Computer Life (CPL) ISSN: 1819-4818
DELIVERING QUALITY SCIENCE TO THE WORLD

Overview of Research on Vehicle Flow Detection Algorithms Suitable for ARM

Shugang Liu [1], Yipeng Zhu [2,*]

[1]School of North China Electric Power University, Baoding, China

[2]School of North China Electric Power University, Baoding, China

*Corresponding author Email: ncepuyipengzhu@163.com

Abstract: In recent years, with the application and development of convolutional neural networks in the field of computer vision, CNN-based target detection algorithms have also developed rapidly. For ARM, lightweight architecture has always been a goal. Girshick et al. [1] proposed R-CNN Later, for the first time, the accuracy of target detection was greatly improved based on the CNN framework, and unlike the inter-frame difference method, it can be completely detected. Subsequently, CNN-based target detection algorithms, such as the improved R-CNN algorithm Fast R-CNN based on candidate regions, as well as YOLO based on regression methods and SSD methods based on single target detectors have been proposed successively and gradually applied to ARM On the detection of vehicle volume recognition. This article introduces the above several common target detection algorithms and compares their performance.
Keywords: Target detection algorithm, SSD, YOLO, R-CNN, Inter-frame difference method.

---

1. INTRODUCTION

In recent years, urban road congestion has become serious and the transportation system has become more complex. Intelligent transportation systems [2] have emerged as the times require. Intelligent transportation systems improve transportation efficiency, alleviate traffic jams, and improve road networks through the harmony and close cooperation of people, vehicles, and roads. Through capacity, reduce traffic accidents, reduce energy consumption, and reduce environmental pollution. One of the important directions is the traffic flow detection technology, which can provide accurate road condition information in real time, help drivers plan routes, and improve road utilization.

CNN is commonly used for image recognition and classification. With the continuous improvement of performance requirements, network architectures such as VGG, GoolgeNet, and ResNet have emerged. Most of them increase the number of network layers and improve accuracy. This also increases a lot of parameters. The computing power and memory of the device have put forward higher requirements. In actual application scenarios, mobile-end embedded devices do not have this feature. Therefore, lightweight network structures have begun to appear. Together with the continuously optimized ARM, traffic flows The detection enters the era of video mode [3]. Therefore, the research and application of ARM-based vehicle flow detection have important value and broad

prospects. In this paper, several target detection algorithms are compared and introduced, and their performance and development prospects are predicted.

2. SEVERAL TARGET DETECTION ALGORITHMS

2.1 Inter-frame difference method

The inter-frame difference method [4] takes advantage of the continuity of the video sequence captured by the camera. When there is no moving target in the scene to be captured, the consecutive frames are basically unchanged. When the moving target appears, there will be obvious differences between the connected frames. Commonly used frame difference methods include two-frame difference method and three-frame difference method [5] , which is to perform difference operation on two or three frames of images connected in time, and subtract the corresponding pixels to determine the absolute value of the gray difference. Whether it is within a certain threshold, when it is exceeded, we consider the target to be in motion, so as to detect the moving target.

For the selection of the threshold, if it is too high, it will cause serious "holes" in the moving target area (the moving object will have similar gray values in a few frames due to insignificant motion) [6], and if it is too low, it will introduce a lot of noise. Therefore, the threshold can be determined. According to the current image gray level to complete. The narrative is not expanded here.

The principle of the frame difference method is very simple, and the calculation amount is smaller than other detection algorithms, so it can quickly detect the moving target, but the experiment shows that the target detected by the frame difference method is not complete. As shown in Fig. 1, it is slower for motion. For the target, it is inevitable that there will be overlaps between several frames, and it is difficult to detect it.

Fig. 1Inter-frame difference method detection results

2.2 Sliding window detector

The sliding window detector is a candidate region-based, and it is also the simplest and most straightforward target detection algorithm. The principle is to determine multiple sliding windows with a fixed aspect ratio as needed. The window slides from left to right and from top to bottom, cutting the image into multiple blocks, converting it into a fixed size image, and giving it to CNN Classifier processing. After the classifier extracts 4096 features, the SVM classifier [7] is used to identify the category and another linear regressor. When there are multiple targets in multiple locations, we need multiple sliding windows to detect, so the amount of calculation will increase, so if we want to improve performance, we must reduce the number of windows[8].

2.3 R-CNN and Fast R-CNN

The R-CNN algorithm draws on the idea of sliding windows and is also a scheme for region recognition [9]. The specific steps of the algorithm are:

1. Use Selective Search to extract 2000 candidate regions from the input image (ROI).
2. These regions are converted into fixed-size images and sent to CNN to extract a fixed-length feature vector.
3. Use SVM to classify each area.

As shown in Fig. 2, the R-CNN algorithm effectively reduces the number of windows and uses fewer and higher-quality ROIs. R-CNN is faster and more accurate than sliding windows.

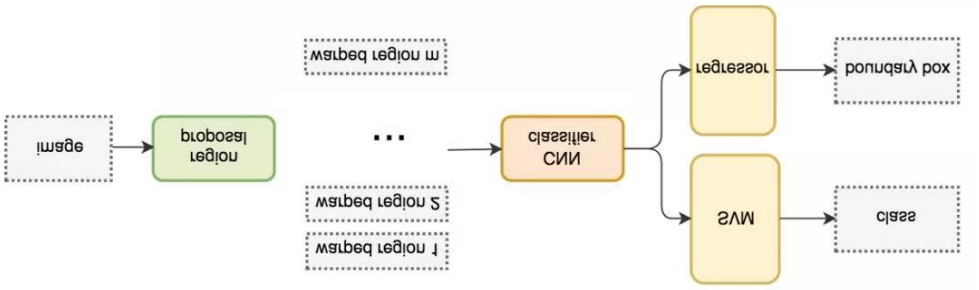

Fig. 2 R-CNN principle

In order to further improve the accuracy of the candidate region and reduce excessive overlap, we can first use CNN to extract the features of the entire image, and then use the R-CNN method on the feature map. This is Fast R-CNN [10]. This method does not repeatedly extract features, so it is faster.The ROI pooling layer extracts a corresponding large-scale feature vector from the convolution feature for each region of interest. All the feature vectors are input to the fully connected layer and the results are shared. Two branches are generated and enter two different layers. One layer is responsible for using softmax regression to calculate the probability estimation of a certain type of target and a "background" category, and the other The layer is responsible for outputting the coordinate value of the detection frame on each frame of image [11].

2.4 SSD

SSD (Single-Shot MultiBox Detector) is a single-stage detector (One-Stage) that uses the VGG-16 network as a feature extractor [12]. Add a custom convolutional layer after VGG-16, and use the convolution kernel for detection.

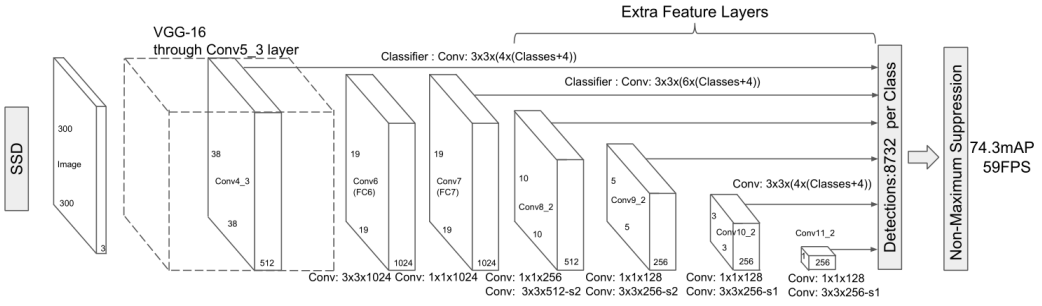

Fig. 3 SSD model framework.

Fig. 3 is the SSD structure model. We can see that the first half is the architecture of VGG-16. At the level of vgg-16, the author transforms the two fully connected layers (fc6, fc7) behind VGG-16 into

convolutional layers. The layer after conv7 is the recognition layer added by the author himself. We can observe in Fig. 3 [12] that in the conv4_3 layer, there is a Classifier layer, which uses a layer of (3,3,(4*(Classes+4))) convolution for convolution (Classes is the number of types of recognized objects, Represents the score of each object, 4 is the x, y, w, h coordinates, the front 4 is the number of default boxes), this layer of convolution is to extract the feature map not only in Conv4_3, Conv7, Conv8_2, Conv9_2, Conv10_2 and Conv11_2 all have one such convolutional layer, and finally a total of 6 feature layers are extracted. The calculation method is as follows:

The feature map size obtained by Conv4_3 is 38*38: 38*38*4 = 5776

The feature map size obtained by Conv7 is 19*19: 19*19*6 = 2166

The feature map size obtained by Conv8_2 is 10*1: 10*10*6 = 600

The feature map size obtained by Conv9_2 is 5*5: 5*5*6 = 150

The feature map size obtained by Conv10_2 is 3*3: 3*3*4 = 36

The feature map size obtained by Conv11_2 is 1*1:1*1*4 = 4

Among them, parameters 4 and 6 are the number of default boxes in each layer of feature maps, and each box corresponds to a position coordinate (loc) and score (conf) [13]. Finally, the number of feature maps obtained in each layer is added together to obtain 8732 results, and the SSD will identify the target among these results.

SSD completes detection through multiple feature maps. However, the bottom layer will not be selected to perform target detection. They have high resolution but insufficient semantic value, resulting in a significant decrease in speed and cannot be used. SSD only uses the upper layer to perform target detection, so the detection performance for small objects is poor.

2.5 YOLO

YOlO (You Only Look Once) [14] is another single-shot target detector that uses DarkNet [15] for feature detection after the convolutional layer. It does not use a multi-scale feature map for independent detection, it smoothes part of the feature map and stitches it with another lower-resolution feature map. For example, YOLO reshapes a 28*28*512 layer into 14*14*2048, and then joins it with a 14*14*1024 feature map. After that, YOLO applies the convolution kernel on the new 14*14*3072 layer to make predictions.

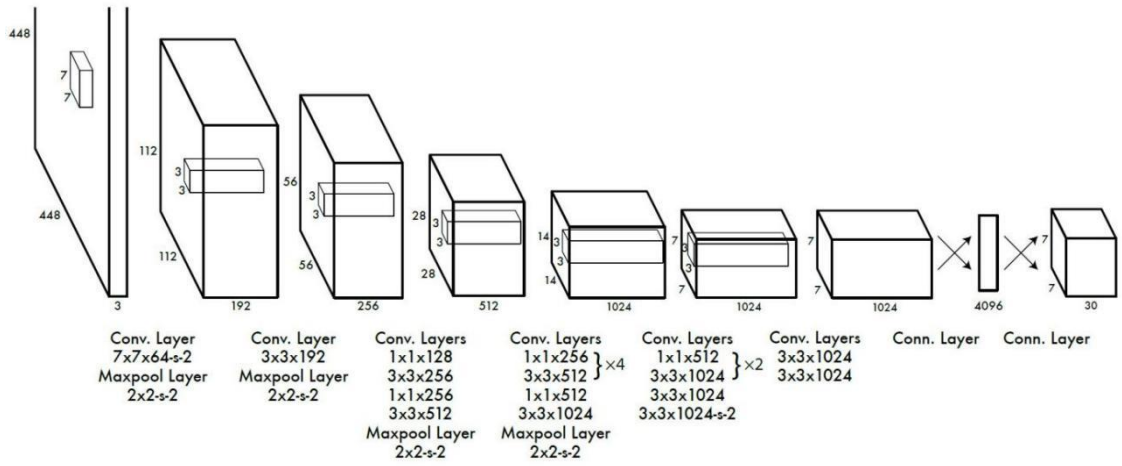

Fig. 4 YOLO model framework

Fig. 4 [14] shows the YOLO model framework. Enter a frame of 448*448 image. YOLO first divides the image into 7*7 grids. For each grid, two bounding boxes will be predicted, which are the confidence of the detection target and the class probabilities of the border area. Perform non-maximum suppression on the obtained bounding box.

It converts a target detection task into a regression problem, which can greatly speed up the detection speed and reduce the error rate. However, YOLO only uses a 7*7 rough grid for regression [16], which makes the target unable to be positioned very accurately, and the detection accuracy is not high. However, the subsequent YOLO v2, YOLO v3 and YOLO 9000 solve the above problems from different angles and greatly improve the performance of YOLO. Among them, YOLO 9000 can identify more than 9000 categories [17].

3. CONCLUSION

According to the development, the article summarizes and introduces several commonly used common or more important target detection methods. Among them, the inter-frame difference method can detect moving targets well, but it cannot extract the complete area of moving targets. It can only extract round or, and visualization is not. ideal. In addition, the inter-frame difference method relies heavily on the selection of inter-frame time intervals and thresholds, so the scope of application is limited; sliding window detectors, R-CNN, Fast R-CNN are all algorithms based on candidate regions, and their optimization is based on extracting features Figure, try to make the window less and more accurate, reduce the amount of calculation, and improve the efficiency of the algorithm; SSD and YOLO are single-shot detectors. The accuracy of the original YOLO is far less than that of R-CNN, because it does not use multi-scale feature maps like SSD to do independent detection, but with continuous development, YOLO v3 has been nearly 1000 times faster than R-CNN and 100 times faster than Fast R-CNN under the same detection accuracy [16]. Therefore, more and more people favor SSD and YOLO, and their lightweight features are also fully demonstrated on ARM [18].

REFERENCES

[1]. Zhang N , Donahue J , Girshick R , et al. Part-based R-CNNs for Fine-grained Category Detection[C]// European Conference on Computer Vision. Springer International Publishing, 2014.
[2]. Huang, W. , et al. "Next-generation innovation and development of intelligent transportation system in China." Science China Information Sciences 60.11(2017):1-11.
[3]. Xue Shiran. ARM is not only about the Internet of Things, but also artificial intelligence[J]. Microcontrollers and Embedded System Applications, 2017, 17(08): 82-83
[4]. Cheng, Y. H. , and J. Wang . "A Motion Image Detection Method Based on the Inter-Frame Difference Method." Applied Mechanics & Materials 490-491(2014):1283-1286.
[5]. Guo, W. J. , and S. D. Qiao . "Obstacle Detection Algorithm Based on Three Inter-Frame Difference Method." Journal of Shanxi Datong University(Natural Science Edition) (2015).
[6]. Weng, M. , G. Huang , and X. Da . "A new interframe difference algorithm for moving target detection." International Congress on Image & Signal Processing IEEE, 2010.
[7]. Joachims, T. . "Making large-Scale SVM Learning Practical. Advances in Kernel Methods– Support Vector Learning." (1999).
[8]. Tang Hongtao. "Multi-target tracking false alarm detection using sliding window detector." Control Engineering 11(2017): 128-132.
[9]. Zhang N , Donahue J, Girshick R , et al. Part-based R-CNNs for Fine-grained Category Detection[C]// European Conference on Computer Vision. Springer International Publishing, 2014.
[10]. Jianan, et al. "Scale-Aware Fast R-CNN for Pedestrian Detection." IEEE Transactions on Multimedia (2017).

[11].Cao Shiyu, Liu Yuehu, and Li Xinzhao. "Vehicle Target Detection Based on Fast R-CNN." Chinese Journal of Image and Graphics 5(2017).
[12].Liu, W. , et al. "SSD: Single Shot MultiBox Detector." European Conference on Computer Vision Springer, Cham, 2016.
[13].Liu Yan, Zhu Zhiyu, and Zhang Bing. "Target detection based on SSD-Mobilenet model." Ship Electronic Engineering 10(2019).
[14].Redmon, J. , et al. "You Only Look Once: Unified, Real-Time Object Detection." IEEE (2016).
[15].Setiyono, B. , D. A. Amini , and D. R. Sulistyaningrum . "Number plate recognition on vehicle using YOLO - Darknet." Journal of Physics: Conference Series 1821.1(2021):012049 (11pp).
[16].Redmon,J.,and A. Farhadi . "YOLOv3: An Incremental Improvement." arXiv e-prints (2018).
[17].Redmon, J., and A. Farhadi . "YOLO9000: Better, Faster, Stronger." IEEE (2017):6517-6525.
[18].Zhang,J. S.,J.Cao , and B. Mao. "Application of deep learning and unmanned aerial vehicle technology in traffic flow monitoring." 2017 International Conference on Machine Learning and Cybernetics (ICMLC) IEEE, 2017.