Research on Face Detection Algorithm Based on Improved YOLOv5

Zhen Mao[1]，Xiaoyan Chen[1*]，Xiaoning Yan[2], Yuwei Zhao[2]

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

[2]Shenzhen Softsz Co. Ltd., China
E-mail: 1037331281@qq.com
www.tust.edu.cn

Abstract

Face detection technology is one of the research hotspots in the field of deep learning in recent years. Aiming at the problems of slow detection speed and low accuracy of various target detection algorithms, this paper proposes an improved target detection algorithm based on YOLOv5. By introducing lightweight network, changing the depth and width of YOLOv5 network structure and reducing the number of model parameters, the network reasoning speed can be greatly accelerated. At the same time, the method uses Acon adaptive activation function to further improve the accuracy of face detection. Experimental results show that the improved algorithm has faster detection speed and higher detection accuracy than the traditional algorithms.

Keywords: Deep learning, Face Detection, Yolov5, Acon ctivation function

1. Introduction

With the rapid development of deep learning, face detection has become a hot research direction in the field of artificial intelligence. In recent years, with the continuous progress of target detection algorithm, the research on face detection is also effective, and even can exceed people's resolution level in some scenes. Face detection technology is widely used in video surveillance, identity recognition, and human-computer interaction and so on, which makes people's life more convenient and safe. However, with the continuous improvement of CPU processing speed and the increasing ability of graphics card image processing, people have a further demand for the detection accuracy and speed of face detection. How to achieve higher accuracy and speed detection based on the existing algorithms has become the research direction of this paper[1].

The research period of face detection algorithm can be divided into three stages. The first stage is the semi-mechanical recognition stage, the second stage is man-machine interactive recognition stage, and the third stage is face detection method based on deep learning. The function and purpose of face detection is to accurately find all faces in a digital picture and mark their positions and information. The traditional detection methods have low detection accuracy due to the lack of classification ability and feature extraction ability of the classifier. With the development and progress of convolutional neural network (CNN)[2], the detection method of deep learning based on neural network gradually replaces the traditional face detection method, and has made great improvement in accuracy, recall and location. In 2015, R. Joseph et al. proposed a new target detection algorithm Yolo. In this paper, the latest yolov5 algorithm is used to study face detection[3]. On the basis of the original model, the depth and width of the network structure are changed to reduce the parameters of the model. At the same time, ACON adaptive activation function is used to improve the speed and accuracy of detection.

2. Yolov5s Model Structure

In this paper, Yolov5 detection algorithm is used for face detection. When Yolov5 network is used for detection, there are deficiencies in detection speed and accuracy. A method to reduce the depth and width of the network is proposed, which reduces the amount of parameters and operations, improves the network speed, and adopts the newly proposed Acon activation function, It makes up for the decline of network performance after reducing network depth and width.

Yolov5 model contains four target detection versions. Among them, Yolov5s has the smallest network scale and the fastest speed, but at the same time, the value of AP, that is, the accuracy, is also the lowest. It is suitable for application scenarios that focus on large target detection and pursue high detection rate[4]. The other three models are based on the Yolov5s model to improve the ability of feature extraction and feature fusion by continuously deepening the depth of the network and broadening the width of the feature map. Yolov5s model consists of input, backbone, neck and prediction. The input adopts adaptive image scaling and mosaic data amplification methods, including random scaling, clipping and layout splicing, adaptive anchor box calculation to adapt to different data sets and adaptive image scaling to improve the reasoning speed. Backbone mainly includes slice structure (focus), convolution module (conv), bottleneck layer (C3) and spatial pyramid pooling (SPP)[5]. The focus structure is unique to Yolov5 model, which is different from Yolov3 and Yolov4. It is mainly used to slice images. The neck network part is composed of csp2_ Composed of X structure and FPN +PAN structure, the feature fusion ability of the network is strengthened and richer feature information is retained. Output layer: bounding box loss using GIOU_ Loss replaced IOU_ Loss, as a loss function, predicts targets of different sizes on characteristic maps of different sizes[6]. The structure diagram of the model is shown in Figure 1.

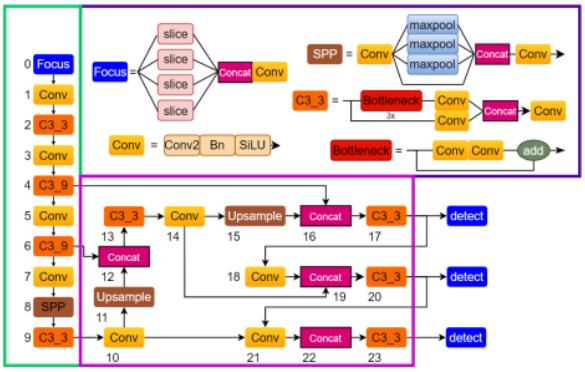

Fig. 1. The structure diagram of YOLOv5

3. Improved Yolov5 Model

Due to the task of face detection is relatively simple, single category, and has a sufficient number of data sets, so the use of native Yolov5s network still exists the phenomenon of network performance surplus. Therefore, the Yolov5 network was modified in this paper, and the depth and width of the network were reduced. As a result, the number of convolution kernels for sampling operations during the call of residual modules in the network and feature extraction were reduced, which greatly reduced the parameters and computation amount in the network operation. At the same time, Acon activation function is used to realize automatic adaptive automatic learning to improve network performance.

3.1. *Adjustment of network parameters*

In the Yolov5s network structure, GW = width_ Multiple is 0.5, while in the standard focus, C2 = 64, which is substituted into the code to calculate the result 32, that is, the number of convolution layers in the first convolution down sampling operation is 32, while in the second convolution down sampling operation, C2 = 128, GW = 0.5, the calculated result is 64, that is, the number of convolution cores in the second convolution down sampling operation is 64. The parameters for training with yolov5s model are: model summary: 283 layers, 7063542 parameters, 7063542 gradients, 16.4 gflops. We can see that there are 7063542 parameters, 283 convolution layers and 16.4g flops. After final adjustment, the depth of the network is 0.2 and the width is 0.3. The parameters of the final model are: model summary: 265 layers, 2621006 parameters, 2621006 gradients, 6.3 gflops.

3.2. *Acon activation function*

In the original Yolov5 model, the activation function in the backbone network is Silu type activation function, which is replaced by Acon activation function in this paper. Acon function learns the parameter switching between nonlinear (active) and linear (inactive) by introducing a switching factor. The activation function can significantly improve the depth model in the tasks of image classification, target detection and semantic segmentation.

4. Analysis of Experimental Results

4.1. *Experimental data set*

The data set in this paper is mainly the face and mask data set. The data set source is the public data set. After sorting, the data set photo is 11396, including 11042 face frames with masks and 17444 label frames with single face. It is divided into 9409 training sets (9483 face frames with masks and 15808 single face frames), 983 test sets (626 face frames with masks and 1229 single face frames), and 1004 verification sets (933 face frames with masks and 407 single face frames).

4.2. *Evaluating indicator*

In this paper, the accuracy P (precision), model size, recall, AP (average precision) and mAP (mean average precision) are used to evaluate the performance of the model for small target detection. Accuracy rate is used to measure the accuracy of model detection, that is, precision rate. Recall rate is used to evaluate the comprehensiveness of model detection, that is, recall rate. The formula used is as follows.

$$R = \frac{TP}{TP + FP} \times 100\% \tag{1}$$

$$P = \frac{TP}{TP + FN} \times 100\% \tag{2}$$

$$AP = \int_0^1 P(\mathrm{r})dr \tag{3}$$

$$mAP = \frac{\sum_{i=1}^{k} AP_i}{k} \tag{4}$$

In the formula, TP represents the number of positive samples predicted, FP represents the number of negative samples predicted, FN represents the number of negative samples predicted,AP represents the area covered under P(r) curve with recall as X-axis and Precision as Y-axis, which measures the recognition accuracy of a certain category. mAP is the average value of AP of each category. It measures the average quality of AP in all categories.

4.3. *Analysis of test results*

Firstly, based on yolov5s network, change the width and depth of the network, and compare the depth and width coefficient of the original network. The adjusted network is 0.2 in depth and 0.3 in width. The comparison of the three performance indicators is shown in Figure 2. Table 1 shows the training results of adjusting network parameters.

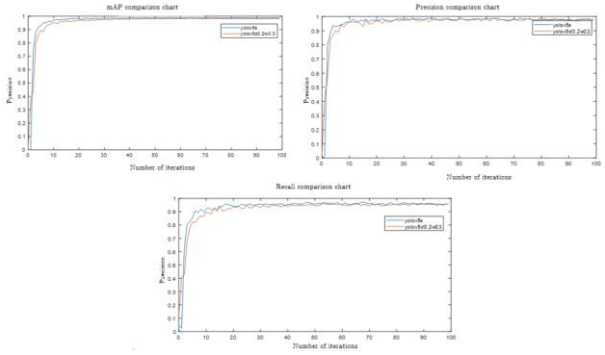

Fig. 2. Yolov5s and Yolov5_ d0. 2_ w0. 3 performance comparison chart

Table 1. Compare the performance of the model.

model	AP	Precision	Recall
Yolov5s	0.9880	0.9800	0.9574
Yolov5_d0.2_w0.3	0.9825	0.9720	0.9566

It can be seen that the AP value of the original network is 0.9880, while the AP value of the changed network is 0.9825, a decrease of 0.0055; The precision value of the original network is 0.9800, while the precision value of the changed network is 0.972, a decrease of 0.008; The recall value of the original network is 0.9574, while the recall value of the changed network is 0.9566, a decrease of 0.0008. While keeping the accuracy basically unchanged, our parameters are less. After changing the network, the parameters of the network are reduced from more than 7 million to more than 2 million, the amount of computation is changed from 16.4 g flops to 6.3 g flops, and the model reasoning speed is faster.

After the activation function is changed in this paper, the performance of the model is further improved. The comparison of the changed indicators is shown in Figure 3. Table 2 shows the training results using Acon activation function.

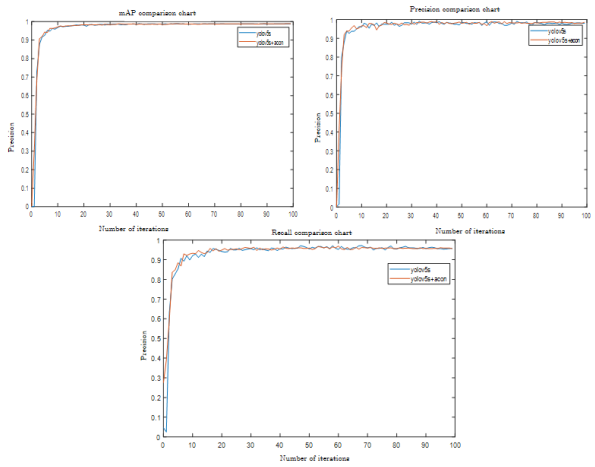

Fig. 3. Yolov5s and Yolov5s+Acon performance comparison chart

Table 2. Compare the performance of the model.

model	AP	Precision	Recall
Yolov5s	0.9880	0.980	0.957
Yolov5s+Acon	0.9883	0.985	0.956

According to the AP, precision and recall comparison data of yolov5s and yolov5s + Acon activation functions in the above chart, the map value of yolov5 + Acon is 0.9883, while the map value of the original yolov5s is 0.988. Because its ap value is very close to 1, it is very difficult to improve all, but Acon activation function finally increases the AP value by 0.0003. In the comparison diagram of precision, the precision of the original network is 0.98, while the precision after using Acon activation function is 0.985, an increase of 0.005. In the recall comparison chart, the original network is 0.957, yolov5s + Acon is 0.956, a decrease of 0.001. In general, adding Acon activation function can improve AP and precision to a certain extent. The performance of the whole network is improved.

The comparison between adjusting model parameters and changing Acon function and yolov5 model indicators is shown in Figure 4 below. Table 3 shows the training results of the comprehensive improved network

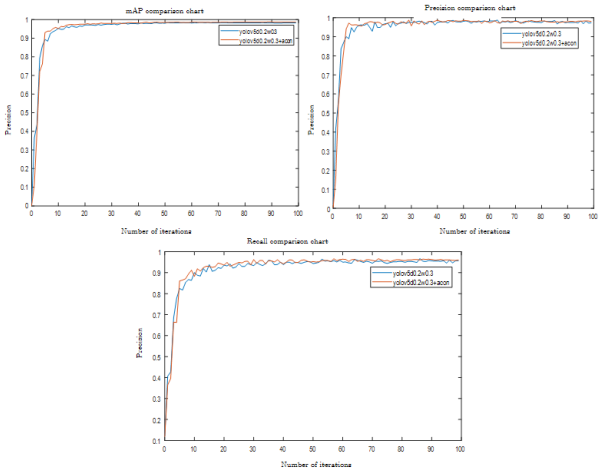

Fig. 4. Yolov5s and Yolov5_ d0. 2_ w0. 3+Acon performance comparison chart

Table 3. Compare the performance of the model.

model	AP	Precision	Recall
Yolov5s	0.988	0.9801	0.957
Yolov5_d0.2_w_0.3 +Acon	0.987	0.9802	0.96

It can be seen from Figure 4 and table 3 that the AP, precision and recall values of yolov5s network are 0.988, 0.9801 and 0.957 respectively, while the AP, precision and recall values of the network finally changed in this paper are 0.987, 0.9802 and 0.96 respectively. Although the final network differs from the original yolov5s network by 0.001 in AP value, the precision and recall are improved by 0.0001 and 0.003 respectively. In general, the performance of the two networks is comparable. However, because the depth and width of the network in this paper are smaller than that of yolov5s network, the amount of parameters and computation are small, especially in terms of computation, the comprehensive performance of the network in this paper is better, Reasoning is faster.

5. Discussion and Conclusion

This paper changes the depth and width of yolov5 network structure, significantly reduces the amount of parameters, and greatly speeds up the reasoning speed of the network. At the same time, the Acon activation function is used to further improve the performance of the network. The yolov5 network with depth of 0.2 and width of 0.3 is used with Acon activation function, so that

the network can not only maintain rapidity but also not lose accuracy.

6. References

1. Zou Z, Shi Z, Guo Y, et al. Object detection in 20 years: A survey. arXiv preprint arXiv:1905.05055, 2019.
2. R. Alex Krizhevsky, Ilya Sutskever, Geoffrey Hinton. ImageNet Classification with Deep Convolutional Neural Networks. Advances in neural information processing systems, 2012, 25: 1097-1105.
3. Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi. You Only Look Once: Unified, Real-Time Object Detection. IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV, USA, 2016.
4. J GAO M H,YANG C.Traffic target detection method based on improved convolutional neural network .Journal of Jilin University(Engineering Edition),2021,45(5):1-9(in Chinese).
5. HE K, ZHANG X, REN S, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition.IEEE transactions on pattern analysis and machine intelligence, IEEE, 2015, 37(9): 1904–1916.
6. LIU S, QI L, QIN H, et al. Path aggregation network for instance segmentation//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 8759–8768.

Authors Introduction

Mr. Zhen Mao

He obtained his bachelor's degree from Nanjing Institute of Technology in China in 2020. He is currently studying for a master's degree at Tianjin University of Science and Technology. The research direction is the edge computing field of deep learning, engaged in the research of model deployment.

Prof. Dr. Xiaoyan Chen

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.2012y.

Mr. Xiaoning Yan

He obtained a bachelor's degree in software engineering from South China University of technology in 2012 and a master's degree in computer science from the University of Texas Dallas in 2015 . Now he is the technical director of Shenzhen Ansoft Huishi Technology Co., Ltd

Dr. Hidekazu Suzuki

He Graduated from Tianjin University of science and technology, majoring in control science and engineering. Now he is an algorithm engineer of Ansoft technology. He is engaged in edge computing and artificial intelligence algorithm application research, mainly involving the research direction of embedded edge computing.