Hindawi
Scientific Programming
Volume 2021, Article ID 4520190, 13 pages
https://doi.org/10.1155/2021/4520190

Research Article

Recognition of Taxi Violations Based on Semantic Segmentation of PSPNet and Improved YOLOv3

Qiong Yang[1] **and Lifeng Yu**[1,2]

[1]*Department of Computer, Zhejiang Industry Polytechnic College, Shaoxing 312000, China*
[2]*School of Computer Science and Technology, Zhejiang University, Hangzhou 310027, China*

Correspondence should be addressed to Qiong Yang; yangqiong525@hdu.edu.cn

Received 28 May 2021; Revised 18 September 2021; Accepted 29 September 2021; Published 29 November 2021

Academic Editor: Jianping Gou

Taxi has the characteristics of strong mobility and wide dispersion, which makes it difficult for relevant law enforcement officers to make accurate judgment on their illegal acts quickly and accurately. With the investment of intelligent transportation system, image analysis technology has become a new method to determine the illegal behavior of taxis, but the current image analysis method is still difficult to support the detection of illegal behavior of taxis in the actual complex image scene. To solve this problem, this study proposed a method of taxi violation recognition based on semantic segmentation of PSPNet and improved YOLOv3. (1) Based on YOLOv3, the proposed method introduces spatial pyramid pooling (SPP) for taxi recognition, which can convert vehicle feature images with different resolutions into feature vectors with the same dimension as the full connection layer and solve the problem of repeated extraction of YOLOv3 vehicle image features. (2) This method can recognize two different violations of taxi (blocking license plate and illegal parking) rather than only one. (3) Based on PSPNet semantic segmentation network, a taxi illegal parking detection method is proposed. This method can collect the global information of road condition images and aggregate the image information of different regions, so as to improve the ability to obtain the global information orderly and improve the accuracy of taxi illegal parking detection. The experimental results show that the proposed method has excellent recognition performance for the detection rate of license plate occlusion behavior DR is 85.3%, and the detection rate of taxi illegal parking phenomenon DR is 96.1%.

1. Introduction

In recent years, with the rapid development of economy and urbanization, the total amount of roads and vehicles in various cities in China has also shown a trend of continuous growth [1]. Among them, as a convenient and efficient way of travel, taxis are also welcomed by the general public. But we have noticed that illegal parking of rental vehicles is also increasing. However, taxis have the characteristics of mobility, dispersion, and nonspecific service groups [2–4]. So relevant managers urgently need a fast and accurate method of judging taxi violations to facilitate the management of urban traffic and support the green and efficient demand for intelligent transportation.

The current existing supervision methods are mainly through manual screening. The staff conduct randomly checks on the past videos in the monitoring center database to observe whether the drivers spotted have any violations [5, 6]. However, due to human eye fatigue and lack of concentration, this method cannot guarantee continuity and reliability, and it also consumes a lot of manpower. Therefore, how to realize the automatic detection and recognition of taxi vehicle violations has become one of the research focuses and difficulties.

With the advent of the era of the Internet of Everything, the urban transportation network is also covered with a variety of heterogeneous detection terminals, forming a powerful Intelligent Transport System (ITS). ITS effectively integrates the Internet of Things, big data, cloud computing, and other high-tech technologies [7, 8]. It constitutes a smart city traffic monitoring system for terminal situation awareness-cloud decision analysis. This also provides new ideas for the detection of taxi violations. The ITS terminal monitor collects road condition images in real time and

uploads them to the cloud for accurate analysis of taxi violation images [9, 10]. Therefore, an efficient vehicle parking violation image detection method is particularly important.

As a fusion product of big data technology and artificial intelligence technology, deep network provides a new solution for the detection of complex traffic image violations [11, 12]. The deep network model uses the multilayer network structure to train and learn the image data set collected by ITS through the continuous training and learning of the multilayer network to realize effective judgment of violations. But it should be noted that due to the multilayer network learning mechanism of the deep network, the network parameter setting is cumbersome, and there are many elements in need of being processed in video surveillance, such as vehicle classification, road segment discrimination, license plate recognition, and vehicle speed detection [13, 14]. However, current taxi violation detection methods can only analyze simple images. For complex images, it is difficult to achieve orderly and effective extraction of image features.

In view of the low performance of current vehicle violation detection methods, this study proposed a new taxi violation detection method based on the improved YOLOv3 network and PSPNet network. The content is as follows:

(1) In this study, the spatial pyramid pooling of semantic segmentation is merged into the traditional YOLOv3 network. Based on the improved YOLOv3 network, a new method for detecting taxi violations is proposed. The problem of information distortion in the collected images is avoided, and accurate detection of taxi and taxi license plate occlusion behavior can be realized.

(2) This study proposed a detection method for illegal taxi parking based on PSPNet semantic segmentation network. It can comprehensively collect the global information of the road condition image to realize the orderly extraction of the image characteristics of the violation behavior. It further enhances the accuracy of taxi parking violation detection.

The rest of this article is organized as follows. The first section introduces corresponding researches on vehicle image detection. The second section shows the taxi recognition method based on the improved YOLOv3 network model. The third section continues to introduce the method of identifying illegal taxi license plate occlusion based on the improved YOLOv3 network model. The fourth section dicusses the PSPNet network model and the detection method of violation behavior. The fifth section realizes the simulation verification of the proposed taxi violation detection method based on the actual traffic collected images. The sixth section concludes the whole article.

2. Related Works

As the main mode of urban travel, taxis have strong mobility. At the same time, due to the randomness of the service target, the phenomenon of illegally rolling over the solid line parking is more likely to occur when taxis pick up passengers. In addition, some taxi owners also hide their license plates in order to avoid legal liability [15]. This has further increased risk factors of urban traffic travel. Therefore, effective taxi detection and analysis methods are particularly important to ensure urban traffic safety.

Traditional vehicle detection methods are based on moving target detection technology, including methods such as ViBe background interframe difference and codebook modelling [16]. Among them, the interframe difference method is one of the most commonly used methods for moving target detection and segmentation. The pixel-based time difference is used between two or three adjacent frames of the image sequence, and the motion area in the image is extracted through the closed value. Cao et al. [17] used unsupervised static recognition and dynamic tracking methods for vehicle dynamic tracking. However, it should be noted that if the color of the moving target is close to the video background color, or the moving speed of the moving target is relatively slow, the frame difference method will often detect a more obvious hole. Even the moving target is regarded as noise, resulting in missed detection [18]. Aiming at jittery traffic video, Xu et al. [19] proposed a hybrid algorithm of codebook algorithm and binary mode (LBP). Among them, the codebook modelling used the codebook to represent the background pixels according to the color distortion degree of the continuous sampling values of the pixels and the brightness range thereof. Then, used the background difference method to compare and judge the new input pixel value and its corresponding codebook. Combined with the advantages of the binary mode, the foreground target pixels can be extracted. But the disadvantage of this method is that it is difficult to balance noise and foreground holes. Moutakki et al. [20] also used the codebook background analysis method to realize the positioning of the vehicle in the image. However, multiple modules have been added, such as vehicle segmentation, vehicle classification, and vehicle counting.

As a product of computer technology in a new generation, deep networks have achieved good applications in many fields due to their powerful image processing capabilities [21, 22]. At present, relevant researchers in the field of intelligent transportation are also paying attention to it. Some improve the YOLOv2 network based on the residual network and use multiscale information to improve the accuracy of target detection [23]. Also, based on the Elu activation function, the Kelu activation function is designed to ensure the accuracy of license plate detection. Reference Tang et al. [24] proposed an improved, single-shot, multiframe detector based on deep learning. The attention mechanism is introduced through the spatial transformation module, so that the neural network can actively perform spatial transformation on the feature map. Also, adding context information transmission in the designated layer can achieve accurate illegal parking detection. Abbas [25] constructed a pretrained convolutional neural network model with a four-layer architecture and used Shenxin network model to detect vehicle overlimit, speed limit overlimit, and yellow line driving and other illegal phenomena. Liu et al. [26] realized effective vehicle tracking in

CNTK toolkit based on Fast-RNN network model and realized the identification of parking violations. Although the above method is based on the multilayer network structure to realize the effective recognition of the vehicle, it does not comprehensively analyze the characteristic information of the collected image. There is a lot of background noise in the actual collected road condition images. If the image features can be extracted effectively and orderly [27], the accuracy of the detection and recognition of illegal taxi behaviors will be more accurate. The technical differences of some literatures are summarized in Table 1.

In response to the above-mentioned problems, this study proposed a new method for detecting violations of taxis based on the improved YOLOv3 network model and the PSPNet network model. It can further extract the effective information in the actual road condition images to realize efficient image detection and analysis.

3. Methodology

The inspiration of this method comes from modular programming, which divides the whole research into several modules, including analyzing each module, considering the reuse technology between modules, and improving and optimizing the module appropriately, so as to obtain better system performance.

The whole framework of this method is shown in Figure 1. As shown in Figure 1, this method is mainly divided into three modules: taxi detection module, license plate occlusion detection module, and taxi illegal parking detection module. Because the research object of this study is one taxi, it is necessary to first detect whether the vehicle is a taxi. In the first module, this article improved the traditional YOLOv3 network structure (as shown in Figure 2), enhanced the expression ability of the network for feature information, and realized the efficient detection of taxis. The second module is based on the first module and realized the behavior judgment of illegal shielding of license plate based on the improved YOLOv3 model. The methods of shielding license plate include full shielding and partial shielding, and the process is shown in Figure 3. The third module is to detect the illegal stop behavior. Its method framework is shown in Figure 4. The detection methods include offline detection and real-time detection.

The main problems solved by the three modules are as follows: (1) how to accurately detect that the vehicle in the current frame is a taxi. (2) How to judge the illegal act of license plate occlusion. (3) How to quickly and accurately detect taxi parking violations. The uniqueness of our research is to detect multiple illegal acts of taxis, and the module can be reused for different situations.

4. Taxi Detection Based on Improved YOLOv3

In this study, the traditional YOLOv3 network structure is improved for the higher accuracy requirements of the test in the actual road scene in order to enhance the network's ability to express feature information and then to implement efficient taxi detection based on the improved YOLOv3.

The YOLOv3 network mainly includes 3 branches. The original image of the video to be detected is used as input, and the result is 3 feature maps with different resolutions. The branch at the highest level processes the original image using a multilayer convolution method, and the resulting feature-map resolution is usually 13×13, and the number of channels is usually 256. The branch in the middle layer will obtain the result of the middle convolution calculation of the high-level branch and use the upsampling method to connect the convolution results of the original image. The resulting feature map has a resolution of 26×26, and the number of channels is 256. After the branch at the bottom layer obtains the convolution result of the middle layer branch, the same operation method as the above middle layer is performed. The final feature-map resolution is usually 52×52, and the number of channels is usually 256. In the YOLOv3 network, the convolutional neural network uses residual connections to effectively alleviate the problem of gradient disappearance in the training phase. In the upsampling process, the intermediate convolution result is enlarged by 2×2, so that the intermediate convolution result can be better connected with the direct convolution result of the original image.

Finally, after obtaining three feature maps with different resolutions, the target frame prediction strategy of anchor points is adopted. That is to predict the location range of abnormal behaviors through a fully convolutional neural network. In this process, s kinds of scale anchor points need to be set for all possible target positions. The length and width (x, y, h, w) of the rectangular box and the coordinates of its upper left corner are the results that need to be output. In addition, the confidence of the target frame is the estimated range of the candidate frame. If the search range of the target is the grid area of $L \times L$, it can be obtained that the resolution of the final output decision diagram is $L \times L$ and the number of channels is $5 \times s$.

In the YOLOv3 model, its network weights are generally obtained through the training and testing of the ImageNet data set. The target of this data set is quite different from the abnormal behavior to be detected. Therefore, the YOLOv3 algorithm and the feature map are improved, the receptive field is expanded, and the loss of semantic features is avoided as much as possible.

In order to strengthen the detection of small targets, YOLOv3 algorithm draws on Feature Pyramid Network (FPN). The high-level feature and the shallow feature information are fused, and the multiple-scale fusion method is used to perform position and category prediction on multiple-scale feature maps [28]. However, the three-scale feature fusion method adopted by the YOLOv3 network structure has an adverse effect on the detection of smaller targets in the surveillance video. The semantic loss of the 13×13 feature map is serious, which can easily cause the loss of small targets, considering that the resolution of the feature will directly affect the detection of small targets and the overall performance index. Therefore, on the basis of Darknet-53, the three-scale resolutions of the original feature map of 13×13, 26×26 and 52×52 are modified to two larger-scale resolutions of 26×26 and 52×52. The network

Table 1: Technical differences in some relevant literature.

Ref.	Types of violations detected	Main technologies used	Need more manual feature selection?
[17]	1	Unsupervised static recognition	No
[20]	1	Code book model	Yes
[23]	1	YOLOv2	Yes
[24]	1	SSD	No
[25]	3	Pretrained convolutional neural network	No
[26]	1	Fast-RNN	No

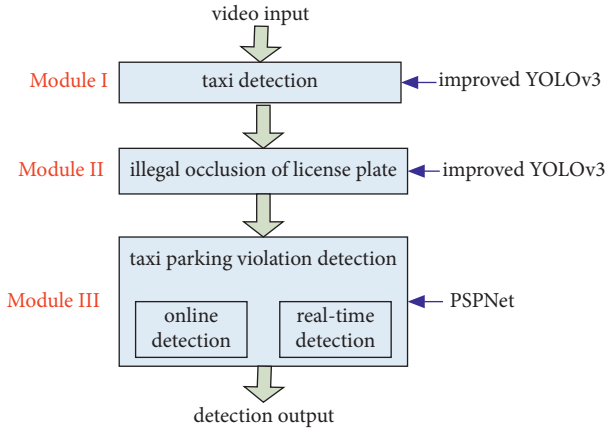

Figure 1: The overall flow chart of the proposed method.

structure is shown in Figure 2. In order to avoid image distortion caused by operations such as image scaling, stretching, and cropping, the Spatial Pyramid Pooling (SPP) of semantic segmentation is introduced into it, and the module structure of SPP is shown in Figure 5. Through the SPP module, the feature map of any resolution can be converted into the designed feature vector of the same dimension as the fully connected layer. At the same time, SPP solves the problem of repeated extraction of image features, so it also improves the calculation speed. Its specific effects and operations can be found in the reference.

It is worth mentioning that the feature map resolution of the original network structure of 26×26 is used as the first scale; the L61 (61 layer) result is subjected to five convolution operations. First, in order to improve computational efficiency, 1×1 convolution operation is performed to reduce dimensionality, then upsampling, and then fused with L36 (36 layers). After the final fusion, a 3×3 convolution kernel is used to convolve the fusion result. The purpose is to eliminate the aliasing effect of upsampling, so a new feature map of 52×52 is obtained as the second-scale feature. The improved YOLOv3 network has the advantages of high resolution and large receptive field, which can effectively improve the reliability of small target detection.

Figure 6 is a schematic diagram of some samples of the training set. The trained classifier is tested. The test video is collected from the traffic law enforcement camera and included 96 hours. The experimental results are shown in Tables 2 and 3.

From Tables 2 and 3, it can be seen that the taxi face classifier has the best detection effect when the positive samples are normalized to 48×36 and the number of iterations is 20, with a precision rate of 94.8%. The taxi body classifier has the best detection effect when the positive sample is normalized to 72×4, and the number of iterations is 20, with a precision rate of 96.8%.

5. Detection of Illegal Acts Concealed by Taxi License Plates

On the basis of successfully inspecting the taxi, we continue to implement the behavior determination of illegal license plate occlusion based on the improved YOLOv3 network model. By analyzing the traffic monitoring video, it can be known that the ways that taxis conceal the license plate include full occlusion and partial occlusion and then the license plate occlusion determination algorithm is correspondingly designed. The flow chart is shown in Figure 3.

Firstly, the license plate area is determined according to the positional relationship between the license plate area and the vehicle face area. Secondly, we can extract the blue pixels of the license plate area and generate the smallest bounding rectangle. Finally, the corresponding judgment method can be set according to different occlusion methods. The details are shown in Table 4.

6. Taxi Parking Violation Detection

As a traditional deep semantic segmentation network, ResNet can achieve effective information extraction from simple images. However, for actual traffic scenes, there are too many elements in the image, and it is difficult to ensure the accuracy of the results using the ResNet network model for semantic recognition. The PSPNet network integrates the pyramid pooling module with the ResNet network, which can aggregate image information from different regions, thereby improving the ability to obtain global information in an orderly manner. Therefore, this article is based on the PSPNet network to test the illegal parking behavior of taxis.

6.1. PSPNet Offline Detection. Figure 4 shows the overall structure of the PSPNet network model. The PSPNet network model uses the convolutional neural network (CNN) model to extract the features of the input image and sends the feature map to the pyramid pooling model.

As shown in Figure 4, in order to extract the multiscale information of the image, the feature map of any size is converted into a fixed-length feature vector. The model combines four parallel pooling features of different scales. In order to extract global features, a 1×1 convolution is used after the pooling operation of each scale to reduce the channel of the corresponding level to 1/4 of the original. It is then restored to

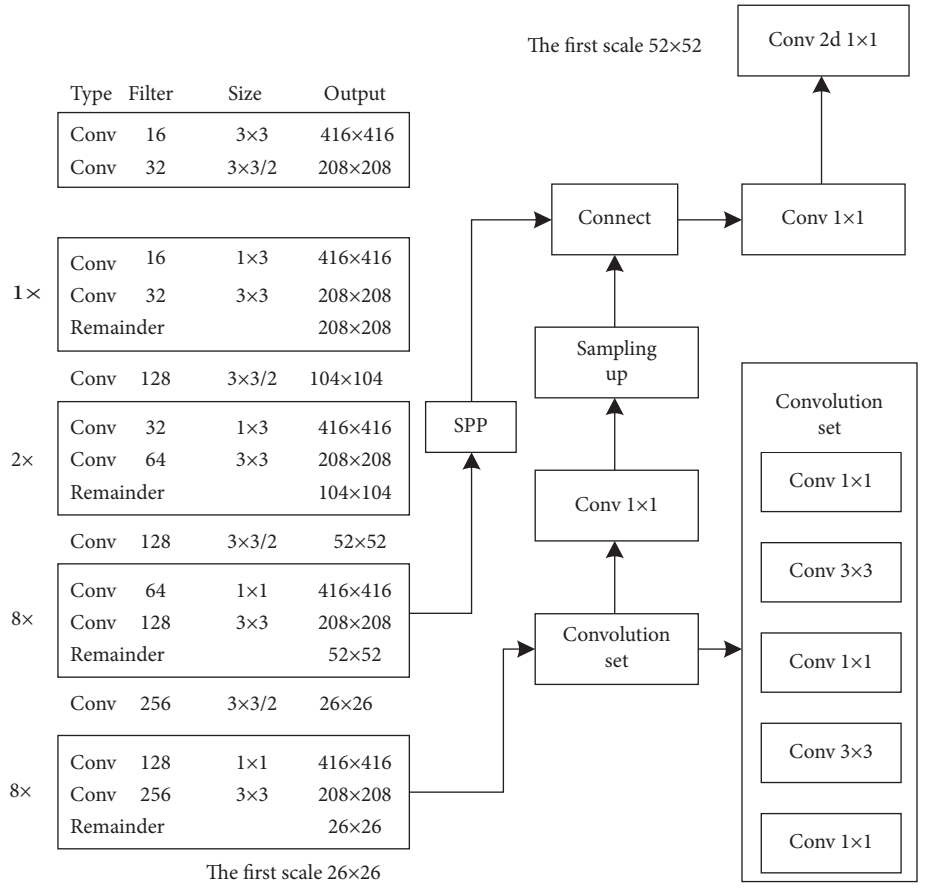

Figure 2: Improved YOLOv3 network structure.

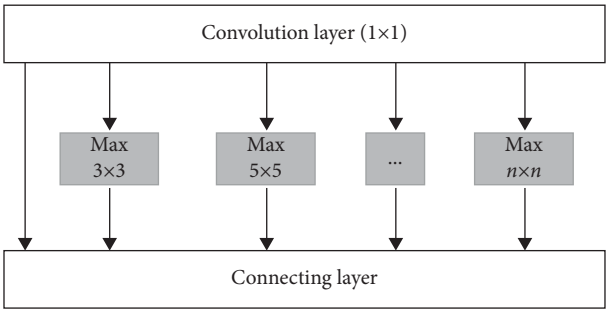

Figure 3: Flow chart of license plate illegal occlusion inspection.

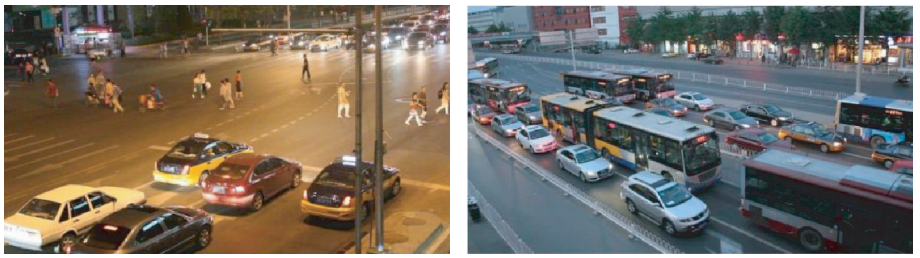

Figure 4: PSPNet network architecture.

the size before pooling through bilinear interpolation and connected with the feature map before pooling. Finally, a convolutional layer is used to generate the final prediction result. The spatial pyramid pooling model integrates local and global information and uses different spatial information to understand the actual road scene as a whole.

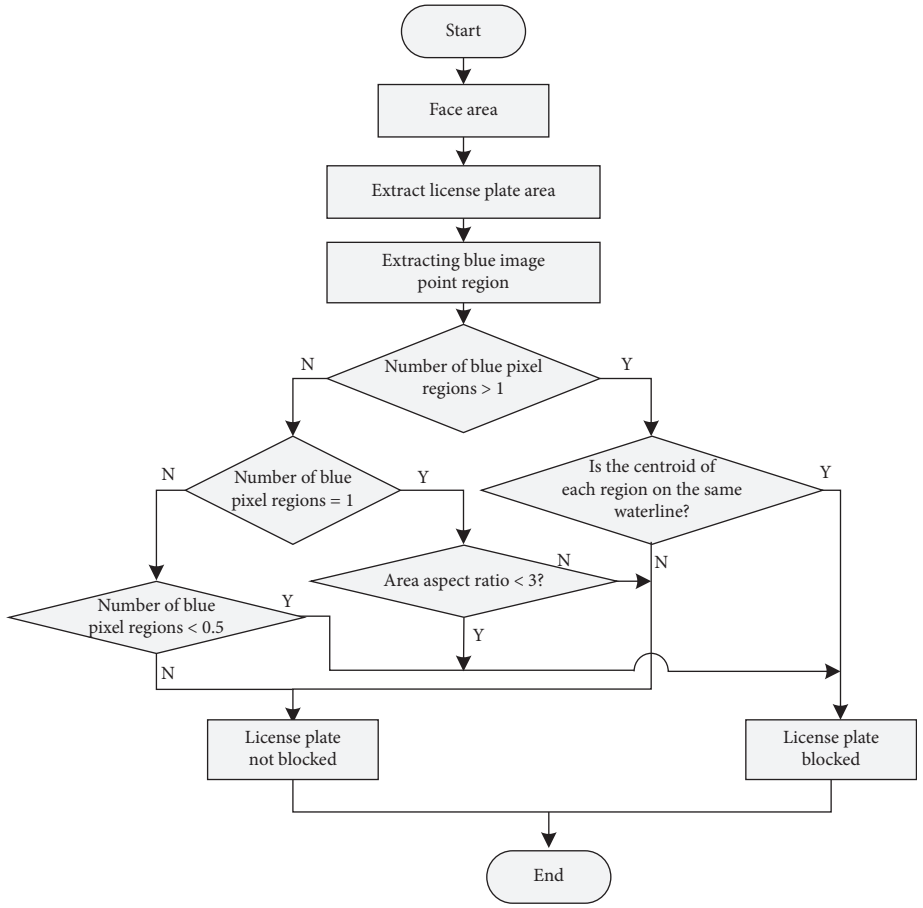

Figure 5: Module structure of SPP.

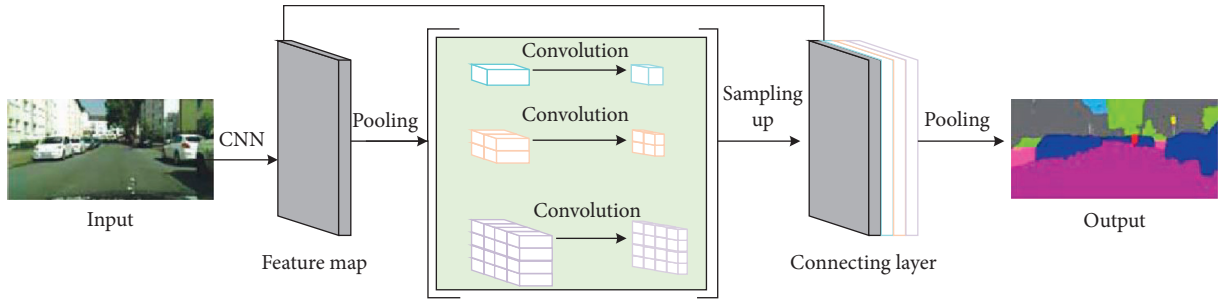

Figure 6: Schematic diagram of training samples.

Table 2: Precision rate of taxi face.

Iterations	Normalized size (%)			
	52×40	48×36	32×32	27×27
12	88.5	91.2	87.9	85.3
14	90.0	92.7	88.1	86.7
16	90.3	93.4	89.3	88.7
18	91.5	94.5	90.4	89.2
20	92.1	94.8	90.9	90.0

Based on target detection and semantic segmentation, this study proposed a sidewalk parking detection method. The same image is sent to the two subnetworks of target detection and semantic segmentation. On one hand, the target detection network outputs a detection image that contains all the vehicles in the image. On the other hand, the semantic segmentation network outputs a semantic segmentation map to divide the sidewalk and the road

Table 3: Precision rate of taxi body.

Iterations	Normalized size (%)			
	52 × 52	72 × 64	48 × 48	27 × 27
12	91.3	92.4	91.0	89.3
14	92.6	93.5	92.3	90.1
16	93.7	95.0	93.1	91.3
18	94.4	96.3	94.6	92.9
20	94.9	96.8	95.1	93.5

Table 4: Illegal occlusion mode and judgment mode of license plate.

Illegal occlusion of license plate	Judgment method
Total occlusion	Proportion of blue pixels
Single area partial occlusion	Aspect ratio of blue area
Multi area partial occlusion	Relative position of centroid in blue area

area in the urban road. For each category in the target category detection frame, in order to highlight its position information, the lower half of it is first intercepted, and then, it is compared with the segmentation map. After that, the overlap area A_{sidewalk} of the lower part of the detection frame with the sidewalk and the overlap area A_{road} with the road, respectively, and the overlap ratio according to the following formula can be calculated. If $P > 1$, it is judged as sidewalk parking [25].

$$P = \frac{A_{\text{sidewalk}}}{A_{\text{road}}}. \tag{1}$$

The traditional cross entropy loss function will cause the network to tend to easy-to-learn samples due to the accumulation of simple pixels and a large number of types of pixels, and the large accumulation of classification loss errors. As a result, the network learns better and better for simple types and a large number of samples, but it learns worse and worse for complex types and a small number of samples. This creates a vicious circle of training, which is not in line with the original intention of model training. In order to solve this problem, we carry out weight control on the types of imbalances, and the hierarchical loss calculation is given as follows [12]:

$$L_{\text{loss}} = \frac{1}{n}\sum_{i=0}^{n} \mu_i \lg S_i, \tag{2}$$

where S_i is the value of the Softmax function. μ_i is an adjustable parameter. i is the number of training sessions.

Changing μ_i controls the proportion of different types of pixel loss error in the total error. However, this approach only controls the balance of the proportions of unbalanced samples, and fundamentally, it is still impossible to distinguish difficult samples. In the middle and later stages of training, the gradient is still updated in the direction of sample types that are easier to learn. Therefore, it cannot promote the neural network to learn from difficult samples. The larger the output probability value of the Softmax function, the greater the probability that it is a certain category. At this time, the network has a relatively high degree of credibility in determining the pixel, that is, the input at this time is a simple sample. Conversely, when the output probability of the Softmax function is small, it means that it is difficult for the network to distinguish the exact category of the input. That is, the input at this time is a difficult sample. Therefore, the distinction between simple and difficult samples can be determined based on the output probability of the Softmax function. With reference to the Focal Loss function, we adopt the form of fusing multilevel loss and take the average and propose an inhibitory cross entropy loss function to solve the problem of sample imbalance in autonomous driving. The function is defined as follows [10].

$$L_{\text{ICEloss}} = -\frac{1}{n}\sum_{i=0}^{n} \mu_i (1 - S_i)^{\gamma} \log S_i, \tag{3}$$

where γ is an adjustable parameter.

When a sample is easier to distinguish, S_i is larger, and then, $(1 - S_i)^{\gamma}$ is smaller. The product of the two is equivalent to suppressing the loss of the sample, and the proportion of the total loss error is also smaller. Relatively speaking, if the loss error value of the difficult sample is amplified to a certain extent, its proportion in the total loss error will increase. The model will also be more inclined to learn difficult samples.

It should be noted that the PSPNet semantic segmentation network cannot achieve real-time detection. Therefore, this article will discuss how to perform real-time sidewalk parking violation detection. The network structure diagram of offline detection proposed in this study is shown in Figure 7.

6.2. Real-Time Detection. We have found that for the same fixed camera, the pictures after semantic segmentation are very similar. Therefore, only one semantic segmentation is required for the road background extracted by each camera. The semantic segmentation map can be provided for all subsequent detections. At the same time, the real scenes in surveillance videos are often intricate and fickle. Background modelling of surveillance scenes is the basis for subsequent processing, such as target detection, segmentation, tracking, classification, and behavior understanding. This article first

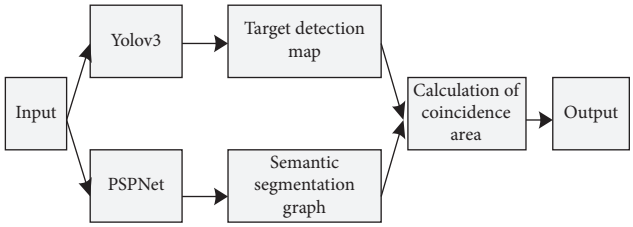

Figure 7: Network structure of illegal parking detection.

uses the Gaussian mixture background model to extract a clean road background from a video. Semantic segmentation of the background image is more conducive to locating the location and space information of the vehicle.

Gaussian mixture background modelling is a background representation method based on the statistical information of pixel samples. It is considered that the color information between pixels is not related to each other, and the processing of each pixel is independent of each other. For each pixel in the video image, the change of its value in the sequence image can be regarded as a random process of continuously generating pixel values, that is, Gaussian distribution is used to describe the color presentation law of each pixel (single peak, multiple peak). So, the mixed Gaussian background model can overcome the problems of image jitter, noise interference, light changes, and moving target movement and extract a clean background from the video stream.

In this study, a mixture of Gaussian background model training is performed on a video of about two minutes, and a relatively clean background image is extracted. Semantic segmentation is performed on the video screenshot and background image, respectively. As shown in Figure 8, the above image is a screenshot of a certain frame of the video and its semantic segmentation diagram. The following figure shows the background extracted by the Gaussian mixture model and its semantic segmentation. It filters all vehicles in motion, and the generated images are more suitable for semantic segmentation. Before using the convolutional neural network to perform semantic segmentation on the image, the mixed Gaussian background modelling is used to preprocess the image, which can achieve a better segmentation effect.

Figure 9 is a diagram of the real-time detection network structure of the Gaussian mixture background model. In the offline part, for a fixed camera, the first step is to input a piece of video into the Gaussian mixture background model to extract a clean background image. Then, the background image is sent to PSPNet for semantic segmentation, and the semantic segmentation map is obtained. In the real-time part, YOLOv3 detects the vehicle in the surveillance video in real time and compares it with the semantic segmentation map to calculate the overlap area of the target and the region. Finally, the result of sidewalk parking detection is output.

7. Experiment and Analysis

In this section, to verify the superiority of the proposed method's recognition performance, based on the works of some authors [23, 25, 26] as a comparison method, a simulation experiment for the identification of taxi violation behaviors is realized under the same experimental scenarios and conditions.

Experimental evaluation indicators are receiver operating characteristic (ROC) curve and Equal error rate (EER).

The ROC curve takes the false-positive rate P_{fp} as the horizontal axis, and the true-positive rate $P_{tp}$is the image obtained on the vertical axis. It can intuitively reflect the relationship between the false-positive rate and the true-positive rate and then judge the pros and cons of the model.

$$P_{\text{fp}} = \frac{\text{FP}}{\text{FP} + \text{TN}},$$
$$P_{\text{tp}} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \tag{4}$$

where TPandFPrepresent the abnormal samples detected correctly and incorrectly, respectively. The true-negative TN and false-negative FN, respectively, represent the negative and normal samples detected correctly and incorrectly.

The frame level standard is evaluated using EER, that is, the value when the error acceptance rate and the error rejection rate are equal. The lower the EER, the better the performance. The pixel-level features are evaluated by the detection rate (DR). In the pixel-level standard, abnormal frames are detected only when 40% of the abnormal behavior area is recognized. The higher the DR, the better the performance.

7.1. Configuration and Parameters. The taxi violation detection experiments are all run in the same environment. The hardware environment configuration is as follows: CPU is Intel Core i7 1165G7 processor, graphics card is NVIDIA Geforce MX450, memory is 16 GB, and video memory is 12 GB. The experimental operating system is the Linux operating system (Ubuntu 16.04.4 version), and the development environment is the Pytorch 1.0 deep learning framework.

The experimental data set comes from the actual images collected by the ITS system in a certain city. The experiment assigns the sample data set randomly. It is divided into violation training data set and experimental data set. The training data set accounts for 80%, and the experimental data set accounts for 20%. In the proposed method, the YOLOv3 network model uses the Darknet-53 network. The experiment uses the weight coefficients trained on ImageNet as the pretraining weights. Also, on the basis of it, transfer learning is realized with the taxi vehicle sample data set. Fine-tuning the network parameters to ensure that the loss function can achieve effective convergence.

Limited by the hardware memory capacity, this article sets the batch size to 20, a total of 120 epochs for training, and all training optimizers use Adam optimizers. The learning rate decline curve adopts a fixed long decline curve. The training process adopts freezing training, which means that in a certain training generation, the parameters of the backbone feature extraction network are not updated, and

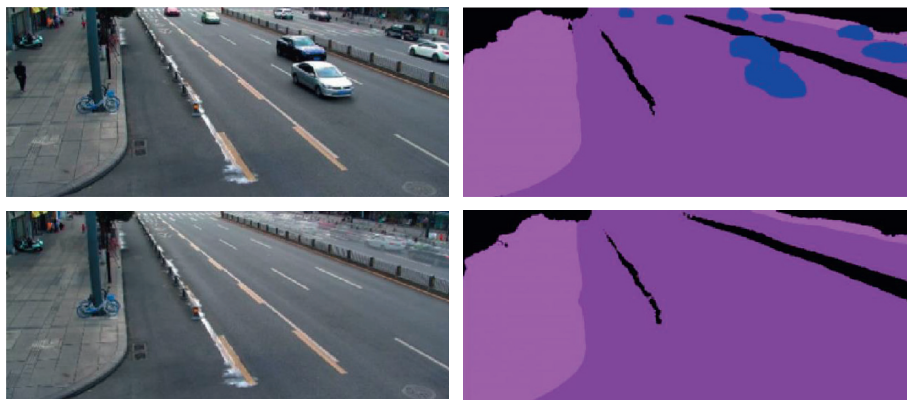

Figure 8: Semantic segmentation of Gaussian mixture model.

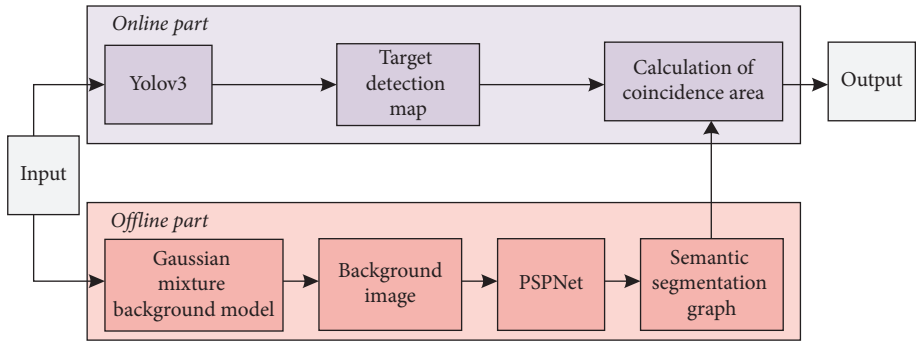

Figure 9: Network structure of taxi violation detection.

only the parameters of the prediction network are updated. After thawing, the next step is to update the parameters of the backbone feature extraction network and the prediction network.

7.2. Taxi Detection. This study first realized the comparison of the taxi detection test based on previous studies [23, 25, 26]. The results are shown in Table 5.

As shown in Table 5, the method proposed in this article can effectively identify and determine taxis. The accuracy rate of vehicle face recognition is 94.8%, which is 2.3% higher than that reported by Abbas [25], which is similar to the accuracy reported by Liu et al. [26]. For vehicle body recognition, the accuracy of all recognition methods is higher than that of the vehicle face. The accuracy of the vehicle body of the method proposed in this article is 96.8%, which is 1.9% higher than repoted by Abbas [25], and is also similar to that reported by Liu et al. [26]. This also shows that the addition of the SSP module can effectively support the YOLOv3 network to achieve accurate taxi identification.

7.3. Taxi License Plate Occlusion Detection. In order to verify the efficiency of the proposed method for the detection of illegal taxi obscuration license plates, this study conducted simulation experiments based on the video collected from actual traffic scenes. The experimental data set contains 700 experimental sequences; of which, a total of 200 data images of illegal occlusion of license plates is included. The results of license plate violation detection under different recognition methods are shown in Table 6.

As shown in Table 6, the method proposed in this study can achieve more accurate license plate occlusion detection, and the accuracy rate of license plate occlusion recognition in actual scenes is 95.1%. Compared with the report of Liu et al. [26], the accuracy is improved by 1.1%, which further proves that the method proposed in this study is superior to the existing vehicle state recognition methods.

At the same time, we also conduct ROC curve analysis on random areas in the license plate violation recognition experiment. The ROC results of different methods are shown in Figure 10.

It can be seen from Figure 10 that the method proposed in this study has a significantly higher true-positive rate than the comparison method in the ROC curve graph, and the convergence speed is faster than the comparison method, showing excellent detection performance. At the same time, the results of EER and DR of the vehicle license plate detection experiment are shown in Table 7.

It can be seen from Table 7 that compared with other methods, the detection performance of the method proposed in this study has been significantly improved. Compared with rate determined by Zhang et al. [23], the detection rate DR of the proposed method increases by 3.5%. This shows that the method in this study is better in the detection of license plate occlusion and can achieve the best effect in the scene of detecting smaller targets. At the same time, because

Table 5: Taxi detection results under different methods.

Method	Item (%)	
	Accuracy rate of vehicle face	Accuracy rate of vehicle body
The proposed method	94.8	96.8
Zhang et al. [23]	93.2	95.3
Abbas [25]	92.5	94.9
Liu et al. [26]	94.1	96.2

Table 6: License plate violation detection results under different methods.

Method	Accuracy rate (%)
The proposed method	95.1
Zhang et al. [23]	93.5
Abbas [25]	92.7
Liu et al. [26]	94.0

some false alarms will be generated in the methods proposed in earlier sudies [25, 26], the obtained equal error rate EER value is relatively higher. The proposed method uses an improved YOLOv3 network to generate candidate regions, and each candidate region contains the entire object. In addition, the addition of the SSP module further improves the accuracy of positioning and eliminates invalid feature information. Therefore, the obtained EER value is the lowest.

7.4. Taxi Parking Violation Detection. In addition, we also have carried out corresponding taxi parking violation detection based on the actual ITS image data set. The parking violation data set contains 850 experimental sequences, of which about 250 violation behavior image data. The results of the illegal stop detection experiment are shown in Table 8.

As shown in Table 8, the PSPNet detection method proposed in this study can guarantee an accuracy of more than 96% for the detection of taxi parking violations. Compared with other methods, it can ensure that accurate violation judgments are maintained under actual road conditions. Figure 11 shows the situation of vehicle parking violation detection under different methods. It can be seen from the figure that the convergence speed reported by Zhang et al. [23] is faster than the method proposed in this study. But for the test accuracy, the method in this study is closer to 1 when the true-positive rate converges. Therefore, it is proved that the detection performance of the method proposed in this study is more outstanding.

At the same time, Table 9 shows the analysis results of the EER and DR indicators of the vehicle parking violation experiment. It can be seen from the table that the detection rate DR of the PSPNet network-based parking violation detection method proposed in this study is 88.2%, which is 17.3% higher than the DR reported by Liu et al. [26]. At the same time, EER is 14.2%, which is also about 20.1% lower than that reported by Liu et al. [26]. The reason is that the PSPNet network model uses pyramid pooling to fuse global and local features in the feature extraction of road condition images. To a certain extent, this makes up for the shortcomings of traditional pooling operations that can only capture fixed window feature information and can further improve its segmentation accuracy.

7.5. Ablation Study. In order to clarify the impact of various network components on the performance of taxi illegal parking detection, ablation experiments are carried out based on ITS image data set, and more complex task of taxi illegal parking detection are considered.

The proposed method includes the following important parts: feature extraction module, YOLOv3 module, SPP module, PSPNet module, and classification module. Because the classification module is necessary for the framework of this study, the classification module is retained. YOLOv3 module is used to detect taxis. Because the research object of this study is taxis, this module must be retained. Then, we delete the feature extraction module, SPP module and PSPNet module, respectively, and then conduct the ablation experiment. The results are shown in Table 10. It can be seen that when one of the modules is deleted or replaced, the detection accuracy decreases to a certain extent compared with the complete framework. Especially, without feature extraction module, the recognition rate decreases the most, to only 52.7%. Generally, the initial features obtained are too rough and directly entering the subsequent processing will seriously affect the subsequent results. Therefore, the feature extraction module is necessary in the network. Deleting the PSPNet module, we use the mixed Gaussian background model to extract the clean road background of a video and then segment the background image semantically. This can achieve the effect of real-time detection, but the detection accuracy is greatly reduced. Deleting the SPP module will reduce the taxi detection accuracy, so as to reduce the accuracy of illegal parking detection. Therefore, each module plays an irreplaceable role in promoting the final result.

7.6. Effects of Different Loss Functions and Training times. Different loss functions and training times will have a direct impact on the whole experiment. Therefore, this section will discuss the relationship between two different loss functions and training times. In the detection of taxi illegal parking, PSPNet uses the ordinary cross entropy loss function (equation (2)), and the suppression cross entropy loss function (equation (3)) proposed in this study aim to carry out two groups of comparative experiments. The results are shown in Figure 12. In Figure 12, CE Loss represents ordinary cross entropy loss, ICE Loss represents inhibitory cross entropy loss, and one epoch represents a complete learning process. It can be seen that ICE Loss converges earlier than CE Loss. The inhibitory cross entropy loss error remains stable after 50 epochs, whereas the

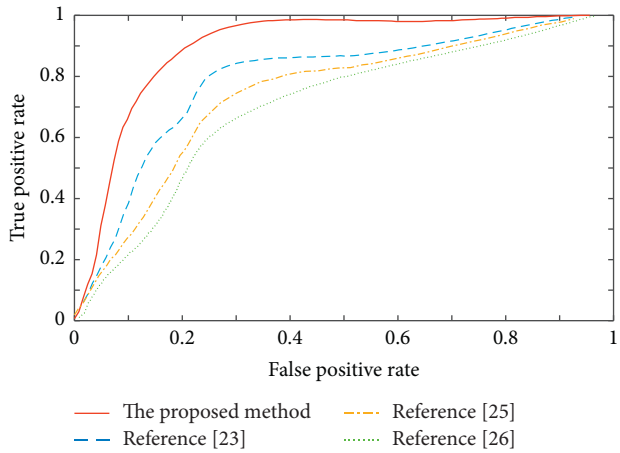

Figure 10: Frame level ROC curve of license plate violation under different methods.

Table 7: Comparison of license plate violation detection performance under different methods.

Method	EER (%)	DR (%)
The proposed method	19.4	85.3
Zhang et al. [23]	25.3	81.8
Abbas [25]	30.6	79.3
Liu et al. [26]	35.9	72.6

Table 8: Detection results of taxi illegal parking under different methods.

Method	Accuracy rate (%)
The proposed method	96.1
Zhang et al. [23]	92.8
Abbas [25]	92.1
Liu t al. [26]	93.5

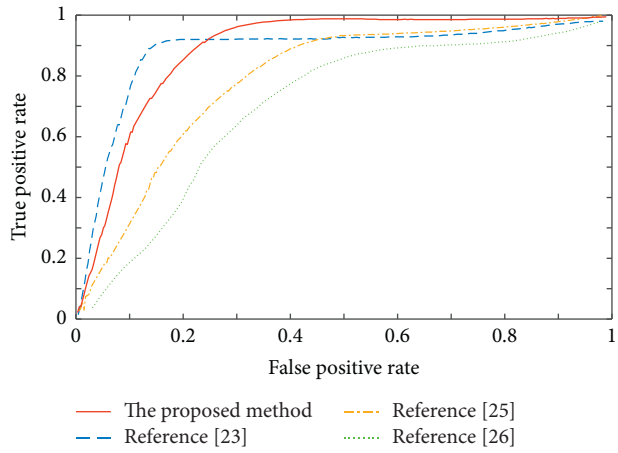

Figure 11: Frame level ROC curve of vehicle illegal stop under different methods.

ordinary cross entropy loss begins to remain stable after more than 80 epochs, which proves that the inhibitory cross entropy loss function can accelerate the training speed of the model and ensure the rapid convergence of the model. At the same time, because each pixel loss error in the consistent cross entropy loss will be multiplied by a factor less than 1, the inhibitory cross entropy loss error is always less than the ordinary cross entropy loss error.

Table 9: Comparison of vehicle stopping detection performance under different methods.

Method	EER (%)	DR (%)
The proposed method	14.2	88.2
Zhang et al. [23]	21.4	78.4
Abbas [25]	27.7	74.8
Liu et al. [26]	34.3	70.9

Table 10: The ablation experimental results.

List of ablation experiments	Average recognition rate (%)
Delete feature extraction module	52.7
Replace PSPNet module	69.7
Delete SPP module	84.5
Complete frame	96.1

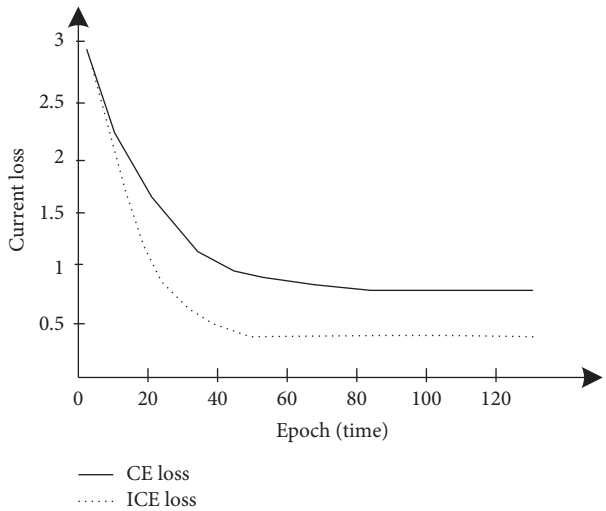

Figure 12: Relationship between training times and loss function.

Table 11: Time complexity analysis of different methods.

Ref.	Time (ms)		fps
	Training	Testing	
Zhang et al. [23]	1791	1280	30
Abbas [25]	9780	3173	20
Liu et al. [26]	11678	1580	15
Proposed	5367	1720	21

7.7. Operation Efficiency Analysis. In order to analyze the time efficiency of the proposed method, the training time, testing time, and video frame rate are considered. As can be seen from Table 11, the training time of the proposed method is 5367 ms, ranking second, and less training parameters are required. The test time was 1720 ms, ranking third. The video frame rate is 21 fps. Also, there are two options of illegal stop detection, offline detection, and real-time detection. Therefore, compared with other methods, the operation efficiency of the proposed method is acceptable.

8. Conclusion and Outlook

Based on the improved YOLOv3 network and PSPNet network, this study proposed a new method for detecting taxi violations. The proposed method can detect two different taxi violations at the same time, including license plate occlusion and illegal parking, and the method can be easily extended to other types of vehicles. Adding SPP module in YOLOv3 can avoid image distortion to a certain extent, solve the problem of repeated extraction of vehicle image features, and effectively distinguish whether the license plate is blocked. In addition, another novelty of this study was to propose a method to give PSPNet network for illegal parking. By aggregating the image information of different regions, we can improve the acquisition ability of global information. To achieve the real-time detection effect, we first use the mixed Gaussian background model to extract the clean road background of a video and then semantically segment the background image Because it is helpful to locate the position space information of the vehicle. The experimental analysis on the ITS collected data set shows that the proposed method has excellent network performance for license plate occlusion and vehicle parking violation behavior. The recognition accuracy of license plate occlusion is 95.1%, and the detection accuracy of taxi illegal parking behavior is more than 96%. However, the actual application scenario will be affected by natural weather conditions (such as fog and rain), which may cause deviations in detection and analysis. Therefore, the next step will be oriented to more complex actual scenes to achieve accurate and efficient road vehicle detection and analysis.

Data Availability

The data included in this paper are available without any restriction.

Conflicts of Interest

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

Acknowledgments

This work was supported by the National Natural Science Foundation of China (51775496), 2020 "Teacher Professional Development" Project for Domestic Visiting Scholars in Colleges and Universities, the "China's 14th Five-Year Plan" (145047) of Shaoxing Philosophy and Social Sciences Research in 2021, and Research on "School Enterprise Cooperation Project" of Visiting Engineers in Zhejiang Province in 2021.

References

[1] F. Mehboob, M. Abbas, and A. Rauf, "Mathematical model based traffic violations identification," *Computational & Mathematical Organization Theory*, vol. 25, no. 3, pp. 302–318, 2019.

[2] M. Özkul and I. Çapuni, "Police-less multi-party traffic violation detection and reporting system with privacy preservation," *IET Intelligent Transport Systems*, vol. 12, no. 5, pp. 351–358, 2018.

[3] L. W. Ma, Y. K. Guo, and J. P. Li, "A vehicle violation detection algorithm using implicit curve family," *Journal of Xi'an University of Engineering Science and Technology*, vol. 43, no. 2, pp. 139–144, 2016.

[4] S. S. Jamiya and P. E. Rani, "LittleYOLO-SPP: a delicate real-time vehicle detection algorithm," *Optik*, vol. 225, no. 1, pp. 1–10, 2021.

[5] P. Waszecki, P. Mundhenk, S. Steinhorst, M. Lukasiewycz, R. Karri, and S. Chakraborty, "Automotive electrical and electronic architecture security via distributed in-vehicle traffic monitoring," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 36, no. 11, pp. 1790–1803, 2017.

[6] N. C. Mallela and R. Volety, "Detection of the triple riding and speed violation on two-wheelers using deep learning algorithms," *Multimedia Tools and Applications*, vol. 80, no. 6, pp. 8175–8187, 2021.

[7] D. Hardiyanto, M. Rojali, R. Iswanto, and R. Muamar, "Pedestrian crossing safety system at traffic lights based on decision tree algorithm," *International Journal of Advanced Computer Science and Applications*, vol. 10, no. 8, pp. 375–379, 2019.

[8] R. Karthikeyan, "Implementation of traffic violation detection and traffic control," *International Journal of Modern Agriculture*, vol. 9, no. 4, pp. 1185–1191, 2020.

[9] T. Kumar and D. S. Kushwaha, "Traffic surveillance and speed limit violation detection system," *Journal of Intelligent and Fuzzy Systems*, vol. 32, no. 5, pp. 3761–3773, 2017.

[10] W. Zhu, S. Yu, X. Zheng, and Y. Wu, "Fine-grained vehicle classification technology based on fusion of multi-convolutional neural networks," *Sensors and Materials*, vol. 31, no. 2, pp. 569–578, 2019.

[11] D. B. Zhang, "Analysis and research on the images of drivers and passengers wearing seat belt in traffic inspection," *Cluster Computing*, vol. 22, no. 1, pp. 9089–9095, 2019.

[12] N. Alam, M. Ahsan, M. A. Based, and H. Julfikar, "Intelligent system for vehicles number plate detection and recognition using convolutional neural networks," *Technologies*, vol. 9, no. 1, pp. 1–18, 2021.

[13] A. Senta, I. Tashiev, F. Kucukayvaz et al., "Performance evaluation of support vector machine and convolutional neural network algorithms in real-time vehicle type and color classification," *Evolutionary Intelligence*, vol. 13, no. 1, pp. 83–91, 2020.

[14] S. Usmankhujaev, S. Baydadaev, and K. J. Woo, "Real-time, deep learning based wrong direction detection," *Applied Sciences*, vol. 10, no. 7, pp. 1–13, 2020.

[15] Z. Ding, L. Xing, and Y. Mo, "Mapping grid based online taxi anomalous trajectory detection," *International Journal of Systems Science*, vol. 51, no. 9, pp. 1589–1603, 2020.

[16] W. Sun, H. Du, G. Ma, S. Shi, X. Zhang, and Y. Wu, "Moving vehicle video detection combining ViBe and inter-frame difference," *International Journal of Embedded Systems*, vol. 12, no. 3, pp. 371–379, 2020.

[17] Y. Cao, J. Lv, Y. Bai, and A. Wu, "Method of unsupervised static recognition and dynamic tracking for vehicles," *Sensors and Materials*, vol. 32, no. 12, pp. 4517–4536, 2020.

[18] J. J. Qu and Y. H. Xin, "Moving target detection method based on continuous frame difference and background difference," *Acta Photonica Sinica*, vol. 43, no. 7, pp. 219–226, 2014.

[19] X.-m. Xu, L.-c. Zhou, Q. Mo, and Q.-y. Guo, "Vehicle detection algorithm based on codebook and local binary patterns algorithms," *Journal of Central South University*, vol. 22, no. 2, pp. 593–600, 2015.

[20] Z. Moutakki, I. M. Ouloul, K. Afdel, and A. Amghar, "Real-time video surveillance system for traffic management with background subtraction using codebook model and occlusion handling," *Transport and Telecommunication Journal*, vol. 18, no. 4, pp. 297–306, 2017.

[21] N. Buch, S. A. Velastin, and J. Orwell, "A review of computer vision techniques for the analysis of urban traffic," *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 3, pp. 920–939, 2011.

[22] P. Qian, K. Yuan, J. Yao et al., "Residual-network-leveraged vehicle-thrown-waste identification in real-time traffic surveillance videos," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 3, pp. 1817–1826, 2021.

[23] C. H. Zhang, B. H. Tong, J. Cheng, B. Zhang, and R. Zhang, "Improved Yolo_v2 illegal vehicle detection method," *Computer Engineering and Application*, vol. 56, no. 20, pp. 104–110, 2020.

[24] H. Tang, A. Peng, D. Zhang, T. Liu, and J. Ouyang, "SSD real-time illegal parking detection based on contextual information transmission," *Computers, Materials & Continua*, vol. 62, no. 1, pp. 293–307, 2020.

[25] Q. V.-I. T. S. Abbas, "Video-based intelligent transportation system for monitoring vehicle illegal activities," *International Journal of Advanced Computer Science and Applications*, vol. 10, no. 3, pp. 202–208, 2019.

[26] Z. H. Liu, D. H. Choi, and Y. M. Jin, "Method of detecting a parking violation using deep running tracking," *Journal of the Institute of Electronics and Information Engineers*, vol. 56, no. 9, pp. 67–74, 2019.

[27] N.-S. Pai, J.-B. Huang, J.-X. Wu, P.-Y. Chen, and Y.-H. Zhou, "Forward collision warning and lane-mark recognition systems based on deep learning," *Sensors and Materials*, vol. 32, no. 6, pp. 1981–1995, 2020.

[28] B. Wang and Y. Gu, "An improved FBPN-based detection network for vehicles in aerial images," *Sensors*, vol. 20, no. 17, pp. 1–20, 2020.