocess the signal. And we use the positioning algorithm to determine the location of the test point.

3. RSS Preprocessing

Kalman filter is a linear minimum variance estimation algorithm. As shown in Figure 2, KF algorithm consists of a gain calculation loop and a filter calculation loop. The gain calculation loop includes filter gain, estimation error, and predic-

tion error. The filter calculation loop includes state prediction and state estimation.

The covariance of the observation noise R is represented by averaging the variance of the RSS at each reference point. The phone remains stationary during the observation at a point. Set the system process noise Q equals to 0.001, the state transition vector Φ equals to 1, and observation vector H equals to 1. During RSS filtering of the reference points, the first estimated error covariance of the point is obtained as

$$P_1 = \frac{1}{N} \sum_{l=1}^N \left(Z_1^l - E[Z^l] \right)^2, \quad (1)$$

where Z_1^l is the first sample RSS of point l . $E[Z^l]$ denotes mathematical expectation of RSS data of the point l . N is denoted by the number of reference points. The following is calculated for point l . The prediction error covariance of the t th sample RSS is expressed as

$$P_{t,t-1}^l = P_{t-1}^l + Q, \quad (2)$$

where P_{t-1}^l represents the $(t-1)$ th sample estimated error covariance. Filter gain of the t th sample RSS, denoted by J_t^l , is

$$J_t^l = P_{t,t-1}^l \left[P_{t,t-1}^l + R \right]^{-1}. \quad (3)$$

The estimated error covariance of the t th sample RSS is expressed as

$$P_{t-1}^l = \left[I - J_t^l \right] P_{t,t-1}^l \left[I - J_t^l \right]^T + J_t^l R J_t^{lT}, \quad (4)$$

where I denote unit vector. The predicted value of the t th RSS is expressed as

$$X_{t,t-1}^l = \Phi X_{t-1}^l. \quad (5)$$

We put the filter gain into the filter calculation loop to get the estimated value of the t th sample of RSS

$$X_t^l = X_{t,t-1}^l + J_t^l \left[Z_t^l - X_{t,t-1}^l \right], \quad (6)$$

where Z_t^l is the t th sample of RSS.

Through the KF, the error generated by the RSS can be reduced in the measurement process so that we can obtain more accurate RSS. For the fingerprinting database, a more accurate offline fingerprinting database is established. At the same time, we ensure RSS real-time accuracy for the points to be located.

4. Spatial Interpolation

When we build the offline fingerprinting database in the fingerprinting positioning system, within a certain resolution range, the positioning accuracy is proportional to the resolution of the offline fingerprinting database. The increase of

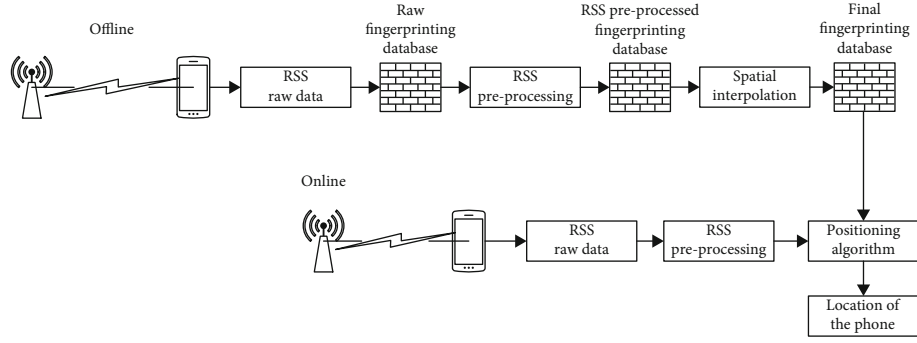


FIGURE 1: Flow chart of fingerprint indoor positioning method.

resolution will lead to a substantial increase in workload. To reduce the time cost while ensuring the positioning accuracy, the spatial interpolation method is used to effectively and correctly improve the resolution of offline fingerprints and reduce the workload. The spatial interpolation method obtains RSS values of interpolation points from those of reference points in the area. As shown in Figure 3, the solid one is an interpolation point, and the hollow ones are the reference points.

In the actual scene, the neighbours of linear interpolation and IDW have a great influence on the result. To solve this problem, we consider Kriging interpolation algorithm. The Ordinary Kriging requires RSS value of point l to meet the second-order stability which is $E[X^l] = C$, where C is constant. However, NR RSS signal cannot satisfy this assumption in indoor room, which means $E[X^l] = m(x_l, y_l)$ is a nonstationary function of the spatial position. Universal Kriging uses a deterministic drift function and residual function to express the RSS value, and the RSS value at any point t , denoted by, X^t , is:

$$X^t = m(x_t, y_t) + r(x_t, y_t), \quad (7)$$

where $m(x_t, y_t)$ represents the drift function of NR RSS related to the position coordinate (x_t, y_t) , $r(x_t, y_t)$ is the residual function of NR RSS expected to be zero. $m(x_t, y_t)$ is used to describe the trend of RSS. And we use a deterministic function to simulate it. According to the distribution characteristics of RSS in two-dimensional space, $m(x_t, y_t)$ is expressed by a quadratic function [26]:

$$m(x_t, y_t) = \sum_{i=0}^L \alpha_i f_i(x_t, y_t) = \alpha_0 + \alpha_1 x_t + \alpha_2 y_t + \alpha_3 x_t^2 + \alpha_4 x_t y_t + \alpha_5 y_t^2, \quad (8)$$

where α_i is the coefficient of the deterministic function $f_i(x_t, y_t)$ and L equals 5. The weight coefficient of UK not only depends on the distance between the interpolation point and reference points but also is related to NR RSS of distribution characteristics in the space. The semivariogram $\gamma(d)$ is related to the distance d between each

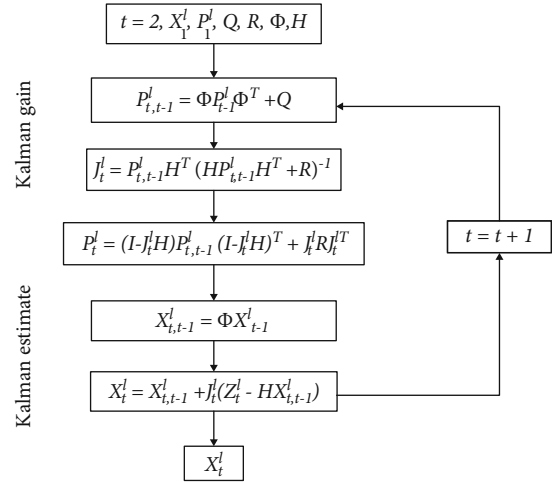


FIGURE 2: Kalman filter structure diagram.

point in space. The semivariogram is equal to half of the mathematical expectation of the square of the difference between NR and RSS of all points separated by a given interval d .

We fit the semivariogram by the RSS value of the known points. Commonly used variation function models include exponential model, spherical model, Gaussian model, and multifunction model. Spherical model is used to fit the function model which has good stability and robustness. The spherical model is defined as

$$\gamma(d) = \begin{cases} 0 & |d| = 0, \\ c_0 + c \left(\frac{3d}{2a} - \frac{d^3}{2a^3} \right) & 0 < |d| < a, \\ c_0 + c & |d| \geq a, \end{cases} \quad (9)$$

where c_0 , c , and a are the coefficient of the semivariogram $\gamma(d)$.

UK algorithm is unbiased and optimal estimation. Unbiasedness means that the expected value of the estimator is equal to the true value. The optimal estimator means that the estimator has the smallest variance among all such linear

unbiased estimators. We need to obtain the weight coefficient λ of each reference point, which is defined as

$$\sum_{u=1}^g \lambda_u = 1, \quad (10)$$

λ_u is the weight coefficient of the point u . Using Lagrange multiplier method to solve the weight coefficient matrix \mathbf{U} :

$$\begin{bmatrix} \mathbf{U} \\ \mathbf{F} \end{bmatrix} = \begin{bmatrix} \mathbf{W} & \mathbf{S} \\ \mathbf{S}^T & \mathbf{O} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{V} \\ \mathbf{G} \end{bmatrix}, \quad (11)$$

where \mathbf{S} represents the coordinate function matrix of the reference point as

$$\begin{bmatrix} f_0(x_1, y_1) & \cdots & f_L(x_1, y_1) \\ \vdots & \ddots & \vdots \\ f_0(x_g, y_g) & \cdots & f_L(x_g, y_g) \end{bmatrix}. \quad (12)$$

\mathbf{U} is the weight coefficient matrix of reference points as $[\lambda_1 \cdots \lambda_g]^T$. \mathbf{G} denotes the coordinate function matrix of the interpolation point h which is $\mathbf{G} = [f_1(x_h, y_h) \cdots f_L(x_h, y_h)]^T$. \mathbf{W} represents the variation function matrix between reference points as

$$\begin{bmatrix} \gamma_{1,1} & \cdots & \gamma_{1,n} \\ \vdots & \ddots & \vdots \\ \gamma_{n,1} & \cdots & \gamma_{n,n} \end{bmatrix}. \quad (13)$$

\mathbf{V} denotes the variation function matrix between the reference point and interpolation point which is $\mathbf{V} = [\gamma_{1,h} \cdots \gamma_{n,h}]^T$. \mathbf{F} is denoted by the Lagrange coefficient matrix $[\eta_0 \cdots \eta_L]^T$, where η is Lagrange coefficient. \mathbf{O} is the $(L+1) * (L+1)$ matrix of zeros.

Then, we get

$$X^*(h) = \sum_{u=1}^g \lambda_u X(u), \quad (14)$$

where $X^*(h)$ is the RSS estimated value at the h th interpolation point, $X(u)$ is NR RSS value of the reference point u , g is the number of reference points of the interpolation point h .

5. Experimental Results and Discussion

Indoor positioning is an indispensable part of human life in the future. Due to the different locations of base stations and the diversity of indoor rooms, indoor positioning in different rooms is considered.

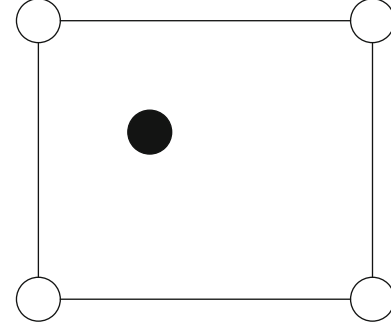


FIGURE 3: Spatial interpolation graph.

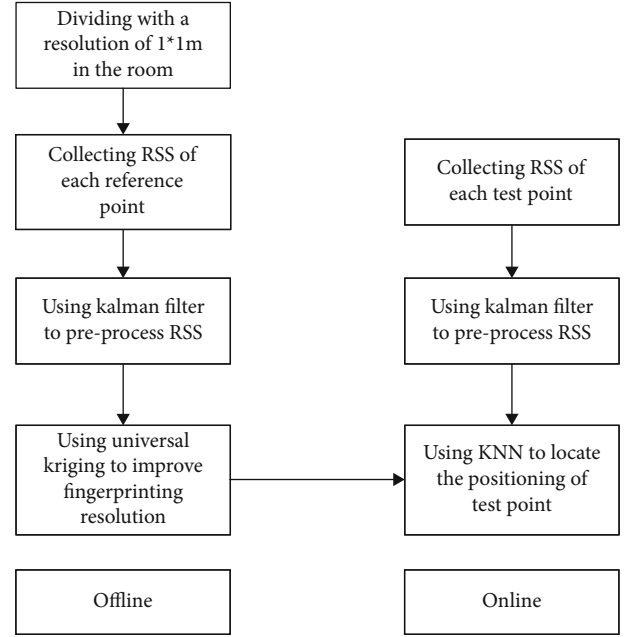


FIGURE 4: Fingerprinting positioning structure.

5.1. Experimental Setup. The experimental system adopts a single 5G base station with fingerprinting offline and online stages as shown in Figure 4.

Positioning accuracy is affected by the placement of BS. This experiment is performed in two different indoor rooms as shown in Figure 5. There were people walking around during measurements. Solid dots are fingerprinting reference points, the hollow ones are spatial interpolation points, and those stars are test points. 5G test phone model is Samsung S20 G9810, and 5G BS is Nokia Aircscale 5G Small Cell.

The BS and room A are located in the same room (as shown in Figure 5(a)). The BS is set up in the corner and 3.5 meters high above the floor. The mobile phone is placed on a one-meter tall tripod. The L-shaped room is divided into 21 squares with a resolution of 1 m * 1 m. In indoor fingerprinting positioning, please note that higher resolution will greatly increase the workload of establishing offline fingerprints. And due to the complexity of the environment, the resolution is chosen to fit for the distribution characteristics of RSS in the room. During the experiments, we analyze the RSS data, which remains stable during two minutes. Hence,

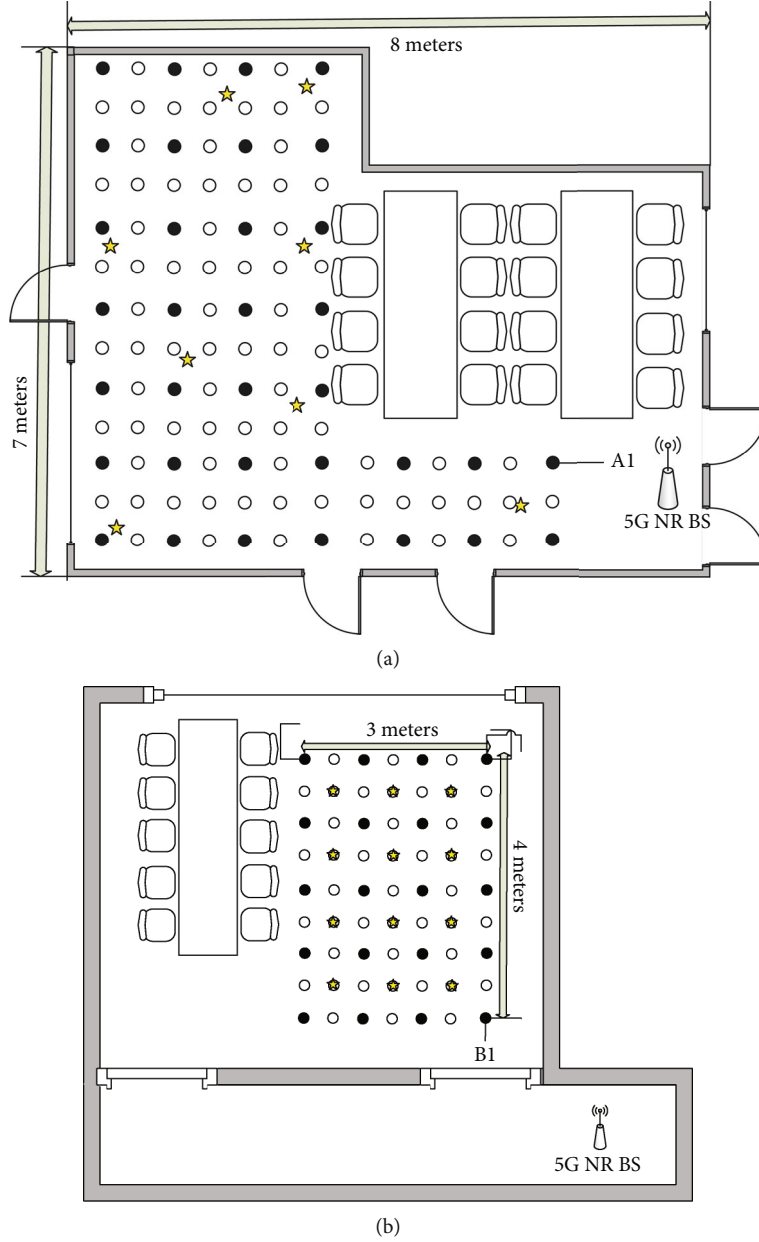


FIGURE 5: (a) Positioning room A. (b) Positioning room B.

we do not measure more time to keep our time. We randomly select 8 test points in the positioning area and statically collect RSS data at 34 reference points and 8 test points for 2 minutes, and the fetch rate of RSS is 100 ms/sample. The inherited value from the last moment is used when raw data is lost. We perform spatial interpolation in room A with a resolution of 0.5 m * 0.5 m.

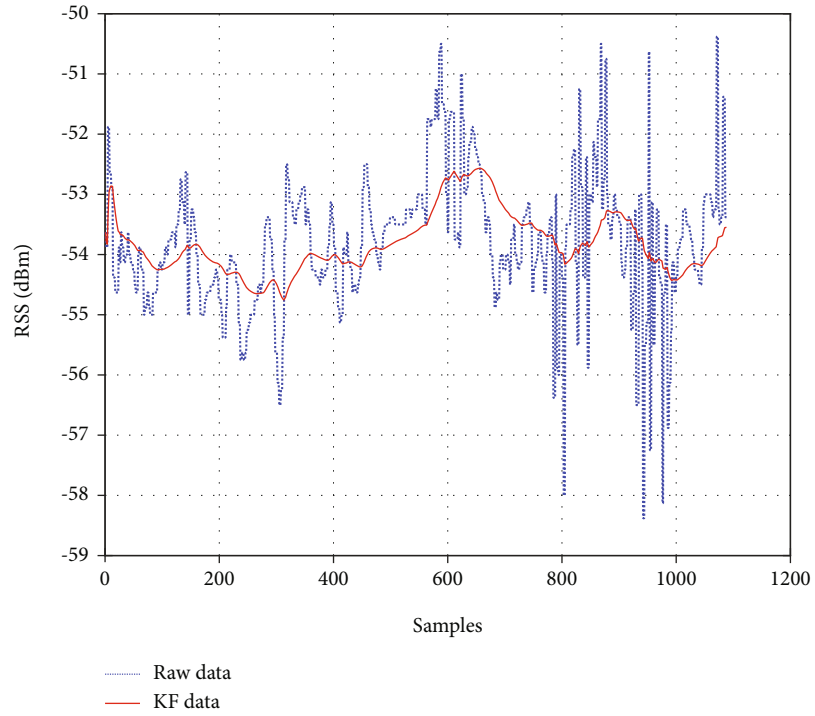
The BS is set up 3.62 meters high above the floor in the corridor adjacent to the room as in room B as shown in Figure 5(b). The doors between them are closed. Mobile phone is placed on a one-meter tall tripod. The distribution of reference points and interpolation points is the same as that of room A. The test points are put in the centre of each grid.

KNN regression algorithm is used as the positioning algorithm in the experiments. In KNN, samples with higher similarity are mapped to close distances. The estimated position is the average of the coordinates of the nearest neighbours.

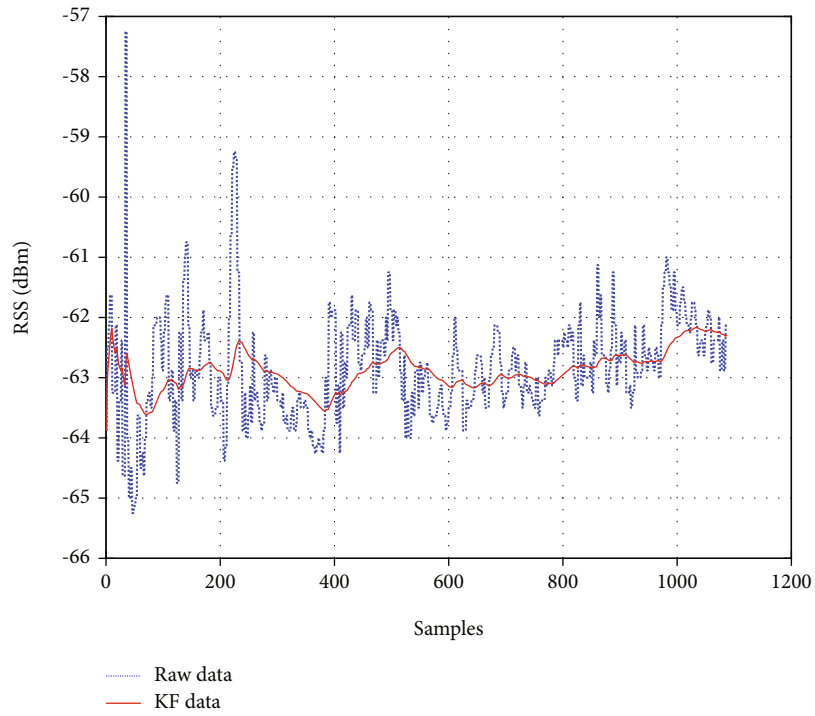
Positioning error is defined as

$$e = \frac{1}{M_k} \sum_{k=1}^{M_k} \sqrt{\frac{1}{M_b} \sum_{b=1}^{M_b} \|q_k - \hat{q}_k(b)\|^2}, \quad (15)$$

where q_k denotes ground-truth of test point k . $\hat{q}_k(b)$ is the estimated position based on the b th sample of test point k .

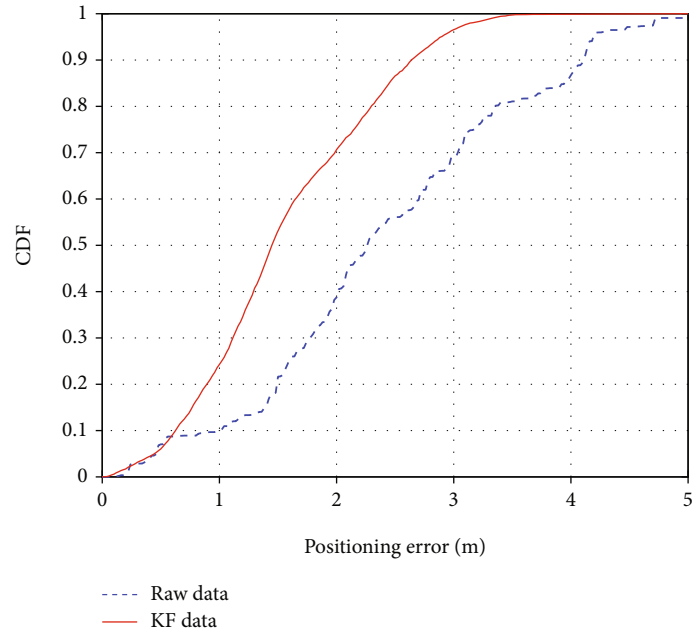


(a)

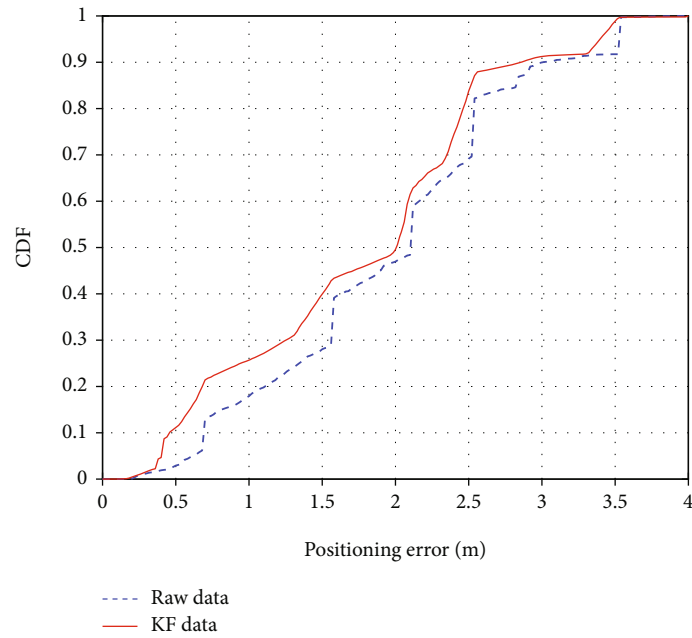


(b)

FIGURE 6: (a) RSS comparison chart before and after Kalman filter of reference point A1 in room A. (b) RSS comparison chart before and after Kalman filter of reference point B1 in room B.



(a)



(b)

FIGURE 7: (a) Comparison of the Cumulative Distribution Function (CDF) of positioning error before and after Kalman filter in positioning room A. (b) Comparison of the CDF of positioning error before and after Kalman filter in positioning room B.

TABLE 1: Positioning accuracy of room A.

Location	Average error (m)	CDF80% (m)	CDF90% (m)
Raw data	2.24	3.36	4.11
KF data	1.58	2.30	2.64

TABLE 2: Positioning accuracy of room B.

Location	Average error (m)	CDF80% (m)	CDF90% (m)
Raw data	1.95	2.54	3.08
KF data	1.77	2.41	2.85

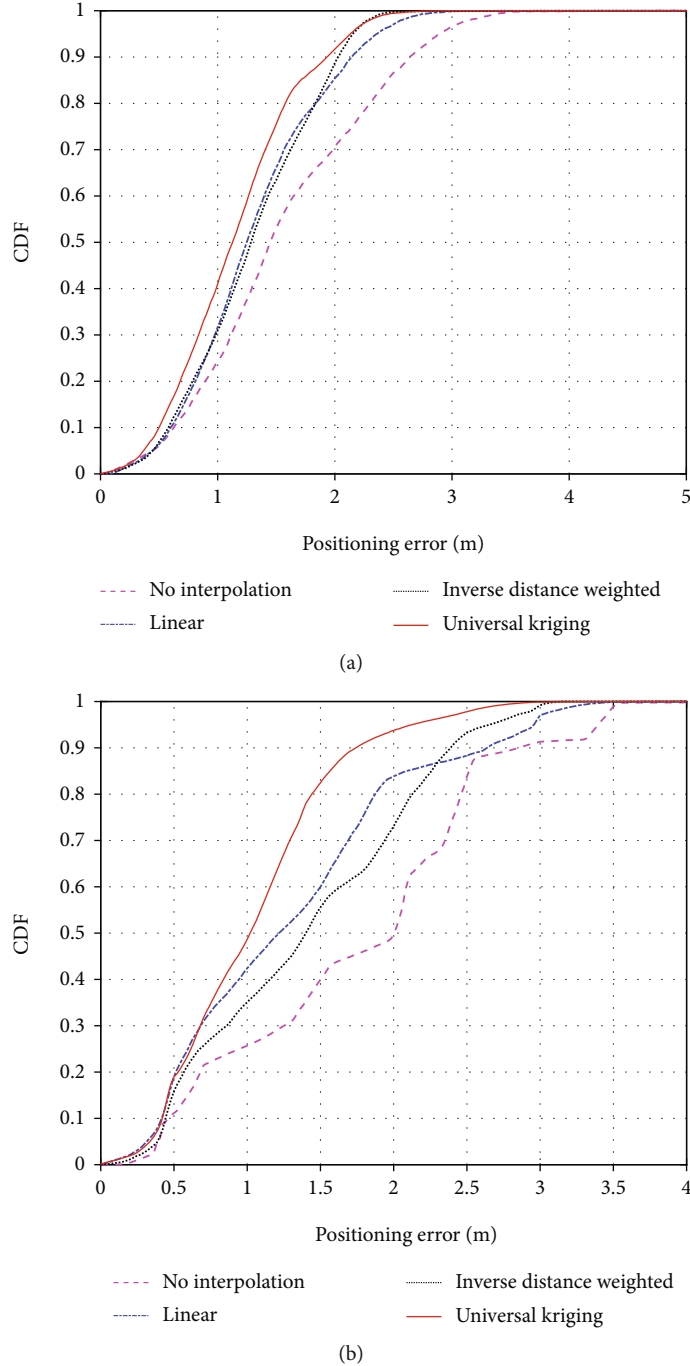


FIGURE 8: (a) CDF of the positioning error using different spatial interpolation algorithms in room A. (b) CDF of the positioning error using different spatial interpolation algorithms in room B.

M_k and M_b represent number of test points and samples over each point, respectively. All test samples equal to $M_k * M_b$.

5.2. RSS Preprocessing. To reduce the influence of varying multipath on RSS, Kalman filter is used to preprocess the NR RSS of the offline fingerprinting database and test points. The preprocessed RSS is more stable, and using KF effectively reduces the disturbance to NR RSS caused by varying multipaths. The RSS value changes slightly around the mean after preprocessing, the variance is smaller, and the data is more

stable. Using KF to preprocess RSS is shown in Figure 6. The RSS comparison before and after preprocessing of reference point A1 in room A and B1 in room B is shown in Figures 6(a) and 6(b), respectively. The distribution range of RSS values narrowed from -58.4 dBm~-50.3 dBm to -54.7 dBm~-52.6 dBm, stabilizing around -54 dBm. The distribution characteristics of RSS are more obvious.

In room A with a relatively simple room, the RSS distribution between each point is relatively close, and the characteristics are not obvious. After preprocessing, the RSS

TABLE 3: Positioning accuracy of room A after different interpolation algorithms.

Location	Average error (m)	CDF80% (m)	CDF90% (m)
No interpolation	1.58	2.30	2.64
Linear	1.32	1.84	2.14
IDW	1.30	1.84	2.03
UK	1.17	1.58	1.94

TABLE 4: Positioning accuracy of room B after different interpolation algorithms.

Location	Average error (m)	CDF80% (m)	CDF90% (m)
No interpolation	1.77	2.41	2.84
Linear	1.42	1.88	2.64
IDW	1.31	2.14	2.38
UK	1.16	1.44	1.76

distribution of each point is more concentrated, so that the RSS cross-term between each point is reduced. The distribution characteristics of RSS between points are more obvious. We use the KNN positioning algorithm to locate the test points. As shown in Figure 7(a), the positioning accuracy is greatly affected by the multipath changes. Kalman filter on the raw RSS data significantly improves the positioning accuracy of the fingerprints. The positioning accuracy improvement effect is shown in Table 1. After using KF, the positioning accuracy has been improved by 31%; we can achieve 2.30-meter positioning error for 80% test samples. The positioning error is the Euclidean distance of test points between the true position and the positioning position.

In room B, NR RSS value between each point varies greatly, and the distribution of the mean value is obvious. Varying multipath has little effect on the positioning accuracy. As shown in Figure 7(b), Kalman filter algorithm can improve the positioning accuracy. The positioning accuracy improvement effect is shown in Table 2. Using KF, the positioning accuracy has improved by 6%. We achieve 2.41-meter positioning error for 80% test samples.

5.3. Spatial Interpolation. When we build the offline fingerprinting database in the fingerprinting positioning system, within a certain resolution range, the resolution of the offline fingerprinting database is proportional to the positioning accuracy. The higher the resolution of the fingerprints, the higher the positioning accuracy. Higher resolution will result in greater workload. In this experiment, a variety of commonly used spatial interpolation methods are used to assign values to each interpolation point separately. To avoid destroying the characteristics of the RSS of the interpolation points, we sort the RSS values of all points in descending order; the interpolation points are interpolated according to the weight of the reference points.

These experiments test different positioning environments in rooms A and B and compare several interpolation methods. We interpolate the preprocessed RSS offline fingerprinting database. The positioning error of various interpolation methods for rooms A and B is shown in Figure 8. In rooms A and B, we use UK that has the best positioning accuracy. The interpolation accuracy in room A is shown in Table 3, and the interpolation accuracy in room B is shown in Table 4. In room A, the positioning error of linear, IDW, and UK has improved by 20%, 20%, and 31%, respectively. Using UK can achieve 1.58-meter positioning error for 80% test samples. In room B, the positioning error of linear, IDW, and UK has been improved by 24%, 11%, and 40%, respectively. Using UK can achieve 1.44-meter positioning error for 80% test samples. After using KF and UK, we can effectively improve the positioning accuracy in both rooms.

6. Conclusions

We use the existing 5G as the positioning base station, which need not rebuild specific positioning equipment. Our intelligent fingerprinting technology optimization adopts KF as preprocessing step to reduce the disturbance of the RSS values caused by multipath. Spatial interpolation method is used to keep fingerprint sampling effort low but still get good resolution. A variety of spatial interpolation methods are compared. UK has the best performance. In room A, compared with that of raw data, KF showed 31% performance improvement. UK provided an additional 26% increase of positioning accuracy. In room B, compared with that of raw data, KF has 6% performance improvement. UK can further reduce the positioning error by 36%. We have achieved a positioning error of more than 80% test samples below 1.6 meters. In the next step, we may research universal fingerprinting, database preprocessing, and spatial interpolation methods in different indoor scenarios. Multiple-base station case in the future is possible. Research on maintaining the fingerprinting database over time is also good to touch.

Data Availability

The experimental data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare no conflict of interest.

Acknowledgments

This work was sponsored by the National Natural Science Foundation of China (No. 61771197) and Science and Technology Commission of Shanghai Municipality (Grant no. 18DZ2270800).

References

- [1] F. Zafari, A. Gkelias, and K. K. Leung, "A survey of indoor localization systems and technologies," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 3, pp. 2568–2599, 2019.
- [2] S. A. Shaikh and A. M. Tonello, "Whitepaper on new localization methods for 5G wireless systems and the Internet-of-Things," in *COST Action CA15104, European Cooperation in Science and Technology (COST)*, Brussels, Belgium, April 2018.
- [3] K. Zhao, T. Zhao, Z. Zheng et al., "Optimization of time synchronization and algorithms with TDOA based indoor positioning technique for Internet of Things," *Sensors*, vol. 20, no. 22, article 6513, 2020.
- [4] X. Li, M. Zhao, M. Zeng et al., "Hardware impaired ambient backscatter NOMA systems: reliability and security," *IEEE Transactions Communications*, vol. 69, no. 4, pp. 2723–2736, 2021.
- [5] C. Hua, K. Zhao, D. Dong et al., "Multipath map method for TDOA based indoor reverse positioning system with improved Chan-Taylor algorithm," *Sensors*, vol. 20, no. 11, article 3223, 2020.
- [6] B. Wang, X. Liu, B. Yu, R. Jia, and X. Gan, "An improved WiFi positioning method based on fingerprint clustering and signal weighted Euclidean distance," *Sensors*, vol. 19, no. 10, article 2300, 2019.
- [7] C.-Y. Chen and W.-R. Wu, "Three-dimensional positioning for LTE systems," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 4, pp. 3220–3234, 2017.
- [8] C.-H. Wang, C.-J. Lee, and X. Wu, "A coverage-based location approach and performance evaluation for the deployment of 5G base stations," *IEEE Access*, vol. 8, pp. 123320–123333, 2020.
- [9] B. el Boudani, L. Kanaris, A. Kokkinis et al., "Implementing deep learning techniques in 5G IoT networks for 3D indoor positioning: DELTA (DeEp Learning-Based Co-operative Architecture)," *Sensors*, vol. 20, no. 19, article 5495, 2020.
- [10] V. Savic and E. G. Larsson, "Fingerprinting-based positioning in distributed massive MIMO systems," in *2015 IEEE 82nd Vehicular Technology Conference (VTC2015-Fall)*, pp. 1–5, Boston, MA, USA, September 2015.
- [11] X. Li, J. Li, Y. Liu, Z. Ding, and A. Nallanathan, "Residual transceiver hardware impairments on cooperative NOMA networks," *IEEE Transactions on Wireless Communications*, vol. 19, no. 1, pp. 680–695, 2020.
- [12] E. Rastorgueva-Foi, M. Costa, M. Koivisto, K. Leppänen, and M. Valkama, "User positioning in mmW 5G networks using beam-RSRP measurements and Kalman filtering," in *2018 21st International Conference on Information Fusion (FUSION)*, pp. 1–7, Cambridge, UK, July 2018.
- [13] X. Li, M. Zhao, Y. Liu, L. Li, Z. Ding, and A. Nallanathan, "Secrecy analysis of ambient backscatter NOMA systems under I/Q imbalance," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 10, pp. 12286–12290, 2020.
- [14] M. A. Khan, N. Saeed, A. W. Ahmad, and C. Lee, "Location awareness in 5G networks using RSS measurements for public safety applications," *IEEE Access*, vol. 5, pp. 21753–21762, 2017.
- [15] F. Wang, J. Chen, and Q. Liu, "SRS-based LTE indoor wireless positioning system," in *2017 IEEE 2nd Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)*, pp. 2356–2359, Chongqing, China, March 2017.
- [16] C. Zhou, J. Yuan, H. Liu, and J. Qiu, "Bluetooth indoor positioning based on RSSI and Kalman filter," *Wireless Personal Communication*, vol. 96, no. 3, pp. 4115–4130, 2017.
- [17] S. Kram, C. Nickel, J. Seitz, L. Patino-Studencka, and J. Thielecke, "Spatial interpolation of Wi-Fi RSS fingerprints using model-based universal kriging," in *2017 Sensor Data Fusion: Trends, Solutions, Applications (SDF)*, pp. 1–6, Bonn, Germany, October 2017.
- [18] S. Sam and C. James, "Reducing the effect of signal multipath fading in RSSI-distance estimation using Kalman filters," in *19th Communications & Networking Symposium (CNS 2016)*, Pasadena, CA, USA, 2016.
- [19] Y. Zhuang, Y. Li, L. Qi, H. Lan, J. Yang, and N. el-Sheimy, "A two-filter integration of MEMS sensors and WiFi fingerprinting for indoor positioning," *IEEE Sensors Journal*, vol. 16, no. 13, pp. 5125–5126, 2016.
- [20] J. Zuo, S. Liu, H. Xia, and Y. Qiao, "Multi-phase fingerprint map based on interpolation for indoor localization using iBeacons," *IEEE Sensors Journal*, vol. 18, no. 8, pp. 3351–3359, 2018.
- [21] A. H. Ismail, H. Kitagawa, R. Tasaki, and K. Terashima, "WiFi RSS fingerprint database construction for mobile robot indoor positioning system," in *2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pp. 1561–1566, Budapest, Hungary, October 2016.
- [22] P.-W. Son, J. H. Rhee, J. Hwang, and J. Seo, "Universal kriging for Loran ASF map generation," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 55, no. 4, pp. 1828–1842, 2019.
- [23] T. Koike-Akino, P. Wang, M. Pajovic, H. Sun, and P. V. Orlik, "Fingerprinting-based indoor localization with commercial MMWave WiFi: a Deep learning approach," *IEEE Access*, vol. 8, pp. 84879–84892, 2020.
- [24] A. Belmonte-Hernandez, G. Hernandez-Penalosa, D. Martin Gutierrez, and F. Alvarez, "SWiBluX: multi-sensor deep learning fingerprint for precise real-time indoor tracking," *IEEE Sensors Journal*, vol. 19, no. 9, pp. 3473–3486, 2019.
- [25] C. Du, B. Peng, Z. Zhang, W. Xue, and M. Guan, "KF-KNN: low-cost and high-accurate FM-based indoor localization model via fingerprint technology," *IEEE Access*, vol. 8, pp. 197523–197531, 2020.
- [26] H. Zhao, B. Huang, and B. Jia, "Applying kriging interpolation for WiFi fingerprinting based indoor positioning systems," in *2016 IEEE Wireless Communications and Networking Conference*, pp. 1–6, Doha, Qatar, April 2016.