Vehicle and Pedestrian Detection Based on Improved YOLOv4-tiny Model

Lingzhe Ma1*, Yu Chen1 and Jilin Zhang1

1School of Electronic and Information Engineering, Changchun University of Science and Technology, Changchun 130022, China
*Corresponding author’s e-mail: 2019100464@mails.cust.edu.cn

Abstract. Target detection is the basic technology of automatic driving system. Deep learning has gradually become the mainstream target detection algorithm because of its powerful feature extraction ability and adaptive ability. How to ensure accuracy and speed is a great challenge in the field of target detection. In order to solve the problems of high miss detection rate of small target and difficult to realize embedded real-time detection in the process of complex environment detection by deep learning method, this paper adds two auxiliary remaining network blocks in the backbone network. So that the backbone network can extract the global and local features of the detected object, and carry out feature extraction based on the feature pyramid network Fusion, adding a scale to form a three scale prediction, to improve the problem of poor detection accuracy of yolov4-tiny network. The simulation results show that: Compared with yolov4-tiny, the accuracy of the improved network structure is improved by 3.3%, and the detection speed is 251 fps, which ensures the requirements of real-time detection. This algorithm has good detection effect in the case of lack of illumination and target occlusion, and its detection accuracy on the mixed data set is better than that of the contrast algorithm, which meets the real-time detection conditions and is suitable for deployment on the embedded system carried by the car.

1. Introduction

With the continuous upgrading of software and hardware platform and the continuous progress of environmental awareness technology, a large number of advanced intelligent auxiliary devices have been applied to automobiles, and the automatic driving system has become a major trend of modern traffic[1-3]. Intelligent assistant driving technology can automatically judge the road situation, remind the possible danger or driver's improper behavior in front of the road, and make a lot of contributions to reduce traffic accident casualties. Target detection is an important part to realize the environment perception of intelligent vehicle and advanced driving assistance system[4]. At present, the detection technology has the most outstanding effect on the large-scale target problem[5,6]. However, the robustness of target detection algorithm is facing challenges in the complex and changeable traffic environment, such as the rapid movement of vehicles, the change of light and the scale change of pedestrians and vehicles at different distances[7].
2. Improvement of yolov4-tiny algorithm

2.1. Introduction of yolov4 tiny model
Yolov4-tiny method is designed on the basis of yolov4 method, which is a simplified version of yolov4. Yolov4 has about 60 million parameters, while yolov4 tiny has only 6 million parameters, which greatly increases the feasibility of deploying object detection method in embedded system.

Because the yolov4-tiny target detection network is relatively simple, although it has high real-time performance and is suitable for the deployment of embedded edge computing platform, its feature extraction operation is simple, the detection scale is small, and the detection accuracy is low, especially in the small target detection, the miss detection rate is high.

2.2. Network model improvement
In this paper, spatial attention mechanism and channel attention mechanism are combined to assign high weight to the features of the targets that are easy to appear in the yolov4 tiny convolution feature map and low weight to the background, so that the output of the model pays more attention to the given targets and solves the problem of missing detection of small targets with local occlusion and natural background confusion.

![Figure 1. Auxiliary network block](image)

For convolutional neural networks, convolution layers with different depths correspond to different levels of feature information. Low level network has higher resolution and more detailed features, while high level network has lower resolution and more semantic features. When the target is occluded, there are only local features of the target. Because the shallow convolutional feature receptive field contains less background noise, it is usually more sensitive to the local part of the target and has better representation ability, so it is necessary to increase the shallow feature output layer of the network to enhance the recognition ability of small targets and occluded targets. Based on the idea of feature fusion, we improved the network structure of yolov4-tiny, added upsample2 to the original model, connected the cspblock layer and upsample2 in the channel dimension, and added a detection scale. The feature pyramid network was upgraded from the original 13 × 13 and 26 × 26 pixels to three detection scales of 13 × 13, 26 × 26 and 52 × 52 pixels, It is helpful to improve the target detection accuracy of the target detection network, and ensure the simplification of the model and the real-time detection.
3. Experiment

3.1. Test environment
The experimental environment of this paper is Ubuntu 16.04 bit operating system; the computer configuration is GPU (NVIDIA GTX 1080ti); CPU (inter core i7-7700hq); 16g memory. In order to make full use of GPU to accelerate network training, CUDA 10.1 and its supporting cudnn are installed in the system. The deep learning framework is pytorch.

3.2. Production of data set
To solve the problem, this paper uses the automatic driving data set Kitti and voc2007 pedestrian data set to mix, and finally gets a total of 5000 images of the mixed data set, in which the proportion of automobile and human is close to 1:1. In order to prevent the over fitting problem caused by the large sample gap in the data set, 70% of them are used for training, 20% for verification and 10% for testing.

3.3. Experimental results and analysis
Map and FPS are used to quantitatively evaluate the performance of different methods. Map is the average of the average accuracy of detecting all categories. FPS represents the number of images that can be successfully detected in one second.

Table 1. Comparison of detection ability of three models

| Algorithm         | Map(%) | FPS  |
|-------------------|--------|------|
| YOLOv4            | 65     | 41   |
| YOLOv4-tiny       | 38.3   | 268  |
| Improved yolov4 tiny | 41.4   | 251  |
For the test data set, this paper uses the model of yolov4 and yolov4 tiny to compare with the model in this paper, and the results are shown in Table 1. It can be seen from Table 1 that although the map of yolov4 algorithm is larger than that of other algorithms, their FPS is much smaller than that of other algorithms. Compared with yolov4 tiny model, the average accuracy of this model increased by 3.3%, but FPS decreased by 19%. After adding auxiliary network module and prediction frame, the FPS of the model is reduced, but it can still meet the requirements of real-time detection.

![Figure 3. Target detection results of YOLOv4-tiny](image)

![Figure 4. Improved target detection results of YOLOv4-tiny](image)

Figure 3 and figure 4 are the real scene detection results of our algorithm, and compared with the detection results of yolov4 tiny. Fig. 3 (a) - Fig. 3 (c) are the detection results of yolov4 tiny, and Fig. 4 (a) - Fig. 4 (c) are the detection results of the improved algorithm in the same case. In Figure 3 (a), the car far away from the left side is missed, and there is no vehicle missed in Figure 4 (a); in Figure 3 (c) and figure 4 (c), there is no missed detection, and pedestrians and vehicles are detected in the case of insufficient light. In Figure 3 (d), the car on the left and a car blocked by the fence in the distance are missed, while in Figure 4 (d), there is no vehicle missed. It can be seen that the improved algorithm has better detection effect on small targets such as vehicles in the distance, alleviates the problem of small target missed detection caused by occlusion and background interference, and still has good detection results in weak light environment, which shows that the improved algorithm has better detection effect. This method has good adaptability to complex real-time environment and can detect targets more accurately.

4. Conclusion

In order to effectively improve the accuracy of target detection on the premise of real-time, this paper proposes an improved algorithm based on yolov4 tiny. In the part of backbone network, a network assistant module combined with attention mechanism is added to extract more global features. In the part of feature fusion, a 52 × 52 scale prediction box is added. Finally, the improved algorithm is used to train the mixed data set, and a comparative test is carried out. The experimental results show that the
target detection accuracy of the improved yolov4 tiny algorithm is 41.4%, which is 3.1% higher than the original algorithm, and the FPS is 251 frame/s, which meets the requirements of real-time detection. However, the algorithm in this paper still has some defects, for the long-distance pedestrian target, the missed detection rate is high, how to improve the detection accuracy of small target is the future research direction.

References

[1] Wang, Y. Jia, K. (2020) Impolite Pedestrian Detection by Using Enhanced YOLOv3-Tiny[J]. Journal on Artificial Intelligence, 2(3) : 113-124

[2] Liu, J. Hou, S. Zhang, K. Zhang, R. (2019) Real-time vehicle detection and tracking based on enhanced Tiny-YOLOv3 algorithm [J]. Transactions of the Chinese Society of Agricultural Engineering, 35(8) : 118-125

[3] Gu, G. Xu, X. (2020) Real-time vehicle detection and information recognition technology based on YOLOv3 improved algorithm. Computer Engineering and Applications, 56(22) : 173-184

[4] Ren, J. Han, J. (2021) A New Multi-scale Pedestrian Detection Algorithm in Traffic Environment [J]. Journal of Electrical Engineering & Technology, 16(2) : 1151-1161

[5] Li, W. Zhang, Y. MO, J. (2020) Detection of Pedestrian and Agricultural Vehicles in Field Based on Improved YOLOv3 tiny. Transactions of the Chinese Society for Agricultural Machinery, 51(S1) : 1-8

[6] Xu X K, Ma Y, Qian X and Zhang Y. (2020) Scale-aware Efficient Det: real-time pedestrian detection algorithm for automated driving. Journal of Image and Graphics, 26(01) : 93-102

[7] Sen, Y. and Yong, Z. and Ming, Y. (2021) Real-time Target Detection Based on Jetson Nano and YOLOv3-Tiny [J]. International Core Journal of Engineering. 2021, 7(2)

[8] Fu, L. (2020) Fast and accurate detection of kiwifruit in orchard using improved YOLOv3-tiny model [J]. Precision Agriculture, 15(3) : 1-23

[9] XU, C. WANG, X. YANG, Y. (2019) Attention-YOLO: YOLO Detection Algorithm That Introduces Attention Mechanism [J]. Computer Engineering and Applications, 55(06) : 13-23