Proceedings of the 2nd International Conference on Computing Innovation and Applied Physics
DOI: 10.54254/2753-8818/19/20230478

The application of convolutional neural networks in face detection

Zhen Huang[1]
[1]Computer Science and Technology, Hunan University, Changsha City, Hunan Province, China

HZ2023VX@163.com

Abstract. Face detection is a popular and challenging issue which is widely studied in the past few decades. Its application includes the identity authentication, human machine interaction, security surveillance and social network. To have a better insight of the application of one of the typical deep learning algorithms called Convolutional Neural Network (CNN) in this field, this paper aims to analyze the current literature and progress about the face detection of low image quality and face detection optimization. The literature of Convolutional Neural Network from 2015 was included in this paper. Past research topics of face detection includes the occlusion, scale, small face cluster, speed, precision and multi-task region proposal network. The comparison between various deep learning-based methods in terms of the performance indicated that there is still no high robustness solution to all problems. The future research agendas of face detection based on the Convolutional Neural Network was also summarized.

Keywords: machine learning, deep learning, face detection, convolutional neural network.

1. Introduction

Face detection is an important research field in computer vision and the first step in facial information analysis. Its application includes the identity authentication, human machine interaction, security surveillance and social network. With the application of face detection, the face detection under unconstrained conditions becomes an important research topic. How to improve the recall rate and reduce the detection time are two main tasks in face detection under unconstrained conditions.

In recent years, the performance of face detection based on CNN is improving. The study of CNN could be traced back to 1994 [1], and with years of research, face detection based on CNN is optimized [2]. Cascade convolutional neural network could exclude the non-face areas and multi-task convolutional neural networks (MTCNN) trained a multitask framework to detect and align face [3]; UnitBox and Grid Loss adopted new loss function to improve the performance [4, 5]. Face detection could be regarded as a single task in target detection and now some target detection framework could be used in face detection after certain improvement. Such frameworks includes R-CNN [6-8], SSD [9], YOLO [10] and Focal Loss [11]. Face detector based on CNN obtained better performance than traditional methods and provides a new benchmark to new methods.

The performance of CNN was excellent, and it became an important method in face detection. Every year, new papers of CNN could be seen in computer vision conference like CVPR, ECCV and ICCV. This paper reviews the development path and future direction of face detection based on CNN.

DOI: 10.54254/2753-8818/19/20230478

2. Low Image Quality Face Detection

Mask, sunglass and hair could block the face during the daily life. Therefore, a good face detection algorithm should have the ability to solve the problem of occlusion. The main problem in this process is the loss of features in blocked areas and the shortage of datasets in blocked face area. The small face detection has certain overlap with scale-invariant detection The method of building an image pyramid and fitting the CNN model to multiple scales will increase the computational load of CNN and slow down the speed of CNN. The single stage headless detector proposed by M.NAJIBI et al. avoids the establishment of image pyramids through a unique design, and becomes a new way of thinking about scaling changes [12]. Small face detection is facing more challenge, the model building using face information to expand the facial features is a new method [13].

2.1. Face blockage

The face blockage could not be avoided during face detection. The randomization in face blockage could change the current face features and bring huge computational load to CNN and requires a large amount of training data. Occlusion becomes a major challenge for face detection tasks due to the lack of features in face occlusion regions and the scarcity of occluded face datasets [14]. At present, the face detection for the occlusion problem is mostly based on the problems faced, in which the face facial feature segmentation and the creation of a dataset are used [5, 14], and good results have been achieved.

Opitz et al. [5] defined a new loss function called Grid Loss:

$$1(\theta) = \max(0,1 - y \cdot \omega^T x + b) + \lambda \cdot \sum_{i=1}^{N} \max\left(0, m - y \cdot \omega_i^T f_i + b_i\right) \tag{1}$$

In this equation, ω and b are connection parameters; λ is to weight the single part detector vs the overall detector, in this experiment it is 1. This method is better than regular CNNs and it is suitable for real time application. Ge SH M et al. built a face occlusion database and proposed LLE-CNNs for facial recognition [14]. LLE-CNNs introduces an embedding module to recover the occluded face features and suppress noise; the verification module uses the repaired face features to verify the face area and fine-tune the face position and scale.

In addition, Tang X et al. assisted detection by learning face-related features, which is robust to partially occluded faces [13]. S.S. FARFADE et al. proposed deep convolutional neural networks (DCNN) based on AlexNet and it is also good for occlusion face detection [15]. Chengji W et al. reduced the probability of missed detection of adjacent faces caused by occlusion by reducing the score by weighting [16].

2.2. Change of Scale

In some practical applications such as video surveillance, it is necessary to detect faces of different scales. Therefore, scale invariance needs to be considered in the design of various face detection systems. It is difficult to use a single network for face detection of different scales. Because the features extracted from a face with a size of 300 pixels are very different from those extracted from a face with a size of 3 pixels [17]. For robust pose, expression, and background occlusion, more convolutional layers are required to learn representative features, but spatial information for detecting small faces is lost due to excessive pooling and convolution operations [18]. There are two methods to solve the scale invariance: 1) Use the image pyramid for multi-scale testing, but it will bring a lot of computational burden; 2) Fit the CNN model to multi-scale, but it will increase the model and the amount of calculation. The high computational cost is not suitable for practical application systems, so it is particularly critical to design algorithms with small computational costs.

Hao Z et al. proposed a scale-aware face detection to deal with the scale problem. Yang SH adopted different network structures for faces of different scales, integrated these discrete networks into one network, and optimized learning in an end-to-end manner [19]. The algorithm can improve the detection speed through model compression. Zhou Anzhong et al. established a multi-scale face detection model by extracting feature forms of different scales in different layers of the convolutional

DOI: 10.54254/2753-8818/19/20230478

neural network [20]; Based on the development of object detection algorithms, the scale-invariant face detector based on the object detection algorithm also achieves good scale invariance, but different algorithms have different processing methods. Based on Faster R-CNN and R-FCN algorithms extract features and train the detector by extracting features on higher-level ROI pooling maps [8, 21, 22]; the face detection algorithm based on SSD object detection framework is based on different convolutions [9]. Layers utilize multi-scale feature maps to train a scale-invariant face detector.

2.3. Small Face Clusters

Face detection has achieved breakthroughs and remarkable results, but detection of small faces is still an open challenge. At present, small face detection mainly faces the following difficulties: 1) the detection rate of small faces is still low under the condition of scale invariance; 2) the overlapping of faces when there are multiple faces greatly reduces the available features of small faces; 3) low resolution and complex backgrounds, etc. Since the spatial information features of detecting small faces will be lost due to excessive pooling and convolution operations, the detection results of small faces can be improved by reasonable adjustment of anchors and effective use of face-related information. Hu P Y et al. trained separate detectors for different scales and supplemented with relevant information to improve the detection results of small faces [17]. Zhang S F et al. proposed a multi-scale face real-time detection method based on a single neural network, which improves the recall rate of small faces and reduces the false positive rate through a scale-compensated anchor matching strategy and maximizing labels [23]. PyramidBox proposed by Tang X et al. utilizes relevant information and Data-anchor-sampling strategy to enhance small face detection [13].

Deep-IR trains a facet detector by extracting features from the bottom convolutional layers to solve the problem of background and facet confusion in high convolutional layers [24], S3FD and FaceBoxes improve the matching strategy and anchor density to Detect small face groups [23, 25], while Scaleface and HR-ER train face detectors by assigning scale-specific convolutional layers [17, 18].

3. Performance Optimization of Face Detection System

A general-purpose face detection system cannot only be aimed at a single problem but must be robust to all possible situations. In-depth research on a single problem is also to better solve this problem. The traditional boosting method and the convolutional neural network have their own advantages in speed and accuracy, respectively. The effective combination of the traditional method and the convolutional neural network is an effective method to solve the speed and accuracy. Multi-task face detection can use the synergy between different characters to assist face detection, but the computation load is a difficult problem.

3.1. Speed Optimization

CNN-based face detection can easily achieve better accuracy than traditional methods, but compared with traditional Boosting methods, it has obvious disadvantages in terms of speed. Nowadays, popular applications such as robots and video surveillance systems have limited computing power and cannot withstand huge amounts of computation when running in real time. Therefore, if CNN is to be practical, the operating speed of the face detection system must be solved. This also makes the speed problem a hot issue for researchers.

Combining with traditional methods is an effective way to increase speed [26], with complementary advantages, the real-time performance of face detection systems was improved while ensuring relative accuracy. The Cascade CNN proposed by Li H X et al. is the representative of the combination of the V-J structure and the convolutional neural network [2]. The algorithm was the fastest operating speed based on the CNN at that time, and was robust to illumination and angle changes to a certain extent, but the first stage in this framework is based on the dense sliding window method for window filtering, which limits the performance limit of the algorithm on images with high resolution and a large number of small faces. Qin H W et al. proposed an improved JTCCNN based on

DOI: 10.54254/2753-8818/19/20230478

Cascaded CNN [3]. JTCCNN uses the BP algorithm for joint training, the latter network will contain the former network, the convolutional layers are shared in different stages, and the concept of cascade is reflected inside the network.

Researchers also adjusted and improved the cascade idea of convolutional neural networks. I. KALINOVSKII et al. proposed a face detection method based on a cascade of compact convolutional neural networks [27]. This algorithm discards the fully connected layers with more parameters in CNN. Due to the CNN modification and simple cascade structure, the speed of the network is ideal, but the accuracy is relatively low, and it can only be used to detect frontal faces. Zhang K P et al. applied the cascade idea to different network layers of CNN [28]. The algorithm completes the simple face detection task and the complex face detection task in the first few layers and the latter part of the CNN network respectively. Through this strategy, simple non-face regions are quickly excluded, and the detection speed is accelerated.

In addition to the use of the cascade structure, researchers have also achieved a substantial increase in speed through the transformation and unique design of the CNN. Zhang SH F et al. proposed FaceBoxes [25], a face detector running on CPU. FaceBoxes consist of rapidly digested convolutional layers (RDCL) and multiple scale convolutional layers (MSCL). RDCL enables FaceBoxes to achieve real-time speed on the CPU; while MSCL enriches receptive fields and discretizations (anchors) at different layers to handle faces of different scales.

M. NAJIBI et al. designed a SSH face detector. SSH detects faces directly from a single stage of the early convolutional layers in the classification network [12]. The optimized design for speed is reflected in three aspects: 1) Eliminate the fully connected layer with a large number of parameters; 2) Replace the image pyramid through a unique design. Multi-scale face detection simultaneously greatly improves the speed; 3) Detects faces with different scales in different layers of a single forward pass of the network.

3.2. Strong Robust Face Detection

Obviously, only solving a single problem in the face detection task is not enough in practical applications, and only a complete and highly robust system has universal practical value.

S.S.FARFADE et al. proposed a deep convolutional neural network model (DCNN) based on the AlexNet network [15]. They leverage the high capacity of deep convolutional neural networks for classification and feature extraction, learn a single classifier for detecting faces from multiple angles, and remove the SVM classifier and bounding box regression module to minimize computational complexity. The algorithm can detect faces from multiple angles, partial occlusions and lighting effects, and scale images to detect faces of different scales. The idea of reconstructing the fully connected layer enables DCNN to accept image input of any size.

Li Y et al. proposed an end-to-end facial key point-based face detection framework based on the combination of Faster R-CNN convolutional network and 3D model [29]. The framework in this paper makes 2 modifications to Faster R-CNN: 1) The 3D average face model is used to replace the predefined anchor boxes in the PRN; 2) The ROI pooling is replaced by a configuration pooling layer according to the face structure information Floor.

Chen D et al. proposed a cascaded convolutional neural network to address large-scale pose changes [30]. The first stage of STN (supervised transformer network) uses multi-task RPN to predict candidate face regions and label key points; the second stage uses R-CNN to classify candidate regions using splicing features. STN introduces the supervised transformer layer to correct the candidate window, uses the Non-top K suppression strategy to improve the recall rate while ensuring the accuracy, and uses the ROI convolution strategy to only calculate the face area to speed up the operation; however, this method has obvious shortcomings and cannot be Real-time performance on raw images.

The Faceness-Net network proposed by Yang S et al. is a typical coarse-to-fine workflow [31]. The detection system is divided into 2 stages: 1) generate a face part map (partness map) according to the attribute-aware deep network, calculate the confidence of the face by combining these components, and then reorder the faces; 2) Train a CNN for face classification and boundary regression to further improve recall. Faceness-Net reduces network parameters by 83% through parameter sharing, greatly speeds up network operation, and realizes multi-tasking in the same network.

Wang Y T et al. proposed a fully convolutional network region face detector Face R-FCN, which uses new technologies such as position-sensitive average pooling, multi-scale training and testing, and online hard example mining strategy to improve detection accuracy [22].

Liu Yingjian et al. combined the Edge Boxes algorithm with a convolutional neural network to enhance the robustness to factors such as occlusion, illumination, and rotation [32]. Wu Suwen et al. used a selective search strategy to screen candidate windows [33], and the convolutional neural network combined with Gabor kernel optimization has better robustness to illumination changes and multi-pose problems in unconstrained environments. Good results have been achieved on the "labled faces in the wild face" dataset. Wang Chengji et al. enhanced the accuracy of face detection under factors such as pose, illumination and scale by means of feature fusion [16].

3.3. Multi-task Face Detection

Multi-task face detection refers to performing tasks such as face alignment, pose estimation, and gender recognition while completing the face detection task. Multi-task face detection can utilize the intrinsic connection and synergy between tasks to improve the effect of face detection [34]. Although multi-task face detection will increase the computational complexity to a certain extent, it is of great reference in some specific applications that require multi-tasking.

4. Face Detection Algorithm Performance

In this study, the occlusion, scale, small face clusters, speed, accuracy, and multi-task face detection in face detection tasks are reviewed in two aspects: low image quality face detection and performance optimization of face detection systems. Table 1 shows the performance of some face detection algorithms on FDDB, where CS means continue score, DS means discrete score, and FP means false positive; Table 2 shows the performance of some face detection algorithms on WIDER FACE; Table 3 is the speed performance of some face detection algorithms.

Table 1. The performance of face detection algorithm on FDDB.

algorithm	CS	DS	FP
SA-RPN	--	0.938	1000
ScaleFace	--	0.960	2000
S3FD	0.856	0.983	1000
PyramidBox	0.860	0.987	1000
Cascade CNN	--	0.857	1000
CasCNN	--	0.872	1000
ICC-CNN	--	0.971	2000
DDFD	--	0.840	1000
Conv3D	0.767	0.912	2000
Faceness-Net-SR-RP	0.744	0.928	2000
Face R-FCN (model-B)	0.760	0.994	2000

DOI: 10.54254/2753-8818/19/20230478

Table 2. The performance of face detection algorithm on WIDER FACE.

algorithm	Easy	Middle	Hard
ScaleFace	0.867	0.866	0.764
Tiny Faces	0.919	0.908	0.823
S3FD	0.937	0.924	0.852
PyramidBox	0.956	0.946	0.887
Faceness-Net-SR-RP	0.717	0.615	0.305
Face R-FCN	0.943	0.931	0.876

Table 3. Speed performance of face detection algorithm.

algorithm	Speed/fps	Notes
S3FD	36	Titan X (Pascal) and cuDNN v5.1
Cascade CNN	100	GPU for VGA-resolution images
CasCNN	10	single CPU core on FDDB
C-CNN Cascade	27	4K Ultra HD video stream
ICC-CNN	40	GPU for VGA-resolution images

5. Conclusion

After systematically analyzing the related literature on face detection based on convolutional neural network since 2015, this paper summarizes the research status and development direction of face detection based on convolutional neural network. Since the V-J structure was proposed, face detection has developed rapidly. In recent years, face detection based on convolutional neural networks has achieved good results. However, the problems faced by face detection still exist, such as the complex background of random transformation, and the strange Face occlusion, unpredictable lighting conditions, changing postures and expressions, different resolutions, and running speed problems brought by CNN, various internal and external factors make face patterns ever-changing, and there is still no high robustness solution to all problems. Face detection based on convolutional neural networks has approached saturation on difficult datasets such as FDDB and WIDER FACE, but the computational cost is still huge. In future research, face detection with robust small faces and low image quality while speeding up the detection speed will become the focus of research, and the research results will be closer to practical applications. In-depth research is required in the following areas:

1) Optimize the network structure. With the advancement of the theoretical analysis of convolutional neural networks, how to reduce and adjust network parameters, how to simplify the network structure while maintaining the detection effect, in order to achieve the purpose of accelerating convergence and operation. In addition, feature extraction and discarding can also be optimized by studying feature extraction and information loss in the middle layer.

2) Combination of traditional methods and convolutional neural networks. The traditional Boosting method has the speed advantage, while the convolutional neural network can achieve high accuracy, and the high speed and high precision of detection can be achieved through the organic combination of the traditional method and the CNN.

3) Auxiliary means to detect complex faces. It is an effective direction to solve the problems of high detection difficulty such as high occlusion rate, low resolution and complex background in face detection by auxiliary means. For example, the application of benchmark, face alignment and other methods improves the efficiency and accuracy of face detection.

4) Optimize the training sample selection algorithm. In some applications, the selection of training samples can directly determine the quality of the model. Selecting effective training samples from a large number of samples (especially negative samples are more complex and changeable) is an important part of optimizing the model. The introduction of the OHEM strategy is convenient played an important role.

DOI: 10.54254/2753-8818/19/20230478

References

[1] Vaillant R 1994 Original approach for the localisation of objects in images Vision, Image and Signal Processing IEE Proceedings 141(4) 245-250.
[2] Li H 2015 A convolutional neural network cascade for face detection In Proceedings of the IEEE conference on computer vision and pattern recognition pp. 5325-5334
[3] Qin H 2016 et al. Joint training of cascaded CNN for face detection In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition pp. 3456-3465
[4] Yu J 2016 et al. Unitbox: An advanced object detection network In Proceedings of the 24th ACM international conference on Multimedia pp. 516-520
[5] Opitz M et al. 2016 Grid loss: Detecting occluded faces In European conference on computer vision pp. 386-402 Springer Cham
[6] Girshick R et al. 2015 Region-based convolutional networks for accurate object detection and segmentation IEEE transactions on pattern analysis and machine intelligence 38(1) 142-158
[7] Girshick R 2015 Fast r-cnn In Proceedings of the IEEE international conference on computer vision pp. 1440-1448
[8] Ren S et al. 2015 Faster r-cnn: Towards real-time object detection with region proposal networks Advances in neural information processing systems 2015:91-99
[9] Liu W et al. 2016 Ssd: Single shot multibox detector In European conference on computer vision (pp. 21-37) Springer Cham
[10] Redmon J et al. 2016 You only look once: Unified, real-time object detection In Proceedings of the IEEE conference on computer vision and pattern recognition pp. 779-788
[11] Lin T Y et al. 2017 Focal loss for dense object detection In Proceedings of the IEEE international conference on computer vision pp. 2980-2988
[12] Najibi M et al. 2017 Ssh: Single stage headless face detector In Proceedings of the IEEE international conference on computer vision pp. 4875-4884
[13] Tang X et al. 2018 Pyramidbox: A context-assisted single shot face detector In Proceedings of the European conference on computer vision (ECCV) pp. 797-813
[14] Ge S et al. 2017 Detecting masked faces in the wild with lle-cnns In Proceedings of the IEEE conference on computer vision and pattern recognition pp. 2682-2690
[15] Farfade S S et al. 2015 Multi-view face detection using deep convolutional neural networks In Proceedings of the 5th ACM on International Conference on Multimedia Retrieval pp. 643-650
[16] Chengji W et al. 2018 A face detection method based on multi-layer feature fusion Journal of Intelligent Systems 13(1):138-146
[17] Hu P et al. 2017 Finding tiny faces In Proceedings of the IEEE conference on computer vision and pattern recognition pp. 951-959
[18] Yang S et al. 2017 Face detection through scale-friendly deep convolutional networks arXiv preprint arXiv:1706.02863
[19] Hao Z et al. 2017 Scale-aware face detection In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition pp. 6186-6195
[20] Anzhong Z et al. 2018 A Face Detection Model Based on Multiscale Convolutional Neural Networks Computer Engineering and Applications 54(14):168-174
[21] Jiang H et al. 2017 Face detection with the faster R-CNN In 2017 12th IEEE international conference on automatic face & gesture recognition (FG 2017) pp. 650-657 IEEE
[22] Wang Y et al. 2017 Detecting faces using region-based fully convolutional networks arXiv preprint arXiv:1709.05256
[23] Zhang S et al. 2017 S3fd: Single shot scale-invariant face detector In Proceedings of the IEEE international conference on computer vision pp. 192-201
[24] Sun X et al. 2018 Face detection using deep learning: An improved faster RCNN approach. Neurocomputing,299 42-50
[25] Zhang S et al. 2017 October Faceboxes: A CPU real-time face detector with high accuracy

In 2017 IEEE International Joint Conference on Biometrics (IJCB) pp. 1-9 IEEE
[26] Kang S et al. 2018 Face Detection Algorithm Based on Cascaded Convolutional Neural Networks Journal of Nanjing University of Science and Technology 42(1):40-47
[27] Kalinovskii I et al. 2015 Compact convolutional neural network cascade for face detection arXiv preprint arXiv:1508.01292
[28] Zhang K et al. 2017 Detecting faces using inside cascaded contextual cnn In Proceedings of the IEEE International Conference on Computer Vision pp. 3171-3179
[29] Li Y et al. 2016 October Face detection with end-to-end integration of a convnet and a 3d model In European Conference on Computer Vision pp. 420-436 Springer Cham
[30] Chen D et al. 2016 Supervised transformer network for efficient face detection In European Conference on Computer Vision pp. 122-138 Springer Cham
[31] Yang S et al. 2017 Faceness-net: Face detection through deep facial part responses IEEE transactions on pattern analysis and machine intelligence 40(8) 1845-1859
[32] Jianyin L et al. 2018 Unrestricted Face Detection Based on Edge Boxes and Deep Learning modern electronic technology 41(13)39-33
[33] Suwen W et al. 2017 Face Detection Based on Selective Search and Convolutional Neural Networks Application Research of Computers 34(9):2854-2857
[34] Chen D et al. 2014 Joint cascade face detection and alignment In European conference on computer vision pp. 109-122 Springer Cham