Hindawi
Journal of Sensors
Volume 2022, Article ID 8445816, 11 pages
https://doi.org/10.1155/2022/8445816

Research Article

Intelligent Intersection Vehicle and Pedestrian Detection Based on Convolutional Neural Network

Senlin Yang,[1,2] **Xin Chong**,[3] **Xilong Li**,[2] **and Ruixing Li**[2]

[1]*Shaanxi Key Laboratory of Surface Engineering and Remanufacturing, Xi'an University, Xi'an, Shaanxi 710065, China*
[2]*School of Mechanic & Material Engineering, Xi'an University, Xi'an, Shaanxi 710065, China*
[3]*Vertiv Technology Ltd., Xi'an, Shaanxi 710075, China*

Correspondence should be addressed to Senlin Yang; linsenyang@xawl.edu.cn

Received 8 November 2021; Revised 4 January 2022; Accepted 26 January 2022; Published 11 March 2022

Academic Editor: Gengxin Sun

The preprocessed images are input to a pretrained neural network to obtain the corresponding feature mapping, and the corresponding region of interest is set for each point in the feature mapping to obtain multiple candidate feature regions; subsequently, these candidate feature regions are fed into a region proposal network and a deep residual network for binary classification and BB regression, and some of the candidate feature regions are filtered out, and the remaining feature regions are subjected to ROIAIign operation; finally, classification, BB regression, and mask generation are performed on these feature regions, and full convolutional nerve network operation is performed in each feature region and output. To further identify the specific model of the vehicle, this paper proposes a multifeature model recognition method that fuses the improved model with the optimized Mask R-CNN algorithm. A vehicle local feature dataset including vehicle badges, lights, air intake grille, and whole vehicle outline is established to simplify the network structure of model. Meanwhile, its detection frame generation process and the adjustment rules of overlapping frame confidence in nonmaximum suppression are improved for coarse vehicle localization. Then, the generated vehicle detection frames after localization are output to the Mask R-CNN algorithm after further optimizing the RPN structure. The localized vehicle detection frames are then output to the Mask R-CNN algorithm after further optimization of the RPN structure for local feature recognition, and good recognition results are achieved. Finally, this paper establishes a distributed server-based vehicle recognition system, which mainly includes database module, file module, feature extraction and matching module, message queue module, WEB module, and vehicle detection module. Due to the limitations of traditional region generation methods, this paper provides a brief analysis of the region generation network in the Faster R-CNN algorithm and details the loss calculation principle of the output layer.

1. Introduction

The number of motor vehicles has exceeded 350 million, cars reached 229 million, motor vehicle drivers exceeded 420 million, including 360 million car drivers, and cars have gradually replaced bicycles and other as one of the main means of transportation for travel, appearing in various scenes such as streets, highways, and communities [1]. The rapid growth of motor vehicles not only brings many conveniences to people's lives but also generates road congestion and criminal cases involving cars, bringing invisible effects to our living environment and travel speed [2]. As the number of motor vehicles grows at a rate of about 10% per year, urban building congestion leads to slow development of road construction. The urbanization of China makes more and more rural population flock to the city, which leading to road congestion, traffic accidents, environmental pollution and other problems [3]. The rapid growth of vehicles not only makes urban traffic overload but also makes the frequent occurrence of criminal cases involving vehicles, bringing new challenges to public safety, and the current management and identification of vehicles basically rely on the existing road traffic management methods and manual judgment [4]. In order to reduce manual operations, automatically detect vehicles, and identify their corresponding areas of interest, so as to make timely responses to traffic

problems occurring in highways, communities, and other environments, the research of intelligent traffic system (ITS) has emerged [5].

The rapid development of artificial intelligence has laid a good foundation for ITS, which integrates technologies such as computer processing, automation, data communication transmission technology, big data, and machine vision into the traffic management system and can replace manual operations with intelligent systems in traffic scenarios such as highways, toll stations, railway stations, and airports to reduce congestion, transportation failures, and other problems, as well as save energy and manpower and reduce economic waste [6]. The core of the vehicle detection and automatic identification system construction lies in the license plate, vehicle color, vehicle brand, and specific model recognition and the matching problem. At present the license plate positioning and recognition system is very mature, the precision and accuracy rate is very high, and it has been widely used in various traffic intersections, neighborhoods, highways, and other places; the body color recognition technology is relatively simple to achieve and also has a good recognition rate; for the specific model recognition, because the similarity of different vehicles may be larger, the vehicle detail recognition aspect has a certain degree of difficulty, the current model recognition technology cannot reach a high industrialization degree, and there is no more perfect model recognition system [7]. In order to solve the above problems, more and more experts and scholars have devoted themselves to the research of vehicle model and vehicle brand recognition in recent years, and certain progress has been made. Due to the difference of application occasions and demand objects, there is also a certain difference in the fine degree and algorithm framework for model recognition [8]. In the general highway, community and parking management system, it generally only needs to determine whether it belongs to large vehicles or small vehicles. In the public security criminal investigation for the search of the set of vehicles or illegal criminal vehicles, it requires for the vehicle detail characteristics that are extremely detailed; model recognition specific to the model and year can provide more effective clues for the public security organs [9].

The improved YOLOv3 coarse vehicle localization method is incorporated in the vehicle detection stage of fine vehicle recognition, and a distributed system is used to assign each local feature to different servers for feature extraction and recognition using the improved Mask R-CNN method, and then, the total server aggregates and outputs the recognition results, which not only improves the generalization ability and robustness of the detection method but also improves the efficiency of detection and recognition. Finally, a model recognition system is established to further verify the feasibility of the algorithm proposed in this paper. For the problem of vehicle-specific model recognition, a fine model recognition algorithm with improved YOLOv3 algorithm is considered as the detection model, while the RPN module in Mask R-CNN that is further optimized and used for recognition is proposed, and the established local feature dataset is introduced. In order to improve the detection efficiency, a method of multithreaded feature recognition using a distributed server system is proposed, and the superiority of this paper compared with other target detection methods is analyzed. Finally, the hardware system for vehicle model recognition built in this paper is introduced, mainly including database module, file module, feature extraction and comparison module, message queue module, WEB module, and vehicle detection module, and the algorithm proposed in this paper is implanted into the system to verify the practical value of the method. This paper mainly focuses on deep learning and convolutional neural network algorithms to optimize the network structure to train the detection and recognition models of large class vehicles and fine vehicles, respectively. Based on the algorithm development of R-CNN and Faster R-CNN and the design of convolutional layer, the superiority of convolutional neural network in target detection and recognition is illustrated, and the advantages and disadvantages of different methods and network frameworks in target detection are analyzed.

2. Related Work

The interframe difference method, background difference method, and optical flow method are the three most traditional methods in target detection [10]. The basic principle of the frame difference method is to determine the moving target area based on the pixel change between frames in the video, and the pixels between adjacent frames are compared by the difference and threshold operation to obtain the moving object [11]. The overall accuracy of the overall model is affected. Based on the three-frame difference method, the researchers binarized the vehicle image after extracting the contour of the moving target vehicle, then applied morphological processing to it, and finally performed line-by-line scanning to obtain the overall binarized image and reconstructed the vehicle image using the connection of contours to obtain the region of the moving vehicle, whose limitation is that it can only detect the moving vehicle, and for the stationary vehicle recognition, there is still a need for further research [12].

Background difference method is the most commonly used method in the early development of target detection. The principle is to first obtain the video or image that does not contain information such as vehicles in the background to generate a background model, and then, the image or video to be measured that inputs and subtracts the information corresponding to the background model can obtain the possible vehicle information you want to identify, finally, the information image for binarization can get more complete vehicle information [13]. Researchers in the background difference method based on the use of background model in obtaining the specific location information of the target vehicle take the labeling technology to give the video frame or image of each vehicle corresponding to the label and then do further processing; the experiment shows that the method has good detection effect in the fixed scene [14]. The optical flow method is very different from the two target detection methods mentioned above, the method is based on giving the velocity vector corresponding to all pixel points in the video frame to achieve the purpose of transforming the

original image into a variable motion field, and each coordinate in the video frame can find its corresponding coordinate on the target to be identified at any moment when training is carried out [15].

Researchers proposed to introduce Lucas-Kanade based on the parallel optical flow method to identify and track moving vehicles in video [16]. The main process is to detect moving targets using optical flow detectors and then perform binarization similar to the background difference method and use the target frame to detect the range of the target, but the limitation of this method is that it can generally only be used for target localization and tracking, not for recognition [17]. In addition to the traditional target detection methods mentioned above, feature-based target detection methods are also used in vehicle model recognition, and more commonly, vehicle detection is performed using features such as histogram of oriented gradient, scale-invariant feature transform, and Haar. The feature-based vehicle recognition methods are generally divided into two major categories: the first one is to directly extract and train features on the whole original image and the other one is to segment the original image into multiple images of appropriate size, perform feature extraction on each small image, and then use classifiers such as SVM to classify the extracted individual features before proceeding to the next step of detection [18]. This method requires a high level of dataset richness and a large number of samples with different environments, angles, and the presence of occlusions for training, which is a huge amount of engineering [19].

As one of the representative algorithms of deep learning, convolutional neural network (CNN) was first proposed in 1987, but it did not have much application at that time; with the hot development of deep learning and the wide application of GPU in recent years, CNN has been used in image processing and target detection [20]. The R-CNN network is derived from the CNN network with improvements, using automatic selection search to obtain the candidate range of the target, then feeding the target candidate range into the convolutional neural network for feature extraction and classification, and finally outputting the recognition results with rectangular boxes. Although RCNN has a great improvement in detection accuracy compared with CNN, it also has the limitation of being more time-consuming [21]. Fast R-CNN changes the convolution of the feature area for each candidate region on the basis of R-CNN, and uses shared whole image for feature extraction, which greatly reduces the detection time. Researchers improve the algorithm on the basis of Fast R-CNN and used a candidate frame extraction network to extract the target range and identify it, which further accelerated the detection speed. In addition to these networks, SPP-Net, R-FCN, GoogLeNet, etc. can be used for target detection and recognition [22].

3. Feature Extraction and Result Output for Vehicles and Pedestrians at Road Junctions

3.1. Feature Extraction. The basic process is shown in Figure 1. Firstly, the key points and key regions of the input image are located using image processing technology, then the feature descriptors in the regions are extracted, and finally the feature descriptors are input to the classifier to realize the classification and recognition of car models. According to the degree of refinement of model recognition, the model recognition technology can be divided into coarse-grained model recognition and fine-grained model recognition. Since different types of vehicles have different appearance shapes, coarse-grained model recognition can be classified mainly based on the appearance shapes of vehicles. In addition, some key parts on the car in the coarse-grained model recognition process also differ greatly (e.g., doors, front end, body, and windows), and these characteristics can also be used as the basis for discriminating coarse-grained models.

The fine-grained model recognition process often requires more detailed features to be considered because of the type and model of the vehicle to be discerned. It is understood that the fine-grained model recognition tends to pay more attention to the vicinity of the license plate as well as the vicinity of the lights and emblem, because these parts are the biggest difference in distinguishing from the same type of vehicles. First of all, the area near the license plate will be the target area, because the license plate is located in the front of the vehicle and is also the main performance part of the appearance design, so it will be the main candidate area of the model fine-grained recognition. Then, in order to carry out model recognition more effectively, the license plate image or the headlight image is usually segmented out separately for processing. Finally, in order to recognize the vehicle type effectively, the extracted features such as edge and color are classified using classifiers (softmax, SVM, etc.). With the increasing ability of feature descriptors to characterize images, such as histogram of gradients (HOG), scale-invariant feature transform (SIFT), and hybrid features, model recognition methods based on such feature extraction have very good robustness. The above vehicle recognition method based on feature extraction is inevitably limited by some external factors, such as fixed image capture view, artificially set feature extraction parameters and so on. In order to effectively solve this problem, some scholars try to apply geometric methods to car model recognition. With the help of computer-aided design (CAD) technology, the authors perform a series of preprocessing (including template matching and selecting viewpoint parameters) before feature extraction of images, so that they can better detect car license plates. The model recognition method based on geometric estimation mainly equates the vehicle area into a rectangle of certain length, width, and height, as shown in Figure 2. Firstly, the vehicle in the image is segmented to generate a rectangle of its smallest outside world, and its length, width, height, and center coordinates are output. Then, the correspondence between the key points in the reconstructed 3D space and the input 2D image is determined according to various parameters (including camera focal length and mapping relationship between coordinate systems). Finally, the inverse projection technique is used to compare the constructed 3D dimensions with the actual dimensions of the vehicle, so as to determine the type of the actual vehicle.

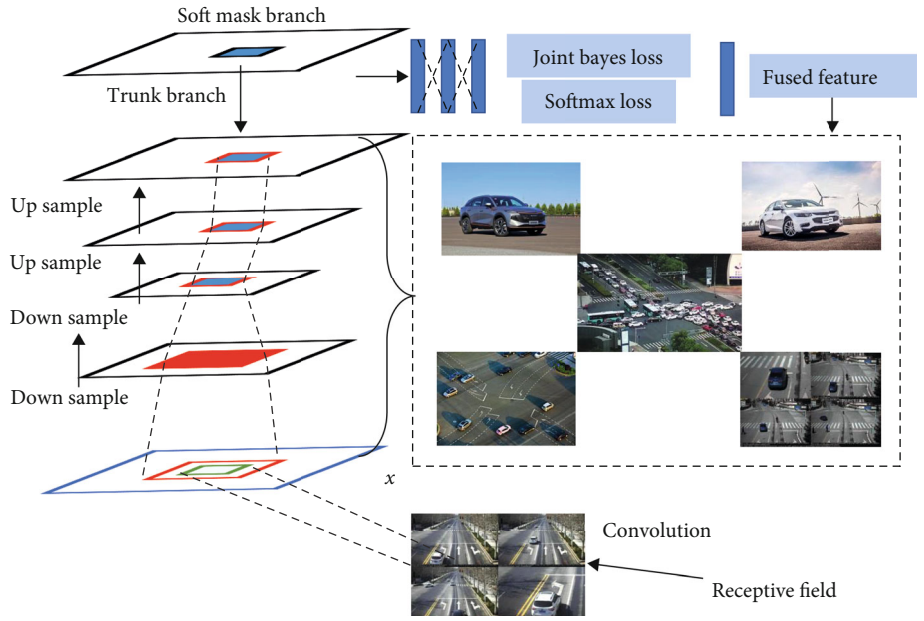

Figure 1: Feature extraction-based model recognition.

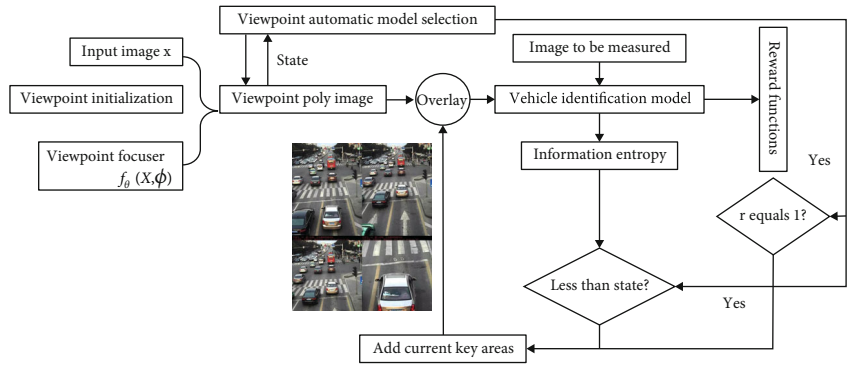

Figure 2: Geometric estimation-based model identification.

3.2. Neural Network to Enhance Recognition. In the fully connected layer, the main role is to reduce the error between the labeled samples in the dataset and the output of the generated model in order to achieve the purpose of continuously fitting the generated network to the original image in this paper.

The neural network-based model recognition method is a method that enables a neural network model to detect and recognize autonomously based on the learned capabilities by constructing it. With the outstanding contributions of neural network technology in the fields of image processing, target detection, and scene analysis, the application of neural networks in model recognition has been gradually promoted. In the research of using neural networks for car model recognition, it is usually done by extracting key frames from surveillance videos as the input of neural networks and then predicting their probabilities. This method, which involves a lot of human intervention, is called supervised training method, and semisupervised and unsupervised methods have also been studied in the literature to train neural networks for car model recognition. However, as far as the current research results are concerned, the neural network-based model recognition technique still does not effectively address the problem of low detection efficiency when dealing with multiangle and complex scenes. In the literature, the authors use images from two specific viewpoints as the infants of the neural network, which makes the neural network model possess higher detection accuracy than a single viewpoint. However, the model still cannot be used for other problems that deviate from the normal viewpoint. The literature uses CompCars, a dataset containing complex scenes and multiple perspectives, to train a neural network model with better robustness, but the model is not very efficient for vehicle detection and recognition because the constructed model is shallow. The literature proposes a car model recognition model with many complex image preprocessing means added to the network, yes the model can be converted from the input two-dimensional image to the unit space for processing, and also small datasets were constructed to verify the effectiveness of the method. However,

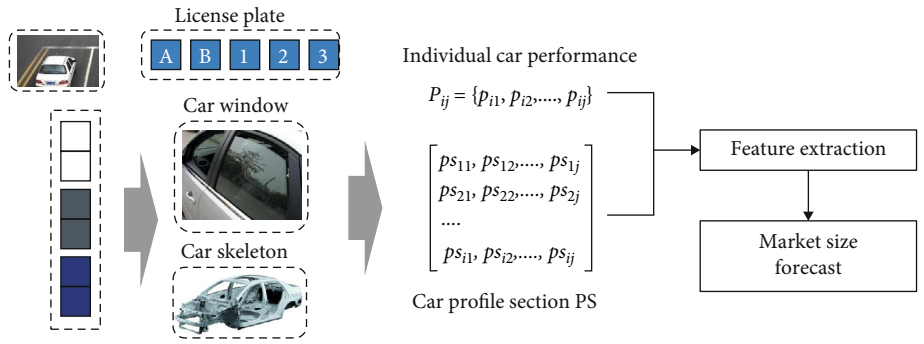

Figure 3: Global VR real estate market size forecast.

this complex preprocessing technique also requires much higher input data and therefore cannot be validated effectively on publicly available datasets, limiting its usage performance and application prospects. Therefore, it is of great interest to investigate a model for vehicle identification that can cope with complex weather, complex scenes, and high robustness.

In this paper, we propose an improved Mask R-CNN-based target detection and recognition method, whose network structure is shown in Figure 3. The original image is preprocessed and input to the pretrained convolutional layer neural network to obtain the corresponding feature map, and the region of interest is set for each point in the feature map to obtain several candidate feature regions, and then, these candidate feature regions are fed into the region suggestion network and the deep residual network (ResNet) for binary classification and BB regression. Finally, the fully convolutional network (FCN) operation is performed in each feature region to classify these feature regions by Mask and predict the target regions.

As a target detection method derived from CNN networks, the Mask R-CNN algorithm originally used feature pyramid networks (FPNs) to achieve efficient use of features at different scales, and FPNs employ top-down lateral connections to fuse (up-sample and sum) features connected at different scales and then perform 3×3 convolution to eliminate the blending phenomenon and then predict the features at different scales, repeating this process continuously until the best resolution is obtained. This feature mapping is shared for the subsequent region recommendation network layer and the fully connected layer. The advantages of FPN are its ability to localize and extract features more accurately for small targets and its shorter detection time, but it has limitations in detecting objects with low pixels or small distinctions. Deep residual network (ResNet) is a deep convolutional network with outstanding performance in target localization, target feature extraction, and target recognition proposed by four scholars from Microsoft Research in 2015, which well solves the problem of network depth and performance degradation. In this paper, we synthesize the special characteristics of vehicle targets and the applicability of other feature extractors in Mask R-CNN. The feature extraction module introduces a deep residual network with ResNet101 to extract vehicle feature information, and ResNet is based on the traditional AlexNex network, adding convolutional layers to achieve the purpose of extracting features more accurately and having stronger learning ability during model training. However, due to the large differences in the proportion of different vehicles in the video or image, background noise, and external contours, in order to better process the samples in the vehicle dataset and make the final generated model extract the vehicle features as much as possible, this paper combines the respective advantages of the deep residual network and the feature pyramid network and fuses the two for the extraction of vehicle features, and the network structure is more concise and modular. The network structure is also more concise and modular, and the convolutional network has fewer manually adjustable hyperparameters to facilitate training.

The activation functions used in this paper are the sigmoid function and the tanh function as follows:

$$\mathrm{sim}(x) = \frac{1}{1 + \exp(-x)}, \tag{1}$$

$$\tan(x) = \frac{\exp(x) - \exp(-x)}{\exp(x) + \exp(-x)}, \tag{2}$$

where 1 is the number of convolution layers, which is set to 5 in this paper; k_{lij} and b_{lj} (for) denote the convolution kernel and the offset of the feature map, respectively; the operation symbol x denotes the convolution operation; M_j is the set of input images. The convolution kernel convolves on the feature map output from the above convolution layer, and then, the new output feature map can be obtained after the sigmoid function and tanh function. The output feature map of each layer in the convolution layer of this paper through the activation function can be represented by multiple preactivation feature maps in the form of a sum, which is calculated as shown in the following equation:

$$x_i = f(q), \tag{3}$$

$$q = \sum_{i=1}^{i \in M_j} x_i * K_{ij}. \tag{4}$$

After preprocessing the original image and passing it through the convolutional layer, a common feature map can be obtained. In the more initial convolutional neural

network target detection frameworks (such as R-CNN and Fast R-CNN), the method of selective search is usually used to extract the candidate frames, which is more time-consuming and takes about 2 s to process an image on the CPU. CNN proposes the RPN method in the part of extracting candidate frames, which only takes about 10 ms to extract the candidate frames of an image, greatly speeding up the detection speed. The regional recommendation network requires less size and pixels of the input image, and its output increases the target frame of classification ratio compared with the convolutional neural network methods such as CNN, which makes the detection results easier to express. The convolutional kernel mentioned in the convolutional layer above is the key to generate the target candidate regions for RPN. The preprocessed image produces the output feature map after the operation of the convolutional layer, and sliding a small window of preset size to this feature map to obtain the corresponding large dimensional feature vector. The window of the sliding operation will generate different candidate regions after the RPN, which will then be input to the fully connected layer for localization and identification, as described below. Simply put, RPN relies on a sliding window on a shared feature map to generate nine target frames with preset aspect ratios and areas for each location, and the Mask R-CNN algorithm is inherited from this network for region prediction.

$$L_i(t, v) = \sum_{i=1}^{i\in\{x,y,w\}} s^2(t_1 - v), \tag{5}$$

$$s_i(x) = \begin{cases} 0.5x^2, & \text{if}, \quad x > 1, \\ |x| - 1, & \text{if}, \quad x \le 1. \end{cases} \tag{6}$$

The training function for training RPN is as follows:

$$L\left(\left\{p_1, p_2, \cdots, p_j\right\}\right) = \frac{p_i\{t_i, t_i^*\}}{L_{ij}}. \tag{7}$$

The network parameters can be determined by the objective function, and the network parameters in the fully connected layer are continuously updated as the objective function decreases. When the objective function reaches convergence, the signal distribution generated by our trained generative model is closest to the label distribution at the time of labeling, and the convolutional neural network for target detection can be well fitted to the original image and data to achieve accurate localization and identification. In the fitting process, the acquisition of the network parameters is essentially the problem of optimizing the nonlinear function, which is simply the problem of finding the best set of parameters $W*$ and $b*$ that can satisfy the following equation.

$$W^*, b^* = \underset{\min}{J}(W, q). \tag{8}$$

The training loss function in this paper is as follows:

$$J_{\text{fix}} = L(\{p_i\}, \{t\}). \tag{9}$$

Linear interpolation for the x-direction is calculated as follows:

$$f\left(R_1 = \frac{x_1 - x_2}{x_1 + x_2}\right) = f(Q_{11}) + f(Q_{12}), \tag{10}$$

$$f\left(R_2 = \frac{x_1 + x_2}{x_1 - x_2}\right) = f(Q_{22}) + f(Q_{12}). \tag{11}$$

Then, linear interpolation for the y-direction is calculated as follows:

$$f(P) = f(x, y), \tag{12}$$

where $f(x, y)$ is the pixel value of the point P to be solved, $f(Q_{11})$, $f(Q_{12})$, $f(Q_{21})$, and $f(Q_{22})$ are the pixel values of the four known points $Q_{11} = (x_1, y_1)$, $Q_{12} = (x_1, y_2)$, $Q_{21} = (x_2, y_1)$, and $Q_{22} = (x_2, y_2)$, respectively, and $f(R_1)$ and $f(R_2)$ are the pixel values obtained by interpolation in the x-direction.

3.3. Dataset Creation. The richness and effectiveness of the dataset is an important part of the car identification research. In this paper, we use the BIT-Vehicle dataset, the Cars dataset, and some data from the CompCars dataset as the basis and expand the dataset by traditional transformation, Gaussian noise, web crawler crawling data, and generative adversarial network (GAN) approach to expand the dataset. In order to ensure the generalization ability of the final model of this experiment, the dataset is expanded in addition to the three car datasets mentioned above, as follows.

(1) Traditional transformation. (a) Random cropping, image flipping, mirror transformation, and image color random dithering are used to change the angle, proportion, brightness, and saturation of vehicles in the original images to achieve the purpose of dataset expansion, and finally, 1800 vehicle pictures are generated

(2) Web crawlers. Web engines (such as Baidu and Google) contain a large amount of vehicle information and images, but manual search and preservation of such images are more time-consuming. Data can be crawled on a specific web page according to user-defined matching rules, parsing and analyzing the acquired page data, parsing out the hyperlinks (URLs) in the page, and downloading the text information, pictures, videos, and other information in the links. In this paper, based on the pyspider crawler framework, we use python to realize the crawler function and finally obtain 1200 vehicle images and keep 600 images of high quality after screening. Since the quality of the

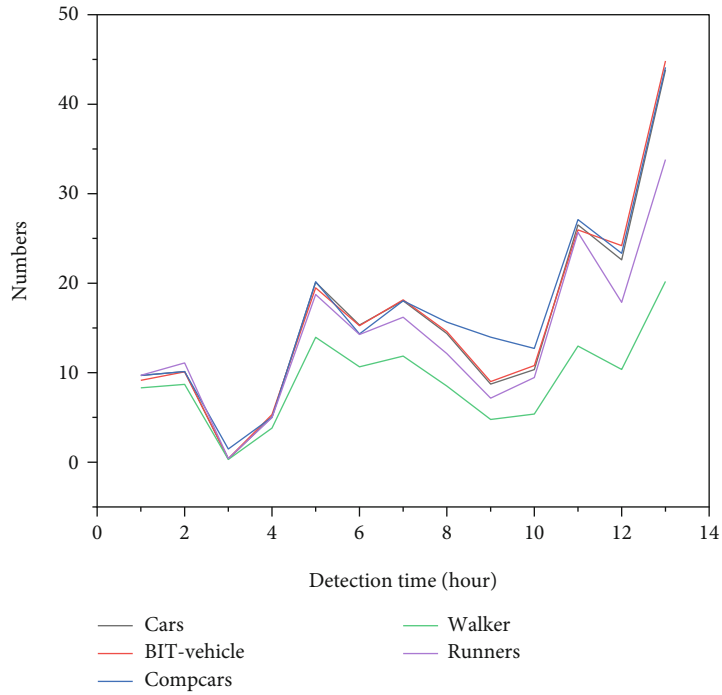

Figure 4: Classification of large categories of vehicle datasets.

images obtained using web crawlers varies, they are directly added to the dataset for use, without extending the data

(3) Generative adversarial network (GAN). In essence, the images generated by traditional transformations and the addition of Gaussian noise do not differ much from the vehicles in the original images, and the web crawler acquires the images slowly and requires manual screening, so this paper proposes to use GAN for data expansion. GAN is a method for training to generate two mutual adversarial models, where a generative model G is used to fit the sample data distribution, and a discriminative model D is used to estimate whether the input samples are from the real training data or the generative model G

(4) This paper uses convolutional neural network to construct generator G and discriminator D. Among them, discriminator D uses 4 convolutional layers with ReLU activation function and 1 fully connected layer to extract features from the input images; generator G uses 4 deconvolutional layers with ReLU activation function to generate false sample images with the same width and height as the input images by deconvolution of the noise generated using Gaussian distribution. Finally, 1500 vehicle images were generated. Based on the above dataset expansion method, this paper finally builds up a dataset including 8600 training set and 4300 test set samples, and the composition of the dataset is shown in Figure 4

4. Experiments and Analysis of Results

Network training requires setting the hyperparameters of the corresponding network, and hyperparameters are the preset values of network training, which are determined manually to achieve the parameters of the specific network training requirements; this experiment is trained from scratch for all networks, in deep learning, epoch represents the number of training steps, and the learning rate controls the learning progress of the model; the smaller the learning rate, the slower the loss gradient decreases and the convergence. The smaller the learning rate, the slower the loss gradient decreases and the longer the convergence time. After debugging, the final number of epochs is set to 50000, the learning rate is set to 0.005, the number of validations after each training step is set to 30, and the learning rate is kept constant at the beginning and decays to 0 in the last 5000 epochs. The weights are randomly initialized with Gaussian distribution, the mean value is 0, and the standard deviation is 0.02, and the specific hyperparameter values are shown in Figure 5. The network parameters can be determined by the objective function, and the network parameters in the fully connected layer are continuously updated as the objective function decreases.

At present, the evaluation indexes for the results in target recognition are precision rate, recall rate, average precision, average precision mean, etc. The average precision is the average of all accurate prediction rates of the car model under different recall rates, which is the best evaluation index of the performance of the target detection algorithm and reflects the comprehensive performance of the algorithm; meanwhile, this paper compares the pixel precision,

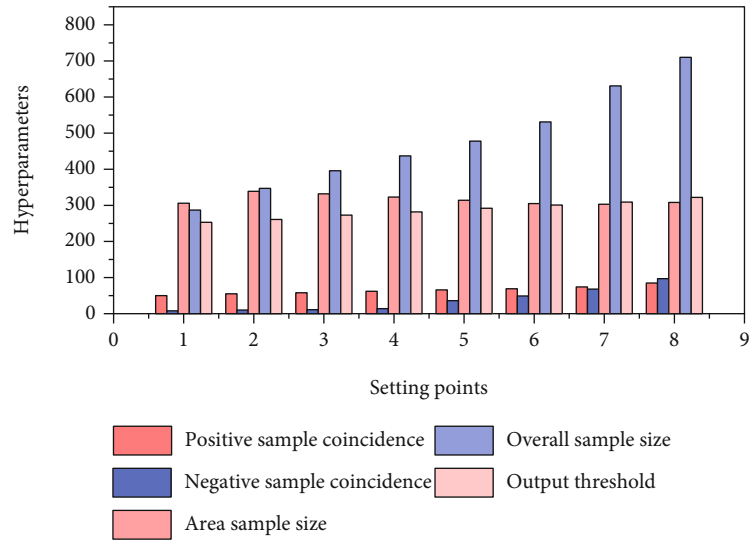

Figure 5: Hyperparameters of the large class vehicle recognition model.

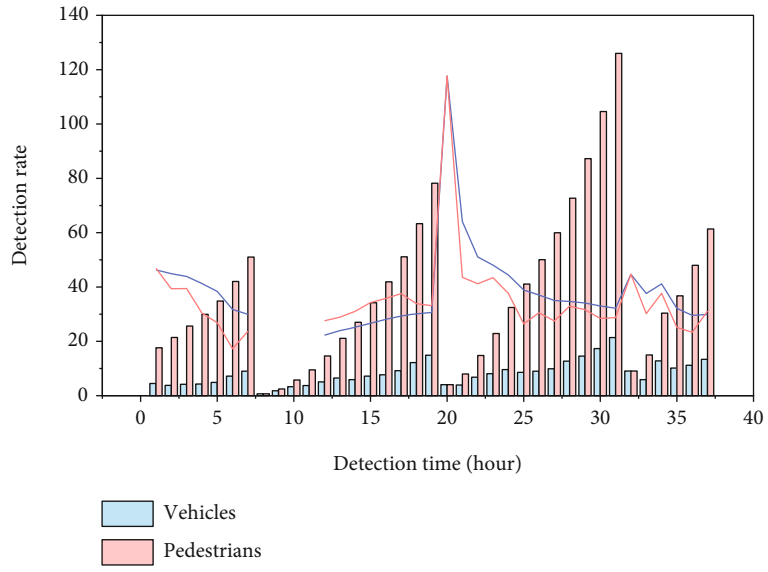

Figure 6: Detection rate of vehicles and pedestrians at smart intersections.

average interaction ratio, and detection recognition. This paper also compares the pixel accuracy, average interaction ratio, and detection recognition speed of this method with the mainstream target detection algorithms to verify the robustness and application value of this model.

In this experiment, the established 12900 datasets are divided into 8600 training sets and 4300 test sets. In common recognition systems, the workload of producing datasets is huge, requiring teamwork and time consumption. However, this system can reduce the time consumption compared with other labeling methods. Different samples can be generated randomly according to the corresponding labels during labeling and unified directly according to the labels during testing, which saves the time of unification processing after the labeling is finished. In order to test the generalization performance of the proposed model, the pictures of vehicles in different environment monitoring and different time and perspective are specially selected for recognition during the test. And the selected scenes also include the case of harsh environment, such as the bad situation of not strong light and too strong light, reflecting the difference between the model of the article and the target detection one-stage mainstream algorithm SSD, YOLO, and other method detection results. As shown in Figure 6, the experimental results show that the detection results of the model in this paper are better when the threshold is set to 0.8, and the improved algorithm has improved about 2.8% compared with the test results before the improvement in the dataset with a total of 50000 images on the KITTI public dataset. As can be seen from the figure, when there are no other occluding objects near the vehicle, the confidence of recognizing the vehicle is all above 92%, and when the

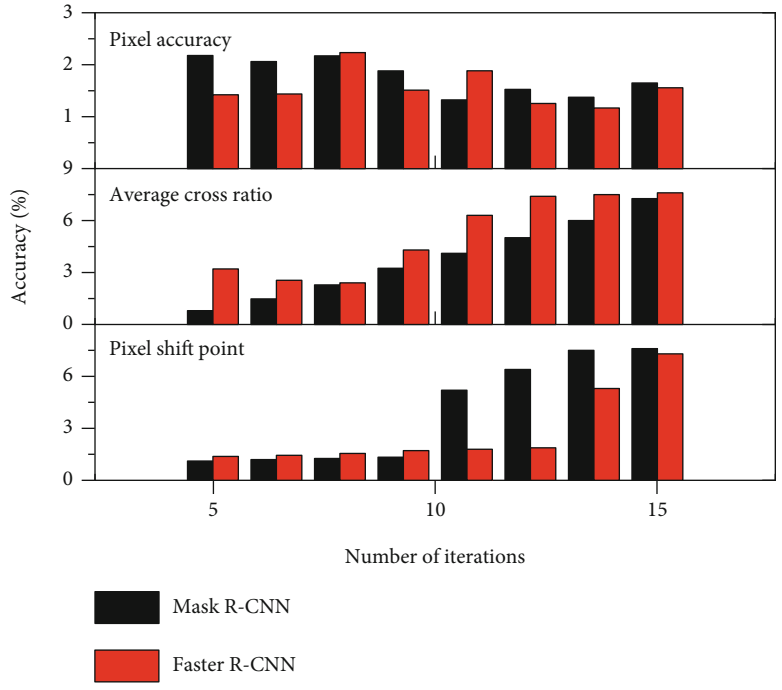

Figure 7: Scores for testing the dataset using different methods.

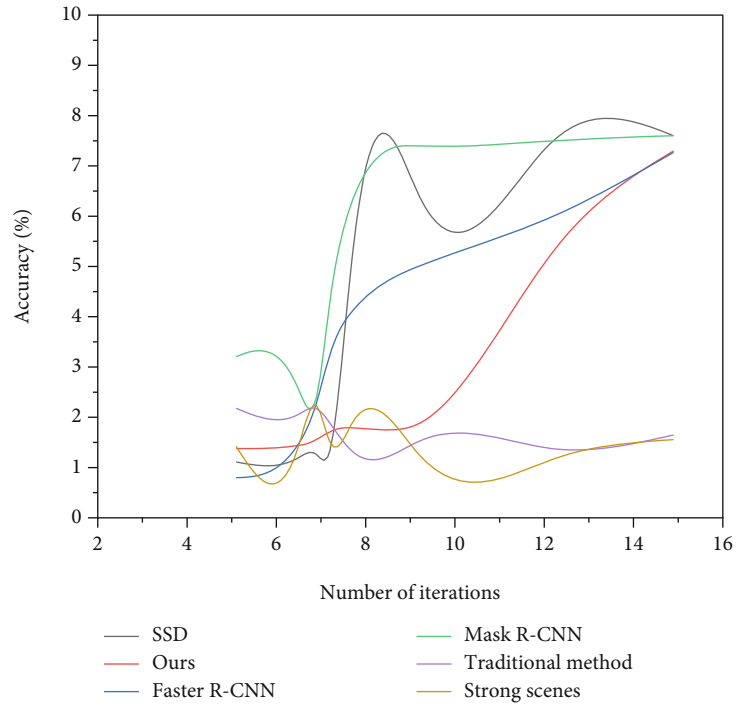

Figure 8: Trend of accuracy of different methods with the number of iterations.

vehicle is partially occluded or the vehicle has more than half of the area within the surveillance, the confidence is also above 82%, and the recognition accuracy can reach above 78%. In addition, the combination of labeled images and recognition results shows that the unlabeled vehicles and vehicles with small pixels in the training set can be recognized well, which again verifies the feasibility of the model.

In order to further verify the generalization ability of this experimental model and the recognition accuracy for different scenes, when testing the model, in addition to the above-mentioned images in the training set, this experiment also selected images in the same scene that were not in the training set and images in other scenes in different scenes for testing, and the test results are shown in Figure 7. It can be seen

that the vehicle recognition accuracy is high, and the recognition confidence for the images with low pixel in the back of the position can still reach 0.843 and the recognition result is accurate, which illustrates the strong generalization ability and high accuracy of the model. In the fully connected layer, the main role is to reduce the error between the labeled samples in the dataset and the output of the generated model in order to achieve the purpose of continuously fitting the generated network to the original image in this paper.

The experiments also selected the current open-source SSD, R-CNN, Faster RCNN, and the improved pre-Mask R-CNN algorithms for vehicle recognition detection. Figure 8 shows the scores of the test on the dataset using different methods, from which it is concluded that the recognition method used in this paper generates more reliable and more realistic results for the images and can get better results for all the scenarios described above. In addition, for the unlabeled vehicles in the training images, the method can still detect them well, which reflects the good robustness of the algorithm in this paper. Although the results of Faster R-CNN algorithm applied to this vehicle recognition also have better recognition results, but the method does not have better robustness, for most of the unlabeled vehicles are not detected, and similar to the traditional convolutional neural network-based CNN method, more postprocessing techniques are required, which increases the complexity of visualization operation, and the authenticity of the detection results is lower. When comparing with the Mask R-CNN algorithm before improvement, we found a more obvious improvement in pixel accuracy, while there is not only little difference in the average interaction ratio, but also a small improvement. Therefore, it can be seen from the above comparison tests that our algorithm has better superiority.

As can be seen from the figure, the accuracy of each method increases and stabilizes with the increase of iterations, among which the SSD method is the fastest to stabilize, and its accuracy stabilizes at about 76% after 10000 iterations, the accuracy of R-CNN is the lowest, and its accuracy stabilizes at about 74.5% after 12500 iterations; the Faster R-CNN method Mask R-CNN algorithm and the improved Mask R-CNN are more effective for car model recognition, and the recognition accuracy of Faster R-CNN method can reach about 84% for seven categories of car models; since Mask R-CNN algorithm requires higher quality of dataset and is more sensitive to pixel extraction, the accuracy of this algorithm is low when the number of iterations is small. However, after the number of iterations reaches 22500, the recognition accuracy of the algorithm for the seven categories of car models is about 86.2%, and the improved algorithm is stable at about 89% after the number of iterations reaches 25000, which is a considerable improvement compared with that before the improvement, further indicating the practical value of this algorithm.

5. Conclusion

This paper mainly focuses on deep learning and convolutional neural network algorithms to optimize the network structure to train the detection and recognition models of large class vehicles and fine vehicles, respectively. Based on the algorithm development of R-CNN and Faster R-CNN and the design of convolutional layer, the superiority of convolutional neural network in target detection and recognition is illustrated, and the advantages and disadvantages of different methods and network frameworks in target detection are analyzed, and the improved Mask R-CNN method is proposed to recognize large classes of vehicles, and the components and functions of the improved algorithm are introduced in detail. In the application of fine vehicles, we propose to use the improved YOLOv3 for detection and optimize Mask R-CNN algorithm for further recognition with good results. To further verify the practicality of the two methods proposed in this paper for engineering applications, a car model recognition system was built based on the existing equipment in the laboratory, and the algorithm was implanted in the server to achieve faster detection and recognition speed. The development of neural networks and the principle of deep learning are explained, and the algorithms related to artificial neural networks, convolutional neural networks, and target detection are introduced, and the advantages and shortcomings of each method are discussed. The speed and accuracy of convolutional neural networks in target candidate region generation, border regression, and feature extraction are discussed in detail, the improvements of new algorithms for target detection in recent years are analyzed, and the network framework of deep learning is introduced. Due to the limitations of traditional region generation methods, this paper provides a brief analysis of the region generation network in the Faster R-CNN algorithm and details the loss calculation principle of the output layer. For fine model recognition, this paper continues to expand on the basis of the CompCars dataset, establishes a vehicle dataset containing 18 common car brands such as Volkswagen, Buick, Audi, and BMW with a total of 76 common models, whose samples include vehicle badges, lights, air intake grilles, and overall contours, which can be trained with different detection models according to different needs, and finally uses labeling. Finally, we use labeling software to label all samples and build a more comprehensive model recognition dataset. In the future, the feature descriptors in the regions are extracted, and finally, the feature descriptors are input to the classifier to realize the classification and recognition of car models.

Data Availability

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

Conflicts of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was partially supported by the Natural Science Fund of China (Grant No. 61401356), the Technological Innovation Guidance Project of Shaanxi Province (Grant 2020CGXNG-015), the Science and Technology Project of Xi'an (Grant 2020kjrc0104), the Youth Innovation Team Building Scientific Research Project of Shaanxi Province (Grant 21JP106), and the Intelligent Perception and Control Research Team of Xi'an University (Grant XAWLKYTD019).

References

[1] R. Arnay, J. Hernández-Aceituno, J. Toledo, and L. Acosta, "Laser and optical flow fusion for a non-intrusive obstacle detection system on an intelligent wheelchair," *IEEE Sensors Journal*, vol. 18, no. 9, pp. 3799–3805, 2018.

[2] L. C. Bento, R. Parafita, H. A. Rakha, and U. J. Nunes, "A study of the environmental impacts of intelligent automated vehicle control at intersections via V2V and V2I communications," *Journal of Intelligent Transportation Systems*, vol. 23, no. 1, pp. 41–59, 2019.

[3] W. Cao, J. Zhang, C. Cai et al., "CNN-based intelligent safety surveillance in green IoT applications," *China Communications*, vol. 18, no. 1, pp. 108–119, 2021.

[4] L. W. Chen and Y. F. Ho, "Centimeter-grade metropolitan positioning for lane-level intelligent transportation systems based on the Internet of vehicles," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 3, pp. 1474–1485, 2018.

[5] A. Daeichian and A. Haghani, "Fuzzy Q-learning-based multiagent system for intelligent traffic control by a game theory approach," *Arabian Journal for Science and Engineering*, vol. 43, no. 6, pp. 3241–3247, 2018.

[6] T. Enokido, D. Taniar, and O. K. Hussain, "Special Issue: Intelligent edge, fog and Internet of Things (IoT)-based services," *Future Generation Computer Systems*, vol. 109, pp. 710-711, 2020.

[7] C. E. Framing, F. J. Heßeler, and D. Abel, "Learning scenario-specific vehicle motion models for intelligent infrastructure applications," *IFAC-PapersOnLine*, vol. 52, no. 8, pp. 111–117, 2019.

[8] L. Hu, J. Ou, J. Huang, Y. Chen, and D. Cao, "A review of research on traffic conflicts based on intelligent vehicles," *IEEE Access*, vol. 8, pp. 24471–24483, 2020.

[9] S. K. Kumaran, S. Mohapatra, D. P. Dogra, P. P. Roy, and B. G. Kim, "Computer vision-guided intelligent traffic signaling for isolated intersections," *Expert Systems with Applications*, vol. 134, pp. 267–278, 2019.

[10] K. Liu, "Bi-level optimisation model for greener transportation with intelligent transport system," *International Journal of Reasoning-based Intelligent Systems*, vol. 10, no. 1, pp. 26–31, 2018.

[11] S. Mohamad, F. M. Nasir, M. S. Sunar et al., "Intelligent agent simulator in massive crowd," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 11, no. 2, pp. 577–584, 2018.

[12] E. Namazi, J. Li, and C. Lu, "Intelligent intersection management systems considering autonomous vehicles: a systematic literature review," *IEEE Access*, vol. 7, pp. 91946–91965, 2019.

[13] H. Namazi and A. Taghavipour, "Traffic flow and emissions improvement via vehicle-to-vehicle and vehicle-to-infrastructure communication for an intelligent intersection," *Asian Journal of Control*, vol. 23, no. 5, pp. 2328–2342, 2021.

[14] I. O. Olayode, L. K. Tartibu, M. O. Okwu, and U. F. Uchechi, "Intelligent transportation systems, un-signalized road intersections and traffic congestion in Johannesburg: a systematic review," *Procedia CIRP*, vol. 91, pp. 844–850, 2020.

[15] Y. Sang, J. Tan, and W. Liu, "Research on many-objective flexible job shop intelligent scheduling problem based on improved NSGA-III," *IEEE Access*, vol. 8, pp. 157676–157690, 2020.

[16] M. Song, R. Li, and B. Wu, "Intelligent control method for traffic flow at urban intersection based on vehicle networking," *International Journal of Information Systems and Change Management*, vol. 12, no. 1, pp. 35–52, 2020.

[17] Q. Wu, F. He, and X. Fan, "The intelligent control system of traffic light based on fog computing," *Chinese Journal of Electronics*, vol. 27, no. 6, pp. 1265–1270, 2018.

[18] H. Xiao, M. A. Sotelo, Y. Ma et al., "An improved LSTM model for behavior recognition of intelligent vehicles," *IEEE Access*, vol. 8, pp. 101514–101527, 2020.

[19] Q. Xin, R. Fu, W. Yuan, Q. Liu, and S. Yu, "Predictive intelligent driver model for eco-driving using upcoming traffic signal information," *Physica A: Statistical Mechanics and its Applications*, vol. 508, pp. 806–823, 2018.

[20] H. Yang and K. Oguchi, "Intelligent vehicle control at signal-free intersection under mixed connected environment," *IET Intelligent Transport Systems*, vol. 14, no. 2, pp. 82–90, 2020.

[21] N. Yao and F. Zhang, "Contention-resolving model predictive control for an intelligent intersection traffic model," *Discrete Event Dynamic Systems*, vol. 31, no. 3, pp. 407–437, 2021.

[22] R. Zhang, A. Ishikawa, W. Wang, B. Striner, and O. K. Tonguz, "Using reinforcement learning with partial vehicle detection for intelligent traffic signal control," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 1, pp. 404–415, 2021.