Received August 3, 2020, accepted August 9, 2020, date of publication August 21, 2020, date of current version September 1, 2020.
Digital Object Identifier 10.1109/ACCESS.2020.3017894

Vehicle Detection in High Resolution Satellite Remote Sensing Images Based on Deep Learning

QULIN TAN[1,2], JUAN LING[3], JUN HU[4], XIAOCHUN QIN[1,2], AND JIPING HU[1,2]
[1]School of Civil Engineering, Beijing Jiaotong University, Beijing 100044, China
[2]Institute of Spatial Information for Line Engineering, Beijing Jiaotong University, Beijing 100044, China
[3]School of Tourism, Resources and Environment, Zaozhuang University, Zaozhuang 277100, China
[4]School of Computer and Information Technology, Beijing Jiaotong University, Beijing 100044, China

Corresponding author: Juan Ling (lj1905@foxmail.com)

This work was supported in part by the National Key Research and Development Program of China under Grant 2018YFC1505505, in part by the National Natural Science Foundation of China under Grant 40801121, in part by the NSFC under Grant 41807427, in part by a Project of Shandong Province Higher Educational Science and Technology Program under Grant KJ2018BBH018, and in part by the Scientific Research Project of Zaozhuang University under Grant 102061901.

ABSTRACT With the exponential growth of the number of vehicles, a series of traffic system problems continue to emerge. Vehicle detection plays an important role in the traffic system. Researchers have invested a lot of energy on it, and have made many achievements. However, there are still some deficiencies in the accuracy and robustness. The development of satellite remote sensing technology and VR panoramic technology provides technical support for this study. In order to detect vehicles more accurately, provide accurate and effective data information to relevant departments, and improve traffic conditions, this study uses deep learning algorithm to detect vehicles in high-resolution satellite remote sensing images. Firstly, the images in the images are classified by using the Alexnet network model, and then the vehicle target detection ability of the Faster R-CNN model algorithm is tested, and the algorithm is optimized by the method of model pruning and quantization, so that the average accuracy rate reaches 70.34%, and has strong robustness. Then the algorithm model is applied to the practical application of vehicle detection at an intersection. The traffic flow and the number of speeding vehicles are detected, and the false detection rate and missing detection rate are calculated. The results showed that the rate of missed detection and false detection was 0% on December 19, and the highest rate was 4.5% and 2.7% respectively. This shows that the vehicle detection based on deep learning algorithm in high resolution satellite remote sensing images has high accuracy, can effectively detect speeding vehicles, and can play a normal role in practical applications.

INDEX TERMS Deep learning, satellite remote sensing images, vehicle detection, high resolution, convolutional neural networks.

I. INTRODUCTION

A. BACKGROUND SIGNIFICANCE

With the continuous development of economy, people's requirements for material life are also improved, and the vehicle has gradually become indispensable. This has brought great pressure to the traffic, even if it is to advocate green travel, it will inevitably cause various traffic problems. On the basis of vehicle detection system, the performance of intelligent transportation system is greatly affected. The detection effect of traditional vehicle detection methods is not very good. Compared with ground sensors, satellite remote sensing is more convenient and economical to obtain urban traffic information, and deep learning technology has been widely used in speech recognition and image recognition. Based on deep learning algorithm and high-resolution satellite remote sensing technology to study efficient and accurate vehicle detection method is an important measure to solve traffic problems. [4].

The associate editor coordinating the review of this manuscript and approving it for publication was Zhihan Lv.

B. RELATED WORK

Due to the important position of vehicle detection in traffic system, many researchers have carried out research on how to effectively carry out vehicle detection, and have made many achievements [5]. Chen Z transforms RGB color space into color name space, embeds color name information into the grid of directional gradient feature histogram, and carries out vehicle detection according to the super-pixel segmentation result of aerial image [6]. His research only focuses on RGB image, which is far from enough in practical application. Wang H proposed an occlusion vehicle detection algorithm based on local connectivity and depth model [7]. It uses AdaBoost classifier to generate suspected occluded vehicles. Any sub images rejected in the last two stages of AdaBoost classifier are regarded as suspected blocked vehicles, and the final decision is made by color histogram matching [8]. In the case of dense vehicles, his algorithm increases the amount of computation and reduces the accuracy. Roxana VP proposed a vehicle occlusion detection method based on monocular vision, and proposed a vehicle tracking system based on monocular vision [9]. Firstly, the moving object was segmented from the background by using adaptive Gaussian mixture model (GMM), and the geometric features of the vehicle were extracted, and the vehicle was tracked by adaptive Kalman filter. The vehicle was divided into three categories: small, medium and large [10]. His detection method can not effectively detect vehicles parked on the side of the road, which affects the detection accuracy of the algorithm.

C. INNOVATIVE POINTS IN THIS PAPER

In order to detect vehicles more accurately and provide accurate traffic information effectively and timely, this study uses deep learning algorithm to detect vehicles in images from high-resolution satellite remote sensing technology. Firstly, the images in the images are classified by using the Alexnet network model. Then the Faster R-CNN network model of deep learning algorithm is introduced to detect the vehicle target, and the precision and recall are obtained. Based on the algorithm, the network model is pruned and quantified, so as to optimize the algorithm, improve the accuracy and robustness of the algorithm, and then put it into practical application to detect the actual effect. In the detection of vehicle overspeed at an intersection, the total correct rate reaches 98.42%. This shows that the Faster R-CNN algorithm based on deep learning optimization can effectively detect vehicles.

II. DEEP LEARNING AND VEHICLE DETECTION FROM SATELLITE REMOTE SENSING IMAGES

A. DEEP LEARNING THEORY

Deep learning is a type of machine learning, which uses artificial neural networks as the architecture to perform characterization learning on data. The application of convolutional neural network in deep learning makes this algorithm show its talents in the field of image processing.

1) THE COMPOSITION OF CONVOLUTIONAL NEURAL NETWORK

Convolutional Neural Network (CNN) is one of the most classic deep learning modes, playing an increasingly important role in computer vision, speech signal recognition and other fields [11]. The calculation process of CNN mainly includes two processes of forward propagation and back propagation. Forward propagation is a process in which input features are input from the input layer, processed in the hidden layer, and then output from the output layer. Backpropagation is the process of sending from the output layer to the input layer through the hidden layer using the chain rule. Convolutional neural networks mainly use local connections, shared weights, and spatial or temporal correlation sampling methods to extract richer functional information. Convolutional neural networks mainly include convolutional layers, pooling layers and fully connected layers [12], [13]

Convolution layer, the most important part of convolution neural network, is always the focus of research. The function of this layer is to extract image features. The convolution layer contains the corresponding convolution kernel. The height and width of convolution kernel are usually very small, and the depth is usually the number of input characteristic graphs. The number of convolution kernels is usually the number of output characteristic graphs of layers. In the forward propagation process, each filter will slide over the width and height of the input feature map, and the convolution between the filter and the corresponding input feature is calculated.

Pooling layer: integrate feature points into small area to generate new features. Pool operation can greatly reduce the number of features of the model, reduce the number of parameters of subsequent operations, effectively reduce the complexity of model calculation, and improve the efficiency of model execution. At the same time, it has the advantage of no deformation in translation. Currently, pooling operations commonly used include maximum pooling and average pooling.

The full connection layer is the same as the conventional neural network. The neurons in this layer are completely connected with all the feature elements of the upper layer. Compared with the local connection of convolution layer, more parameters are needed. Therefore, the full connection layer usually takes up most of the parameters of convolutional neural network.

Convolution neural network works by simulating the function of human brain and is used to fit nonlinear functions. Deep convolution neural network can extract rich features of high dimension and low dimension, and the effect of image processing is better than that of shallow neural network, but the more network layers, the more accurate the model is. The number of layers and the number of gradients in the training process will disappear.

2) NORMALIZATION AND COST FUNCTION

In convolutional neural networks, in order to process data more conveniently, data are usually mapped between 0 and 1 [14]. Dimension expressions can be transformed into non normalized expressions. It has two main functions: one is to speed up the convergence of the model. If the range of the two eigenvalues is inconsistent in the iteration process, the gradient descent direction will appear a zigzag, and the iteration will be very slow; the second is to improve the accuracy of the model [15].

Batch normalization can converge the model faster, solve the problem of gradient dispersion, and improve the stability and accuracy of the network model. However, the accuracy of batch normalization will be affected by the batch size. If the batch size is too small, the accuracy will be reduced, and if the batch size is too large, it is easy to be restricted by the display memory. Group normalization is not affected by the batch size. The channel is divided into several groups and calculated in groups.

In deep learning, the loss function is applied to a single sample to represent a sample error, while the cost function is applied to the entire data set to represent the average of all sample errors. In other words, the cost function is the average of the sum of all loss functions. For the algorithm, after selecting the network structure, the model framework is determined. In order to make the training results more accurate, it is necessary to choose the appropriate cost function. The cost function optimizes the model by comparing the predicted value with the actual value. The general cost function includes: mean square error, which is evaluated by calculating the sum of squares of predicted value and actual value; average absolute error, which takes absolute value for the difference between predicted value and actual value; cross entropy, which describes the difference information between two probability distributions [16].

3) OPTIMIZATION METHOD

The so-called model optimization is to constantly adjust the network parameters, so that the input parameters can perform a variety of nonlinear transformation, so that the output is more suitable. In other words, the optimization model is the process of finding the best solution of the function [17]. In deep learning, there are errors between the predicted values and the actual values. The problem of finding the minimum error is transformed into the problem of finding the optimal value of the cost function. It is very important to choose an appropriate optimization algorithm when searching for the optimal value related to the convergence speed and accuracy of the model.

Gradient descent, the parameters along the gradient descent direction forward a step longer, to get the desired objective function descent. The computational cost of gradient descent is low, which can be divided into batch gradient descent and random gradient descent. These two algorithms have their own advantages and disadvantages. The small batch gradient descent algorithm combines the advantages of batch and random, which can not only ensure the training speed, but also ensure the accuracy of convergence [18].

Momentum optimization method includes momentum optimizer and Newton accelerated gradient algorithm. Momentum optimizer is an improvement of stochastic gradient algorithm, which can accelerate the fast convergence speed and reduce the oscillation [19]. Newton accelerated gradient algorithm is an improvement of Momentum optimizer, which can predict the direction of parameter update and calculate the gradient of next position parameter [20].

The adaptive learning rate optimization algorithm improves the problem of different learning rate. Different learning rate is used in different training stages, which can accelerate the convergence in the early stage, and quickly reach the optimal value in the later stage. Common adaptive learning rate algorithms include AdaGrad algorithm, RMSProp algorithm, AdaDelta algorithm and Adam algorithm.

B. HIGH RESOLUTION SATELLITE REMOTE SENSING IMAGE

1) REMOTE SENSING IMAGE CLASSIFICATION

Remote sensing image is to use the difference and spatial variation of brightness value or size of pixel value to represent the difference between ground objects. This is the basis of segmentation of various types of image features. At present, the classification methods of remote sensing images can be divided into two types. One is manual visual interpretation method, the other is machine interpretation method based on information technology [21]. The second type of methods can also be divided into supervised classification and unsupervised classification. The features of the above two methods are inconsistent, and they are widely used in various scenes, but the basic idea is to classify all pixels in the image into corresponding feature categories.

Manual visual interpretation is the most basic method of remote sensing image interpretation. After careful observation and interpretation of remote sensing image, employees can manually extract information according to their own knowledge and experience. The principle is simple and easy to operate, but it has higher requirements for the professional knowledge of interpretation staff, and the workload is also relatively large.

The computer interpretation method uses the difference of different spectral and spatial distribution characteristics, uses the pattern recognition theory to divide the pixels in the remote sensing image into different types, and identifies various objects after calculation and analysis using probability and statistics theory [22]. Supervised classification requires the use of known sample pixels to identify unknown pixels, which generally includes maximum likelihood method, nearest distance method and spectral angle classifier. Unsupervised classification divides pixels in an image into a certain category without knowing the category and class name in advance. Based on cluster theory, the main implementation method is clustering algorithm.

2) IMAGE SEGMENTATION

Traditional image segmentation methods are mainly based on low-level visual information such as gray, color, texture, shape and so on. The main methods include threshold segmentation method, segmentation method based on pixel clustering, segmentation method based on edge detection and region growth method [23]. All of these methods use big features of different fields to distinguish them. If the segmented object is more complex, the features of each part are more obvious, or the current scene and background features are similar, most of the conventional methods are not very accurate. The method based on graph theory can solve the above problems, and usually has excellent segmentation effect, but this type of algorithm requires user interaction, and some parts need to be manually labeled [24].

Multiple texture images can be obtained by multi-scale image decomposition, which can be divided into two types: Based on different resolutions and based on different filter parameters. Decomposition methods based on different resolutions include Gaussian pyramid and Laplacian Pyramid [25]. The decomposition based on different filter parameters needs to be performed at the same resolution. The multi-scale decomposition can also extract texture, and the texture image of the image can be obtained by fusing the multiple detail layers obtained after image decomposition in a certain way.

3) DATA PREPROCESSING

In the process of data acquisition, due to the function limitation of remote sensing system, there are many errors. In order to improve the quality of remote sensing image preprocessing and image enhancement.

In order to eliminate sensor error and obtain accurate radiation value, it is necessary to conduct radiation correction on the original remote sensing image, and convert the prominent digital quantization value into physical quantity, namely radiation calibration [26]. In order to eliminate the influence of atmospheric reflection on the ground, we need to correct the atmospheric reflectance. In order to eliminate the influence of terrain or the distortion caused by camera azimuth angle and produce planar Orthophoto Image, orthorectification is required, which is also the highest level of geometric correction. Not only traditional geometric correction is carried out, but also the image caused by terrain fluctuation is corrected, and the elevation information is added to the image [27]. In order to match and overlap multiple images acquired at different times, image registration is needed. Image fusion can resample low spatial resolution multispectral images and high spatial resolution panchromatic images into high-resolution remote sensing images, so that they have high spatial resolution and multispectral features. Finally, we need to mosaic the image to form the edge line and then cut it to obtain the satellite image.

C. VEHICLE DETECTION METHOD

Vehicle detection methods are generally divided into vehicle feature-based method, template matching method and vehicle motion based method. This paper studies the deep learning algorithm of machine learning method.

The method based on vehicle features uses the explicit external characteristics of the vehicle to find the consistent features in the image to determine the vehicle in the image. Among them, the main vehicle characteristics that can be used include symmetry, color, shadow, texture, horizontal / vertical edge, lamp, etc. [28]. The method based on template matching is mainly to model the vehicle shape model in advance, and carry out correlation matching operation between the established vehicle shape model and the detected image, and the vehicle position is determined according to the degree of correlation matching. The traditional vehicle detection algorithm based on vehicle motion mainly has three stages: first, the candidate region is formed, then the feature is extracted, and finally the target object is classified by the classifier [29]. At present, the common vehicle detection algorithms are frame difference method, background difference method and optical flow method.

1) INTER FRAME DIFFERENCE METHOD

Inter frame difference method is widely used in moving vehicle detection. Two frame difference method needs to observe two consecutive images. The current frame subtracts the previous frame's gray level to get the difference image of the two images. Then, the threshold is selected to segment the foreground and background of the image, in which the pixel greater than the threshold is the front scenic area, that is, the candidate vehicle.

The threshold value t is set, and the threshold value of the difference image $G_i(x, y)$ is binarized, and the binary image $f_i(x, y)$ is obtained. The calculation formula is as follows:

$$f_i(x, y) = \begin{cases} 1, & G_i(x, y) > t \\ 0, & G_i(x, y) \leq t \end{cases} \tag{1}$$

When the difference between frames is greater than the threshold t, it is the foreground, and less than the threshold t is the background.

The inter frame difference method has the advantages of small amount of calculation, simple algorithm, not easy to be affected by light and lens factors, and can better remove static objects. However, it has inevitable disadvantages. It cannot be used in moving cameras and cannot identify stationary vehicles or slow moving vehicle targets. When the position change of moving target between two adjacent frames is not obvious, there will be holes, and the complete target cannot be detected [30], [31].

2) BACKGROUND DIFFERENCE METHOD

Background subtraction is one of the common sensing technologies. Its principle is to obtain the road background image of the vehicle moving area, and then use the current input image and the saved background image for differential operation to obtain the difference image. The binary image can be distinguished by differential image processing. In the background difference method, the accuracy of background image selection is directly related to the accuracy of the final detection results. In other words, in an ideal situation, the background image is completely static.

In the background difference method, firstly, the current frame $N_i(x, y)$ and the background image $D_i(x, y)$ are differential operated to obtain the gray image $M_i(x, y)$ of the target moving area, and then the gray image is processed by threshold T binarization to extract the moving region $S_i(x, y)$, and the background image will be updated adaptively according to the current frame. The calculation formula of background difference method is shown in Formula 2:

$$S_i(x, y) = \begin{cases} 1, & M_i(x, y) > t \\ 0, & M_i(x, y) \le t \end{cases} \tag{2}$$

Background subtraction method is simple and easy to use, with relatively small amount of calculation, high speed, powerful real-time performance and relatively complete feature data. However, this method is weak in terms of interference prevention and robustness, so it is very sensitive to various environmental interference and noise (light change, lens jitter, etc.) in the scene [32]. In practical applications, in order to adapt to the changes of the actual scene, it is often necessary to update the background in real time.

3) OPTICAL FLOW METHOD

The optical flow of the algorithm is based on the motion of the target and the camera in the traffic video. The movement of moving target can be regarded as the set of displacements on the screen. The displacement represents the moving distance of the moving target in the traffic scene. The process of projection movement is called optical flow. Optical flow, also known as the instantaneous speed of moving pixels in an image, can reflect the shape and action of moving objects.

Several assumptions are needed to calculate the optical flow. Firstly, the most basic assumption is that the brightness is constant, that is, the brightness between two adjacent frames is constant, which is the basis of the basic equation of optical flow method. Secondly, the motion between adjacent frames is very small, through which the optical flow method can approximately calculate the partial derivative of the gray change. Finally, it is necessary to keep the spatial consistency and have the same motion on the pixels of the same sub image, but this is the unique premise of Lucas Kanade optical flow method, which is not available in other optical flow methods [33].

The advantage of optical flow method is that it can detect and track moving targets better in moving cameras, and has higher accuracy, and can solve the problems of occlusion and overlap to a certain extent [34]. Its disadvantages are large amount of calculation, complex calculation process, long time-consuming, poor robustness, not very good anti noise, and very sensitive to light.

III. EXPERIMENTS OF VEHICLE DETECTION BASED ON DEEP LEARNING ALGORITHM

A. VEHICLE IMAGE CLASSIFICATION AND DETECTION

1) IMAGE CLASSIFICATION

Image classification is the basis of target detection. Therefore, before vehicle detection, it is necessary to classify the target in high-resolution satellite remote sensing image. In this study, the Alexnet neural network model is used for image classification.

There are eight layers in the Alexnet network structure, including five volume layers and three full connection layers. Finally, the output transfer of full connection layer corresponds to 1000 categories of Imagenet. Because alexnet has two GPUs for training, the network structure diagram is divided into two parts. The upper level of one GPU training chart and the lower level of the other GPU training chart. The two GPUs communicate only between the third convolution layer and the three full connection layers.

2) VEHICLE TARGET DETECTION

In this study, Faster R-CNN model is used for vehicle detection, which is a more accurate and high-speed algorithm model based on convolutional neural network. Faster R-CNN is a general target detection network. It is not accurate to detect small vehicle targets directly. Therefore, it is necessary to adjust parameters to adapt to the vehicle target detection task in satellite remote sensing images.

First, set the initial preselects of the model to 32, 64, and 128. Firstly, the two-dimensional convolution with 3×3 convolution kernel size and 64 output feature dimension is carried out, and the two-dimensional maximum pooling is performed. The size of pooling kernel is 2×2. Then the convolution kernel is 3×3, and the two-dimensional convolution of feature dimension 128 is output, and the two-dimensional maximum pooling of pooling kernel is 2×2.

The images detected in this study are from high-resolution satellite remote sensing images, and the size of the images is consistent. There are 160 training sets and 45 test sets. Among them, the background to be detected includes roads, green plants and buildings.

In order to get more training data, we need to expand the training set, and add random number and random angle vehicle images to the training set images. Finally, there are 7239 images in the training set. In order to detect better, it is necessary to balance the training set. First of all, the positive and negative samples are used. After the vehicle in the picture is cut down, the vehicle is added artificially in the picture to achieve the balance of positive and negative samples. Then copy the less part in the positive sample to ensure the balance of different color vehicles in the sample.

First of all, image preprocessing and convolution feature calculation are needed for vehicle detection. The image size should be cut to 1024×512, and the value should be adjusted to 0 to 1. According to the intersection ratio between candidate frame and vehicle target, binary label is assigned to candidate frame. The positive and negative of binary label need to be distinguished according to the size of region frame. Network training is carried out and parameters are adjusted to meet the requirements. Then, the detection network is used for classification and border regression, and a fixed size feature map is generated through pooling layer, which is input into the full connection layer, and the network is extracted and optimized alternately by sharing features. Finally, the maximum suppression is performed. Q is the set of detection frames, D is the set of confidence scores corresponding to the detection frames, and L is the final set of detection frames. Q is sorted in descending order according to the score corresponding to D. the first detection frame is placed in L, and the detection frame is removed from Q. The IoU value of the remaining detection frame and the first detection frame is calculated. The detection frame whose IOU value is greater than the threshold t is removed from Q, and the operation is repeated until there is no detection frame in Q.

B. NETWORK MODEL OPTIMIZATION

The design of Faster R-CNN takes into account the general situation of target detection, so there are many layers in the feature extraction network, including multiple pooling layers, so it is difficult to detect small objects. Besides adjusting the parameters, the algorithm needs to be optimized.

1) MODEL PRUNING

Pruning the light weight connections in the network can reduce the complexity of the network model and reduce the over fitting. The specific steps of pruning are as follows: first, the original network model is trained normally, then all the links corresponding to the weight less than the threshold are pruned, and finally the pruned network model is trained again.

2) MODEL QUANTIFICATION

Model quantization is the quantization of weight matrix and gradient matrix, which can reduce the number of bits of data representation and calculation [35].

The ownership in the matrix is regrouped and classified, and the offsets of four cluster centers with different colors and 16 weights relative to the cluster centers are obtained. When the network calculates the gradient matrix, the gradient matrix is also clustered in the same cluster center. The clustering results of the four centers are summed and multiplied by the learning rate to update the clustering of the four weight matrices.

By pruning and quantifying the model, the network model can be compressed to obtain a good sparse representation, and the detection accuracy of the network model will not be affected.

C. PRACTICAL APPLICATION OF VEHICLE DETECTION

After testing the optimized network model of Faster R-CNN, it is proved that the algorithm model has high accuracy and robustness, which can be used in practical application for further detection experiments.

In this study, a section of intersection in a city is selected as the detection site, and the vehicle target detection is carried out on the local image transmitted by satellite remote sensing, and the traffic flow and speeding vehicles at the intersection in different periods from December 18 to 20, 2019 are recorded. Compared with the actual number of vehicles, the algorithm can effectively detect vehicles in the actual road section.

TABLE 1. Confusion matrix of classification results.

The true situation	Forecast results	
	Positive example	Counter example
Positive example	TP	FN
Counter example	FP	TN

IV. DISCUSSION ON EFFECT OF VEHICLE DETECTION BASED ON DEEP LEARNING ALGORITHM

A. AVERAGE ACCURACY OF VEHICLE DETECTION

Taking the average accuracy rate as the evaluation index of network detection performance, the samples are divided into true positive (TP), false positive (FP), true negative (TN) and false negative (FN). The total number of all types is the total number of samples. The "confusion matri" of classification results is as follows:

As shown in Table 1, these are four cases of the sample, and the precision ratio n and recall ratio w can be calculated. The calculation formula is as follows:

$$N = \frac{TP}{TP + FP} \tag{3}$$

$$M = \frac{TP}{TP + FN} \tag{4}$$

Taking the precision and recall as the vertical and horizontal axes respectively, an N-W curve is obtained. The average accuracy R is the area under the curve of the N-W curve. The calculation formula is as follows:

$$R = \sum_{x=1}^{i} N(x) \Delta W(x) \tag{5}$$

where x is the x-th sample after sorting, i is the number of samples, $N(x)$ is the precision ratio to the x-th sample, and $\Delta W(x)$ is the change of recall from the x-1 sample to the x-th sample. Under the detection of optimized Faster R-CNN algorithm, the N-W curve is as follows:

As shown in Figure 1, the N-W curve is relatively full, and the average accuracy R value of Faster R-CNN for vehicle target detection in remote sensing images is 70.34%, and the detection effect is good. The detection method based on convolutional neural network has good segmentation accuracy, which makes the detection method has high accuracy and robustness.

B. ACTUAL RESULTS OF VEHICLE DETECTION

1) TEST RESULTS OF TRAFFIC FLOW

The network model after the test is put into the practical application of vehicle detection at an intersection. The traffic flow data of the intersection in three different periods from December 18 to December 20, 2019 are recorded as follows:

As shown in Figure 2, no matter which day, the traffic flow at different time periods varies greatly. The peak traffic flow in the morning and evening is larger, but after 10:00 p.m., the traffic flow decreases greatly. In the same period of time, the three-day traffic flow is relatively close, probably because it is working days, and the number of vehicles to and from work and pick up children to and from school is relatively stable.

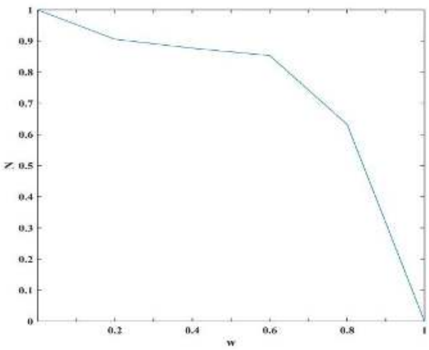


FIGURE 1. N-W curve.

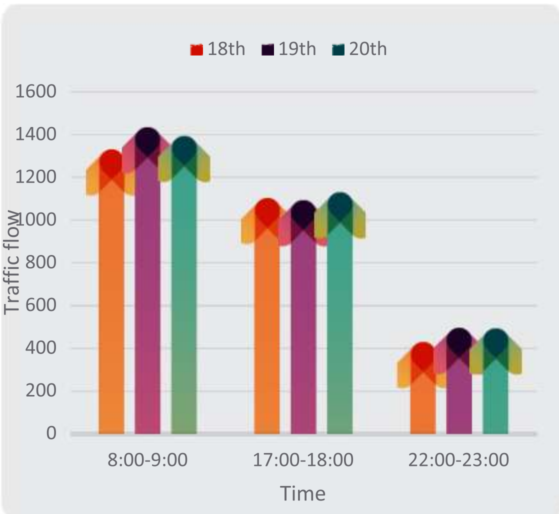


FIGURE 2. Traffic flow statistics.

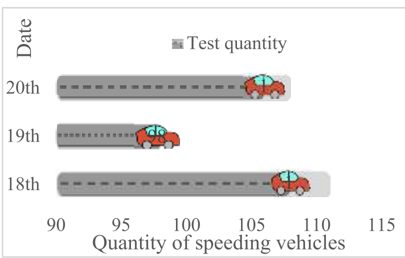


FIGURE 3. Number of detected and actual speeding vehicles.

2) DETECTION RESULTS OF SPEEDING VEHICLES

The number of speeding vehicles in the intersection from December 18 to 20, 2019 is detected, the number of speeding vehicles in three days is recorded, and compared with the real situation, the false detection rate and missing detection rate are calculated. The detected and actual speeding vehicles under the Faster R-CNN model are as follows:

As shown in Figure 3, the number of speeding vehicles detected under the Faster R-CNN optimization model is almost the same as the actual number, and the number on the 19th day is the same, and there is a slight deviation in the other two days, but only 2 and 3 vehicles are missing. The total correct rate of detection is as high as 98.42%. The results are as follows:

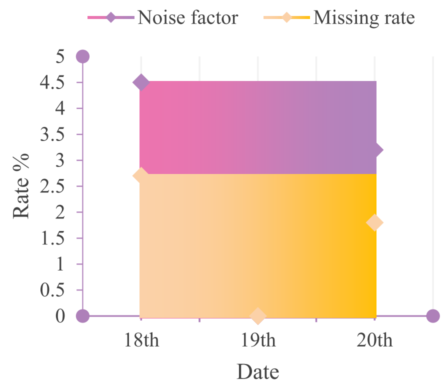


FIGURE 4. Noise factor and missing rate.

As shown in Figure 4, the noise factor and missing rate of Faster R-CNN optimization model are very low. On December 18, they are relatively high, 4.5% and 2.7% respectively, but this is also a very low level, which may be affected by the rainy weather on that day. On the 19th day, there was no missed and false detection. On the 20th day, the noise factor was 3.2% and the missing rate was 1.8%. This shows that the Faster R-CNN optimization model can be used for vehicle detection with high accuracy.

V. CONCLUSION

Vehicle detection can obtain the information of road condition and traffic flow, which is the basic guarantee of traffic management system. Traditional vehicle detection methods, such as inter frame difference method, background difference method and optical flow method, are all based on vehicle motion detection methods, and their detection effect is poor. The application of deep learning algorithm in all walks of life has a very huge role, its application in intelligent transportation system is icing on the cake.

Satellite remote sensing technology can provide accurate traffic image for deep learning algorithm. Based on this, the optimized faster-cnn model in this study has strong accuracy in vehicle target recognition and detection, and is more accurate in detecting traffic flow and speeding.

Due to the limited time and knowledge, there are some deficiencies in the process of this study. Although the accuracy of the optimization algorithm in this study has been improved to a certain extent, compared with the detection target with complete accuracy and zero error, there is still a big gap. In the next research work, the accuracy of the algorithm needs to be further improved.

REFERENCES

[1] G. Xiao, R. Wang, C. Zhang, and A. Ni, "Demand prediction for a public bike sharing program based on spatio-temporal graph convolutional networks," *Multimedia Tools Appl.*, Mar. 2020, doi: 10.1007/s11042-020-08803-y.

[2] S. Wan, X. Li, Y. Xue, W. Lin, and X. Xu, "Efficient computation offloading for Internet of vehicles in edge computing-assisted 5G networks," *J. Supercomput.*, vol. 76, no. 4, pp. 2518–2547, Apr. 2020.

[3] L. Wu, C.-H. Chen, and Q. Zhang, "A mobile positioning method based on deep learning techniques," *Electronics*, vol. 8, no. 1, p. 59, Jan. 2019.

[4] A. K. Dutta, M. Elhoseny, V. Dahiya, and K. Shankar, "An efficient hierarchical clustering protocol for multihop Internet of vehicles communication," *Trans. Emerg. Telecommun. Technol.*, vol. 31, no. 5, p. e3690, Jul. 2019.

[5] Z. Lv, S. Zhang, and W. Xiu, "Solving the security problem of intelligent transportation system with deep learning," *IEEE Trans. Intell. Transp. Syst.*, early access, Mar. 20, 2020, doi: 10.1109/TITS.2020.2980864.

[6] X. Chen, S. Xiang, C.-L. Liu, and C.-H. Pan, "Vehicle detection in satellite images by hybrid deep convolutional neural networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 10, pp. 1797–1801, Oct. 2014.

[7] S. R. V. Kittusamy, M. Elhoseny, and S. Kathiresan, "An enhanced whale optimization algorithm for vehicular communication networks," *Int. J. Commun. Syst.*, p. e3953, Apr. 2019.

[8] Z. Chen, C. Wang, H. Luo, H. Wang, Y. Chen, C. Wen, Y. Yu, L. Cao, and J. Li, "Vehicle detection in high-resolution aerial images based on fast sparse representation classification and multiorder feature," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 8, pp. 2296–2309, Aug. 2016.

[9] M. Elhoseny, "Multi-object detection and tracking (MODT) machine learning model for real-time video surveillance systems," *Circuits, Syst., Signal Process.*, vol. 39, pp. 611–630, Aug. 2019.

[10] H. Wang, Y. Cai, X. Chen, and L. Chen, "Occluded vehicle detection with local connected deep model," *Multimedia Tools Appl.*, vol. 75, no. 15, pp. 9277–9293, Aug. 2016.

[11] T. Tang, S. Zhou, Z. Deng, L. Lei, and H. Zou, "Arbitrary-oriented vehicle detection in aerial imagery with single convolutional neural networks," *Remote Sens.*, vol. 9, no. 11, p. 1170, Nov. 2017.

[12] R. Velazquez-Pupo, A. Sierra-Romero, D. Torres-Roman, Y. Shkvarko, J. Santiago-Paz, D. Gómez-Gutiérrez, D. Robles-Valdez, F. Hermosillo-Reynoso, and M. Romero-Delgado, "Vehicle detection with occlusion handling, tracking, and OC-SVM classification: A high performance vision-based system," *Sensors*, vol. 18, no. 2, p. 374, Jan. 2018.

[13] D. Shen, G. Wu, and H. Suk, "Deep learning in medical image analysis," *Annu. Rev. Biomed. Eng.*, vol. 19, pp. 221–248, Jun. 2017.

[14] Y. Chen, Z. Lin, X. Zhao, G. Wang, and Y. Gu, "Deep learning-based classification of hyperspectral data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2094–2107, Jun. 2014.

[15] S. Levine, P. Pastor, A. Krizhevsky, J. Ibarz, and D. Quillen, "Learning hand-eye coordination for robotic grasping with deep learning and large-scale data collection," *Int. J. Robot. Res.*, vol. 37, nos. 4–5, pp. 421–436, Apr. 2018.

[16] A. Rajkomar *et al.*, "Scalable and accurate deep learning for electronic health records," *npj Digit. Med.*, vol. 1, no. 1, 2018, Art. no. 18.

[17] S. K. Lakshmanaprabu, M. Elhoseny, and K. Shankar, "Optimal tuning of decentralized fractional order PID controllers for TITO process using equivalent transfer function," *Cognit. Syst. Res.*, vol. 58, pp. 292–303, Dec. 2019.

[18] S. Pan, R. Hu, S.-F. Fung, G. Long, J. Jiang, and C. Zhang, "Learning graph embedding with adversarial training methods," *IEEE Trans. Cybern.*, vol. 50, no. 6, pp. 2475–2487, Jun. 2020.

[19] X. Sun, H. Zhang, W. Meng, R. Zhang, K. Li, and T. Peng, "Primary resonance analysis and vibration suppression for the harmonically excited nonlinear suspension system using a pair of symmetric viscoelastic buffers," *Nonlinear Dyn.*, vol. 94, no. 2, pp. 1243–1265, Oct. 2018.

[20] T. O'Shea and J. Hoydis, "An introduction to deep learning for the physical layer," *IEEE Trans. Cognit. Commun. Netw.*, vol. 3, no. 4, pp. 563–575, Dec. 2017.

[21] X. Wang, L. Gao, S. Mao, and S. Pandey, "CSI-based fingerprinting for indoor localization: A deep learning approach," *IEEE Trans. Veh. Technol.*, vol. 66, no. 1, pp. 763–776, Jan. 2017.

[22] K. Sirinukunwattana, S. E. A. Raza, Y.-W. Tsang, D. R. J. Snead, I. A. Cree, and N. M. Rajpoot, "Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1196–1206, May 2016.

[23] C. E. Moore, T. Brown, T. F. Keenan, R. A. Duursma, A. I. J. M. van Dijk, J. Beringer, D. Culvenor, B. Evans, A. Huete, L. B. Hutley, S. Maier, N. Restrepo-Coupe, O. Sonnentag, A. Specht, J. R. Taylor, E. van Gorsel, and M. J. Liddell, "Reviews and syntheses: Australian vegetation phenology: New insights from satellite remote sensing and digital repeat photography," *Biogeosciences*, vol. 13, no. 17, pp. 5085–5102, Sep. 2016.

[24] J. L. Peña-Arancibia, M. Mainuddin, J. M. Kirby, F. H. S. Chiew, T. R. McVicar, and J. Vaze, "Assessing irrigated agriculture's surface water and groundwater consumption by combining satellite remote sensing and hydrologic modelling," *Sci. Total Environ.*, vol. 542, pp. 372–382, Jan. 2016.

[25] H. J. Lee, R. B. Chatfield, and A. W. Strawa, "Enhancing the applicability of satellite remote sensing for $PM_{2.5}$ estimation using MODIS deep blue AOD and land use regression in California, United States," *Environ. Sci. Technol.*, vol. 50, no. 12, p. 6546, 2016.

[26] H. Zhang, G. Zhang, J. Yong-Hua, and T. Wang, "A SRTM-DEM-controlled ortho-rectification method for optical satellite remote sensing stereo images," *Acta Geodaetica et Cartographica Sinica*, vol. 45, no. 3, pp. 326–331 and 378, 2016.

[27] A. D. Richardson, K. Hufkens, T. Milliman, and S. Frolking, "Intercomparison of phenological transition dates derived from the PhenoCam dataset V1.0 and MODIS satellite remote sensing," *Sci. Rep.*, vol. 8, no. 1, Dec. 2018, Art. no. 5679.

[28] A. Kundu, S. Dwivedi, and D. Dutta, "Monitoring the vegetation health over India during contrasting monsoon years using satellite remote sensing indices," *Arabian J. Geosci.*, vol. 9, no. 2, p. 144, Feb. 2016.

[29] X. Yang, Y. Zheng, G. Geng, H. Liu, H. Man, Z. Lv, K. He, and K. de Hoogh, "Development of $PM_{2.5}$ and NO_2 models in a LUR framework incorporating satellite remote sensing and air quality model data in pearl river delta region, China," *Environ. Pollut.*, vol. 226, pp. 143–153, Jul. 2017.

[30] K. Mu, F. Hui, and X. Zhao, "Multiple vehicle detection and tracking in highway traffic surveillance video based on SIFT feature matching," *J. Inf. Process. Syst.*, vol. 12, no. 2, pp. 183–195, 2016.

[31] N. Audebert, B. Le Saux, and S. Lefèvre, "Segment-before-detect: Vehicle detection and classification through semantic segmentation of aerial images," *Remote Sens.*, vol. 9, no. 4, p. 368, Apr. 2017.

[32] T. Tang, S. Zhou, Z. Deng, H. Zou, and L. Lei, "Vehicle detection in aerial images based on region convolutional neural networks and hard negative example mining," *Sensors*, vol. 17, no. 2, p. 336, Feb. 2017.

[33] J. Sang, P. Guo, Z. Xiang, H. Luo, and X. Chen, "Vehicle detection based on faster-RCNN," *Chongqing Daxue Xuebao/J. Chongqing Univ.*, vol. 40, no. 7, pp. 32–36, 2017.

[34] H. Kuang, L. Chen, F. Gu, J. Chen, L. Chan, and H. Yan, "Combining region-of-interest extraction and image enhancement for nighttime vehicle detection," *IEEE Intell. Syst.*, vol. 31, no. 3, pp. 57–65, May 2016.

[35] H.-B. Zeng, X.-G. Liu, W. Wang, and S.-P. Xiao, "New results on stability analysis of systems with time-varying delays using a generalized free-matrix-based inequality," *J. Franklin Inst.*, vol. 356, no. 13, pp. 7312–7321, Sep. 2019.

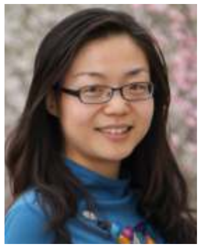
[36] T. Yang, X. Wang, B. Yao, J. Li, Y. Zhang, Z. He, and W. Duan, "Small moving vehicle detection in a satellite video of an urban area," *Sensors*, vol. 16, no. 9, p. 1528, Sep. 2016.

[37] C. Wang, Y. Fang, H. Zhao, C. Guo, S. Mita, and H. Zha, "Probabilistic inference for occluded and multiview on-road vehicle detection," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 1, pp. 215–229, Jan. 2016.

QULIN TAN was born in Ningdu, Jiangxi, China, in 1975. He received the Ph.D. degree in cartography and geographical information system from the Institute of Remote Sensing Applications, Chinese Academy of Sciences, Beijing, China, in 2003. He is currently working with the School of Civil Engineering, Beijing Jiaotong University, Beijing. His current research interests include surveying, mapping, remote sensing, and geographical information system applications in transportation field.

JUAN LING was born in Linyi, Shandong, China, in 1981. She received the master's degree from the Chengdu University of Technology, China. She is currently working with the School of Tourism, Resources and Environment, Zaozhuang University. Her research interests include remote sensing, GIS, and big data analysis.

JUN HU was born in Jiujiang, Jiangxi, China, in 1964. He received the Ph.D. degree from Beijing Jiaotong University, China, in 2005. He is currently working with the School of Computer and Information Technology, Beijing Jiaotong University. His research interests include visualization technology, data mining, and data processing.



XIAOCHUN QIN was born in Baotou, China, in 1982. She received the Ph.D. degree in highway and railway engineering from the South China University of Technology, Guangzhou, China, in 2007. She is currently working with the School of Civil Engineering, Beijing Jiaotong University, Beijing, China. Her current research interests include transportation network planning and highway and railway landscape environment.

JIPING HU was born in Xiangtan, Hunan, China, in 1967. He received the master's degree from Southwest Jiaotong University, China, in 1994. He is currently working with the School of Civil Engineering, Beijing Jiaotong University. His research interests include surveying engineering, RS, and GIS applications.