Chapter 32
Vehicle Detection Based on Area and Proportion Prior with Faster-RCNN

Hao Yuan, Bin Zhang, and Ming Liu

Abstract With the development of neural networks, detection accuracy and speed constantly improved. However, the detection effect is still insufficient in some special scenarios such as traffic environment. Therefore, we combine neural network with prior knowledge to improve its performance in vehicle detection. In this paper, we propose two effective prior: proportion and area prior to enhance the vehicle detection ability of neural network in traffic environment. The proportion and area prior is the statistical data of the vehicle at different angles and distances from the camera. In the traffic monitoring video, the proportion of most vehicles is mainly divided into several values. The area of all vehicles is also included between the thresholds. Experimental results demonstrate the effect of prior. Detection effect for vehicle in traffic environment of sample network in this paper increase by 6.56%.

32.1 Introduction

Vehicle detection is the technique that uses computer to simulate human eyes to acquire vehicles in images captured in different scenarios. This technology is highly desired in intelligent transport system and intelligent video surveillance. First, detecting vehicle can significantly increase work efficiency of traffic department and save the manpower. If traffic condition is represented by accurate statistics, more useful information can be read from traffic monitor video. Second, unmanned vehicles [1] also need vehicle detection technology. When unmanned vehicle is running, how to find the barrier ahead is the essential factor. During the last year, unmanned vehicle have developed much, vehicle equipped with a variety of sensors can travel well without driver. But when comes to the environment of road monitoring, only a few

H. Yuan · B. Zhang (✉)
School of Software, Xi'an Jiaotong University, 710049 Xi'an, China
e-mail: bzhang82@mail.xjtu.edu.cn

M. Liu
Command and Dispatch Section of Science and Technology Branch of Traffic Police, Branch of Xi'an Public Security Bureau, 710049 Xi'an, China

S.-L. Peng et al. (eds.), *Sensor Networks and Signal Processing*, Smart Innovation, Systems and Technologies 176, https://doi.org/10.1007/978-981-15-4917-5_32

methods try to detect vehicle based on aerial image. In recent research, most vehicle detection methods are based on object detection which are mainly divided into traditional methods and deep learning methods.

In this paper, we propose two prior knowledge: proportion and area prior aiming at improving detection effect on traffic environment. Two priors are based on statistics of vehicles in traffic monitor video. We find that, different to images from unmanned vehicle or aerial photography, vehicles in traffic monitor video have multi-proportion and size based on height and angle of cameras on road. Different angles and different distances generate different values of prior. Therefore, vehicles in every part on image have a different prior. If we conclude these values and combine them with network training, we can improve the detect effect of vehicle in on-road monitor video.

This paper has applied method which introduced on the above. We use faster-RCNN object detection framework as example, as experimental data shows, our approach is valid and able to detect distant vehicle in image, which shows the enhancement of detect ability.

The rest of this paper is organized as follows: In Sect. 32.2, it introduces the background of the object detection in past years from two aspects: one is traditional method and the other is object detection by neural network. Recent research on vehicle detection is also introduced. In Sect. 32.3, definition and extraction of prior knowledge are discussed in this part. In Sect. 32.4, the training and prediction method is evaluated and analyzed by the experiment. Section 32.5 draws the conclusion of this paper.

32.2 Related Work

32.2.1 Background on Object Detection

Recent years, the way to detect target in image mainly is divided into traditional method and deep learning method. Two methods are different in many aspects.

In implementation of object detection, traditional algorithm mainly detects the target through hand-designed feature extractor and classifier. When image is inputed, pre-processing such as denoising, image enhancement is performed, and then the possible region of target is selected by sliding window. For each region, its feature is extracted by extractor based on factors like target shape, illumination and background. Specific process is shown in Fig. 32.1a. Feature extractor commonly used are SIFT, HOG, etc. Finally, features extracted were used to trained the classifier (like SVM, AdaBoost) to obtain the category and location information of detected target. In addition, background modeling is also a common method used in video object detection. As shown in Fig. 32.1b, this method firstly obtains the background model through image sequence training, then obtains the target by subtracting new image from model. However, in practical applications, traditional methods still have the following drawbacks:

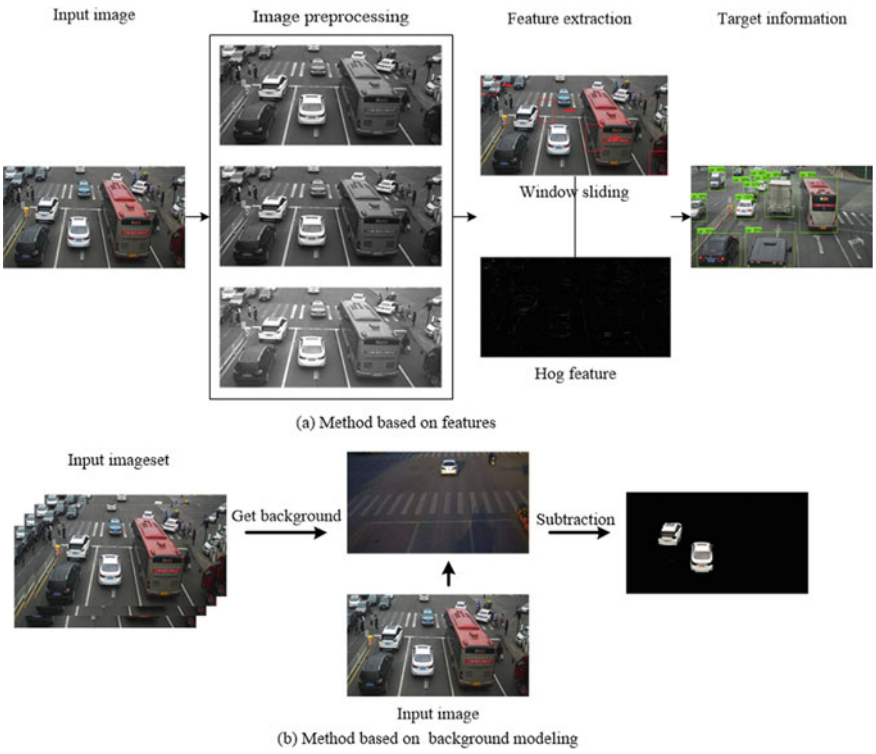

Fig. 32.1 Traditional object detection method

1. Area selection strategy based on sliding window is not targeted for identifying objects which leads to high time complexity and window redundancy.
2. Traditional methods use artificially designed feature extractors, which need to switch between different targets. Similarly, different designs have different effects. Therefore, traditional algorithms are often not robust enough. For example, in 2003, Paul Viola and Michael-Jones use Viola-Jones (VJ) [2] detector in paper published on CVPR. This method uses sliding window to select target region, then extracts Haar feature and uses AdaBoost to classify targets. In pedestrians detection [3, 4], Dalal and Triggs proposed HOG feature [5] and classified it by SVM classifier. In addition, for targets that are generally detachable into multiple different components, the deformable parts model (DPM) algorithm [6] is used. This algorithm divides the object into several components, for example, vehicles are divided into the body, window, wheel and so on.

In addition to traditional algorithms, deep learning methods [7] in real-world object detection [8] are also an important research direction in field of computer vision. In 2006, Hinton proposed deep learning [9], which uses deep neural networks to automatically get high-level features from large volumes of data. Compared with

traditional method, neural network obtains more comprehensive and accurate features, the detection effect improve greatly. Subsequently, Professor LeCun proposed convolutional neural network [10], which further developed deep learning. This network is a hierarchical model consisting of input layer, convolution layer, pooling layer, fully connected layer and output layer. It is specifically designed for image processing. The features received by convolutional layer in network are localized by previous layers. The feature is obtained by convolving shared weights. In convolutional neural network, image is extracted by initial convolution and pooling layers. As the maturity deepens, feature map generated by each layer becomes smaller and the extracted feature level becomes higher. The level of abstraction increases continuously. These features are then classified by fully connected layers and output layers to produce a one-dimensional vector which represents the current picture category. According to function of each layer, convolutional neural network can be divided into two parts: feature extraction part and classifier part.

Later, Girshick et al. proposed the RCNN model [11], which uses selective search to obtain candidate regions from input image, then the candidate regions are converted into uniform sizes and extracted feature by convolutional neural networks. Finally, multiple SVM classifiers are used to achieve multi-object detection.

In order to further improve the detection speed, faster-RCNN [12] adds a network for finding candidate boxes after convolutional layer, named region proposal network (RPN) [13]. By training this network, faster-RCNN can directly get the candidate area. We use faster-RCNN as an example to introduce how prior knowledge [14] improve the detection effect of vehicles [15, 16].

32.2.2 Background of Vehicle Detection

In recent years, due to the success of neural network, vehicle detection has developed rapidly, a variety of detect methods based on neural network sprung out. Some methods aim at detecting vehicle in on-board camera in order to build the driver assistance system [17] or intelligent vehicle [18]. This kind of methods are mainly used in driving environment. In addition, some methods focus on vehicle detection in aerial images like [19, 20], these methods mainly enhance the ability of neural network in detecting small vehicle. Some researchers also improve the detection speed by optimizing the structure of network.

32.3 Prior Knowledge

32.3.1 Definition of Prior Knowledge

In this paper, we use faster-RCNN to detect vehicles in images taken from road surveillance cameras to prove that using prior knowledge in neural network can enhance its effect on vehicle. Following types of features can be used as prior knowledge when detecting vehicles:

Prior knowledge 1: If size of image has not been clipped or scaled by network during training, the aspect ratio of rectangular frame which can contain the entire vehicle is considered as the prior knowledge. If it has been cropped, ratio is calculated based on cropped image.

Prior knowledge 2: When range of target size in training set is large, area of rectangular frame which can contain the target is also the prior knowledge. Size of area of rectangular frame corresponding to minimum target is taken as lower limit, and size of area of largest target corresponds to the rectangular frame is the upper limit.

Above are the proportion and area prior. All prior knowledge can be described by digital quantification. Prior knowledge one can optimize the interest region which generated by RPN in faster-RCNN and then improve the detection accuracy of vehicle. Prior knowledge two improved the size range of vehicle in detection. Through this prior, neural network can detect extremely small or large target better. By introducing features of detection target as prior knowledge, we can enhance the network's tendency to vehicle in different states.

Different values of area and proportion prior are mainly caused by the following factors: (a) Distance between shooting device and vehicle. When distance is large, vehicle area in image is small, vice versa. (b) Angle and height, when shooting device is in different directions of vehicle or at different shooting heights, proportion of vehicle changes greatly. For example, image taken from rear of font has quite different ratios.

32.3.2 Extracting Prior Knowledge from Object

When extracting prior experience as mentioned above, we divide vehicles in image set into five parts: vehicles far away from camera which result in small vehicle area in image, this kind of vehicles represents the minimum area threshold in detection; vehicle close to camera having the largest area in image represents the maximum area threshold; vehicles have same or opposite direction to camera; vehicle which is turning at 45 or 90° to camera. These five types of vehicles basically represent all kind of vehicles in image set.

Therefore, when neural network is training, we first get proportion of vehicle in different situations. Then obtain maximum and minimum values of vehicle area

in image. If data set has more priori value, we can set multiple scale values. As for neural network, automatically acquiring the prior knowledge by algorithm in training process still needs further research.

32.4 Prediction Method

32.4.1 Neural Network Used in Experiment

This paper combines prior knowledge to the training process of faster-RCNN-inception-v2 as example. Specific training process is shown in upper part of Fig. 32.2. During training, prior knowledge is added as a prior condition to the entire learning process, adding a different prior knowledge may generate different models.

In faster-RCNN, network firstly obtains original features from input image through feature extraction network (such as VGG [21] and resnet [22]). Then feeds the feature maps into RPN to get candidate boxes. Rest of network may extract the feature of candidate box and classify them into two outputs: recommended borders and categories (foreground and background). The processing flow is shown in lower half of Fig. 32.2.

This paper mainly modifies the regional proposal network (RPN) with prior knowledge in training process. In faster-RCNN, RPN network generates N candidate boxes in different shapes and sizes for each pixel in feature map through sliding window and then performs subsequent processing. Where N is determined by number of proportion prior and multiple. Calculation method is shown as follows:

$$N_{\text{boxes}} = N_{\text{ratios}} \times N_{\text{scales}} \tag{32.1}$$

Changing the size and shape of candidate boxes generated by regional proposal network by combining two prior knowledge can enhance the detection effect of neural network on vehicle and improve its ability on vehicles whose size is too large/small.

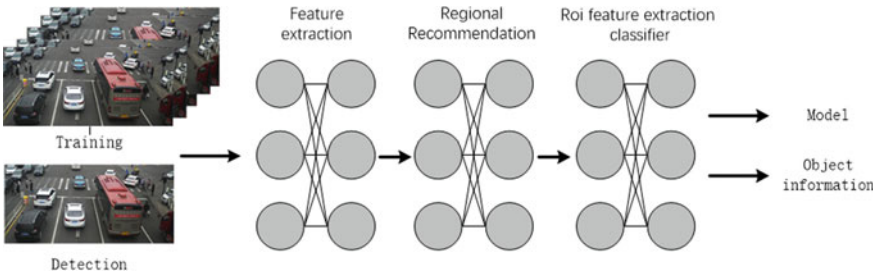

Fig. 32.2 Faster-RCNN training and detection process

32.4.2 Anchors Generation Affected by Area and Size Prior

When starting training, regional proposal network obtains feature map of input image from feature extraction network. Due to multiple convolution and pooling operations, each pixel of feature map represents a considerable area of original image. For each pixel, RPN network generates N candidate boxes corresponding to the corresponding positions of original image according to the set correlation parameters. Area prior is combined in this step to change the area size of candidate boxes. B_s parameter in following formula is the side length of generated minimum square candidate boxes, S is the area of generated candidate boxes, S_a is the magnification of other size of candidates based on B_s. When candidate box is enlarged in area, default candidate shape of boxes is square, and the area of different candidate boxes is the square of B_s, multiplied by the multiple. Amplification process is shown in Fig. 32.3b.

$$S = B_s^2 \times S_a \tag{32.2}$$

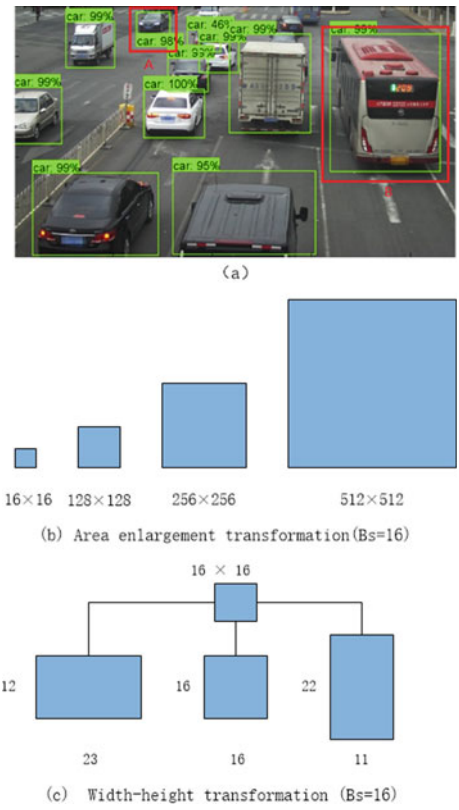

Fig. 32.3 Size and area transformation

Taking Fig. 32.3a as an example, vehicle contained by red box in upper left corner has the minimum area, size of this area is set to be the minimum value of area prior. Vehicle in red box on the right side of picture has the maximum area, so we set it to be the maximum value in area prior. Assuming that, minimum area of image is 250 and the maximum is 250,000, we can make B_s, 16 and S_a, 8, 16, and 32, respectively. Thus, after the area is enlarged, the largest candidate box can contain an area having the value of 262,144, which is larger than maximum value in area prior. By combining area priors, network can better detect vehicles that are too small/large.

The proportion prior mainly affects width and height of candidate boxes when the area has been determined. Transformation manner is shown in the following formula. P_r is proportion prior, which determines the aspect ratio of square box after the value of area has been determined by B_s, L_w represents the width of candidate boxes, and L_h is the height. Figure 32.3c is an example of width-height transformation.

$$L_b = \frac{S_a \times B_s}{\sqrt{P_r}} \tag{32.3}$$

$$L_h = L_b \times P_r = \sqrt{P_r} \times S_a \times B_s \tag{32.4}$$

Formula (32.3) represents width of candidate boxes and Formula (32.4) represents the height. S is value of area.

Then, the candidate boxes were sent to subsequent network for classification and scoring.

32.4.3 Neural Network Training

In this paper, sample network is trained with manually annotated vehicle data from road monitoring video. The algorithm is implemented on tensorflow which is an advanced deep learning framework. In network config, we set momentum of 0.9, a weight decay of 0.0005 and learning rate of 0.0001. Hardware used in experiment is a 1080ti GPU and an i7-7700 3.6 GHz CPU. In the RPN part of the network structure, we modified its P_r, B_s, and S_a parameters based on a different prior knowledge.

32.4.4 Experimental Result

In the experiment, there are 1200 images in total to be trained and tested. Among the images, there are 1000 images used for training, and 200 images are used for test. The number of steps for training is 200,000 in total. Through GPU acceleration, the network training has been finished in 20 h. The test result is shown in Tables 32.1 and 32.2. Except parameters combined with prior knowledge, the other parameters in neural network which participated in comparison are consistent. Different network's

Table 32.1 Detection effect of network under different ratios

Network	Ratio	Identification quantity
	0.5, 1.0, 2.0	1083
	1.8, 1.0, 1.1	1150
Faster-RCNN-inception-V2	0.7, 1.2, 1.5	1086
	0.5, 1.5 0.5, 1.0, 1.5, 2.0	1113 1103
	0.5, 0.75, 1.0, 1.25, 1.5, 1.75, 2.0	1099

Table 32.2 Detection quantity comparison

Network	Identification quantity	Increased percentage
Resnet-101	1089	5.60
Resnet-50	1115	3.14
Faster-RCNN-inception-V2	1150	–

detection effects under different S_a and P_r are evaluated. We run each experiment four times on same data with each parameter and take the average value as final data.

The result of experiment to tests the detection effect with different is in Table 32.1. First row in Table 32.1 is the result of unmodified network. The other lines are results with different ratios. In the test results, average detection quantity of unmodified network is less than other comparative network. In addition to comparison with different values, we also test on other networks, the result is shown in Table 32.2. In Table 32.2, we can see that on same test set, our network is 3.14 and 5.60% more detect than the other two networks. Combing the results of Tables 32.1 and 32.2, we can draw the following conclusions: By adding prior knowledge, the detect effect of the neural network can be enhanced, and the detect ability of distant vehicle can be improved.

As for P_r, we can also find that when value of P_r is closer to actual proportion, network has better performance. When the number of P_r is increased to fit all possible proportions, performance of network is also promoted. In addition, according to prior knowledge acquisition rules, appropriate value of B_s can improve the ability of network on detecting too small/large targets. Figure 32.4 is the detection result

Fig. 32.4 P_r are 0.5, 1.0, 2.0

Fig. 32.5 P_r are 0.8, 1.0, 1.1

of network whose P_r value has not been modified. Figure 32.5 shows the result of network whose P_r value has been modified with prior knowledge. From two results, we can find that network combined with prior knowledge detect more cars in small size. In other words, the network have better performance on extremely small or large vehicle.

In addition, another conclusion can be obtained through the experiment. Compared to using actual proportion values of vehicle as prior knowledge directly, increasing the number of P_r only enhances the detection effect in a small range. Data represented by histogram in Fig. 32.6 shows the influence of different S_a on detection result under the same test set. When P_r is approximately or included in 0.7–1.5, value of detection result is higher. When P_r deviates from actual ratio, network detection ability drops significantly.

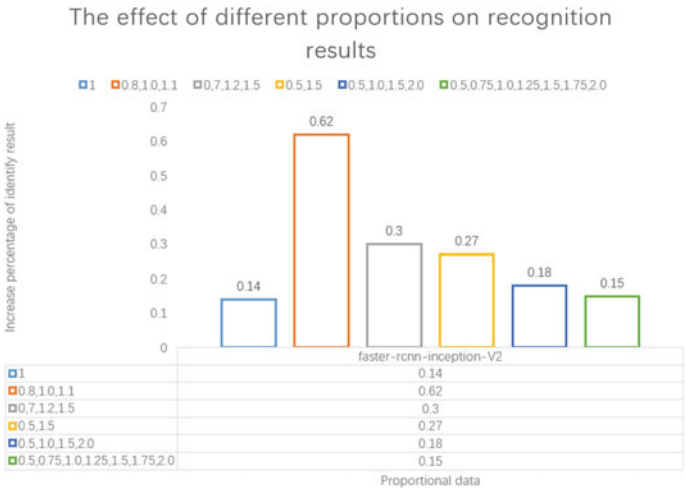

Fig. 32.6 Effect of different proportions on detection results

32.5 Conclusion

In this paper, we propose the method which adds prior knowledge into training process of neural network to enhance its detection ability for on-road vehicle. Proportion and area information of vehicle at different angles are manually extracted from training set and feed to neural network as prior knowledge. When tested on same test set, number of targets detected by modified network increased by 6.56% compared to initial network. From this result, we can reach the conclusion that the ability of network on detecting vehicle can be improved by adding prior knowledge.

In latter work, we will try to make the neural network automatically acquire prior knowledge and optimize it. Attempt will be made to train the network with prior knowledge as training data alone. Other prior knowledge should be introduced more than proportion and area prior. In addition, we will combine prior knowledge with other neural networks to test whether the proposed method is equally valid for most networks. Similarly, pedestrian and multi-object detection will also serve as our future research direction.

Compared with other algorithms, neural networks have very considerable potential advantages in task of object detection. We will try to reduce the network structure when detecting single target later to shorten the training time and improve the detection speed. At the same time, for specific goals, definition and extraction method of prior knowledge are also the future research content. Although the neural network improves the ability to detect vehicles by adding prior knowledge, we are still not sure whether the method is effective for non-vehicle targets. Therefore, this method still has great room for development.

References

1. Sun, Z., Bebis, G., Miller, R.: On-road vehicle detection: a review. IEEE Trans. Pattern Anal. Mach. Intell. **5**, 694–711 (2006)
2. Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. CVPR **1**(511–518), 3 (2001)
3. Zhao, X., Li, W., Zhang, Y., et al.: A faster RCNN-based pedestrian detection system. In: 2016 IEEE 84th Vehicular Technology Conference (VTC-Fall), pp. 1–5. IEEE (2016)
4. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05). IEEE (2005)
5. Dalal, N., Triggs, B., Schmid, C.: Human detection using oriented histograms of flow and appearance. In: European Conference on Computer Vision, pp. 428–441. Springer, Berlin, Heidelberg (2006)
6. Felzenszwalb, P.F., Girshick, R.B., McAllester, D.: Cascade object detection with deformable part models. In: 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 2241–2248. IEEE (2010)
7. Schmidhuber, J.: Deep learning in neural networks: an overview. Neural Netw. **61**, 85–117 (2015)
8. Ren, S., He, K., Girshick, R., et al.: Object detection networks on convolutional feature maps. IEEE Trans. Pattern Anal. Mach. Intell. **39**(7), 1476–1481 (2016)

9. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. Nature **521**(7553), 436 (2015)
10. Sermanet, P., Chintala, S., LeCun, Y.: Convolutional neural networks applied to house numbers digit classification. arXiv preprint arXiv:1204.3968 (2012)
11. Girshick, R., Donahue, J., Darrell, T., et al.: Rich feature hierarchies for accurate object detection and semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 580–587 (2014)
12. Ren, S., He, K., Girshick, R., et al.: Faster R-CNN: towards real-time object detection with region proposal networks. Adv. Neural Inf. Process. Syst. 91–99 (2015)
13. Ma, J., Shao, W., Ye, H., et al.: Arbitrary-oriented scene text detection via rotation proposals. IEEE Trans. Multimed. **20**(11), 3111–3122 (2018)
14. Kohara, K., Ishikawa, T., Fukuhara, Y., et al.: Stock price prediction using prior knowledge and neural networks. Intell. Syst. Account. Financ. Manage. **6**(1), 11–22 (1997)
15. Huo, Z., Xia, Y., Zhang, B.: Vehicle type classification and attribute prediction using multi-task RCNN. In: 2016 9th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), pp. 564–569. IEEE (2016)
16. Jun, S., Pei, G., Zhili, X., et al.: Vehicle detection based on faster-RCNN. J. Chongqing Univ. **40**(7), 32–36 (2017)
17. Huang, D.Y., Chen, C.H., Chen, T.Y., et al.: Vehicle detection and inter-vehicle distance estimation using single-lens video camera on urban/suburb roads. J. Vis. Commun. Image Represent. **46**, 250–259 (2017)
18. Zhang, R.H., You, F., Chen, F., et al.: Vehicle detection method for intelligent vehicle at night time based on video and laser information. Int. J. Pattern Recognit. Artif. Intell. **32**(04), 1850009 (2018)
19. Tang, T., Zhou, S., Deng, Z., et al.: Vehicle detection in aerial images based on region convolutional neural networks and hard negative example mining. Sensors **17**(2), 336 (2017)
20. Chen, X., Xiang, S., Liu, C.L., et al.: Vehicle detection in satellite images by hybrid deep convolutional neural networks. IEEE Geosci. Remote Sens. Lett. **11**(10), 1797–1801 (2014)
21. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)
22. He, K., Zhang, X., Ren, S., et al.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)