

## Research Article

# An Online Semisupervised Learning Model for Pedestrians' Crossing Intention Recognition of Connected Autonomous Vehicle Based on Mobile Edge Computing Applications

Shicai Ji,<sup>1</sup> Ying Peng,<sup>1</sup> Hongjia Zhang<sup>1,2</sup> , and Shengbo Wu<sup>2</sup>

<sup>1</sup>School of Vehicle Engineering, Shandong Transport Vocational College, Weifang, Shandong 261206, China

<sup>2</sup>School of Automobile, Chang'an University, Xi'an, Shaanxi 710064, China

Correspondence should be addressed to Hongjia Zhang; zhanghongjia@chd.edu.cn

Received 25 November 2020; Revised 7 January 2021; Accepted 19 January 2021; Published 5 February 2021

Academic Editor: Shaohua Wan

Copyright © 2021 Shicai Ji 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.

One of the major challenges that connected autonomous vehicles (CAVs) are facing today is driving in urban environments. To achieve this goal, CAVs need to have the ability to understand the crossing intention of pedestrians. However, for autonomous vehicles, it is quite challenging to understand pedestrians' crossing intentions. Because the pedestrian is a very complex individual, their intention to cross the street is affected by the weather, the surrounding traffic environment, and even his own emotions. If the established street crossing intention recognition model cannot be updated in real time according to the diversity of samples, the efficiency of human-machine interaction and the interaction safety will be greatly affected. Based on the above problems, this paper established a pedestrian crossing intention model based on the online semisupervised support vector machine algorithm (OS<sup>3</sup>VM). In order to verify the effectiveness of the model, this paper collects a large amount of pedestrian crossing data and vehicle movement data based on laser scanner, and determines the main feature components of the model input through feature extraction and principal component analysis (PCA). The comparison results of recognition accuracy of SVM, S<sup>3</sup>VM, and OS<sup>3</sup>VM indicate that the proposed OS<sup>3</sup>VM model exhibits a better ability to recognize pedestrian crossing intentions than the SVM and S<sup>3</sup>VM models, and the accuracy achieves 94.83%. Therefore, the OS<sup>3</sup>VM model can reduce the number of labeled samples for training the classifier and improve the recognition accuracy.

## 1. Introduction

According to the annual road traffic accident statistics report released by the Traffic Administration Bureau of the Ministry of Public Security of the People's Republic of China, the number of people who died in traffic accidents in China in 2019 was 62,763 and the number of people injured was 256,101. Among them, accidents between pedestrians and vehicles resulted in 17,473 deaths and 45,495 injuries [1].

Pedestrians are vulnerable road users and require active protection. Both autonomous driving and connected cars are designed to provide greater safety benefits [2, 3]. Autonomous vehicles need to have the ability to determine the

intentions of other road users and communicate with them. This interaction is crucial between vehicles and pedestrians. However, pedestrian crossing intention is a relatively complex process, which depends on various factors, such as pedestrian status and traffic environment [4].

According to a recent report by Google's autonomous vehicles, 90% of the failures of autonomous vehicles occur on busy streets, and 10% of them are due to misrecognition of pedestrian intention. Among the various behaviors of pedestrians, street crossing is the most important one, which is related to the safety of pedestrians. Through visual communication, gesture communication, and even auditory communication with pedestrians, human drivers can easily

recognize the pedestrian's crossing intention. However, for autonomous vehicles, such communication with pedestrians is quite challenging [5, 6].

Most of the current researches mainly reveal pedestrians' crossing intentions from pedestrian movement data and pedestrian posture data. Pedestrian's intention to cross the street is actually a classification problem of behavior sequence data. There is a strong front and back dependency between sequence data [3].

Schulz and Stiefelhagen [7] proposed a pedestrian crossing intention recognition method combining multiple interactive multiple model filters and latent-dynamic conditional random field model. The input parameters are mainly position and speed. Varytimidis et al. [8] used the convolutional neural network (CNN) algorithm to extract the features of the pedestrian's posture and then recognized the pedestrian's crossing intention based on the pedestrian's head posture. The algorithms used are the support vector machine (SVM) and artificial neural network (ANN). Völz et al. [9] used SVM, CNN, and long- and short-term memory networks (LSTM) to identify pedestrian crossing intentions. The input feature parameters are mainly the distance between the pedestrian and the zebra crossing, and the distance between the vehicle and the zebra crossing. Park and Lee [10] obtained the EMG signal during pedestrian movement and used CNN for learning. It was found that the features can effectively decode the pedestrian's movement intention. Zhang et al. [11] proposed a pedestrian crossing intention recognition model based on the attention mechanism of long- and short-term memory networks (AT-BiLSTM). The input feature parameters are mainly pedestrian speed, distance between pedestrian and vehicle, and distance between vehicle and zebra crossing. Schneemann and Heinemann [12] proposed a pedestrian crossing intention model based on contextual features, using a support vector machine algorithm. The descriptor captured the movement of pedestrians relative to the road and the spatial layout of other scene elements in a generic manner. It showed that context-based data are good indicators for crossing prediction. Zhao et al. [13] proposed a pedestrian crossing intention model based on improved naive Bayesian networks. The input feature data source is Lidar. Camara et al. proposed an intention heuristic model. The input feature parameters include pedestrian trajectory, vehicle trajectory, and relative position. It was found that the model recognition accuracy was high, reaching 96%. Fang et al. [14] established a pedestrian crossing intention model based on SVM and found that the model can accurately identify pedestrian crossing intentions with an accuracy rate of 93%. The input of the model is the feature parameters of pedestrian body posture. Škovierová et al. [15] collected the position, speed, and orientation information of all traffic participants of pedestrians, and realized the recognition of pedestrian intention through the Bayesian network. Quintero et al. [16, 17] used pedestrian posture features to recognize pedestrian crossing intentions and retrogrades. Although the recognition accuracy is high, the recognition time lags to a certain extent.

Through the above review, it can be seen that the current pedestrian crossing intention recognition is mainly based on

data-driven and pedestrian posture feature-driven, and recognition algorithms are traditional supervised learning algorithms. However, the shortcomings of the current pedestrian street crossing intention model are also obvious. First of all, some pedestrian street crossing intention models are established based on pedestrian posture data. When the pedestrian's head is blocked or the sunlight is too strong, it will seriously affect the recognition accuracy of the intention model. Secondly, the current pedestrian street crossing intention model is mainly based on supervised learning. A large number of data labels need to be manually labeled. In the era of big data, this method is very time-consuming. Finally, the current pedestrian street crossing intention model cannot perform online self-learning. When faced with some special situations, it cannot perform self-learning based on data, which greatly affects the generalization performance of the model.

Samples are known to have certain features as a training data set, build a model, and then use this model to classify unknown samples. This method is called supervised learning and is the most commonly used machine learning method. Generally speaking, each set of feature data corresponds to a specific label when the classification model is trained [18, 19]. Semisupervised learning (SSL) [20] is a key issue in the field of pattern recognition and machine learning. It is a learning method that combines supervised learning and unsupervised learning. Semisupervised learning uses a large amount of unlabeled data and simultaneously uses labeled data for pattern recognition. When using semisupervised learning, it will require as few people as possible to do the work, and at the same time, it can bring relatively high accuracy. Online learning can quickly make model adjustments based on online feedback data and improve the accuracy of online predictions. The process of online learning includes presenting the prediction results of the model to the user and then collecting the user's feedback data, which is then used to train the model to form a closed-loop system [21, 22]. Online semisupervised learning is the product of the fusion of semisupervised learning and online learning. While the labeled and unlabeled samples can be stored, it also has the characteristics of online learning. The online semisupervised learning algorithm is a sequence of continuous learning cycles ongoing. In each learning cycle, the learner is given a training sample and is required to predict the label of the sample in the case of training unlabeled samples [23, 24].

A large number of sensors, cameras, millimetre wave radars, and Lidar devices are installed on the connected autonomous vehicles (CAVs). These devices will generate abundant data, which will lead to the lack of storage resources, computing resources, and communication resources. Due to the limitations of the equipment in computing power and storage performance, the CAVs cannot perform computationally intensive tasks. At this point, the task needs to be sent to the server, which will process the task, and then the processing results will be fed back to the vehicle. The data processing based on the traditional cloud computing model will not only lead to long task execution delays but also increase the energy consumption. Centralized cloud servers are far away from terminal devices, which leads to inefficiency in computing-

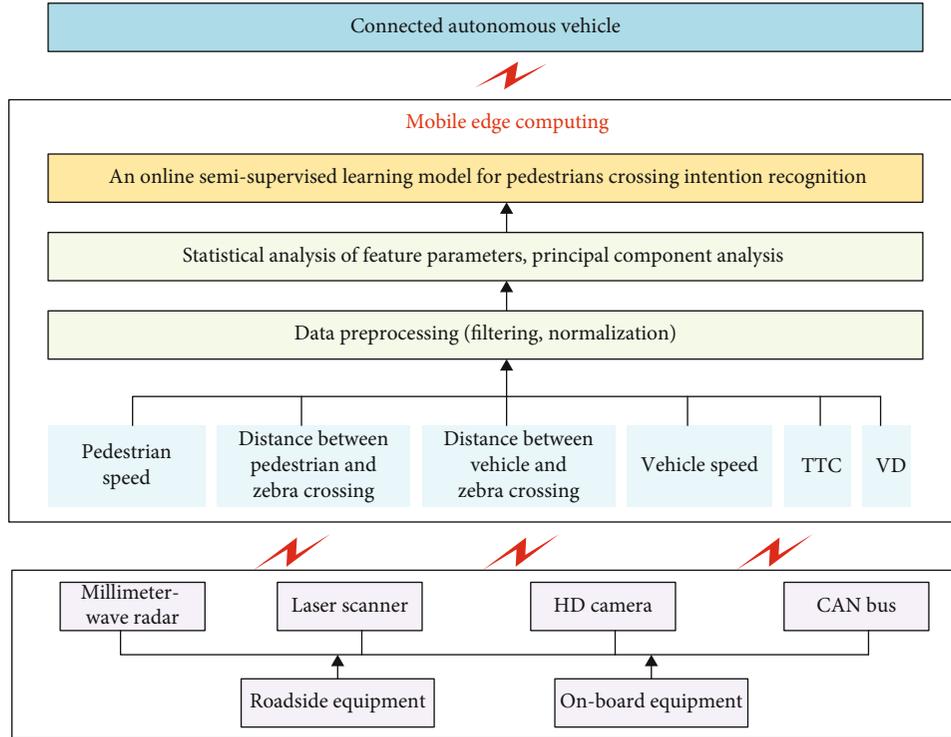


FIGURE 1: The structure of pedestrian intention recognition based on mobile edge computing.

intensive environments. At the same time, the transmission of computing resources to the cloud consumes energy, which may also reduce the service life of mobile batteries. In addition, the cloud computing patterns make it difficult to provide mobile users with complex memory utilization applications and higher data storage capacity. Mobile edge computing (MEC) is regarded as an effective way to solve the above problems. By deploying computing resources at the edge of the network, the delay-sensitive tasks such as collision prediction, surrounding vehicle and pedestrian intention prediction, vehicle avoidance control, and other tasks are assigned to the edge server for calculation, which can greatly reduce the communication delay and also can effectively improve the data security.

In this study, an online semisupervised learning model for pedestrians' crossing intention recognition based on mobile edge computing technology was established. The MEC technology is an ideal choice for CAVs, which can play a key role in assisting the intelligent vehicles. Therefore, edge intelligence was employed to acquire and process pedestrian and vehicle data at the edge of the network, and the pedestrian intention recognition result was fed back to the decision-making system of CAVs in time. By comparing the characteristics of supervised learning, semisupervised learning, online learning, and online semisupervised learning, we chose an online semisupervised learning algorithm to identify pedestrian crossing intentions. Existing supervised learning algorithms need to manually label data, and the model cannot be updated in real time. In view of this, this work proposed an online semisupervised support vector

machine algorithm ( $OS^3VM$ ). Based on the semisupervised support vector machine classification model, the local concave-convex process optimization (LCCCP) algorithm was employed to remark the soft labels of unmarked samples in an iterative way. Then, the greedy promotion algorithm was used to further update the dual variables to realize the online learning process of the  $S^3VM$  model. The proposed pedestrians' crossing intention recognition model has the structure shown in Figure 1.

## 2. Experimental and Data Collection

*2.1. Experiment Site and Equipment.* The experimental road is two-way four-lane, and the length of the zebra crossing is 12 m. The center of the road is separated by a double yellow line, and there is no fence or refuge island in the middle. The experimental site is a section without signal light control. The experimental site is relatively common in China and has a certain representativeness. The transportation elements are mainly cars and buses. Traffic flow is about 450 veh/h. Figure 2 shows the experiment site.

The main experimental equipment is a 4-layer laser scanner and high-definition (HD) camera. The laser scanner model was an IBEO LUX with a scanning frequency of 12.5 Hz, a detectable range of 0.3–200 m, and a vertical viewing angle of  $3.2^\circ$ FOV. It was mainly used to collect some objective parameters such as the vehicle speeds, crossing speeds of the pedestrians, and the distances between vehicles and pedestrians. The video capture equipment used a mini HD monitor with a video resolution of  $1920 \times 1080$ , which ensured the



FIGURE 2: Experiment site.

definition of the video and met the experimental requirement. HD cameras are mainly used to determine the age, gender, and group attributes of pedestrians. In order to avoid the observer effect, the equipment is installed 15 m away from the zebra crossing. Figure 3 shows the experiment equipment.

**2.2. Intention Feature Extraction.** In this work, pedestrians' crossing intentions are mainly divided into two categories, namely, "stopping" and "crossing". When the pedestrian's crossing intention is "stopping", it means that the pedestrian's speed at a certain distance from the zebra crossing is relatively large, and when the pedestrian reaches the curb, the pedestrian's crossing speed is 0. When the pedestrian's crossing intention is "crossing", it means that the pedestrian crosses the zebra crossing at the original speed, and the speed before crossing the street is not much different from the speed when crossing the street.

All experimental data were collected in sunny weather. Avoid the weather's interference with pedestrians' intention to cross the street. The experiment was carried out for about one month, and mainly collected the movement data of pedestrians and vehicles before crossing the street. Through data extraction, the pedestrian crossing intention feature parameters selected in this work include the pedestrian speed before crossing the street (PS), the distance between pedestrian and zebra crossing (DPZ), the distance between vehicle and zebra crossing (DVZ), vehicle speed (VS), time to collision (TTC), and vehicle deceleration (VD). In addition, the age, gender, and group attributes of pedestrians have a greater impact on pedestrians' intention to cross the street. There, it is also considered in this work. The detailed analysis of feature parameters is described as follows. This study focused on two pedestrian street crossing decisions: crossing and stopping. Two different locations were selected for the experiment. The experiment period is 22 days. We collected

900 samples of pedestrians who intended to stop and 900 samples of pedestrians who intended to cross. Pedestrian crossing intention recognition is actually a kind of sequence data recognition. In this work, we intercepted the data of pedestrians and vehicles 2 s before crossing the street for analysis. In other words, the input length of the feature parameters is 2 s.

The data was preprocessed before statistical analysis, mainly including data filtering and data normalization. Pedestrian speed and vehicle speed collected by laser scanner may have a step phenomenon in a short time. In order to minimize this phenomenon, the paper used Gaussian smoothing filter for data processing. In addition, since the vehicle speed and the distance between the vehicle and the zebra crossing are relatively large, in order to improve the recognition accuracy and training speed of the model, the data was normalized.

**2.2.1. PS.** When the pedestrian's crossing intention is "stopping", the mean PS before crossing the street is 2.48 km/h. When the pedestrian's crossing intention is "crossing", the mean PS before crossing the street is 4.15 km/h. As shown in Figure 4. The one-way analysis of variance (ANOVA) test found that the mean PS before crossing the street under the two kinds of crossing intentions was significantly different ( $F(1.1776) = 375.31, p < 0.05$ ).

**2.2.2. DPZ.** When the pedestrian's crossing intention is "stopping", the mean DPZ before crossing the street is 0.63 m. When the pedestrian's crossing intention is "crossing", the mean DPZ before crossing the street is 0.99 m. As shown in Figure 5, the one-way ANOVA test found that the mean DPZ before crossing the street under the two kinds of crossing intentions was significant difference ( $F(1.1776) = 160.64, p < 0.05$ ).

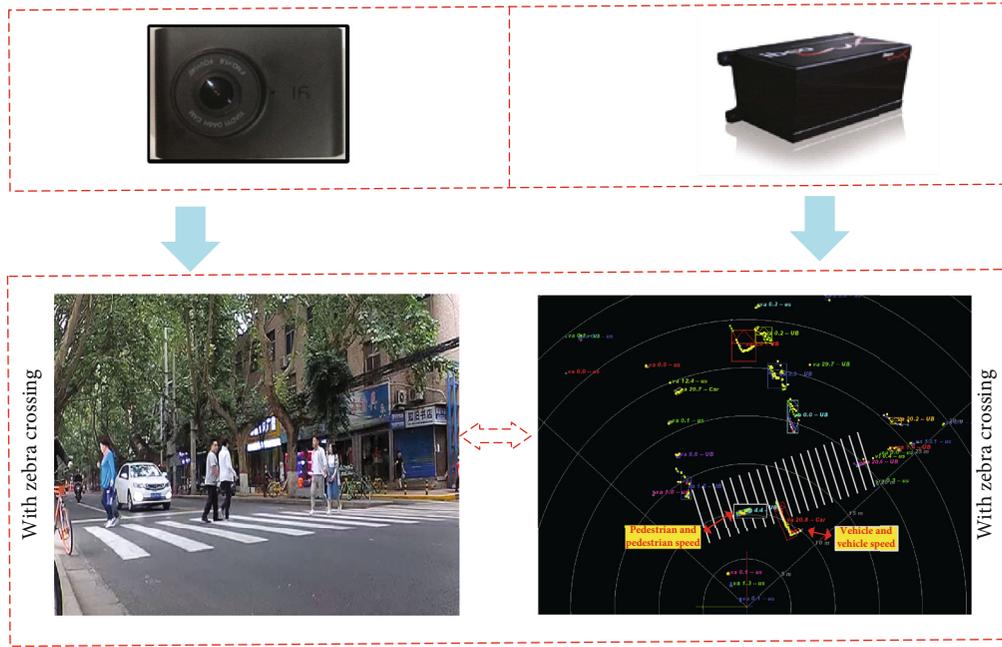


FIGURE 3: Experiment equipment.

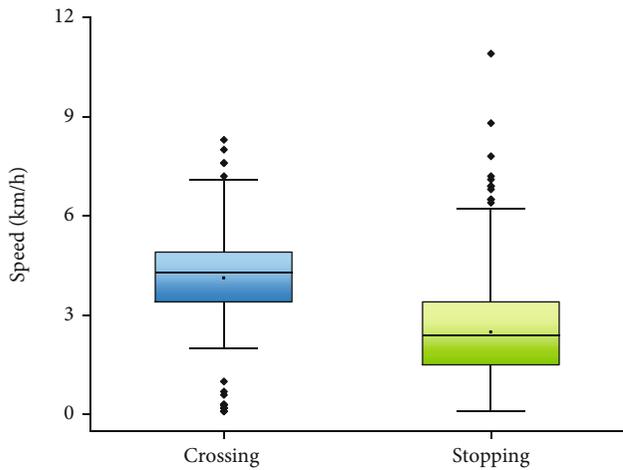


FIGURE 4: Comparison of PS in the intention of “crossing” and “stopping”.

2.2.3. *DVZ*. When the pedestrian’s crossing intention is “stopping”, the mean DVZ before reaching the street is 19.59 m. When the pedestrian’s crossing intention is “crossing”, the mean DVZ before reaching the street is 38.56 m. As shown in Figure 6, the one-way ANOVA test found that the mean DVZ before crossing the street under the two kinds of crossing intentions was significant difference ( $F(1.1776) = 660.64, p < 0.05$ ).

2.2.4. *VS*. When the pedestrian’s crossing intention is “stopping”, the mean VS before reaching the street is 28.91 km/h. When the pedestrian’s crossing intention is “crossing”, the mean VS before reaching the street is

28.90 km/h. As shown in Figure 7, the one-way ANOVA test found that the mean VS before crossing the street under the two kinds of crossing intentions was no significant difference ( $F(1.1776) = 0.96, p > 0.05$ ).

2.2.5. *TTC*. When the pedestrian’s crossing intention is “stopping”, the mean TTC before arriving the street is 2.45 s. When the pedestrian’s crossing intention is “crossing”, the mean TTC before arriving the street is 4.50 s. As shown in Figure 8, the one-way ANOVA test found that the mean TTC before crossing the street under the two kinds of crossing intentions was no significant difference ( $F(1.1776) = 410.45, p < 0.05$ ).

2.2.6. *VD*. When the pedestrian’s crossing intention is “stopping”, the mean VD before arriving the street is 2.26  $m/s^2$ . When the pedestrian’s crossing intention is “crossing”, the mean VD before arriving the street is 1.20  $m/s^2$ . As shown in Figure 9, the one-way ANOVA test found that the mean VD before crossing the street under the two kinds of crossing intentions was no significant difference ( $F(1.1776) = 406.43, p < 0.05$ ).

2.2.7. *Age, Gender, and Group*. It is well known that the age, gender, and group attributes of pedestrians can significantly affect pedestrians’ intention to cross the street. Middle-aged pedestrians tend to take risks, while the elderly is relatively conservative. Compared with male pedestrians, female pedestrians are also relatively conservative. In addition, group pedestrians are more radical than single pedestrians [25–27]. In order to improve the recognition accuracy of pedestrian crossing intention recognition model, we take pedestrian age and pedestrian gender as input variables to train the model. The pedestrians’ age was divided according

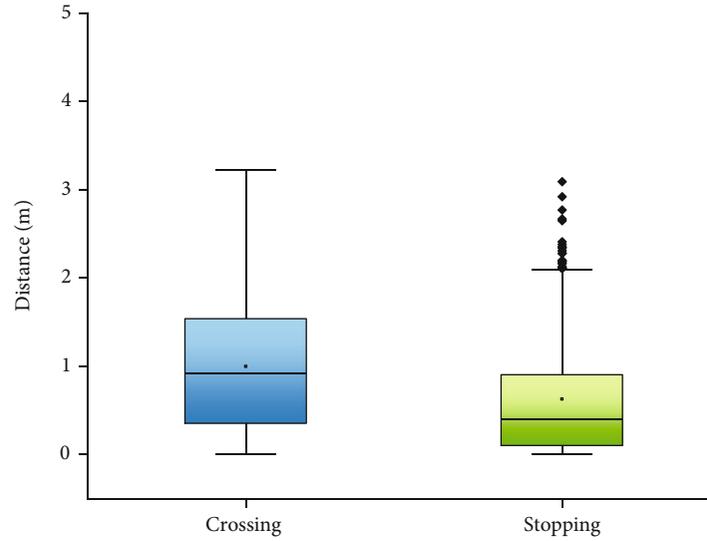


FIGURE 5: Comparison of DPZ in the intention of “crossing” and “stopping”.

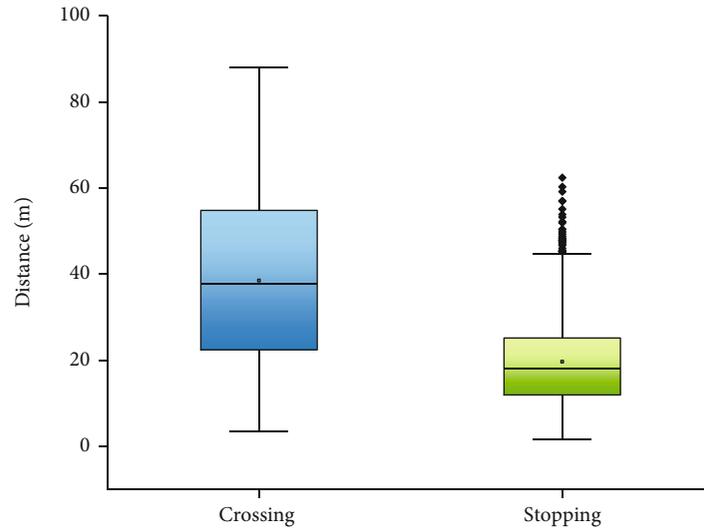


FIGURE 6: Comparison of DVZ in the intention of “crossing” and “stopping”.

to natural observation, using the classification method mentioned in the references which define 18–30 as a youth, 30–59 as middle age, and >60 as old age [28, 29]. Hashimoto et al. [30] found that individuals or groups have great differences in pedestrian crossing behavior and use this attribute as input variable to train the intention recognition model.

**2.3. Principal Component Analysis (PCA).** Through feature analysis, it can be seen that only the vehicle speed of the above six feature parameters has nothing to do with pedestrian crossing intention. Although multiple features contain rich information, omission of features can be avoided. However, the long input feature parameters will slow down the recognition speed and reduce the recognition accuracy. In addition, the model may also be overfitting. Therefore, this

paper used PCA to reduce the dimension of the feature parameters. On the basis of retaining the original feature parameter information, reduce the dimension of the parameters [31]. The PCA algorithm is used to reduce the dimensionality of five feature parameters of crossing intention, and the correlation between the variables and the principal components is shown in Table 1.

From Table 1, we can see that VD, PS, and PC1 have a strong correlation. TTC, DVZ, and PC2 have a strong correlation. TTC, DVZ, VD, and PC3 have a strong correlation. Figure 2 shows PCA feature extraction. It can be seen from Table 2 that the eigenvalues of the first three principal components are all greater than 1, so the first three principal components are selected to replace the original variables. The corresponding cumulative variance contribution rate is

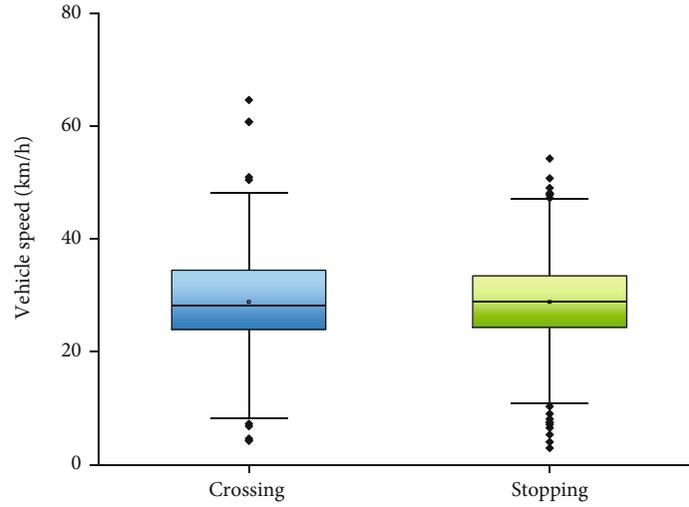


FIGURE 7: Comparison of VS in the intention of “crossing” and “stopping”.

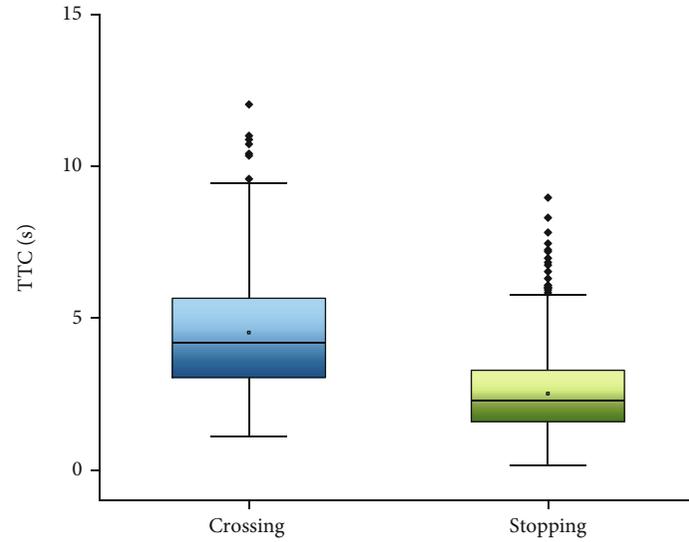


FIGURE 8: Comparison of TTC in the intention of “crossing” and “stopping”.

91.92%, which shows that extracting the first 3 principal components to replace the original variables only loses 8.08% of the information. Therefore, this experiment selected the first three principal components (PC1~PC3) as the feature input of the pedestrian crossing intention model. In addition, the input feature parameters of the intention model also include pedestrian age, gender, and group attributes.

**2.4. Sample Label.** The experiment collected a total of 1800 data sets. *K*-means clustering is widely used in pattern recognition and sample labeling [32, 33]. In addition, *K*-means clustering also plays an important role in semisupervised learning. In this work, the labeled samples are divided into two categories according to certain characteristics through *K*-means clustering. When unlabeled samples enter the model, they are first clustered and labeled, and then further trained by a semisupervised learning algorithm. In this work,

the intention of pedestrians to cross the street is divided into two categories, namely, “crossing” and “stopping”.

### 3. Recognition Model Design

The online semisupervised learning (OSSL) algorithm possesses the advantages of both semisupervised learning and online learning algorithms [34, 35]. The semisupervised data stream containing marked samples and unmarked samples can be learned online simultaneously, and then, the recognition model can be updated in real time. In this paper, an online semisupervised support vector machine (OS<sup>3</sup>VM) algorithm was established to identify the pedestrian crossing intentions. Based on the semisupervised support vector machine classification model, the local concave-convex process optimization (LCCCP) algorithm was employed to remark the soft labels of unmarked samples in an iterative

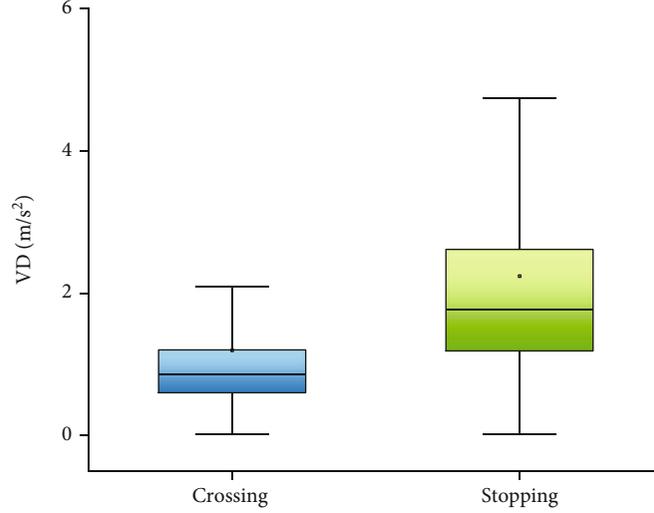


FIGURE 9: Comparison of VD in the intention of “crossing” and “stopping”.

TABLE 1: Correlation between the variables and the principal components.

Variable	PC1	PC2	PC3	PC4	PC5
TTC	-0.11	0.79	0.45	-0.07	-0.02
DVZ	0.02	-0.69	0.67	-0.21	-0.45
VD	0.74	-0.15	0.46	-0.52	-0.07
PS	-0.48	-0.08	-0.07	-0.01	-0.02
DPZ	-0.24	-0.15	0.06	-0.22	-0.07

TABLE 2: PCA feature extraction.

PC	Feature value	Variance contribution rate (%)	Cumulative variance contribution rate (%)
PC1	3.753	36.958	36.958
PC2	2.241	29.505	66.463
PC3	1.273	25.458	91.921
PC4	0.673	4.802	96.723
PC5	0.528	3.277	100

way. Then, the greedy promotion algorithm was used to further update the dual variables to realize the online learning process of the S<sup>3</sup>VM model. This section focuses on the introduction of the OS<sup>3</sup>VM algorithm based on the dual lifting process.

**3.1. S<sup>3</sup>VM Model.** The objective optimization function of traditional SVM can be expressed as follows [36]:

$$J(m) = \frac{1}{2} \|m\|_2 + c \sum_{t=1}^T \sigma_t [1 - y_t \langle m, x_t \rangle], \quad (1)$$

where  $c$  represents the weight parameter,  $x_t$  represents the sample data,  $y_t$  represents the sample label,  $\sigma_t = 1$  represents

the marked sample, and  $\sigma_t = 0$  represents the unmarked sample.

S<sup>3</sup>VM extends the idea of maximizing the classification interval to semisupervised learning and comprehensively considers the role of marked samples and unmarked samples when maximizing the interval. The hat loss function is usually used in S<sup>3</sup>VM to describe the loss caused by unmarked data:

$$L_t(m) = [1 - |\langle m, x_t \rangle|]. \quad (2)$$

Then, the objective function of S<sup>3</sup>VM can be expressed as follows:

$$J(m) = \frac{1}{2} \|m\|_2 + \sum_{t=1}^T (c\sigma_t [1 - y_t \langle m, x_t \rangle] + \gamma(1 - \sigma_t) [1 - |\langle m, x_t \rangle|]), \quad (3)$$

where  $c$  and  $\gamma$  represent the weight parameters.

In order to adapt to online learning, this paper adopted a balanced penalty function to relax the constraint of the objective function on the decision boundary, and the objective function of S<sup>3</sup>VM can be described as follows:

$$J(m) = \frac{1}{2} \|m\|_2 + \sum_{t=1}^T (c\sigma_t [1 - y_t \langle m, x_t \rangle] + \gamma(1 - \sigma_t) [1 - |\langle m, x_t \rangle|]) + \mu \sum_{t=1}^T \left| \frac{1}{u_t} \sum_{i=t-\tau+1}^t (1 - \sigma_i) \langle m, x_i \rangle - \frac{1}{l_t} \sum_{i=t-\tau+1}^t \sigma_i y_i \right|, \quad (4)$$

where  $c$ ,  $\gamma$ , and  $\mu$  represent the weight parameters,  $P_t \subseteq \{1, 2, \dots, t\}$ ,  $u_t$ , and  $l_t$  are the number of unlabeled samples and labeled samples subscripted in  $P_t$ , respectively, and  $\tau$  represents the size of  $P_t$ .

Equation (4) can be further simplified as follows:

$$J(m) = \frac{1}{2} \|m\|_2 + \sum_{t=1}^T w_t(m), \quad (5)$$

where  $\sum_{t=1}^T w_t(m)$  represents the loss function of the sample  $(x_i, y_i, \sigma_i)$ .

The sample loss function  $w_t(m)$  can be decomposed into the sum of convex function  $w_t^{\text{vex}}(m)$  and concave function  $w_t^{\text{ave}}(m)$ , and the objective function of  $S^3\text{VM}$  can also be decomposed into the sum of convex function  $J^{\text{vex}}(m)$  and concave function  $J^{\text{ave}}(m)$ :

$$J(m) = \frac{1}{2} \|m\|_2 + \sum_{t=1}^T w_t(m) = J^{\text{vex}}(m) + J^{\text{ave}}(m) + \text{con}, \quad (6)$$

where con is a constant term, which will not affect the solution of minimizing  $J(m)$ , and this term can be ignored.

The optimal boundary vector can be defined as  $m^* = \underset{m}{\text{argmin}} J(m)$ ; by combining Equations (4) and (6), it can be obtained as follows:

$$(J^{\text{ave}})'(m^*) = \sum_{t=1}^T (w_t^{\text{ave}})'(m^*). \quad (7)$$

The predicted label based on the boundary vector  $m^*$  of the unlabeled sample  $x_t$  was defined as  $\bar{y}_t = \text{sign}(\langle m, x_t \rangle)$ . If there was a soft label  $\tilde{y}_t$  approaching the predicted label  $\bar{y}_t$ , then the  $S^3\text{VM}$  model can be achieved by minimizing the following equation:

$$J^{\text{vex}}(m) - \gamma \sum_{t=1}^T \tilde{y}_t \langle m, x_t \rangle. \quad (8)$$

In this paper, the latest boundary vector obtained in the previous learning process was used as the soft label of the unmarked sample.

**3.2.  $OS^3\text{VM}$  Algorithm Based on the Dual Lifting Process.** The  $OS^3\text{VM}$  algorithm mainly includes two processes. The first is the prediction of the soft labels of unlabeled samples, and the second is the lifting process of the dual function [37]. The dual function corresponding to Equation (8) can be expressed as follows:

$$\begin{aligned} D(\varphi_1, \varphi_2, \dots, \varphi_T, \theta) &= -\frac{1}{2} \left\{ \sum_{t=1}^T \left[ c\sigma_t \varphi_{t1} y_t x_t + \gamma(1 - \sigma_t)(\varphi_{t2} - \varphi_{t3} + \tilde{y}_t) x_t - \mu \theta_t \frac{1}{u_t} \sum_{i \in P_t} (1 - \sigma_i) x_i \right] \right\}^2 \\ &+ \sum_{t=1}^T \left[ c\sigma_t \varphi_{t1} + \gamma(1 - \sigma_t)(\varphi_{t2} + \varphi_{t3}) - \mu \theta_t \frac{1}{l_t} \sum_{i \in P_t} \sigma_i y_i \right], \quad (9) \end{aligned}$$

where  $\varphi_t = [\varphi_{t1}, \varphi_{t2}, \varphi_{t3}]$  ( $t \in \{1, 2, \dots, T\}$ ) represents the system vector corresponding to the loss function, and satisfies the constraint conditions  $\varphi_{t1}, \varphi_{t2}, \varphi_{t3} \in [0, 1]$ ;  $\theta = [\theta_1, \theta_2, \dots, \theta_T]$  represents the coefficient vector corresponding to the penalty function and satisfies constraint conditions  $\theta_1, \theta_2, \dots, \theta_T \in [-1, 1]$ .

Equation (9) indicates that the variables in the dual function corresponding to Equation (8) are actually a set of coefficient variables with constraints, and the function value of the dual function can be determined according to the value of the coefficient vector group  $(\varphi_1, \varphi_2, \dots, \varphi_T, \theta)$ . In addition, since the coefficient variables in the dual function  $D(\varphi_1, \varphi_2, \dots, \varphi_T, \theta)$  are independent of each other, the function value of the whole dual function can be improved only by changing the value of some coefficient variables, so as to solve the  $OS^3\text{VM}$  model update problem.

$(\varphi_i)_t$  and  $(\theta)_t$  represent the value coefficient vectors  $\varphi_i$  ( $i \in \{1, 2, \dots, t\}$ ) and  $\theta$  in the learning period  $t$ . According to the characteristics of the data flow in the online semisupervised learning process, the coefficient vector group  $(\varphi_1, \varphi_2, \dots, \varphi_T, \theta)$  should meet the following four conditions in addition to its own constraints during the learning cycle update process [38]:

- (1) For any  $i \in \{1, 2, \dots, T\}$ ,  $(\varphi_{i1})_t = 0$
- (2) For any  $i \in \{t+1, t+2, \dots, T\}$ ,  $(\varphi_{i2})_t - (\varphi_{i3})_t + \tilde{y}_i = 0$
- (3) For any  $i \in \{t+1, t+2, \dots, T\}$ ,  $(\theta_i)_t = 0$
- (4) The new dual vector group can improve the function value of the dual function, namely,

$$D((\varphi_1)_t, (\varphi_2)_t, \dots, (\varphi_t)_t, (\theta)_t) \geq D((\varphi_1)_{t-1}, (\varphi_2)_{t-1}, \dots, (\varphi_{t-1})_{t-1}, (\theta)_{t-1}). \quad (10)$$

According to the above analysis, the boundary vector  $m_t$  of the learning period  $t$  in the  $OS^3\text{VM}$  algorithm can be expressed as follows:

$$\begin{aligned} m_t &= \sum_{i=1}^t \left[ c\sigma_i (\varphi_{i1})_i y_i x_i + \gamma(1 - \sigma_i) (\varphi_{i2} - \varphi_{i3} + \tilde{y}_i) x_i \right. \\ &\quad \left. - \mu (\theta_i)_i \frac{1}{u_i} \sum_{j \in P_i} (1 - \sigma_j) x_j \right]. \quad (11) \end{aligned}$$

Considering the computational complexity and punishment function design, the  $OS^3\text{VM}$  algorithm proposed in this paper only used the samples marked in  $I_t = \{t - \tau + 1, t - \tau + 2, \dots, t\}$  to achieve dual promotion process in the learning cycle  $t$ . Therefore, the set of dual variables that can be updated in the learning period was  $\{\varphi_{t-\tau+1}, \varphi_{t-\tau+2}, \dots, \theta_t\}$ . At the same time, considering that the greedy promotion process would bring a greater degree of dual function promotion in each learning cycle, the paper proposed an  $OS^3\text{VM}$  algorithm based on greedy promotion. In the learning period  $t$ , the value of the dual function can be

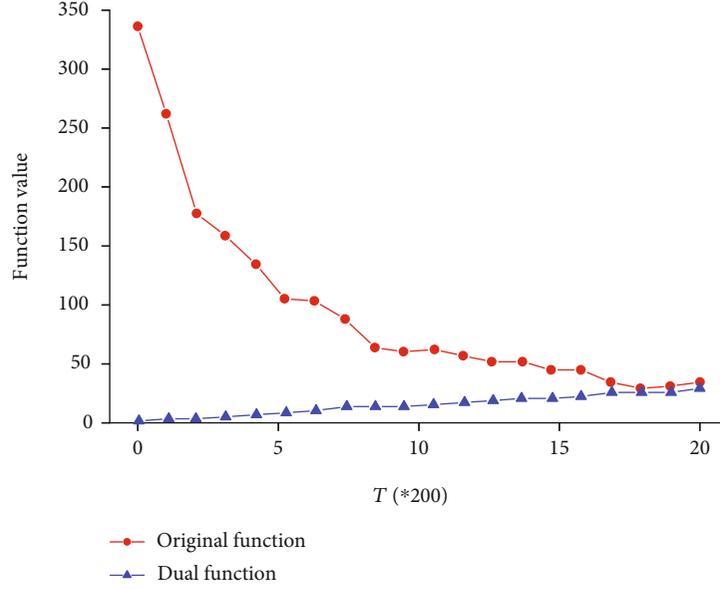


FIGURE 10: The change curve of the original function and the dual function.

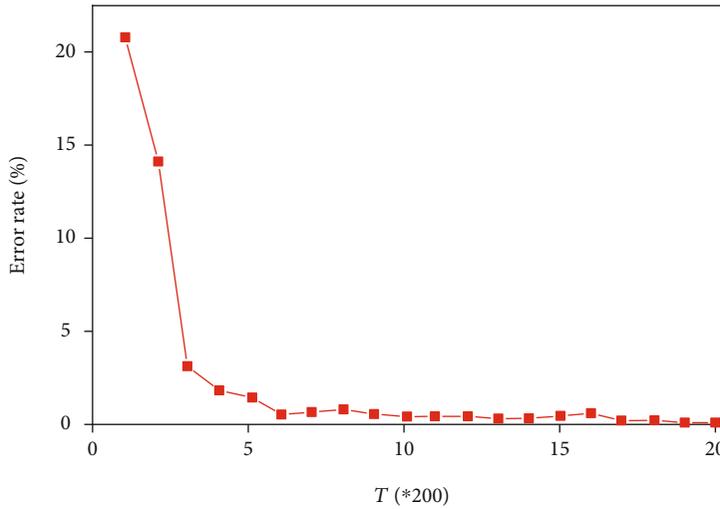


FIGURE 11: The error rate of boundary vectors in the whole data set.

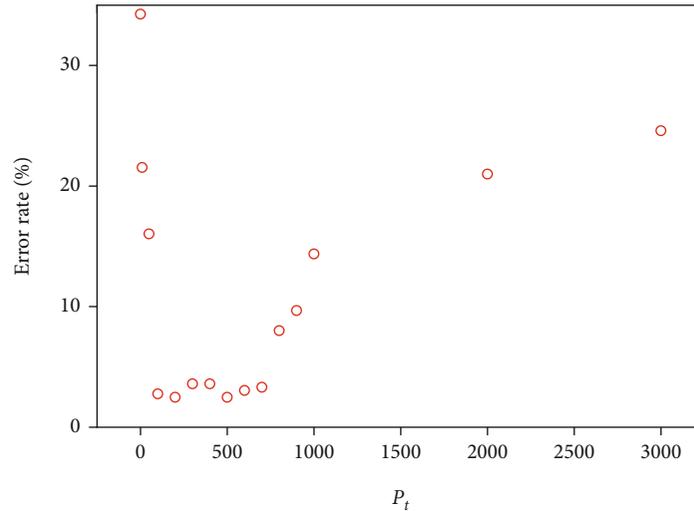
improved by solving the following quadratic programming (QP) problem [24]:

$$\begin{aligned}
 ((\varphi_1)_t, (\varphi_2)_t, \dots, (\varphi_i)_t, (\theta)_t) &= \underset{\varphi_2 \in [0,1]^3, \theta_i \in [-1,1]}{\operatorname{argmax}} D(\varphi_1, \varphi_2, \dots, \varphi_i, \theta) \\
 \text{s.t. } \forall i \notin P_t, \varphi_i &= (\varphi_i)_{t-1}, \forall i \neq t, \theta_i = (\theta_i)_{t-1}.
 \end{aligned} \tag{12}$$

Therefore, Equation (12) would have a maximum dual promotion in the dual variable set  $\{\varphi_{t-\tau+1}, \varphi_{t-\tau+2}, \dots, \theta_t\}$  after given a soft label, thus achieving the semisupervised support vector machine online learning process.

## 4. Simulation Results

In this article, we used MATLAB language for modelling; a total of 1800 groups of experimental data samples were obtained and randomly selected 10% to 70% of the samples (in 10% increments) without replacement as labeled samples  $l_t$  and then the remaining samples as unlabeled samples  $u_t$ . In addition, 30% of the unmarked samples  $u_t$  were randomly determined as the test sample. The SVM algorithm and  $S^3VM$  algorithm were employed to compare with the  $OS^3VM$  algorithm proposed in this study. The difference between a supervised model and a semisupervised model is that the SVM only uses the labeled samples  $l_t$  for training, while the  $S^3VM$  uses both marked samples  $l_t$  and unmarked samples

FIGURE 12: The impact of the size of  $P_t$  on the error rate of the algorithm.

$u_t$  for training. Based on the  $S^3VM$  model, the  $OS^3VM$  can update the trained model in real time. In order to process the large data streams in practical applications, it is also necessary to sparse the boundary vectors in the process of the  $OS^3VM$  algorithm. Therefore,  $k$ -maximum dual coefficient ( $k$ -MC,  $k = 400$ ) method was selected as the sparsification method in the experimental process. Since the  $OS^3VM$  algorithm only contains the inner product operation between sample points, the kernel function could also be introduced to find the linear classification surface. In this study, the standard RBF kernel function was determined to find the linear classification surface and its form can be expressed as follows:

$$K(x_i, x_j) = e^{-\|x_i - x_j\|^2 / 2\sigma_k^2}. \quad (13)$$

**4.1. Dual Lifting Process.** In order to further illustrate the effect of the dual lifting process, Figure 10 compares the change curves of the original function and the dual function in the  $OS^3VM$  algorithm. It can be seen from Figure 11 that the two curves are constantly approaching each other along with the algorithm process. The value of the dual function increases during the  $OS^3VM$  algorithm. In contrast, the value of the original function tends to decrease, and there will be some small fluctuations during the descending process. The results demonstrate the correctness of the proposed  $OS^3VM$  algorithm and the feasibility and effectiveness of the algorithm based on the dual promotion process. In addition, Figure 10 further reveals the error rate of boundary vectors  $m(t)$  in the process of the  $OS^3VM$  algorithm in the whole data set. The simulation result indicates that the proposed algorithm is a process of improving the performance of the predictor. Since the algorithm only uses the information of local samples to update the predictor, the effect of the predictor will inevitably produce some small fluctuations during the learning process.

In the  $OS^3VM$  algorithm,  $P_t$  not only controls the number of local sample points used in the balanced penalty func-

TABLE 3: Accuracy of pedestrian crossing intention recognition based on SVM,  $S^3VM$ , and  $OS^3VM$ .

Labeled sample proportion	SVM	$S^3VM$	$OS^3VM$
10	82.43	85.93	89.16
20	84.57	87.42	91.38
30	86.12	88.98	92.47
40	87.86	90.50	94.83
50	88.32	91.07	94.66
60	88.45	91.21	94.72
70	89.27	91.43	94.65

tion to punish the unbalanced division but also determines the range of sample points used in the dual promotion process, and therefore also controls the computational time complexity of the algorithm process in each learning cycle. Obviously, the larger the size of  $P_t$  is, the higher the computational time complexity of the algorithm process in each learning cycle is. Figure 12 shows the impact of the size of  $P_t$  on the error rate of the algorithm. The results demonstrate that if the  $\tau$  value is too small or too large, the classification effect of the algorithm will be worse. If the  $\tau$  value is too large, the large number of sample points used in the dual lifting process will hardly express the sparse region of the current sample distribution, thus making the learning effect worse. In other words, too large a value of  $\tau$  will make the algorithm unable to respond to changes in the data stream in time.

Based on the above analysis, the reasonable choice of  $P_t$  in the  $OS^3VM$  algorithm will be beneficial to improve the computational time complexity and the classification accuracy of the algorithm. Therefore, in the paper, the value of  $\tau$  was determined as 200.

**4.2. Comparison of Recognition Accuracy.** The comparison results of pedestrian crossing intention recognition accuracy of SVM,  $S^3VM$ , and  $OS^3VM$  are shown in Table 3 and Figure 13. The results indicated that, in the case of the same

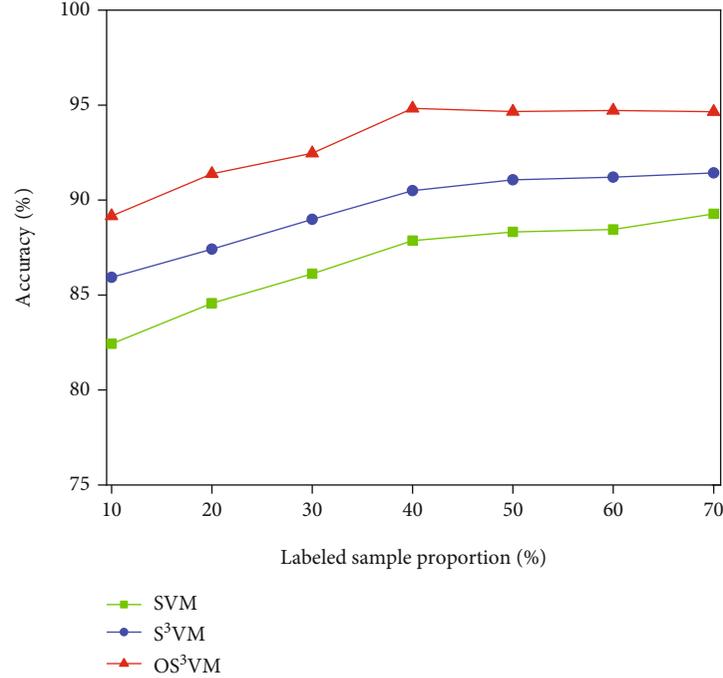


FIGURE 13: The comparison results of recognition accuracy of SVM, S<sup>3</sup>VM, and OS<sup>3</sup>VM.

proportion of labeled samples, the OS<sup>3</sup>VM algorithm exhibits a better ability to recognize pedestrian crossing intentions than the SVM and S<sup>3</sup>VM models. Under the condition of 10% marked sample ratio, the accuracy of pedestrian crossing intention recognition of OS<sup>3</sup>VM is higher than of the SVM model with 60% marked sample ratio and the S<sup>3</sup>VM with 30% marked sample ratio, and the accuracies are 89.16%, 88.45%, and 88.98%, respectively. It can be seen that the OS<sup>3</sup>VM algorithm can improve the recognition ability of pedestrian crossing intention by using the unmarked samples. In addition, when the proportion of labeled samples is greater than 40%, the established model tends to converge and the accuracy rate converges to 94%. As the proportion of labeled samples increases, the training time of the model will greatly increase. Therefore, comprehensively considering the training time and recognition accuracy, the paper determines that the best-labeled sample ratio of the OS<sup>3</sup>VM recognition model is 40%.

## 5. Conclusions

Autonomous vehicles need to understand pedestrian behavior in order to achieve better performance. Recognizing the pedestrian intention is one of the most critical capabilities for autonomous vehicles to ensure the safe operation of the urban environment. However, for autonomous vehicles, it is quite challenging to accurately identify pedestrians' crossing intentions, because they are affected by their emotions, traffic environment, road environment, and weather. At present, pedestrian crossing intention models have the problem that the models cannot be updated online in real time, which limits their applicability and generalization. To accurately identify pedestrians' crossing intentions, the model

needs to be able to update the model online in real time according to the diversity of the samples. To achieve this goal, this paper proposes a OS<sup>3</sup>VM. In order to verify the effectiveness of the model, this paper uses laser scanner to collect a large amount of pedestrian crossing data and vehicle movement data, and determines the input feature parameters of the model through statistical analysis and PCA feature extraction.

The semisupervised support vector machine is a type of semisupervised learning method based on low-density region segmentation. This type of method believes that the decision-making area should be located in some areas with low data density. Since the solution of S<sup>3</sup>VM algorithm is a nonconvex problem and difficult to handle, there is relatively little research work on online semisupervised support vector machine. This study proposes an OS<sup>3</sup>VM model based on dual promotion process to identify the pedestrian crossing intentions. Firstly, the hat loss function is employed to describe and define the basic learning problem of the S<sup>3</sup>VM model. Then according to the inspiration in the process of concave and convex, the nonconvex problem would be transformed to a convex problem. Therefore, an OS<sup>3</sup>VM model based on dual promotion process is established with using the greedy algorithm to achieve the improvement of the dual function. Finally, the online update of the classifier is completed. In order to verify the validity of the proposed algorithm, the SVM and S<sup>3</sup>VM models are established, respectively, and the accuracy of pedestrian crossing intention recognition of different models is compared under the different labeled sample proportions. The results demonstrate that the proposed OS<sup>3</sup>VM model can reduce the number of labeled samples for training the classifier and improve the recognition accuracy.

## Data Availability

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

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This work was supported in part by the Key Research and Development Program of Shaanxi under Grant 2020GY-173 and in part by the Fundamental Research Funds for the Central Universities, CHD 300102220220.

## References

- [1] Traffic Administration Bureau of the Ministry of Public Security of the People's Republic of China, *Road Traffic Safety Development Report*, Traffic Administration Bureau of the Ministry of Public Security of the People's Republic of China, Wuxi, China, 2019.
- [2] W. Chen, X. Zhuang, Z. Cui, and G. Ma, "Drivers' recognition of pedestrian road-crossing intentions: performance and process," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 64, pp. 552–564, 2019.
- [3] C. Chen, B. Liu, S. Wan, P. Qiao, and Q. Pei, "An edge traffic flow detection scheme based on deep learning in an intelligent transportation system," *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–13, 2020.
- [4] A. Rasouli and J. K. Tsotsos, "Autonomous vehicles that interact with pedestrians: a survey of theory and practice," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 3, pp. 900–918, 2020.
- [5] B. Yang and R. Ni, "Vision-based recognition of pedestrian crossing intention in an urban environment," in *2019 IEEE 9th Annual International Conference on CYBER Technology in Automation, Control, and Intelligent Systems (CYBER)*, pp. 992–995, Suzhou, China, 2019.
- [6] S. Kalantarov, R. Riemer, and T. Oron-Gilad, "Pedestrians' road crossing decisions and body parts' movements," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 53, pp. 155–171, 2018.
- [7] A. T. Schulz and R. Stiefelhagen, "A controlled interactive multiple model filter for combined pedestrian intention recognition and path prediction," in *2015 IEEE 18th International Conference on Intelligent Transportation Systems*, pp. 173–178, Las Palmas, 2015.
- [8] D. Varytimidis, F. Alonso-Fernandez, B. Duran, and C. Englund, "Action and intention recognition of pedestrians in urban traffic," in *2018 14th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS)*, pp. 676–682, Las Palmas de Gran Canaria, Spain, 2018.
- [9] B. Völz, K. Behrendt, H. Mielenz, I. Gilitschenski, R. Siegwart, and J. Nieto, "A data-driven approach for pedestrian intention estimation," in *2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*, pp. 2607–2612, Rio de Janeiro, Brazil, 2016.
- [10] K.-H. Park and S.-W. Lee, "Movement intention decoding based on deep learning for multiuser myoelectric interfaces," in *2016 4th International Winter Conference on Brain-Computer Interface (BCI)*, pp. 1–2, Yongpyong, South Korea, Feb. 2016.
- [11] H. Zhang, Y. Liu, C. Wang, R. Fu, Q. Sun, and Z. Li, "Research on a pedestrian crossing intention recognition model based on natural observation data," *Sensors*, vol. 20, no. 6, p. 1776, 2020.
- [12] F. Schneemann and P. Heinemann, "Context-based detection of pedestrian crossing intention for autonomous driving in urban environments," in *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 2243–2248, Daejeon, South Korea, 2016.
- [13] J. Zhao, Y. Li, H. Xu, and H. Liu, "Probabilistic prediction of pedestrian crossing intention using roadside LiDAR data," *IEEE Access*, vol. 7, pp. 93781–93790, 2019.
- [14] Z. Fang, D. Vázquez, and A. López, "On-board detection of pedestrian intentions," *Sensors*, vol. 17, no. 10, 2017.
- [15] J. Škovierová, A. Vobecký, M. Uller, R. Škoviera, and V. Hlaváč, "Motion prediction influence on the pedestrian intention estimation near a zebra crossing," in *Proceedings of the 4th International Conference on Vehicle Technology and Intelligent Transport Systems*, pp. 341–348, Madeira, Portugal, 16–18 March 2018.
- [16] R. Quintero Mínguez, I. Parra Alonso, D. Fernández-Llorca, and M. Á. Sotelo, "Pedestrian path, pose, and intention prediction through Gaussian process dynamical models and pedestrian activity recognition," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 5, pp. 1803–1814, 2018.
- [17] R. Quintero, I. Parra, J. Lorenzo, D. Fernández-Llorca, and M. A. Sotelo, "Pedestrian intention recognition by means of a hidden Markov model and body language," in *2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)*, pp. 1–7, Yokohama, 2017.
- [18] W. Shaohua, X. Xiaolong, W. Tian, and G. Zonghua, "An intelligent video analysis method for abnormal event detection in intelligent transportation systems," *IEEE Transactions on Intelligent Transportation Systems*, 2020.
- [19] R. Caruana and A. Niculescu-Mizil, "An empirical comparison of supervised learning algorithms," in *Proceedings of the 23rd International Conference on Machine Learning - ICML '06*, New York, 2006.
- [20] X. Zhu and A. B. Goldberg, "Introduction to semi-supervised learning," *Synthesis Lectures on Artificial Intelligence and Machine Learning*, vol. 3, no. 1, pp. 1–130, 2009.
- [21] C. Wang, Q. Sun, Y. Guo, R. Fu, and W. Yuan, "Improving the user acceptability of advanced driver assistance systems based on different driving styles: a case study of lane change warning systems," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 10, pp. 4196–4208, 2020.
- [22] S. Ding, S. Qu, Y. Xi, and S. Wan, "A long video caption generation algorithm for big video data retrieval," *Future Generation Computer Systems*, vol. 93, pp. 583–595, 2019.
- [23] A. B. Goldberg, M. Li, and X. Zhu, "Online manifold regularization: a new learning setting and empirical study," in *Machine Learning and Knowledge Discovery in Databases*, pp. 393–407, Springer, Berlin, Heidelberg, 2008.
- [24] B. Zeisl, C. Leistner, A. Saffari, and H. Bischof, "On-line semi-supervised multiple-instance boosting," in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, San Francisco, CA, 2010.
- [25] J. Zhao, J. O. Malenje, Y. Tang, and Y. Han, "Gap acceptance probability model for pedestrians at unsignalized mid-block

- crosswalks based on logistic regression,” *Accident; Analysis and Prevention*, vol. 129, pp. 76–83, 2019.
- [26] B. R. Kadali and P. Vedagiri, “Proactive pedestrian safety evaluation at unprotected mid-block crosswalk locations under mixed traffic conditions,” *Safety Science*, vol. 89, pp. 94–105, 2016.
- [27] D. Herrero-Fernández, P. Macía-Guerrero, L. Silvano-Chaparro, L. Merino, and E. C. Jenchura, “Risky behavior in young adult pedestrians: personality determinants, correlates with risk perception, and gender differences,” *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 36, pp. 14–24, 2016.
- [28] K. V. R. Ravishankar and P. M. Nair, “Pedestrian risk analysis at uncontrolled mid-block and unsignalized intersections,” *Journal of Traffic and Transportation Engineering (English Edition)*, vol. 5, no. 2, pp. 137–147, 2018.
- [29] X. Zhuang and C. Wu, “Modeling pedestrian crossing paths at unmarked roadways,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, no. 3, pp. 1438–1448, 2013.
- [30] Y. Hashimoto, Y. Gu, L. Hsu, and S. Kamijo, “Probability estimation for pedestrian crossing intention at signalized crosswalks,” in *2015 IEEE International Conference on Vehicular Electronics and Safety (ICVES)*, pp. 114–119, Yokohama, 2015.
- [31] S. Wan, R. Gu, T. Umer, K. Salah, and X. Xu, “Toward offloading Internet of vehicles applications in 5G networks,” *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–9, 2020.
- [32] H.-P. Deutsch, M. W. Beinker, H.-P. Deutsch, and M. W. Beinker, “Principal component analysis,” in *Derivatives and Internal Models: Modern Risk Management*, pp. 793–804, Palgrave Macmillan, Cham, 2019.
- [33] C. M. Martinez, M. Heucke, F. Wang, B. Gao, and D. Cao, “Driving style recognition for intelligent vehicle control and advanced driver assistance: a survey,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 3, pp. 666–676, 2018.
- [34] X. J. Zhu, *Semi-supervised learning literature survey*, University of Wisconsin-Madison Department of Computer Sciences, 2005.
- [35] L. Wang, H. Zhen, X. Fang, S. Wan, W. Ding, and Y. Guo, “A unified two-parallel-branch deep neural network for joint gland contour and segmentation learning,” *Future Generation Computer Systems*, vol. 100, pp. 316–324, 2019.
- [36] O. Chapella and A. Zien, “Semi-supervised classification by low density separation,” *Machine Learning*, 2004.
- [37] S. Shalev-Shwartz and Y. Singer, “A primal-dual perspective of online learning algorithms,” *Machine Learning*, vol. 69, no. 2–3, pp. 115–142, 2007.
- [38] S. Shalev-Shwartz, “Online learning and online convex optimization,” *Foundations and Trends in Machine Learning*, vol. 4, no. 2, pp. 107–194, 2011.