

PAPER • OPEN ACCESS

Face Detection Algorithm Based on Deep Residual Network

To cite this article: Shifan Tao *et al* 2021 *J. Phys.: Conf. Ser.* **1802** 032142

View the [article online](#) for updates and enhancements.

You may also like

- [TRL-Net: a tiny object recognition and localization method in the artemisinin manufacturing process](#)
Ling Guo, Ping Wan, Jie Li et al.
- [Research on Multi-Angle Face Detection Method Based on Improved YOLOV2 Algorithm](#)
Huanping Zhang, Hanhua Cao, Zhendan Liu et al.
- [Multi-View Faces Detection Using Viola-Jones Method](#)
Edy Winarno, Wiwien Hadikurniawati, Ahmad Ainun Nirwanto et al.



The Electrochemical Society
Advancing solid state & electrochemical science & technology

242nd ECS Meeting

Oct 9 – 13, 2022 • Atlanta, GA, US

Early hotel & registration pricing
ends September 12

Presenting more than 2,400
technical abstracts in 50 symposia

The meeting for industry & researchers in

BATTERIES
ENERGY TECHNOLOGY
SENSORS AND MORE!



Register now!



ECS Plenary Lecture featuring
M. Stanley Whittingham,
Binghamton University
Nobel Laureate –
2019 Nobel Prize in Chemistry



Face Detection Algorithm Based on Deep Residual Network

Shifan Tao, Yufeng Li, Yufeng Huang and Xiaoyu Lan*

School of Electronic and Information Engineering, Shenyang Aerospace University, Shenyang, China

*Corresponding author: xiaoyulan@sau.edu.cn

Abstract. With the rapid development of artificial intelligence, face detection technology is widely used in our daily lives, such as mobile payment, video conferencing and personal identification. However, face the scenario while the face been blocking or crowding, the face detection accuracy would be greatly reduced. Therefore, in this paper, a high-precision face detection algorithm based on deep residual network has been proposed to solve this issue. Firstly, adding neural framework branches based on the Resnet-50 to improve the detection accuracy. Then import the soft-NMS method to enhance the robustness of algorithm. Experimental test on the public data set FDDB, the results indicate that the accuracy of this algorithm can reach 94.2% with good robustness, both accuracy and speed are better than the previous algorithm.

Keywords: Face Detection, Soft Non-Maximum Suppression, Residual Network

1. Introduction

Artificial intelligence has developed very fast in recent years, face detection has gradually been applied to all aspects of daily life. However, in certain scenarios, while the face is occluded or the faces are dense, the probability of detection success will be greatly reduced. Therefore, how to achieve efficient and accurate face detection under special circumstances with incomplete information is an urgent research topic.

In recent years, a large number of intelligent algorithms for face detection have emerged on the basis of target detection algorithms. In terms of face detection, Face++ [1] is a paper on face recognition algorithm research proposed by Konshi Science and technology team. CascadeCNN [2] is a deep convolutional network implementation of the classic Viola-Jones [3] method, and proposes a cascaded structure to detect faces from coarse to fine. Face R-CNN [4] is based on the Faster R-CNN [5] framework for face detection, the center loss is added to the original R-CNN structure. Although the face detection technology has made great progress, the detection effect of the occluded face is still not satisfactory, because there is no prior knowledge about the occlusion part, the occlusion part can be in any position of the face image, any size or shape.

In order to solve the problems of low accuracy of occlusion and dense face detection, this paper proposes an algorithm model based on deep residual network [6] and non-maximum value suppression [7] technology.

Our main contributions are listed as follows:



The full convolutional layer of the deep residual network is replaced by two full convolutional layer branches in order to achieve accurate prediction of face target pixel-level bounding boxes and classification scores to improve the accuracy of face detection.

The improved non-maximum suppression technology is used in the detection process to overcome the false detection and missed detection caused by the dense or overlapping face targets, and further improve the efficiency of the algorithm.

2. Face detection algorithm based on deep residual network

This paper proposes a face detection algorithm importing ResNet-50 as the backbone network and improved NMS (Non-Maximum Suppression) algorithm. The algorithm is described as follows.

2.1. Improved deep residual network

ResNet introduces residual learning to solve the problem that deep network is difficult to optimize. Its module structure is shown in Figure 1. Let $H(x)$ denotes the optimal mapping, and let the stacked nonlinear layer fit another mapping $F(x) = H(x) - x$. At this time, the optimal mapping can be expressed as $H(x) = F(x) + x$. Residual mapping adds shortcut connection in the feedforward network to perform simple identity mapping, which will not increase additional parameters and computational complexity, and is easier to optimize than the original mapping, as shown in Figure 1.

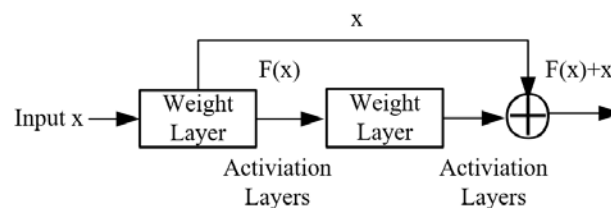


Figure 1. Diagram of residual network

In this paper, ResNet-50 is used as the backbone network, and the fully connected layer in ResNet-50 is deleted and two fully convolutional layers are added to predict the pixel-level bounding box and confidence score respectively. The network structure is shown in Figure 2.

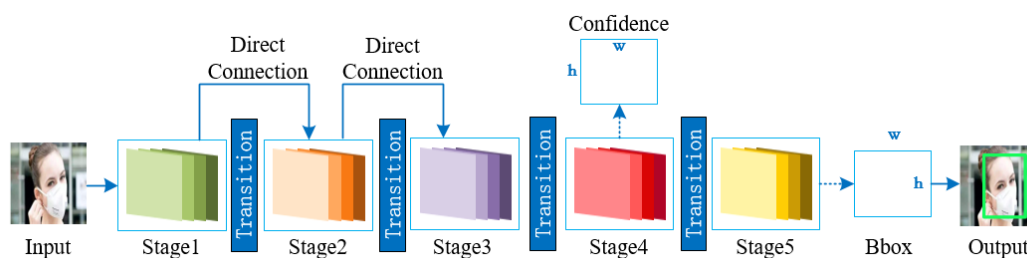


Figure 2. Add branch to deep residual network

As shown in Figure 2, a convolutional layer is added at the end of the fourth stage of ResNet-50 with a step size of 1, and a kernel size of $512 \times 3 \times 3 \times 1$, and then linear interpolation is performed to adjust the feature map to the original image after aligning the feature map with the input image. The S-shaped cross entropy loss is used on the feature map to regress the generated confidence heatmap. In order to predict the bounding box heatmap, a convolutional layer with a convolution kernel size of $512 \times 3 \times 3 \times 4$ is added at the end of the fifth stage of ResNet-50. Similar to the fourth stage, the feature map is adjusted to the original image size and matched with the input image alignment. The final loss is calculated as the weighted average of the losses of the two branches.

The confidence branch is connected at the end of the fourth stage of ResNet-50, and the bounding box branch is inserted at the end of the fifth stage, because the bounding box for loss function calculation is a whole, so a larger receptive field is required, and can intuitively predict the bounding box of the object from the confidence heatmap. In this way, the bounding box branch can be regarded as a bottom-up strategy, and the bounding box can be abstracted from the confidence heatmap.

2.2. Improved NMS algorithm

For detection tasks, NMS is a necessary component, a post-processing algorithm for performing redundant removal operations on detection results. The standard NMS performs greedy clustering based on a fixed distance threshold, that is, greedily selecting high-scoring detection results and deleting those adjacent results that exceed the preset threshold, so that a trade-off between recall and precision can be achieved. The result of this will cause the face bounding box generated by the detection of closely adjacent face objects to be deleted because of the high overlap. The original NMS can be described as formula 1:

$$s_i = \begin{cases} s_i, IoU(N, x_i) < T \\ 0, IoU(N, x_i) \geq T \end{cases} \quad (1)$$

N represents the bounding box with the highest current score, which is a to-be-processed frame. The candidate window x_i overlapping with N has an overlap degree IoU less than the preset threshold T . The score is retained. For windows greater than the preset threshold T , the scores are all set to zero.

The improved form adopts Gaussian weighted processing method. For the candidate bounding box x_i that overlaps the bounding box N with the highest current score, when their overlap IoU is greater than preset threshold T , their score will not be set to 0, but the score add a weight to avoid violent deletion of closely adjacent face objects due to their high overlap. Instead, the larger the overlap IoU between the candidate face bounding box and N , the faster the score of the candidate bounding window will drop. The formula is as follows:

$$s_i = \begin{cases} s_i, IoU(N, x_i) < T \\ s_i, e^{-\frac{IoU(N, x_i)^2}{\sigma}}, \forall x_i \notin D, IoU(N, x_i) \geq T \end{cases} \quad (2)$$

3. Analysis of experimental results

This article uses the FDDB (Face Detection Dataset and Benchmark) dataset to test the experimental results, which is one of the current authoritative face detection evaluation sets and contains a total of 2845 pictures with various face states, including occlusion, dense, low resolution and so on.

The face detection result and the confidence heat map are shown in Figure 3. It can be seen that the algorithm in this paper can accurately find the pixels with high confidence in the face as shown in Figure 3 (b), can accurately generate the corresponding face detection the bounding box is shown in Figure 3 (a).



(a) Face detection boundary block diagram



(b) Face detection confidence heat map

Figure 3. Face detection result

The main purpose of applying the improved non-maximum suppression (NMS) algorithm is to find the best face detection position and select the box that most accurately characterizes the face, which can effectively distinguish two faces that are too close, as shown in Figure 4. It shows that the faces in Figure 4 (a) that are too close can be effectively distinguished after adding NMS, as shown in Figure 4 (b).

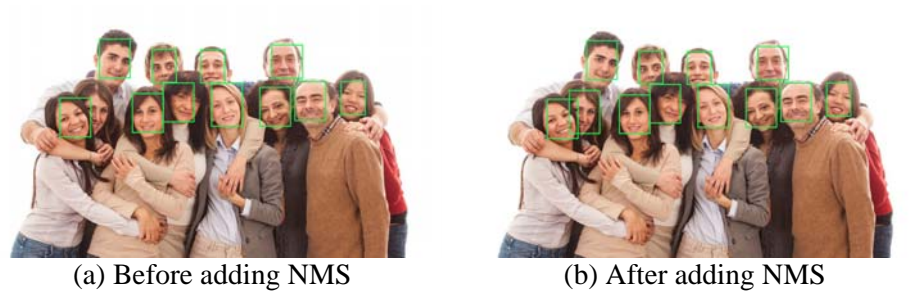


Figure 4. Improved NMS algorithm can distinguish dense faces

This paper uses the commonly used ROC (Receiver Operating Characteristic) curve and FPS (Frame Per Second) in the field of target detection to objectively evaluate the ability of the algorithm in this paper to detect faces.

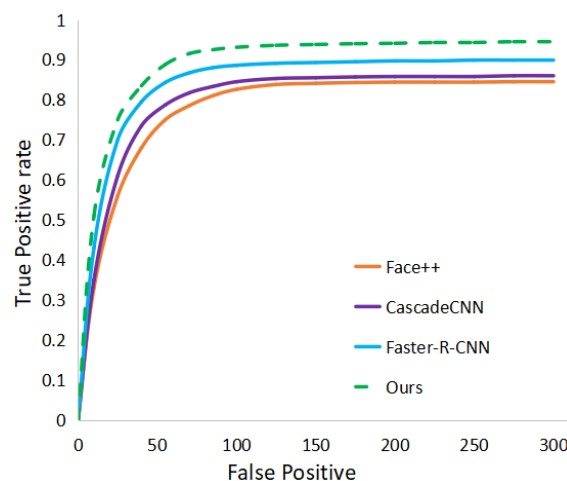


Figure 5. ROC curve comparison

By comparing the algorithm in this paper with the algorithms Face++, CascadeCNN, Faster-R-CNN, which perform better in face detection, it can be seen that the detection performance of this algorithm is better than other face detection methods, as shown in Figure 5. The accuracy of Face++ and CascadeCNN algorithms is poor, about 85%; Faster-R-CNN algorithm has better accuracy, reaching about 90%; the algorithm in this article has higher accuracy, basically above 90%. Further statistics on the accuracy rate when the false positive number is 300 are shown in Table 1.

FPS represents the number of pictures processed per second and is used to measure the detection efficiency of the algorithm. The experimental results of this article are shown in Table 1.

Table 1. Face detection performance

	Face++	CascadeCNN	Faster-R-CNN	Ours
Precision/%	85.0	85.6	89.9	94.5
FPS/(Frame/s)	5	7	8	10

It can be seen from the table that the face detection accuracy of the algorithm in this paper is 94.5%, which is the highest algorithm in other common algorithms. In terms of detection speed, the detection speed of the algorithm in this paper is significantly better than other algorithms, reaching 10FPS.

4. Conclusions

In this paper, the Resnet-50 is used as the backbone network, and a confidence branch and a face detection branch are added at the same time, and then an improved non-maximum suppression method is added in the detection process to make the target location when the face is occluded more accurately. The experiment of this algorithm on the Fddb dataset proves the improvement of face detection performance, and the ideal detection effect can also be obtained when detecting occluded faces. At the same time, its detection speed is better than other networks, and it can be better used in real scenes that require real-time detection of human faces.

Acknowledgments

This work was financially supported by National High Score Special Project (No.70-Y40-G09-9001-18/20), Liaoning Natural Fund Project (No.20180550334), Liaoning Provincial Department of Education Project (No.L201701) fund.

References

- [1] Zhou E, Cao Z, Yin Q. Naive-Deep Face Recognition: Touching the Limit of LFW Benchmark or Not? [J]// Computer ence, 2015.
- [2] H. Li, Z. Lin, X. Shen, et al. A convolutional neural network cascade for face detection [C]// 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston, MA, USA: IEEE Computer Society, 2015: 5325-5334.
- [3] Viola P A, Jones M J. Rapid Object Detection using a Boosted Cascade of Simple Features [C]// Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001. Kauai, HI, USA, USA: IEEE Computer Society, 2001: I-I.
- [4] Huaizu Jiang, Erik Learned-Miller. Face Detection with the Faster R-CNN [C] //2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017). Washington, DC, USA: IEEE Computer Society, 2017: 650-657.
- [5] Shaoqing Ren, Kaiming He, Ross Girshick, et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, Volume (39): 1137-1149.
- [6] He K, Zhang X, Ren S, et al. Deep Residual Learning for Image Recognition [C]// IEEE Conference on Computer Vision & Pattern Recognition. IEEE Computer Society, 2016.
- [7] Jan Hosang, Rodrigo Benenson, Bernt Schiele. Learning Non-maximum Suppression [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, HI, USA: IEEE Computer Society, 2017: 6469-6477.