A new face detection method based on Faster RCNN

He Yan, Xiaotang Wang, Yuhan Liu, Yuning Zhang and Huan Li
College of Liangjiang Artificial Intelligence, Chongqing University of Technology, Chongqing 401147, China
E-mail: tangtang@2020.cqut.edu.cn

Abstract. The Non-Maximum Suppression (NMS) method based on Faster RCNN uses hard threshold to identify candidate face frames which to be detected. In complex scenes with partial occlusion of the face and uneven lighting, the phenomenon of missing and false detection of the face is prone to occur. Aiming at this problem, this paper proposes a new face detection method to extract face features through CNN, and generates a large number of face candidate frames which to be detected by the Region Proposal Network (RPN); the hard threshold of NMS is improved by linear weighting method, and the face candidate frame is screened by the linearly weighted NMS. Comparison experiment results show that on the FDDB dataset, the new face detection method in this paper can effectively avoid the false detection and missed detection of multiple faces under partial occlusion and uneven illumination, also has high detection accuracy and robustness.

1. Introduction
Face detection refers to the use of a certain method to determine the position of the face in any picture or video. However, the accuracy of face detection is affected by various factors to a certain extent in complex scenes, such as uneven lighting, changes in human posture and occlusion. In the early stages of face detection, the method of combining artificial design features and machine learning technology was adopted [1], while the artificially designed features can only be used in specific scenarios, they cannot be stable in different scenarios. In the era of deep learning, Jonathan Brandt et al. [2] used a multi-cascade structure Cascade CNN [3] composed of three-layer convolutional neural networks for face detection. The RCNN model proposed in [4] and the Fast RCNN model proposed in [5] are still very complicated because the selective search is still used. On this basis, In [6], proposed a Faster RCNN real-time detection model, using the Region Proposal Network (RPN) instead of the traditional selective search method to directly generate candidate windows, and using non-maximum suppression (NMS) [7] method classifies the target and regresses to screen the candidate window, which greatly improves the accuracy and speed of target detection. The traditional NMS algorithm is a greedy algorithm [6] and its central idea is to search for elements with local maximums to suppress non-maximum values and eliminate a large number of redundant candidate windows to find the best target detection position [8]. If it is in a complex scene with uneven lighting or occlusion, it will easily cause false detection and missed detection. In view of the still unresolved problem, this paper proposes a new Faster RCNN face detection method, which is to improve the linear weighting of the traditional NMS algorithm used in the RPN network. When the IOU value of the candidate frame is greater than the hard core threshold, it is not deleted directly but reduced its confidence through the improved NMS algorithm, so that the position of the high classification confidence frame is more accurate,
retain more real frames, and improve the accuracy of face detection to a certain extent, which can effectively reduce the missed and false detection of target faces in complex scenes.

2. Faster RCNN face detection model

As is shown in Figure 1, the Faster RCNN face detection model is mainly composed of RPN layer and detection network [7].

![Figure 1. Faster RCNN detection process](image)

The Faster RCNN is mainly divided into four steps:

- Convolutional layer: Input a face image, extract facial features through a conv+relu+pooling multi-layer network and use the obtained features for the RPN network.
- RPN layer: Replace the traditional selective search method to generate more accurate candidate regions, classify and regress the candidate boxes to obtain the proposal [9].
- ROI layer: It pools proposals of different sizes into the same size, and then sends them to the subsequent fully connected layer for face classification and position adjustment regression.
- Classification layer: Use the NMS method to remove the redundant part of the candidate frame, and classify and regress the candidate frame again to obtain real position of the face.

3. Improved Faster RCNN

The RPN network is the most critical detection network in the Faster RCNN model, and its essence is generate candidate windows and train high-quality suggested regions through an end-to-end approach. Then uses NMS algorithm to remove redundant and duplicate candidate windows to get the final face area. The traditional NMS algorithm process is first to set up a set \( B = \{b_1, \ldots, b_i, \ldots, b_n\} \) containing \( n \) target face candidate windows and a set \( M \) for storing the selected optimal window, Secondly, according to the score of the candidate frame (\( s_i \) represents the score of the \( i \)-th) candidate frame sort in the set \( B \), then select the candidate box \( m \) with the highest target score, and remove it from set \( B \) and add it to \( M \). Thirdly, calculate the IOU (Intersection and Union Ratio) of all the remaining candidate frames in the set \( B \) and \( m \) respectively. Set a threshold \( N_t \) (In this article, \( N_t = 0.7 \)), if \( \text{IOU} \geq N_t \), it can be determined to overlap with frame \( m \), and then this frame is removed from the set \( M \). Lastly, repeat the first step until the set \( B \) is an empty set, and finally the candidate box in the set \( M \) is what we want to get.

IOU is used as an evaluation function for face detection. The specific formula is as follows:

\[
\text{IOU} = \frac{P \cap G}{P \cup G}
\]

(1)

Where \( P \) represents the predicted face candidate frame, and \( G \) represents the manually labeled face candidate frame.

However, in a complex scene, such as multiple faces partially overlap, the generated candidate frame will easily cause misdetection due to improper selection of the threshold [11]. The traditional NMS algorithm selects the candidate frame with the higher current IOU value according to the greedy strategy, and directly removes the face frame with lower value, which leads to the missed detection of
adjacent target faces. Aiming at the unsolved problems of the traditional NMS algorithm, this paper proposes a linear weighted NMS algorithm on this basis to reduce the confidence of the target frame [8], and make the position of the highly classified candidate frame more accurate. The RPN network screens candidate frames for face detection according to the NMS algorithm. The traditional method and the improved method are shown in equation (2):

\[
\begin{align*}
    s_i &= \begin{cases} 
        s_i & \text{IOU}(M, h_i) < N_i, \\
        0 & \text{IOU}(M, h_i) \geq N_i
    \end{cases} \\
    s_i &= \begin{cases} 
        s_i & \text{IOU}(M, h_i) < N_i, \\
        s_i \cdot (1 - \text{IOU}(M, h_i)) & \text{IOU}(M, h_i) \geq N_i
    \end{cases}
\end{align*}
\]

Where \( s_i \) represents the score of the \( i \)-th candidate frame, when the IOU value is greater than the set threshold, \( s_i \) is not set to 0 and deleted directly, but the final score is calculated by linear weighting. If the IOU of \( h_i \) and \( M \) is larger, the value of \( s_i \) is smaller, and the confidence is lower.

As shown in Figure 2, in the case of adjacent overlapping faces, traditional NMS algorithm (a) only detected one face frame and the detection result is inaccurate, but our method used the improved NMS algorithm (b), both faces are accurately detected, the problem of missing detection is solved to some extent by our method in this paper, and the accuracy of face detection is significantly improved.

4. Experimental results and analysis

4.1. Data source and Test comparison

The Wider Face [12] dataset contains 32203 face images and 393703 labeled faces, most of which contain occlusions, different tilt angles, uneven lighting, and other faces. The FDDB [13] dataset selects 2485 pictures taken in an outdoor environment, including 5171 facial images including various occlusion, multi-posture, low-resolution, and blurry facial images. In this paper, the improved faster face detection model is trained on the Wider Face dataset and tested on the FDDB dataset.

It can be seen from Figure 3 that the accuracy of the Faster RCNN face detection model using the traditional NMS algorithm on the FDDB dataset has reached 93.56%, while under the improved NMS
algorithm, the accuracy has been reached 94.12%. Figure 4 is Recall and Precision curves. It can be seen that under the same recall rate, improved linear weighted NMS algorithm in this paper has a higher accuracy rate than the traditional Faster RCNN face detection method.

In order to further verify that the improved algorithm in this paper improves the accuracy of face detection in complex scenes, 300 images were randomly selected from the FDDB dataset, there are 400 faces, including single face and multiple face pictures in different conditions, such as occlusion and uneven lighting. Comparison experimental results are shown in Table 1, the accuracy of face detection under the traditional Faster RCNN has reached 88%, while our method able to reach 91%. After verification, our method reduces the missed detection and false detection to a certain extent, and effectively improves the accuracy of face detection.

| The Network | Total number of faces | Number of people detected | Number of missed inspections | Number of false detections | Accuracy(%) |
|-------------|-----------------------|---------------------------|----------------------------|----------------------------|-------------|
| Faster RCNN | 400                   | 365                       | 35                         | 13                         | 88          |
| ours        | 400                   | 372                       | 28                         | 8                          | 91          |

4.2. Actual effect

Two multi-face photos are randomly selected on the Internet, and the improved Faster RCNN is used conduct face detection comparison experiments in complex scenes such as occlusion and uneven lighting.

![Figure 5. Occlusion condition; traditional NMS algorithm (a); ours (b).](a) ![Figure 6. Uneven lighting condition; traditional NMS algorithm (a); ours (b).](b)

It can be seen from Figure 5 and Figure 6, we used the traditional NMS algorithm (a) and the improved linear weighting method (b) to conduct comparative experiments under the conditions of occlusion and uneven illumination. Results show that the overlapping target face part will be misdetected and the bright part will be missed by Traditional Faster RCNN face detection model. While our method is a linear weighted improvement of the traditional NMS algorithm, which can detect face frames in complex scenes and effectively solve the problem of missed detection and false detection, then improve the accuracy of face detection to a certain extent.

5. Conclusion

This paper proposes an improved face detection model based on Faster RCNN, compared with the traditional NMS algorithm, the linear weighted NMS algorithm used in this article can effectively reduce the problem of missing and false target face detection in complex scenes such as occlusion and uneven illumination. Future research focuses on how to improve the face detection rate in low-resolution images and further improve the accuracy of face detection.
Acknowledgments
The National Key R&D Plan “Intelligent Robots” Key Project of P.R. China (Grant No.2018YFB1308602), the National Natural Science Foundation of P.R. China (Grant No. 61173184), the Chongqing Natural Science Foundation of P.R. China (Grant No. cstc2018jcyjAX0694).

References
[1] Hua, Y., Baozong, Y., (2000) Feature extraction in human face recognition system. In: Proceedings of 2000 5th International Conference on Signal Processing Volume., Beijing. pp. 1273-1276.
[2] Li, H., Lin, Z., Shen, X., et al. (2015) A convolutional neural network for face detection. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR.), Boston. pp.5325-5334.
[3] Li, Q., Zhou, X., Gu, A., et al. (2018) Nuclear norm regularized convolutional Max Pos @Top machine. Neural Comput and Applic 30, 463–472.
[4] Girshick, R., Donahue, J., Darrell, T. and Malik, J. (2014) Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. In: IEEE Conference on Computer Vision and Pattern Recognition., Columbus, pp. 580-587.
[5] Girshick, R. (2015) Fast R-CNN. In: Proceedings of the IEEE International Conference on Computer Vision., USA. pp. 1440-1448.
[6] Jiang, H., Learned-Miller, E. (2017) Face Detection with the Faster R-CNN. In: 2017 12th IEEE International Conference on Automatic Face and Gesture Recognition., Washington. pp. 650-657.
[7] Zhang, H., Du, Y., Ning, S., Zhang, Y., et al. (2017) Pedestrian Detection Method Based on Faster R-CNN. In: 2017 13th International Conference on Computational Intelligence and Security(CIS)., Hong Kong. pp. 427-430.
[8] Zhao, W.Q., Yan, H., Shao, X.Q., (2018) Object detection based on improved non-maximum suppression algorithm. Journal of Image and Graphics., 23(11):1676-1685.
[9] He, K., Zhang, X., Ren, S., Sun, J. (2016) Deep Residual Learning for Image Recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition., Las Vegas. pp. 770-778.
[10] Ren, S., He, K., Girshick, R. et al (2017) Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(6):1137-1149.
[11] Stewart, R., Andriluka, M. et al (2016) End-to-end people detection in crowded scenes. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition., USA. pp. 2325-2333.
[12] Yang, S., Luo, P., Loy, CC., et al (2016) WIDER FACE: A Face Detection Benchmark.In:2016 IEEE Conference on Computer Vision and Pattern Recognition., Las Vegas. pp. 5525-5533.
[13] Jain, V., Learned-Miller, E. (2010) FDDB: A benchmark for face detection in unconstrained settings. Technical Report UM-CS-2010-009, University of Massachusetts, Amherst.