ISSN: 2664-9640

Traffic Volume Detection Method Based on Video Image Processing

Yujiong Tan

College of Traffic & Transportation, Chongqing Jiao Tong University, Chongqing 400074, China

Abstract

with the accelerating process of China's modernization and the rapid increase of vehicle ownership, vehicle monitoring has become an important technical research and development direction in the field of transportation. In order to enable drivers to better understand the traffic flow of the road section before travel, so as to better plan the travel route, reduce the waiting time and avoid a large number of vehicles. In traditional image recognition, vehicle counting is generally based on target detection, but its effect is poor in scenes with dense vehicles and serious occlusion. In addition, for scenes with multiple targets, the accuracy of ordinary algorithms is low and large errors often occur. This paper takes the driving vehicles in the road traffic scene as the research object, uses yolov5 neural network to realize the multi-target tracking of road vehicles, extracts the characteristics and position information of different kinds of vehicle images, and realizes the tracking and quantity statistics of driving vehicles. Through image processing and statistics, the traffic flow density of the road section can be obtained, which can provide real-time and dynamic reference for the driver's travel, and judge whether to go to the road section, so as to make other path decisions. The experiment is carried out on the first section of Xuefu Avenue, Nan'an District, Chongqing. The experimental results show that the yolov5 detection algorithm can meet the tracking and statistics of driving vehicles in the actual road scene, and has high accuracy in vehicle multi-target tracking and counting, and has achieved good detection results.

Keywords

Image processing;Multitarget tracking;YOLOv5;Traffic flow; Path planning.

1. Introduction

The continuous growth of car ownership in China has put more and more pressure on the urban highway transportation system, with a sharp increase in traffic flow and frequent traffic jams, and various social problems have become increasingly prominent. The most important part of the road scene is the car, so the research of car recognition and tracking technology in the road scene is becoming more and more important. Real time vehicle counting not only helps to quickly find the accident section, but also helps to predict the duration of traffic accidents, and provides a basis for the timely release of induced and predictive traffic information under accidents and the rapid elimination of the impact of accidents. In recent years, many scholars have done a lot of research on how to accurately estimate the number of vehicles in the image, and put forward many methods and schemes for vehicle counting based on deep learning[1] technology.

One class of methods is based on object detection. These methods usually try to identify and locate each vehicle in the image, and then add the number of detected objects to get the result. Another kind of method is multi-target tracking algorithm detection, which is divided into traditional methods and deep learning based methods. Traditional methods mainly include optical flow method[2], kernel correlation filtering[3-4], etc. their tracking speed is fast, but they

lack the processing of scale changing targets, resulting in poor tracking effect. Driven by the pedestrian re identification (Reid) technology[5-6], the tracking algorithm based on deep learning has achieved unprecedented development, realizing the double improvement of speed and accuracy. Reid based target tracking algorithms are mainly divided into two types: one-stage[7] method and two-stage. These two types of models are tracking by detection methods. Among them, the two stage algorithm divides the tracking process into two independent processes: detection and matching, and uses two different networks to extract features respectively. Although this can improve the accuracy of the algorithm, the calculation cost of using two network models is too high, and it is not suitable for the scene with high real-time requirements such as automatic driving. Typical two stage algorithms include deepsort[8] and hogm[9].With the development of multi-target tracking algorithm and the problem that the speed of two stage algorithm is too slow, one shot algorithm is proposed. This kind of algorithm embeds Reid into the detector, takes two independent tasks as a multi task learning model, and outputs the target detection frame and Reid apparent features at the same time by sharing weights, so as to achieve an approximate real-time speed. For example, Wang et al. [10] proposed the JDE (joint detection and embedding) algorithm. By using yolov3[11] as the detector, the Reid is embedded in the detector to jointly learn the feature representation, and the speed and accuracy are improved. Zhang et al. [13] proposed the fairmot algorithm by analyzing the deficiencies in JDE, and improved the accuracy of the algorithm by using the DLA (deep layer aggregation) [13] model and anchor free algorithm [14].Yan Kang[15] and others improved the performance of JDE tracking algorithm to a certain extent by combining space and channel attention. Xuejuntao et al. [16] used mobilenet[17] to replace the backbone feature extraction network of the yolov3 detector, significantly improving the real-time performance of the tracking algorithm, but reducing the tracking accuracy of the tracking algorithm. Mayongjie et al. [18] improved the detection accuracy of the algorithm by adding a detection head to the yolov3 algorithm and combining it with the deepsort algorithm. Although these advances have carried out preliminary research on lightweight networks, it is still a difficulty in the field of vision to pursue the best accuracy and speed compromise in a very limited computing budget. This paper will use the yolov5[19] detection method to realize multi-target tracking of road vehicles, and mark the target vehicles with horizontal borders, so as to reduce the complexity of the traditional algorithm model and effectively improve the tracking performance and accuracy of multi-target detection.

2. Key Technologies

2.1. Target detection

As the first step of intelligent traffic management system, traffic road detection technology directly affects the accuracy and effectiveness of intelligent traffic management. If the detection of traffic targets such as vehicles, traffic lights, obstacles and so on is missed or mistakenly detected, the later analysis of vehicle operation status is impossible. At present, the existing vehicle detection technologies are mainly based on radar, ultrasonic, infrared and ring sensors, which have their own advantages and disadvantages [20].

In recent years, with the rapid development of image processing technology and artificial intelligence technology, video image-based detection technology has also begun to develop[21].This technology uses the image acquisition equipment to collect the real-time image of the detected target, and processes and analyzes the collected image through the intelligent data mining algorithm, so as to realize the real-time state monitoring of the moving target. Among them, the key technologies involved mainly include three aspects: first, the selection of image acquisition equipment, most of which are cameras with low maintenance cost; Secondly, the development of intelligent data processing algorithm, adaptive genetic programming, self

coding neural network and other feature extraction methods provides a favorable tool for image processing; Thirdly, target recognition technology, support vector machine, neural network and other classification algorithms are mature and have been widely used in the field of image recognition. Therefore, the use of video detection technology not only makes it more convenient to monitor the traffic situation in real time, but also has the advantages of wide detection range, low maintenance cost and high reliability .At present, most computer-based target detection methods are used. A successful example is the well-known Viola Jones face detector around 2000, which makes target detection a relatively mature technology[22].At present, the more popular target detection methods include deformable component model (DPM), YOLO&SSD, cascade CNN, etc.

At present, some achievements have been made in the research of video detection technology at home and abroad, such as the long-distance face recognition project in the United States, which can recognize the face of people far away, expand the scope of target detection and recognition, and make the detection no longer limited by distance; In addition, through the research on real-time video surveillance technology, Maryland University has realized the recognition and segmentation of human motion, so that more elaborate target detection tasks can be carried out. In addition, international journals such as PAMI provide an interactive communication platform for domestic and foreign scholars committed to the research of video detection technology, which promotes the rapid development of video detection moving target technology. Some scientific research institutions and institutions in China have gradually begun to study video detection technology, including Tsinghua University, Chinese Academy of Sciences and Xi'an Jiaotong University. They have made fruitful achievements in vehicle real-time tracking technology, traffic real-time monitoring system and visual monitoring system.

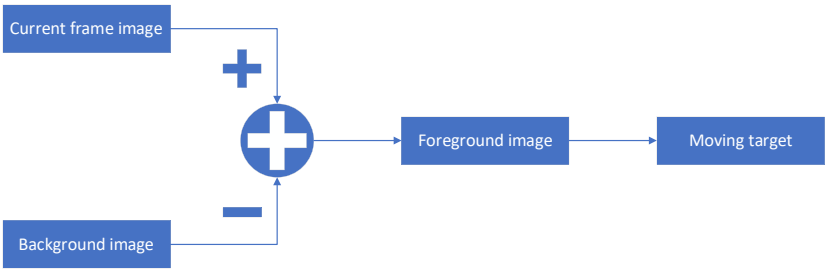

Figure 1: flow chart of background difference method

2.2. Matlab background difference method

Background difference method[23-24]is the simplest method to realize among the commonly used moving target detection methods. Its basic idea is to compare the background model with the current frame image, and use histogram and other methods to count the changes of image information, so as to segment the moving target and judge the abnormal situation. The background subtraction method can provide complete characteristic data, but it is particularly sensitive to the changes of the external environment. At present, many scholars are studying background modeling, hoping to reduce the impact of dynamic scene changes on target segmentation.McKenna et al. [25]proposed a background model combining pixel color information and gradient information to eliminate the influence of shadows on moving object segmentation. Ismail et al. [26]counted the inter frame difference values and maximum and minimum intensity values of all pixels in the scene, and then carried out background modeling and corresponding background update. When the background is known, the background subtraction method can completely segment the moving object, so the background subtraction method is a simple and effective moving object detection algorithm. In the background subtraction method, the background image should be updated in real time, because the accuracy of the background determines the performance of the target detection result and also

affects the subsequent image processing.The flow chart of extracting target by background method is shown in Figure 1.

The mathematical description of the background difference method is as follows: let the gray value at the pixel point (x, y) of the kth frame image in the video image sequence be I (x, y), and the gray value at the same position in the background image be B: (x, y), then the expressions of the difference image D, (x, y) and the binarization result T: (x, y) are shown in formulas (1-1) and (1-2):

$$D_k(x,y) = |I_k(x,y) - B_k(x,y)| \tag{1-1}$$

$$T_k(x,y) = \begin{cases} 1 & D_k(x,y) \geq T \\ 0 & D_K(x,y) < T \end{cases} \tag{1-2}$$

2.3. Yolo detection algorithm

2.3.1 YOLOv1

Target detection is usually divided into two tasks, classification and location. The target detection method before Yolo often separates the two tasks. A large number of regions of interest (ROI) are generated through the region proposal method, that is, the candidate box that may contain the target. Then the classifier is used to classify, judge whether the candidate box contains the target to be detected, and finally calculate the probability of the category to which the target belongs. YoLo is an end-to-end target detection framework. Through one image input, the coordinates of the prediction frame, the confidence of the frame containing the target, and the probability of the category can be obtained at the same time. Because Yolo realizes target detection in a neural network, it is faster to detect targets.Finally, the position of the prediction frame, the confidence of the target and the possibility of belonging to the target category are obtained.

The core idea of Yolo V1 is to use the whole picture as the input of the network, and directly return the location of the bounding box (bounding box, i.e. detection box, abbreviated as bbox) and the category of the bounding box in the output layer.The implementation of Yolo V1 is to divide an image into SXS grid cells.If the center of the target object falls in the grid, the grid is responsible for predicting the target.Yolo V1 target detection consists of three steps: image size - input network, output result - NMS. The first step, resizing, belongs to the conventional operation of deep learning, which is to resize images of different sizes to the same network structure. The IOU is the intersection over union, that is, the intersection ratio and union of the two box regions, which is used to determine the position pixel distance of the two boxes. NMS is to remove the redundant boxes by calculating the IOU. The innovation of yolov1 is that it regards detection as a regression problem and realizes a unified system with a network output location and category. From the perspective of detection, it is one-stage. Its advantage is high efficiency, because the regression problem has no complex process; Good generalization ability. When the types of training set and test set are different, Yolo performs much better than DPM and r-cnn, and there are few crashes when applied to new fields.The disadvantage is that because of the detection mechanism of yolov1, a grid can only predict one target. At this time, if two objects fall into one grid at the same time, the missed detection rate will be relatively high. Moreover, an image can only predict 98 bounding boxes, and the target positioning error is also large. A cell can only predict two boxes and one category, This spatial constraint will inevitably limit the number of predictions;In addition, the model predicts the bounding box according to the data, which is difficult to be extended to objects with new or unusual aspect ratio or configuration.

2.3.2 YOLOv2

Yolov2[27]is a revision based on yolov1. It refers to the network structure of yolov1 and SSD (single shot Multibox detector) [28], adopts a network structure similar to vgg16 [29], uses 3x3 convolution core for many times, and doubles the number of channels after each pooling

operation. The network uses global evaluation pooling, and puts 1x1 convolution kernel between 3x3 convolution kernels to compress features.Finally, the basic network model of darknet-19 is obtained, which includes 19 convolution layers and 5 maximum pooling layers. However, the amount of calculation of darknet-19 is much smaller than that of vgg16. In Imagenet[30], the accuracy of top-1 and top-5 can reach 72.9% and 91.2%.Yolov2 improves the resolution of the initial input image from 224x224 to 448x448, which improves the high-resolution training model map by 4%.Secondly, yolov1 finally uses the full connection layer to predict the frame and classification, resulting in the loss of a lot of spatial information and inaccurate positioning. Yolov2 uses the idea of anchor in RPN for reference and uses lower sampling in the convolution layer, so that the 416x416 input image finally gets a 13x13 feature map, and finally predicts 13x13x9 frames, which greatly improves the recall rate of target detection.Yolov2 also improved the method of predicting the border, using k-means clustering method to train the border. The traditional K-means method uses the Euclidean distance, which means that a large frame is more prone to error than a small frame. Therefore, the author of Yolo proposed to use the IOU (intersection union ratio) score to judge the distance, making the predicted frame more representative, so as to improve the detection accuracy.Yolov2 adopts the idea that SSD uses different feature maps to adapt to different scale targets, adds a transfer layer to the original network, and superimposes the 26x26x512 shallow feature map into a 13x13x2048 deep feature map. Such fine-grained features are very helpful for small-scale object detection. Finally, yolov2 also combines the word vector tree method to detect thousands of targets. Although it has little reference significance for the detection task in this paper, it is also a great breakthrough for the multi-target detection task. By analyzing the characteristics of traffic target detection task, this paper requires to achieve a real-time detection effect.Therefore, from the aspect of detection speed, Yolo is selected as the algorithm model of traffic target detection.Yolov2 is the most popular real-time target detection method.

2.3.3 yolov3

Yolov3 is a very, very classic algorithm of Yolo series of target detection. It is the latest version of Yolo series of target detection proposed by Joseph Redmon in 2018. The network structure of yolov3 is shown in Figure 2.

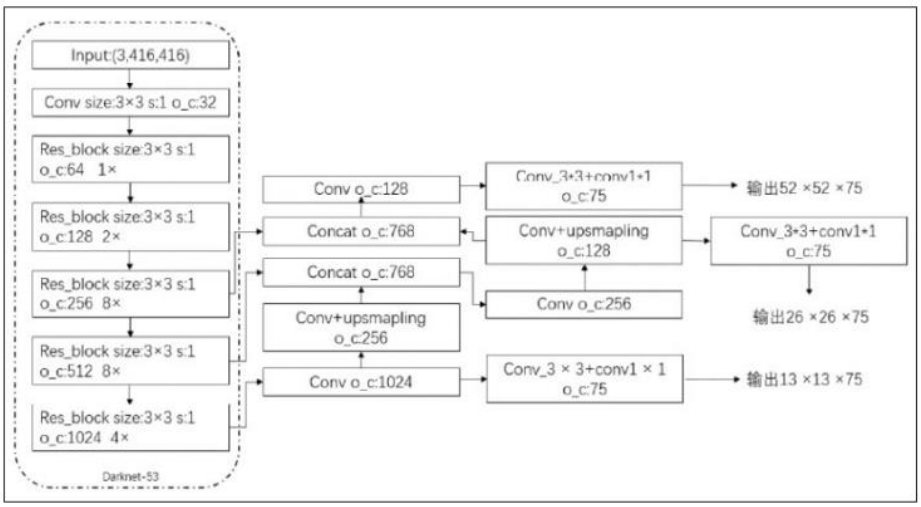

Figure 2: yolov3 network structure

Compared with yolov2, its improvement is mainly reflected in multi-scale prediction.The dimension clustering used for coordinate prediction is the same as that used for anchor boxes to predict the bounding box.During the training period, the sum of the square error losses is used, so the calculation is faster.For category prediction, yolov3 predicts the score of each border belonging to a category target through logistic regression, so as to detect that a target belongs to two label categories.For the cross-scale prediction, the main purpose is to adapt to

ISSN: 2664-9640

the objectives of different scales and make the model more universal.Especially for the detection of small targets, the accuracy has been greatly improved.Yolov3 adopts darknet-53. Compared with yolov1 and yolov2, the network structure is slightly larger, but the accuracy is much improved, especially for the detection accuracy of small targets. When achieving similar performance, yolov3 is 3 times faster than SSD and nearly 4 times faster than retinanet.Therefore, for the detection tasks with high real-time requirements such as traffic target detection, the detection effect based on Yolo model is the best.

2.3.4 yolov4

As an upgraded version of yolov3, the core structure of Yolo V4 is similar to that of yolov3. Each sub structure is improved through new algorithm ideas, and a variety of network design ideas are integrated to further improve the detection accuracy of the Yolo series of target detection algorithms. Its network structure and innovation points are shown in Figure 3:

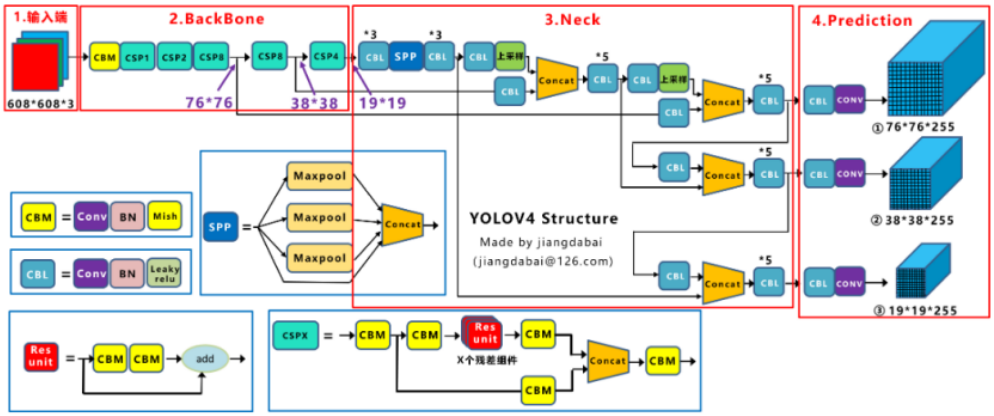

Figure 3 : yolov4 network structure diagram

The innovation lies in the improvement of the input terminal, mainly including mosaic data enhancement, cmbn,sat self confrontation training; Backbone network combines various new methods, including CSPDarknet53 、 Mish activation function and Dropblock;The target detection network often inserts some layers between the BackBone and the last output layer, such as the SPP module and FPN+PAN structure in yolov4;The anchor frame mechanism of prediction output layer is the same as that of yolov3. The main improvement is the loss function CIOU_Loss, and nms filtered by the prediction box becomes Diou_NMS.

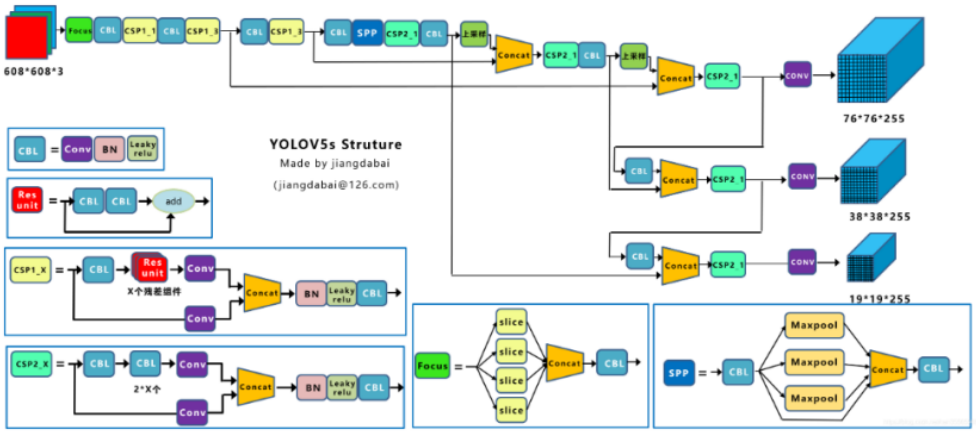

Figure 4 : yolov5s network structure

2.3.5 yolov5

There are four versions of the target detection network given by yolov5, namely yolov5s, yolov5m, yolov5l and yolov5x. Yolov5s is the network with the smallest depth and the smallest width of the feature map in the yolov5 series. The latter three types are deepened and widened on this basis. The AP accuracy is also improved, but the speed consumption is also increasing.

ISSN: 2664-9640

In this paper, yolov5s is used for image recognition, and its network structure is shown in Figure 4.

The above figure is the network structure diagram of yolov5s. It can be seen that it is still divided into four parts: input, backbone, neck and prediction. Compared with Yolo V4, Yolo V5 is slightly weaker in performance, but much stronger in flexibility and speed than Yolo v4. It has a strong advantage in the rapid deployment of models, and its image reasoning speed is up to 0007 s, that is, it can process 140 frames per second, which meets the real-time detection requirements of video images, and has a smaller structure.

3. Experiment

3.1. Preparation before experiment

In this experiment, the Xuefu Avenue section is taken as the research object, the vehicles running in this section are recorded (as shown in Figure 5), and then the vehicles in the image are detected, tracked and counted based on the traditional MATLAB algorithm and yolov5 detection algorithm. The first is the traffic image video collection. Appropriate research samples are selected according to the traffic road images captured by the video. These data are used for subsequent high-quality image selection, training models, etc. Then the yolov5 detection algorithm is trained with a large number of traffic flow samples to reduce the recognition error in the later stage. Finally, by comparing and analyzing the experimental results of MATLAB and yolov5, the advantages of yolov5 detection algorithm are summarized.

3.2. Traditional vehicle identification and detection

The traditional vehicle detection is carried out by using MATLAB background difference method: the binary segmentation of the image is realized by setting the threshold value, and the combination of gray mathematical morphology and binary mathematical morphology algorithm is applied to vehicle target detection, so as to obtain vehicle targets with different gray ranges. The results of moving target detection using the background difference method are shown in the figure below.

Figure (a) current frame

figure (b) background image

Figure (c) gray processing

figure (d) difference result

The above image a is the current frame image in the video sequence, and image B is the background image obtained by the time average method, but the background effect is not ideal. Subtracting the current frame image a from the background image B, and then binarizing and preprocessing the difference image such as image C, image D can be obtained.

The final traffic flow statistical results are shown in Figure 5, Figure 6 and table 1:

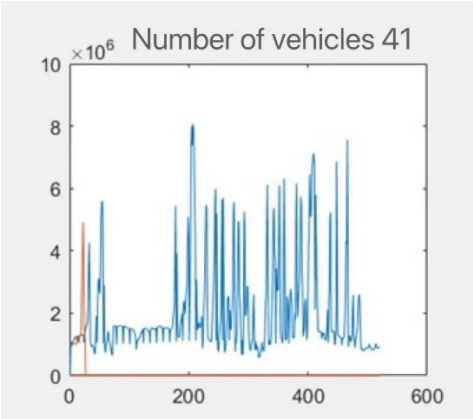

Figure 5:statistical results (1)

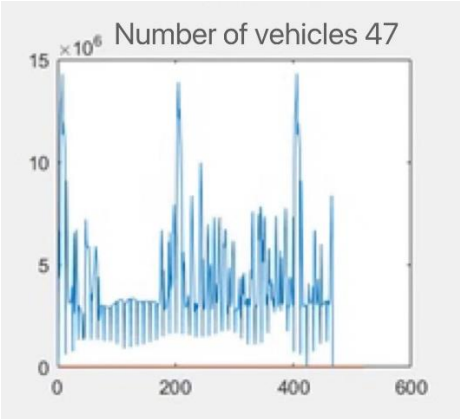

Figure 6:statistical results (2)

Table 1: Statistics of vehicle types by background difference method

	Car	Large truck	Non motor vehicle	total	Accuracy
Background difference method	34	4	3	41	83.6%
	37	4	6	47	95.9%
actual	40	5	4	49	

3.3. Detection results of vehicles and pedestrians based on yolov5 network model

The first step is to build a python running environment, read in the video image, and identify the image with the yolov5 detection algorithm. Read the video code and the original video as shown in Figure 7 and figure 8.

Figure 7: reading video code figure 8 : original video

The second step is to detect the target. The code and test results are shown in Figure 9 and figure 10.

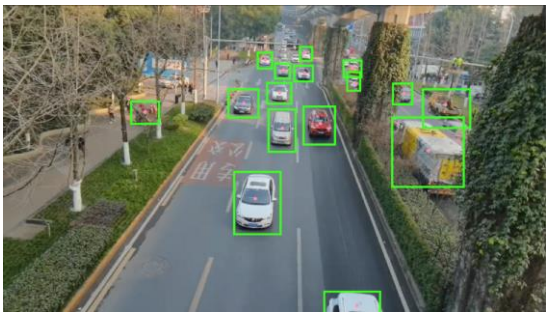

Figure 9 :detection object code Figure 10 :detection object running result

ISSN: 2664-9640

The third step is to track the target. Detect the existence of the target and mark the target;Compare the distance between the center points of the detection targets of the two frames. If it is less than the specified value, it is considered as the same target, and the mark remains unchanged ;When the target disappears on the screen, delete the mark of the target ;When a new target appears in a certain frame of image, the center point coordinates of the object determined to be the target in the previous frame are updated, and then the remaining detected targets (i.e. new targets) are marked .The running results of code and target tracking are shown in Figure 11 and Figure 12.


```
    else:
        track_objects_new = track_objects.copy()

        for object_id, pt2 in track_objects_new.items():

            object_exist = False

            for pt1 in center_points_current:

                distance = math.hypot(pt2[0]-pt1[0], pt2[1]-pt1[1])

                if distance < 20:

                    track_objects[object_id] = pt1

                    object_exist = True
```

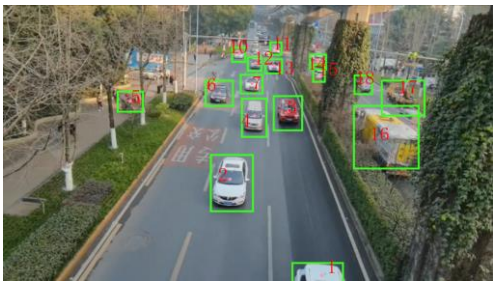

Figure 11: tracking target code Figure 12: tracking target running results

Table 2: Statistics of vehicle types of yolov5 detection algorithm

	Car	motorcycle	big truck	Non-Motor Vehicle	the sum	Accuracy
Yolov5 detection algorithm	48	11	8	6	73	97.3％
actual	48	12	8	7	75	

3.4. Result analysis

The background difference method based on MATLAB is used for traffic flow statistics. Its advantage is that it can detect the objects that stop moving in the video sequence, but the accuracy of the above algorithm is not high, the implementation process is complex, and if the background cannot be updated in real time, the detection results will be affected. Through the comparison of experimental results, it can be found that the probability and accuracy of all target objects are improved.

4. Conclusion

Taking the traffic flow of traffic sections as the research object, this paper realizes the recognition and detection of traffic targets through an image-based idea .Compared with traditional traffic target detection methods and image-based traffic target detection methods, a better method applied to traffic target detection task based on yolov5 algorithm is found, which can meet the needs of traffic target detection to a certain extent, and provide reliable theoretical basis and experimental data for the research of vehicle driving ,This provides a good basic research for traffic target detection in the future.

The next step is to consider how to balance the detection accuracy of different targets, whether collecting more data or virtual some target images through image preprocessing, which will be a very important factor to improve the detection accuracy .Secondly, consider generalizing the model so that it can be applied to more fields .Finally, if we can improve the network model, reduce the amount of calculation of network training and improve the speed of model training, it will be a great contribution to the research and experiment of target detection task.

References

[1] Liu L, Ouyang W, Wang X, et al. Deep learning for ge-neric object detection: A survey[J]. International Journal of Computer Vision, 2020, 128: 261-318.

[2] Zhang Ke, Yang cankun, Zhou Chunping, et alOverview of moving target detection algorithms in UAV video image [J]Liquid crystal and display, 2019, 34 (1): 98-109

[3] Zhang Canlong, Su Jiancai, Li Zhixin, et alInfrared and visible target tracking based on AdaBoost confidence map [J]Journal of Guangxi Normal University (NATURAL SCIENCE EDITION), 2018, 36 (4): 42-50

[4] Zhang Canlong, Li yanru, Li Zhixin, et alKernel based tracking and correlation algorithm [J]Journal of Guangxi Normal University (NATURAL SCIENCE EDITION), 2020, 38 (5): 12-23

[5] Feng Xia, Du Jiahao, Duan Yinong, et alReview on pedestrian re recognition based on deep learning []Computer application research, 2020, 37 (11): 3220-32263240

[6] Yin Zirui, Zhang Sofi, Zhang Lei, et alDesign of two branch efficientnet network suitable for pedestrian re recognition [J]Signal processing, 2020, 36 (9): 1481-1488

[7] Zhao Q, Sheng T, Wang Y, et al. M2Det: A single-shot object detector based on multi-level feature pyramid network[C]//Proceedings of the AAAI Conference on Artificial Intelligence, 2019:9259-9266.

[8] WOJKE N, BEWLEY A, PAULUS D. Simple online and realtime tracking with a deep association metric[C]// 2017 IEEE International Conference on Image Processing (ICIP). Piscataway, NJ: IEEE Press, 2017: 3645-3649.

[9] ZHOU Z W, XING J L, ZHANG M D, et al. Online multi-target tracking with tensor-based high-order graphmatching[C]// 2018 24th International Conference on Patterm Recognition(ICPR). Piscataway, NJ: IEEE Press, 2018:1809-1814.

[10] WANG Z D, ZHENG L, LIU YX, et al. Towards real-time multi-objecettracking[EB/OL]. (2020-07-14)[2021-08-13].

[11] REDMO J, FARHADI A. YOLOv3: An incremental improvement[EB/OL].(2018-04-08)[2021-08-13].

[12] ZHANG Y F, WANG C Y, WANG X G, et al. FairMOT:on the faimess of detection and re-identification in multiple objecttracking[J]. International Journal of Computer Vision, 2021, 129(11): 3069-3087..

[13] YU F, WANG D Q, SHELHAMER E, et al. Deep layer aggregation[C]// 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos, CA:IEEE Computer Society, 2018: 2403-2412.

[14] ZHOU X Y, WANG D Q, KRAHENBUHL P. Objects as points[EB/OL].(2019-04-25)[2021-08-13].

[15] Yan Kang, Zeng Fengcai, he Ning, et alJDE multi-target tracking method with attention mechanism [J / OL]Computer engineering and application, 2021 [2021-08-13]

[16] Xue Juntao, Ma Ruohan, Hu chaofangDeep learning algorithm for multi-target tracking based on mobilenet [J]Control and decision making, 2021, 36 (8): 1991-1996

[17] SANDLER M, HOWARD A, ZHU M L, et al. MobileNetV2: inverted residuals and linear bottlenecks[C]// 2018IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos, CA: IEEE Computer Society, 2018:4510-4520.

[18] Ma Yongjie, Ma Yunting, Cheng Shisheng, et alRoad vehicle detection method based on improved Yolo V3 model and deep sort algorithm [J]Journal of transportation engineering, 2021,21 (2): 222-231

[19] Liu Yifan. Research on the use of YOLOv5 object detec-tion algorithm in mask wearing recognition[J]. World Scientific Research Journal, 2020, 6(11).

[20] Sivaraman S, Trivedi M M. Looking at vehicles on the road: A survey of vision-based vehicle detection, tracking, and behavior analysis[J]. IEEE Transactions on Intelligent Transportation Systems, 2013, 14(4): 1773-1795.

[21] HeK，ZhangX，RenS，etal.Deep1 residual learning for image recognition[C]/Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.

[22] Cheng G, Han J. A survey on object detection in optical remote sensing images[J].ISPRS Journal of Photogrammetry and Remote Sensing, 2016, 117: 11-28

[23] P G Michalopoulos. Vehicle Detection Video through Image Processing: The Autoscope System[J]. IEEE Trans on Vehicular Technology, 1991, 40(1)，pp: 21-29.

[24] Murugas T, Peplow R，Tapamo J R. Video object extraction[J]. Proceeding of the 3rd International Symposium on Image and Signal Processing and Analysis, 2003， 9(2)， pp: .599-604.

[25] Stephen J.McKenna, Sumer Jabri, Zoran Duric et al. Tracking groups of people[J].Computer Vision and Image Understanding, 2000， 80(1), pp:42-56.

[26] Ismail H, David H, Larry S. A Fast Background Scene Modeling and Maintenance for Outdoor Surveillance[C]. Proceedings. 15th International Conference on Patterm Recognition, 2000，vol. 4, pp:179-183.

[27] Redmon J, Farhadi A. YOLO9000: better, faster, stronger{J]. arXiv preprint, 2017.

[28] Liu W，Anguelov D, Erhan D， et al.Ssd: Single shot multibox detector[C]/European conference on computer vision. Springer, Cham, 2016: 21-37.

[29] Liu N, Han J. Dhsnet: Deep hierarchical saliency network for salient object detection[C]/Computer Vision and Pattern Recognition (CVPR), 2016 IEEE Conference on. IEEE, 2016: 678-686.

[30] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[C]/Advances in neural information processing systems. 2012: 1097-1105.