An Improved Small Target Detection Method Based on YOLOv4*

Xia Miao[1], Xiaoyan Chen*[1], Keying Ren[1], Zichen Wang[1], Xiaoning Yan[2], Yue Sun[2]

[1]Tianjin University of Science and Technology, 300022, China;

[2] Shenzhen Softsz Co. Ltd., China
E-mail: miaoxia@mail.tust.edu.cn
www.tust.edu.cn

Abstract

In order to improve the efficiency and accuracy of small target detection in current traffic flow, this research proposes an improved YOLOv4 framework and applies it to small target detection task. A new small target-friendly 4-fold down-sampling residual is added between the second and third residual blocks of CSP Darknet-53 block to improve the detection accuracy of small target. The novel YOLOv4 model is optimized by above strategy. Compared with the original network, the modified framework can significantly improve the recall rate and average detection accuracy of small target.

Keywords: YOLOv4, small target detection, traffic flow

1. Introduction

With the continuous development of urbanization, traffic problems have gradually become an inevitable problem. Traffic congestion is a very important problem; it restricts people's travel. Traffic flow detection is an important part of solving the problem of traffic congestion. At present, target detection has been widely used in military and civilian fields. Among these target detections, small target detection, as an important target detection technology, has become a hotspot and focus of research. Long-range targets are usually small targets. Compared with large targets, small targets have the shortcomings of fewer pixels and unobvious features. Therefore, a series of low detection rate problems will occur in detection. Therefore, small target detection is still a topic worthy of research in target detection.

With the development of artificial intelligence in recent years, more and more researchers have begun to pay attention to vehicle detection algorithms based on deep learning. Among them, a variety of target detection and target tracking algorithms based on deep learning are proposed. Compared with traditional methods, methods based on deep learning can learn more target features. Currently widely used target detection algorithms based on deep learning can be divided into two categories: Two-step target detection algorithms, such as Fast R-CNN[1] (Regional Conventional Neural Network), R-CNN[2], etc. These algorithms divide target detection into two stages, that is, first use the regional candidate network (RPN) to extract candidate target information, and then use the detection network to complete the prediction and identification of the candidate target location and category; Single-step target detection algorithms, such as SSD (Single Shot Multi-box

© The 2022 International Conference on Artificial Life and Robotics (ICAROB2022), January 20 to 23, 2022

Detector), YOLO[3] (You Only Look Once), YOLO 9000[4], YOLO v3[5], YOLOv4[6]. This algorithm does not need to use RPN, and directly generates target location and category information 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.

This research proposes an improved YOLOv4 framework and applies it to small target detection task. A new small target-friendly 4-fold down-sampling residual is added between the second and third residual blocks of CSP Darknet-53 block to improve the detection accuracy of small target. The novel YOLOv4 model is optimized by above strategy. Compared with the original network, the modified framework can significantly improve the recall rate and average detection accuracy of small target.

2. Improved YOLOv4 Network Structure

2.1. *YOLOv4 algorithm introduction*

YOLO is a single-step detection algorithm based on regression methods. This article uses YOLOv4 feature extraction network is CSP(Cross Stage Partial)Darknet53 + PANet-SSP structure. The backbone feature extraction network CSPDarknet53 strengthens the feature extraction networks SPPNet[7] and PANet, and finally uses the YOLO header to convert the extracted features into prediction results. CSPDarknet53 consists of a convolution block and five Resblock_body, which contains a convolution, normalization and Mish activation function.

CSPDarknet reduces the amount of data transmission and calculation in the network, and reduces the amount of calculation and memory consumption required by the CNN network without losing accuracy and light weight. The input feature map is divided into two paths in the channel dimension, one is passed directly backwards, the other is passed backwards through multiple residual blocks, and finally merged with the CSP end. This cross-stage splitting and merging effectively reduces the possibility of gradient replication, increases the diversity of gradients, and reduces the amount of data transmission and calculation in the network. In YOLOv4, each CPSX contains 3+2*X convolutional layers, so the backbone of the entire backbone network contains 72 convolutional layers. In this paper, a new 4 times down-sampling residual block is added between the second and third residual blocks of CSPDarknet-53 block to improve the detection accuracy of small targets.

2.2. *Data set anchor box based on cluster analysis*

The anchor frame mechanism of the output layer of YOLOv4 is the same as that of YOLO v3. The main improvement is the loss function CIOU_ Loss during training (Complete Intersection Over Union_ Loss), which uses the anchor frame idea in Fast R-CNN, and the initial choice of anchor frame The frame size is a set of fixed frames, the selection of which directly affects the accuracy and efficiency of target detection. YOLOv4 uses K-means clustering algorithm to cluster the width and height of the internal target frame of coco data set, and takes the average overlap degree as the required measure for the target clustering analysis.

2.3. *Improved YOLOv4 network*

CSPDarknet-53 is the backbone network structure adopted by YOLOv4, in which the original intention of the CSP structure is to reduce the amount of calculation and enhance the performance of the gradient. The characteristics of the CSP structure are Strengthen the learning ability of CNN; Reduce computing bottlenecks; Reduce memory consumption. Darknet-53 down-sampled the input image 5 times, and predicted the target in the last 3 down-sampling. The last three down-sampling includes the feature map of three-scale target detection. Small feature maps provide in-depth semantic information, and large feature maps provide target location information. The sampled small feature map and the large feature map are fused, so that the model can detect both large targets and small targets. Darknet-53 draws on the idea of residual network-work (RN), which consists of five residual blocks. Each residual block is composed of multiple residual units (R), in which targets of different sizes correspond to different residual modules. In order to give full play to the advantages of Yolov4 network structure detection, a large amount of small target feature information can be obtained to improve detection efficiency and Accuracy, we can improve the detection of small targets on the basis of the original network structure. Add a residual module to the residual module 2 and residual module 3 of CSPDarknet53. The added residual module is a down-sampled target detection layer 4 times. Since the 4x down-sampling feature map contains a lot of small target

position information, the 4x up-sampling feature map output by Yolov4 can be realized. The obtained feature map is mapped to the 4x down output of the second residual block of Darknet-53 Sampling feature mapping connection, building a 4x down-sampling feature fusion target detection layer, applied to small target detection. The structure diagram is shown in Fig. 1.

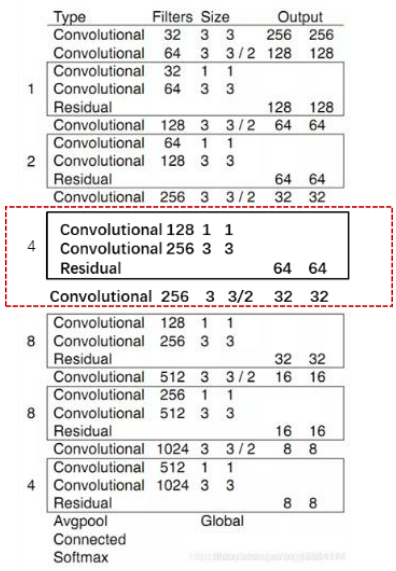

Fig. 1. Figure 1. New backbone network structure diagram

3. Experimental Results and Analysis

In the current detection field, YOLOv4 is one of the representative algorithms. It has the characteristics of detecting large, medium and small targets, and has good detection performance in small target detection. Therefore, the improved YOLOv4 detection algorithm is compared with the original YOLOv4 algorithm. By using the VEDAI data set, the implementations of the two algorithms are compared. Divide the original satellite image into 1024*1024 images, including vehicles and background. Each image in the VEDAI dataset involves 5.5 vehicles, and the target only accounts or 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 1. Number of Targets in the Dataset

Class name	Boat	Camping	Car	Others	Pick up	Tractors	Truck	Airplane
Total number	160	400	1430	200	850	200	350	53

This paper uses the following method to verify the small target detection ability of the improved algorithm: the image data is set to a resolution of 512*512, the smallest three categories of data sets (cars, trucks, and trucks) are divided into one category, and 80% of them are randomly selected The sample of the data set is used as the training set, and the remaining data set is used as the test set. Then use the YOLOv4 algorithm and the improved YOLOv4 algorithm for training and testing respectively. After repeated iterations, all parameters tend to stabilize, and the final loss value decreases. The performance comparison between the original YOLOv4 algorithm and the improved YOLOv4 algorithm is shown in Table 2.

Table 2. Performance comparison table of original algorithm and improved Yolov4.

	Algorithm Performance		
	Recall	Precision	mAP
YOLOv4	86.5%	89.3%	56.87%
NEW-YOLOv4	87.6%	92.1%	63.25%

The performance of various algorithms is calculated by the following formula.

$$\mathrm{Re}call = \frac{TP}{TP + FN} \cdot \tag{1}$$

$$\mathrm{Precision} = \frac{TP}{TP + FP} \tag{2}$$

$$mAP = \frac{\sum` AP}{NC} \tag{3}$$

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; mAP represents multi class average precision.

4. Conclusion

This paper proposes an improved YOLOv4 algorithm and applies it to small target detection. Mainly optimize and improve the network structure and data set. In the optimization of the network structure, in the second and third residual modules of the original network, a 4-fold down-sampling residual module is added to detect small

targets, and then merged with the original output module. In the data set, all samples are clustered and a cluster center is found. Through the above optimization and improvement, the final experimental results show that the improved algorithm has a significant improvement in accuracy, recall and average accuracy compared to the original algorithm.

References

1. Girshick, R. . "Fast R-CNN." arXiv e-prints (2015).
2. Girshick, R. , et al. "rich feature hierarchies for accurate object detection and semantic segmentation tech report (v5)." (2017).
3. Redmon, J. , et al. "You Only Look Once: Unified, Real-Time Object Detection." IEEE (2016).
4. Redmon, J. , and A. Farhadi . "YOLO9000: Better, Faster, Stronger." IEEE Conference on Computer Vision & Pattern Recognition IEEE, 2017:6517-6525.
5. Redmon, J. , and A. Farhadi . "YOLOv3: An Incremental Improvement." arXiv e-prints (2018).
6. Bochkovskiy, A. , C. Y. Wang , and H. Liao . "YOLOv4: Optimal Speed and Accuracy of Object Detection." (2020).
7. He, K., et al. "Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition." IEEE Transactions on Pattern Analysis & Machine Intelligence 37.9(2014):1904-16.

Authors Introduction

Ms. Xia Miao

She is currently studying in Tianjin University of Science and Technology, majoring in electronic information, and main research field is computer vision.

Prof. Xiaoyan Chen

She, professor of Tianjin University of Science and Technology, graduated from Tianjin University with PH.D (2009), worked as a Post-doctor at Tianjin University (2009.5-2015.5). She had been in RPI, USA with Dr. Johnathon from Sep.2009 to Feb.2010 and in Kent, UK with Yong Yan from Sep-Dec.2012. She has researched electrical impedance tomography technology in monitoring lung ventilation for many years.

Mr. ZiChen Wang

He was born in Tianjin, China in 1997. He received the B.E. degrees from Tianjin University of Science and Technology, Tianjin, China, in 2019. He is currently pursuing the M.S. degree in Information and Automatic College at Tianjin University of Science and Technology, Tianjin, China.

Mr. Yue Sun

He is a master's degree in computer application technology from Tianjin University of Science and Technology, is currently an algorithm engineer of Ansoft Technology, engaged in deep learning and computer vision algorithm research, mainly related to the research direction of target detection.

Mr. Xiaoning Yan

He graduated from South China University of Technology with a bachelor's degree in software engineering. 2013.8-2015.5 studied for a master's degree in computer science at the University of Texas at Dallas.