An improved small target detection method based on Yolo V3

Zhang Gongguo[1]
School of Communication and Information Engineering
Chongqing University of Posts & Telecommunications
Chongqing, China

Wei Junhao[*2]
School of Communication and Information Engineering
Chongqing University of Posts & Telecommunications
Chongqing, China
1248565984@qq.com

***Abstract*—Aiming at the efficiency and accuracy of small target detection in current traffic flow, this paper proposes an improved Yolo V3 method and applies it to small target detection. The method is to first optimize the network structure of Yolo V3, and add a new small target-friendly 4-fold down-sampling residual between the second residual block and the third residual block of Darknet-53 Block, improve the detection accuracy of small targets; perform 2-fold up sampling on the 8-fold down-sampling feature map output by the original network, and perform the 2-fold up-sampling feature map with the feature map output by the newly added third residual block Splicing, build a feature fusion target detection layer whose output is 4 times down sampling. The improved Yolo V3 algorithm is compared with the unimproved algorithm, and it is concluded that the improved algorithm can significantly improve the recall rate of small target detection and the average detection accuracy.**

Keywords- traffic flow; small target detection; Yolo V3;

I. INTRODUCTION

With the continuous expansion of urbanization, traffic problems have gradually become an inevitable problem. Traffic congestion is a very important problem, which restricts people's travel. Traffic flow detection is an important part to solve the problem of traffic congestion. At present, target detection has been widely used in military and civil fields. In these target detection, small target detection as an important target detection technology, it has become a research hotspot and focus. Long range targets usually have the characteristics of small targets. Compared with large targets, small targets have the disadvantages of less pixels and less obvious features, so there will be a series of problems in the detection of low detection rate, so small target detection in target detection is still a very worthy research topic.

With the development of artificial intelligence in recent years, more and more researchers begin to pay attention to vehicle detection algorithms based on deep learning. Among them, a variety of target detection algorithms based on deep learning [1-2] and target tracking algorithms [3] are proposed. Compared with the old methods, the method based on deep learning can learn more target features. At present, target detection algorithms based on deep learning and widely used can be divided into two categories: 1) two-step target detection algorithms, such as Fast R-CNN (regional conventional neural network) [4], Faster R-CNN [5], Mask R-CNN [6], etc. These algorithms divide the target detection into two stages, that is, first use the regional candidate network (RPN) to extract the candidate target information, and then use the detection network to complete the prediction and recognition of the location and category of the candidate target; 2) Single step target detection algorithms, such as SSD (single shot Multibox detector) [7], Yolo (you only look once) [8], Yolo 9000 [9], Yolo V3 [10]. this kind of algorithm does not need to use RPN, and directly generates the location and category information of the target through the network. It is an end-to-end target detection algorithm. Therefore, the single-step target detection algorithm has the advantage of faster detection speed.

In order to improve the accuracy and recall rate of small target detection in traffic flow, this paper chooses the structure of single-step target detection algorithm (Yolo V3) with faster detection speed to optimize. This method is first optimized on the network structure of Yolo V3, that is, a new small target friendly 4-fold down-sampling residual is added between the second residual block and the third residual block of Darknet-53 Block, improve the detection accuracy of small targets; perform 2-fold up sampling on the 8-fold down-sampling feature map output by the original network, and stitch the 2-fold up-sampled feature map with the feature map output by the third residual block Establish a feature fusion target detection layer whose output is 4 times down sampling. Finally, the improved algorithm and the original algorithm are used for comparison experiments. The results show that the improved algorithm has a significant improvement in the accuracy and recall rate of small target detection.

II. IMPROVED YOLO V3 NETWORK STRUCTURE DETECTION MODEL

The direction of improvement in this paper is the detection of small targets, so cluster analysis and network structure improvements are made for small target detection in the data set.

A. Target box of data set based on cluster analysis

Because the idea of anchor boxes in Faster R-CNN is used by Yolo V3, the initial selection box of anchor boxes is a set of fixed size boxes, and its selection directly affects the accuracy and efficiency of target detection. Yolo V3 uses K-means clustering algorithm to cluster the width and height of the

978-1-6654-1869-0/21/$31.00 ©2021 IEEE
DOI 10.1109/ECIE52353.2021.00054

internal target frame of coco data set, and takes the average overlap degree as the required measure for the target clustering analysis. Finally, the clustering analysis of VEDAI dataset is carried out, and the AVG IOU objective function is obtained. The relationship between K and AVG IOU objective function can be obtained by selecting K value to analyze the internal samples of data set. If the value of K increases, the objective function will change more smoothly, and the inflection point is the most ideal number of anchor boxes. When k value is greater than 3, the curve tends to be stable, so choosing the number of anchor boxes as 3 can not only realize the convergence of loss function as soon as possible, but also reduce the error caused by candidate box to the point of elimination.

B. *Improved Yolo V3 network structure*

The network structure of Darknet-53 is the network structure adopted by Yolo V3, including 53 convolution layers. The input image is down sampled five times by Yolo V3, and the target is predicted in the last three down samplings. The last three down sampling include the feature maps of three scale target detection. The small feature map provides deep semantic information, while the large feature map provides the location information of the target. The small feature map is fused with the large feature map after up sampling, so the model can detect both large and small targets.

Darknet-53 uses the idea of residual net - work (RN) for reference, which is composed of five residual blocks (RB). Each residual block is composed of multiple residual units (RU), in which different sizes of targets correspond to different residual modules. However, Yolo V3 obtains the feature map through 8 times down sampling output, and detects small targets. The feature map is the target detection layer of 8 times down sampling, which can not achieve ideal results in small target location information detection. Therefore, in order to give full play to the advantages of Yolo V3 network structure detection, obtain a large number of small target feature information, and improve the detection efficiency and accuracy, we can improve the small target detection on the basis of the original network structure, add a residual module in the residual module of 2 and 3 of darknet53, and the added residual module is 4 times of down sampling target detection layer. Since the 4x down sampling feature map contains a lot of small target location information, the 4x up sampling feature map output by Yolo V3 can be realized. The obtained feature map is connected with the 4x down sampling feature map output by the second residual block of Darknet-53, and the 4x down sampling feature fusion target detection layer is constructed, which is applied in small target detection. The improved Yolo V3 network structure is shown in Figure 1.

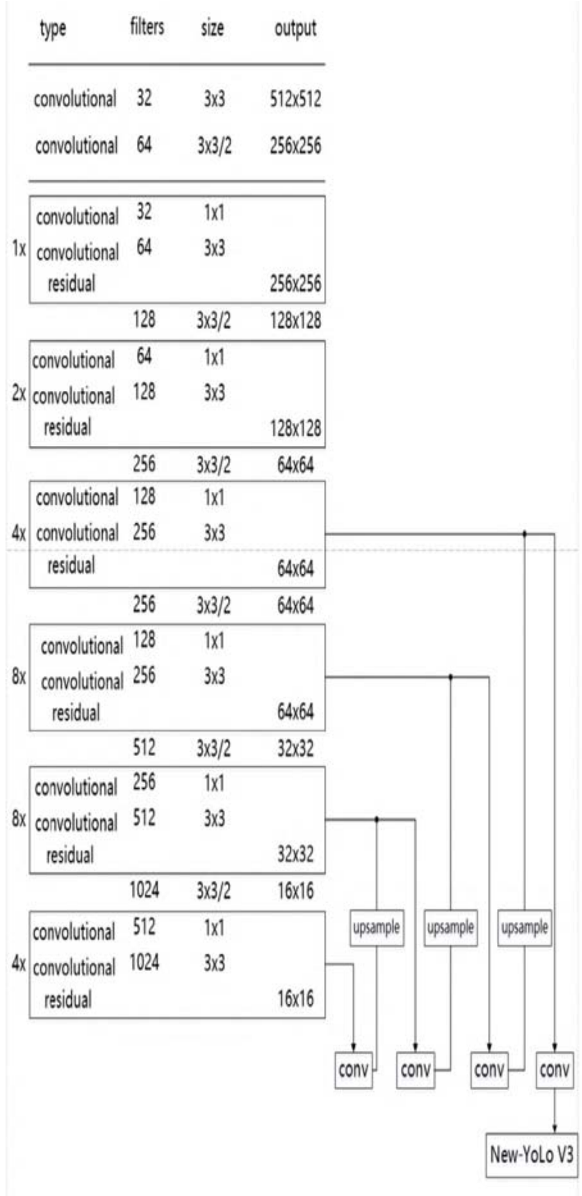

Figure 1. Improved Yolo V3 network structure

III. Experimental Results and Analysis

In the current detection field, Yolo V3 is one of the representative algorithms. It has the characteristics of all kinds of large, medium and small targets, and has good detection performance in the detection of small targets. Therefore, the improved Yolo V3 and the original Yolo V3 detection algorithm are compared.

Through the use of VEDAI datasets, the implementation of Yolo V3 algorithm is compared with the improved Yolo V3 algorithm. The original satellite image is segmented into 1024 × 1024 images, including vehicle and background. In addition, there are visible image and infrared image in the data set. The image distance in the data set is the same as that on the ground, which is the application scene of investigation and surveillance. The data set also includes 1024 × 1024 images, and the resolution of 512 × 512 images can be obtained by down sampling, which can get the target image with low resolution. Each image in the VEDAI dataset involves 5.5 vehicles, and the target only accounts for 0.7% of the total image pixels, which belongs to the small target detection dataset. There are nine types of goals, as shown in Table 1.

TABLE I. Number of Targets in the Dataset

Class name	Boat	Camping	Car	Others	Pickup	Tractors	Truck	Vans	Airplane
Total number	170	390	1340	200	950	190	300	100	47

In this paper, two methods are used to verify the performance of the improved algorithm for small target detection: Method 1, in the image data set with a resolution of 512 × 512, the three smallest categories in the data set (car, pickup, truck) are divided into one category, and 80% of the data in the data set are randomly selected for training, and the remaining data are used as test data; method 2, in the image data set with a resolution of 512 × 512, the three smallest categories in the data set (car, pickup, truck) are divided into one category In the image data set, nine kinds of targets in the data set are detected, and the training of Yolo V3 and improved Yolo V3 are carried out respectively. By rotating the image and improving the contrast, the internal image of the dataset is enhanced and expanded. The improved Yolo V3 draws the loss value collection curve graph, target box and AVG IOU objective function graph during training. After repeated iterations, all parameters tend to be stable, the final loss value decreases, and AVG IOU also tends to be stable. According to the comparison between method 1 and method 2, the performance comparison between the original and improved Yolo V3 algorithm is shown in Table 2. The performance of various algorithms is calculated by the following formula.

TABLE 2. Performance comparison table of original algorithm and improved Yolo V3

Comparison Algorithm	**Algorithm Performance**		
	Recall	***Precision***	***mAP***
YOLO V3	84.5%	86.1%	55.81%
New-YOLO V3	87.5%	91.3%	62.45%

$$Recall = \frac{TP}{TP+FN} \quad (1)$$

$$Precision = \frac{TP}{TP+Fp} \quad (2)$$

$$AP = \frac{\sum Precision}{N} \quad (3)$$

$$mAP = \frac{\sum AP}{NC} \quad (4)$$

In the above formula, recall is the recall rate, which indicates the proportion of the real target identified in the algorithm return result to the total target; precision is the accuracy rate, which indicates the proportion of the real target in the algorithm return result; TP represents that the detection is a positive sample, which is actually a positive sample; FN represents that the detection is a negative sample, which is actually a positive sample; FP represents that the detection is a positive sample Samples are actually negative samples; AP represents average precision; mAP represents multi class average precision.

According to table 2, it can be seen that compared with the original algorithm, the improved algorithm has a significant improvement in the detection accuracy, recall rate and multi category average precision, in which the accuracy is improved by 3%, the recall rate is improved by 5.2%, and the multi category average precision is improved by 6.64%. The comparison results of the above algorithms are drawn as the accuracy call rate curve (PR curve) according to the correlation function. Draw the image as shown in Figure 2. We can visually compare the improved algorithm with the original algorithm. Obviously, from the curve, we can see that the overall detection performance of new Yolo algorithm is better than the previous Yolo v3 algorithm.

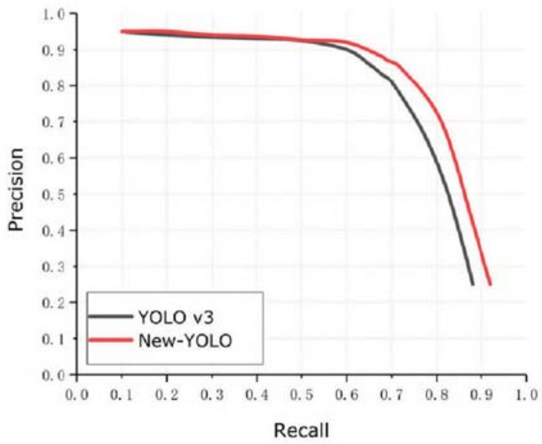

Figure 2. PR curve

IV. CONCLUSIONS

In this paper, an improved Yolo V3 algorithm is proposed and applied to small target detection. It mainly optimizes and improves the data set and network structure. Firstly, all the samples of the data set are clustered to find out the corresponding cluster center, and then the network structure is improved. In the second and third residual module of the original network, a 4-fold down sampling residual module is added to detect small targets, and then it is fused with the original output module. The final experimental results show that the improved algorithm has obvious improvement in accuracy, recall and average accuracy compared with the original algorithm.

REFERENCES

[1] Hua X, Wang X Q, Wang D, et al. Multi-objective detection of traffic field based on improved SSD[J]. Acta Optica Sinica, 2018, 38 (12): 1215003.

[2] Wang W X, Fu Y T, Dong F, et al. Infrared ship target detection method based on deep convolution neural network [J]. Acta Optica Sinica, 2018, 38 (12): 0712006.

[3] Luo H B, Xu L Y, Hui B, et al. Status and prospect of target tracking based on deep learning[J]. Infrared and Laser Engineering, 2017, 46 (5): 052002.

[4] Girshick R. fast R-CNN [C] // 2015 IEEE International Conference on computer vision (ICCV), December 7-13, 2015, Santiago, Chile. New York: IEEE, 2015:1440-1448.

[5] Ren S Q, He K M, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks [J]. IEEE Transactions on Pattern Analysis and Machine intelligence, 2017: 39 (6): 1137-1149.

[6] He K M, Gkioxari G, Dollar P, et al. Mask R-CNN [C] // 2017 IEEE International Conference on Computer Vision (ICCV), October 22-29, 2017, Venice, Italy. New York: IEEE, 2017:2980-2988.

[7] Liu W, Anguelov D, Erhan D. SSD: single shot multibox detector [M] / / Leibe B, Matas J, Seben, et al. Lecture notes in computer science. Cham: Springer, 20169905:21-37.

[8] Redmon J, Divvala S, Girshick R, et al. You only look once: unified, real-time object detection [C] // 2016 IEEE Conference on computer Vision and Pattern Recognition (CVPR), june 27-30, 2016, Las Vegas, NV, USA. New York: IEEE, 2016:779-788.

[9] Redmon J, Farhadi A.YOLO9000: better, faster, stronger [C] //2007 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), july 21-26, 2017, Honolulu, HI, USA. New York: IEEE, 2017:6517-6525.

[10] Redmon J, Farhadi A. YOLOv3: an incretental improvement [EB / OL]. (2018-04-08) [2018-12-25]. https: / / arxiv.org/abs/1804.02767.