Research Article

Yue Qi, Yiqin Wang*, and Yunyun Dong

Face detection method based on improved YOLO-v4 network and attention mechanism

https://doi.org/10.1515/jisys-2023-0334
received December 26, 2023; accepted January 29, 2024

Abstract: Due to insufficient information and feature extraction in existing face-detection methods, as well as limited computing power, designing high-precision and efficient face-detection algorithms is an open challenge. Based on this, we propose an improved face detection algorithm. First, through 1 × 1's common convolution block (CBL) expands the channel for feature extraction, introduces a depthwise separable residual network into the YOLO-v4 network to further reduce the amount of model computation, and uses CBL to reduce the dimension, so as to improve the efficiency of the subsequent network. Second, the improved attention mechanism is used to splice the high-level features, and the high-level features and the shallow features are fused to obtain the feature vectors containing more information, so as to improve the richness and representativeness of the feature vectors. Finally, the experimental results show that compared with other comparative methods, our method achieves the best results on public face datasets, and our performance in personal face detection is significantly better than other methods.

Keywords: YOLO-v4, big data, deep learning, face detection, deep separable residual network, attention mechanism

1 Introduction

Text, photos, videos, and other types of information are becoming increasingly common in everyday life [1,2]. As the main research object of computer vision, pictures, videos, and other information forms are everywhere in our lives, such as movies, animations, face brushing, and sign-in, making our lives more rich and convenient [3–6]. Computer vision uses modern artificial intelligence technology to detect, recognize, and track the video and pictures collected by the monitoring equipment. Face recognition, as an important component of computer vision, is the main method of identity recognition in daily life. Because of its convenience and speed, it is widely used in public places where identity recognition is required, bringing great convenience to staff, consumers, and other personnel [7–10]. Face detection is one of the important applications in the field of computer vision, which involves automatic recognition and verification of individual identity information. The research objects are mainly objects that are mainly divided into static pictures and video streams [11–13]. Among them, video streams are more real-time, rich, and contain more information than static pictures, which increases the difficulty of face detection technology research for dynamic video [14–17].

At present, most of the images used in research are in standard format. However, images in real life do not necessarily have standard faces, and existing face detection technologies have robustness problems [18–21].

*** Corresponding author: Yiqin Wang,** Department of Information Technology and Engineering, Jinzhong University, Jinzhong, Shanxi, 030619, China, e-mail: wangyiqin@jzxy.edu.cn
Yue Qi: Computer Network Center, Taiyuan Open University, Taiyuan, Shanxi, 030024, China, e-mail: 413740189@qq.com
Yunyun Dong: Software College, Taiyuan University of Technology, Taiyuan, Shanxi, 030600, China, e-mail: dongyunyun312902@126.com

This is currently also a challenge in face detection. Deep learning is a branch of machine learning that is based on neural networks and automatically extracts features through training large amounts of data, thereby achieving recognition and understanding of complex patterns [22]. The current deep learning-based face detection algorithms have drawbacks such as sensitivity to environmental factors [23]. Deep learning face detection algorithms may fail under the influence of factors such as lighting, shadows, angles, occlusion, etc., resulting in the algorithm being unable to accurately detect faces [24]. Deep learning algorithms heavily rely on training data, so collecting and processing a large amount of high-quality facial image data is a huge challenge [25]. In addition, deep learning algorithms require a large amount of computing resources and time for training and inference, which limits their promotion in practical applications. Nevertheless, deep learning algorithms still have broad development prospects in the field of facial detection [26]. In the future, deep learning-based facial detection algorithms will develop towards a more intelligent, real-time, and efficient direction, and more innovative application scenarios will also emerge [27]. Based on this, the model proposed in this article includes modifications to the model and the introduction of attention mechanisms. The improved depthwise (DW) separable residual network is introduced in the YOLO-v4 network, and the attention mechanism is introduced.

In this article, the YOLO-v4 backbone network is improved to improve the efficiency of model detection. The advanced features are spliced with the improved attention mechanism, and then the advanced features are fused with the shallow features. In this way, feature vectors containing more information can be obtained, thus improving the richness and representativeness of feature vectors.

This article first introduces the definition of face detection, the research status at home and abroad, and the development process of CNN, then analyzes the difficulties of face detection technology, and finally introduces the innovation work of this article. Second, the DW separable residual network structure and attention mechanism are introduced, and experiments are carried out on the WiderFace dataset and MAFA occluded face dataset. Finally, the article summarizes the full text and points out the shortcomings of this work and the research prospects based on this method.

2 Related research

The purpose of face detection is to detect images, determine whether there are faces in the image, and locate the faces in the image. Face detection algorithms mainly include knowledge-based methods, statistic-based methods, and deep learning methods. Farfade et al. [28] proposed a deep-learning network model based on the Deep Sense Face Detector. This model does not need posture or key point annotation and can capture faces in all directions with a single model. It has strong resistance to various poses. However, the greater the deflection angle and posture change, the lower the accuracy will be. Li et al. [29] construct a cascade structure to detect facial features from rough to fine based on the implementation of a deep convolution network of Cascade CNN. Jiang and Learned-Miller [30] add the center loss function to the original R-CNN structure, which can better detect small-size images. Yu et al. [31] proposed a new intersection over union loss function to replace the commonly used L2 loss function and improve the accuracy of face detection. Woo et al. [32] propose a lightweight convolution block attention module for object recognition. Peng et al. [33] proposed an object partial attention model for fine-grained image classification. This model combines attention mechanism and residual network module and has been successfully applied to fine-grained image classification with good performance. Zeng et al. [34] used the Mish function to improve the original residual network to obtain the improved residual network and introduced the attention mechanism to obtain the final network model to make the extracted facial features more discriminative. Yan et al. [35] designed a method of combining L2 loss and triplet loss to form a loss function and improved the face residual network by using deep separable convolution to reduce the number of network parameters.

However, the above methods can only mine shallow information. The extracted features are weak in the expression of human faces, often resulting in low accuracy of face recognition. To overcome the above problems, the innovation of the proposed method lies in the following:

(1) An improved DW separable residual network is introduced to improve YOLO-v4, and a 1 × 1 common convolution block (CBL) is used to expand the channel for feature extraction. The DW separable convolution is introduced to further reduce the model computation, and CBL is used for dimensionality reduction to improve the computational efficiency of subsequent networks.

(2) The improved attention mechanism is used to fuse advanced features and shallow features to obtain feature vectors containing more information.

3 Proposed method

3.1 Network framework

The complex network structure often makes the detection model slow to run, difficult to train, and difficult to achieve real-time detection speed on equipment with low computing costs. The detection speed based on the lightweight face detection method is fast, which can well meet the requirements of fast face detection in practical applications. However, the precision of face detection does not meet standards, and face position is not accurate, especially for complex faces. Therefore, we propose a lightweight and efficient face-detection algorithm. The backbone network introduces a more balanced and efficient DW separable residual network based on a lightweight network. In addition, our method integrates an attention mechanism. The method is mainly composed of three parts, namely, a lightweight feature extraction network, an improved attention module, and an output layer. Figure 1 depicts the entire network framework.

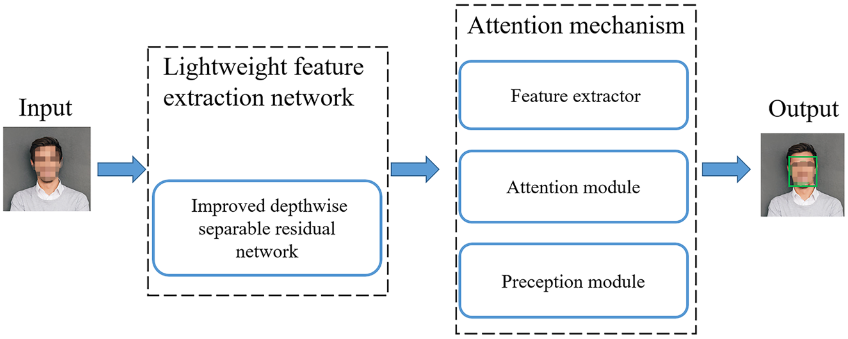

Input
Lightweight feature extraction network
Improved depthwise separable residual network
Attention mechanism
Feature extractor
Attention module
Preception module
Output

Figure 1: Network framework.

3.2 Depthwise separable residual network

YOLO-v4 backbone network adopts residual network structure. The specific process is as follows: The input data first passes through the 2 × 2 basic volume layer and then is divided into two parts. One part is used as the main part (Resblock) to iterate in the loop to obtain the operational relationship between the weight and the input data. The other part is used to establish an independent residual edge, and the input data are directly output after a small amount of processing. Finally, the output data of the two parts are added across layers, and the sum result is used as the output of this layer. This structure is used to make gradient flow spread in different paths by separating gradient flow, so that the network can learn more correlation differences of gradient flow. At the same time, it can reduce computational power consumption and improve the computing speed and network learning ability by reducing the amount of cyclic stacking computation. In this article, the remaining network structure in the YOLO-v4 backbone network is improved to a deeply separable remaining network. Figure 2 depicts the specific structure. This structure continues the idea of YOLO-v4 separating

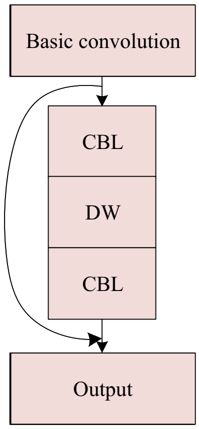


Figure 2: Depthwise separable residual structure.

gradient flow and continues to iterate some input data while the other part jumps to the end. Then, the ordinary convolution residual block of the original cyclic iteration is replaced with the deep separable residual block. First, pass 1 × 1's CBL expands the channel for feature extraction, then introduces DW to further reduce the amount of model computation, and finally uses 1 × 1 CBL to reduce dimension to improve the efficiency of the subsequent network.

In the DW separable residual network structure of this article, the process of convolution is as follows: assuming that the number of input image channels is C_{in} and the output channel is C_{out}; first, the input image is multiplied by C_{in} different channels through the $N \times N$ deep convolution, and the result of the first step is obtained. The width and height of the image vary here, but the number of channels remains unchanged. Then, use $1 \times 1 \times C_{\text{in}}$ common convolution to check the result of the first step for the convolution operation. Then, the calculation formula for the parameter amount of a convolution point of a deep separable convolution is [36]

$$M = C_{\text{in}} \times N \times N + C_{\text{out}} \times C_{\text{in}} \times 1 \times 1. \tag{1}$$

To sum up, the number of output channels of DW separable residual network modules used in the backbone network can be controlled. It is composed of 1 CBL and 17 DW separable residual network structures (I-Resblock) with different steps. This design can reduce the overall computational load of the model in this article, thus improving the speed of face detection in the natural environment, and is conducive to rapid real-time face detection.

3.3 Improved attention module

Attention map focusing on visual interpretation is an important field in image recognition. In the unsupervised learning mode, the previous attention model only uses the response value of the convolution layer to extract the weight of the attention mechanism in the feedforward propagation process. The attention mechanism in this article is shown in Figure 3, including the feature extractor, enhanced attention module, and perception module.

The construction of an enhanced attention module and perception module further extracts the deep features of the picture. The enhanced attention module pays attention to important features, strengthens training on important features, and extracts deep features. The enhanced attention module has a $K \times 3 \times 3$ convolution layer, a global average pooling (GAP) layer, and a full connection (FC) layer. In convolutional neural networks, K represents the size of the convolutional kernel. In feedforward processing, the $K \times 1 \times 1$ convolution layer is simulated on the last FC layer of the attention branch. After the $K \times 1 \times 1$ convolution layer, note that the branch uses the response of GAP and softmax functions to output the class probability. Finally, the attention branch generates an attention graph from a feature graph. In order to aggregate K characteristic graphs, these characteristic graphs are convolved by $1 \times 1 \times 1$ convolution.

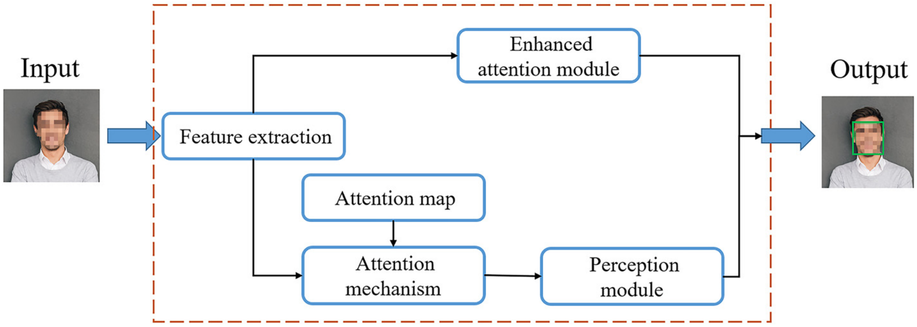


Figure 3: Improved attention mechanism structure chart.

The perception branch structure is the same as that of traditional image classification models such as VGGNet and Res-Net. Attention mapping is applied to feature mapping through the attention mechanism [32]. The specific calculations are as follows:

$$g'(X_i) = M(X_i) \times g(X_i), \tag{2}$$

$$g'(x_i) = (1 + M(X_i)) \times g(X_i), \tag{3}$$

where $g(X_i)$ is the feature map extracted during feature extraction, $M(X_i)$# is an attention map, and $g'(X_i)$ is the output of the attention mechanism.

In general, the parameters in the fully connected layer will be quite large, so directly adding multiple fully connected layers will increase parameter quantity and complexity, making model training slow and prone to over-fitting. Generally, Mirror and Crop are selected for data conversion; that is, large images are cut into small images according to a fixed scale and input into the convolution network. However, this kind of method has the problem of too much parameter calculation and is easy to over-fit. Other methods, such as the dropout method, can reduce the over-fitting phenomenon, but the parameter of this method is the problem of too much calculation and is not easy to implement. Therefore, replace the FC layer in the network with 1 × 1 convolution.

4 Experiment and analysis

4.1 Experimental environment and evaluation index

The designed system involves a lot of calculations during operation, so it has certain requirements for the hardware environment. Table 1 shows the detailed configuration of our system.

Table 1: Experimental configuration

Project	Specific information
Operating system	Windows
Raphics card	GTX 1080Ti
Memory	64GB
Language	Python3.5
Development platform	Pytorch
Tensorflow	2.12
Video camera	720p HD
Solid state drive	1T

To verify the precision of our method, the face detection methods in the studies of Zeng et al. [34] and Yan et al. [35] are selected for comparative experiments. During training, random initialization is used for all convolution layer parameters. The model optimization method uses random gradient descent. The batch size is set to 32, and the weight attenuation is set to 0.0005. The momentum is set to 0.9, and the maximum number of iterations is set to 12×10^4. The first 9×10^4 iterations and learning rate is set to 10^{-3}. The last 3×10^4 iterations and learning rate is set to 10^{-4}.

The method's performance metrics were evaluated using accuracy, recall, and average accuracy (AP) [37]:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \tag{4}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \tag{5}$$

$$\text{AP} = \sum_{k=1}^{n} J(\text{Precision, Recall}). \tag{6}$$

Here, TP denotes the tally of positive samples correctly identified as positive by the classifier, TN represents the count of negative samples correctly identified as negative by the classifier, FP signifies the count of negative samples misclassified as positive by the classifier, and FN signifies the count of positive samples misclassified as negative by the classifier.

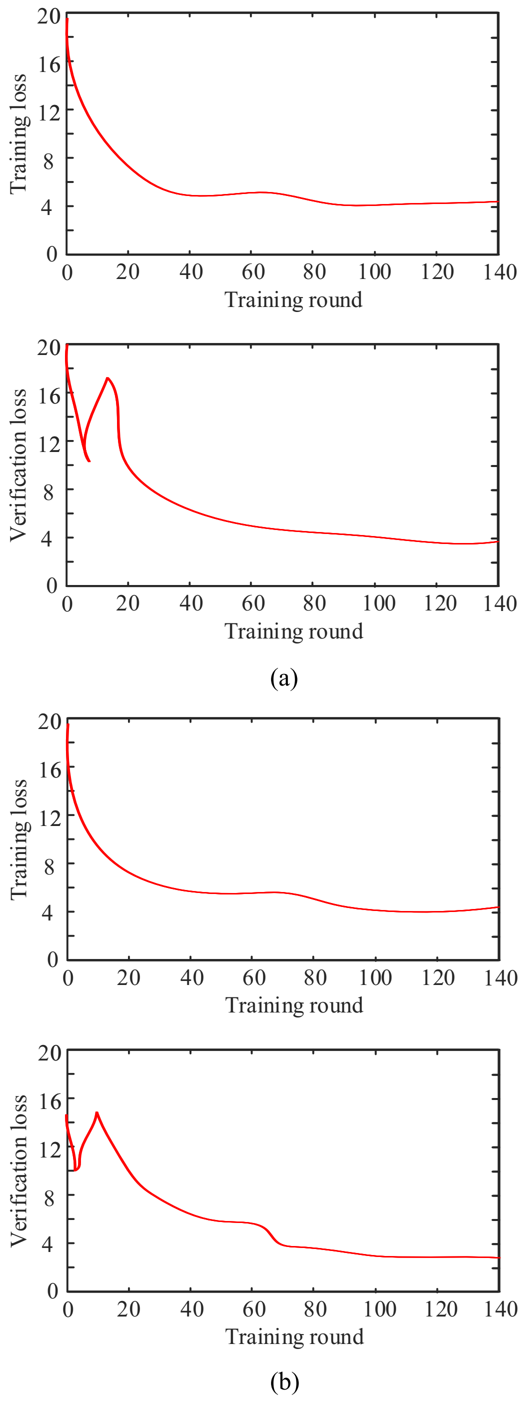
4.2 Training process

Figure 4 shows the training and validation loss curves for different improved networks. The final fluctuation of the training and validation loss curves of each network is small, which indicates that the network stability is good. It can be seen from Figure 4(a) that the original YOLO-v4 algorithm has a small degree of loss reduction and a slow speed during the training process. As can be seen from Figure 4(b), the proposed algorithm has a large degree of loss reduction and a fast speed during the training process. The loss curve of this method no longer decreases at the 80th iteration, and the model convergence is completed. Therefore, the attention mechanism can improve the overall accuracy and convergence rate of our model.

4.3 WiderFace dataset

The WiderFace dataset was first released as a benchmark dataset for human face detection in 2015. It includes 32,203 images and 393,703 labeled faces. Each subset contains three detection difficulty levels: easy, medium, and difficult.

To prove the effectiveness of our method, this section compares the studies of Zeng et al. [34] and Yan et al. [35] with the face detection method. Table 2 shows that average precision values of 0.953, 0.928, and 0.910 for the three difficulty subsets of WiderFace are higher than those of the reference literature. The methods in the studies of Zeng et al. [34] and Yan et al. [35] have weak feature extraction ability and can use limited facial features, resulting in the detection accuracy of only 0.923 and 0.926 in simple subsets. The proposed method improves the deep separable residual network and introduces DW to further reduce the computational load of the model. An improved attention mechanism is used to concatenate advanced features, and advanced features are fused with shallow features to obtain feature vectors that contain more information. Therefore, the proposed method has high detection performance (Figure 5).



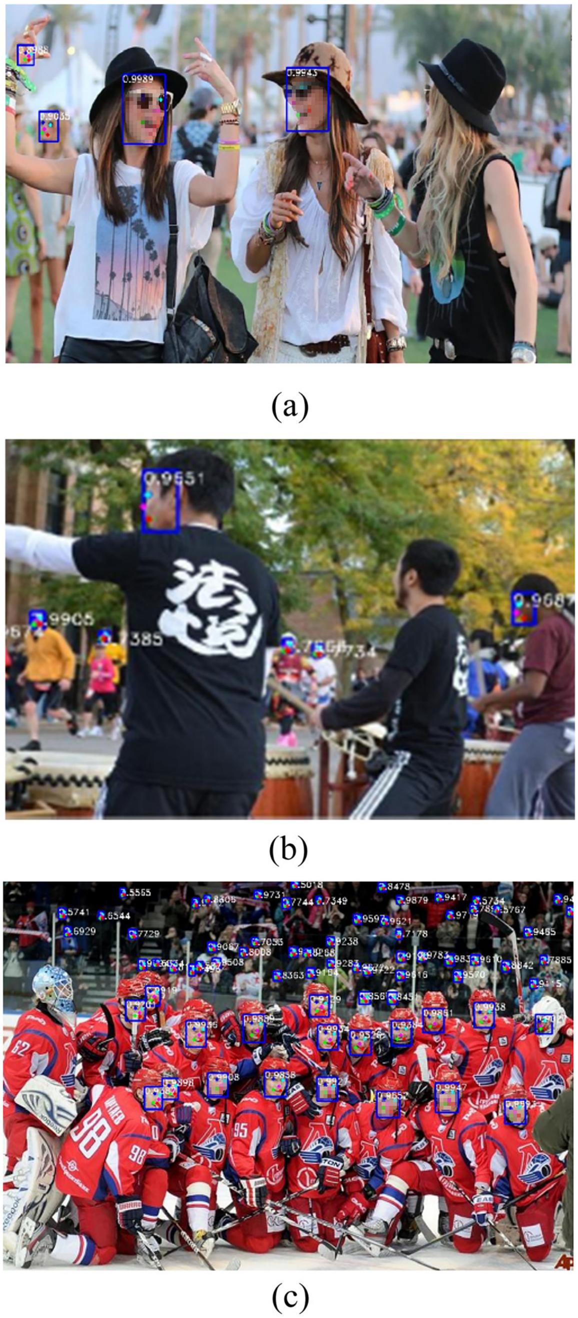
(a)

(b)

Figure 4: Training and validation loss curve: (a) YOLO-v4 and (b) proposed method.

Table 2: Comparison between this method and the comparison method under WiderFace

Model	Easy (AP)	Medium (AP)	Hard (AP)
Zeng et al. [34]	0.923	0.912	0.886
Yan et al. [35]	0.926	0.906	0.872
Proposed method	0.953	0.928	0.910



(a)

(b)

(c)

Figure 5: Test results of different difficulty subsets: (a) easy, (b) medium, and (c) hard.

4.4 MAFA occluded face dataset

The MAFA dataset labeled 35,806 rectangular frames of faces in 30,811 face images containing multiple blocked scenes and mask types. To verify the effectiveness of our method in detecting occluded faces, we conducted experimental comparisons on the MAFA dataset under the same evaluation criteria.

To prove the effectiveness of our method, this section compares the studies of Zeng et al. [34] and Yan et al. [35] with the face detection method. From Table 3, it can be seen that our method has the highest average accuracy on the MAFA dataset. In Table 3, the first five attributes correspond to the five specific directions of the face. As the face deflection angle increases, the detection accuracy in five directions decreases, but our

Table 3: Comparison between this method and the comparison method under MAFA data

Model	Zeng et al. [34]	Yan et al. [35]	Proposed method
Left	0.886	0.912	0.923
Left-front	0.872	0906	0.926
Front	0.910	0.928	0.953
Right-front	0.802	0.821	0.834
Right	0.785	0.876	0.883
Simple	0.615	0.714	0.897
Complex	0.602	0.703	0.881
Body	0.596	0.698	0.864
Hybrid	0.587	0.676	0.856

method achieves the highest detection accuracy. Taking into account the detection results under various attributes, the average precision of this method is 0.891 (Figure 6).

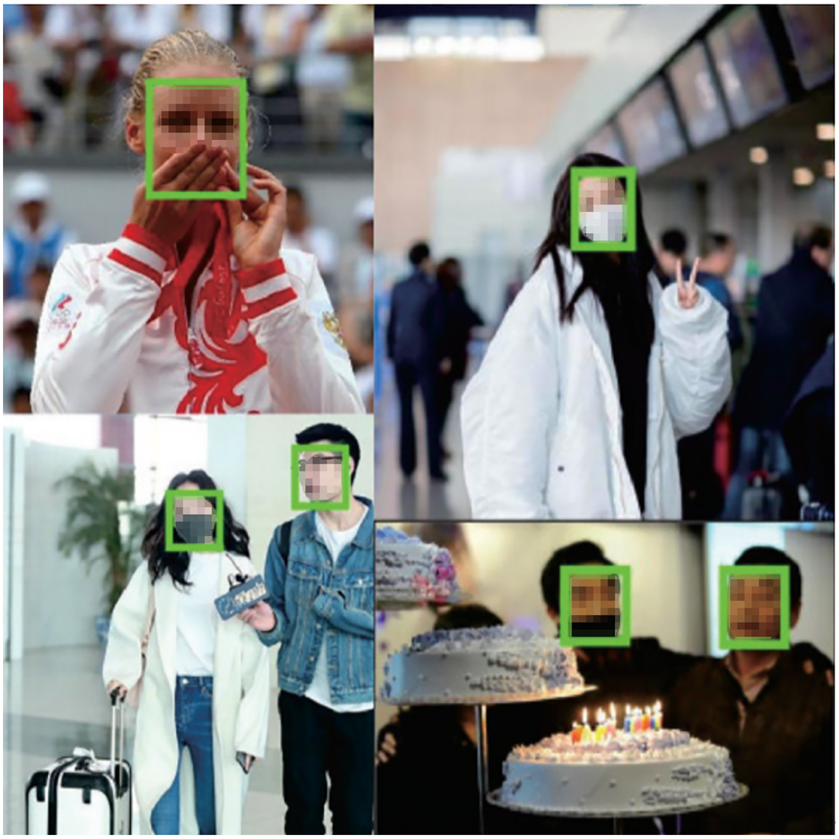


Figure 6: Some test results of the MAFA test set.

To further demonstrate the accuracy and effectiveness of our method, self-comparison experiments were conducted on the MAFA dataset. The experiment takes YOLO-v4 as the baseline method. YOLO-v4 + Idsrn means that the improved DW separable residual network (Idsrn) is added on the basis of YOLO-v4. YOLO-v4 + Idsrn + Am indicates that the improved DW separable residual network and attention mechanism are added on the basis of YOLO-v4. The results of the self-comparison experiment of the above methods in the MAFA test set are described in Table 4. It can be seen that introducing attention networks can significantly improve the detection accuracy of multi-scale occluded faces.

Table 4: Self-comparison experiment results%

Method	AP
YOLO-v4	0.814
YOLO-v4 + Idsrn	0.853
YOLO-v4 + Idsrn + Am	0.902

4.5 Limitation and discussion

To more intuitively verify the effectiveness of our method, visual analysis was conducted using the GradCAM tool. Figure 7 shows the visualization of SeNet and our method. The redder the color, the higher the attention of the model. From Figure 7, it can be seen that the red area range of our method is more concentrated, with darker colors around the eyes and nose, which will pay more attention to the facial area, verifying that our method is more effective in focusing on important facial regions.

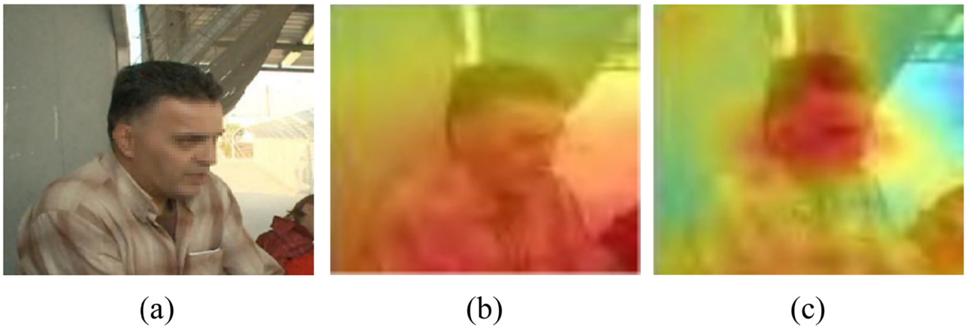


Figure 7: Different attention visualization results: (a) original image, (b) SeNet, and (c) proposed method.

Although the model proposed in this article has improved the detection accuracy, we also noticed some areas that need to be improved. First, the number of parameters in the new model has increased compared to the original model, resulting in an increase in the size of the model, increasing the complexity of the calculation and the need for storage. The next step is to use model compression and model pruning methods to reduce the number of parameters, so as to better balance model size and detection accuracy. Second, there is room for improvement in the detection speed of our model. Compared with other methods, our model does not achieve the fastest detection speed. In order to improve the real-time performance of the model, we will investigate the use of more efficient feature extraction and matching methods or adopt more optimized hardware acceleration techniques. The experimental test in this article only verifies the generalized complex face detection. Although we have achieved some results in the experiment, our next step is to investigate the specific challenges in harsh conditions.

5 Conclusion

This article proposes a face detection method based on improved DW separable residual network and attention mechanism. The improved DW separable residual network is introduced to improve YOLO-v4, and an improved attention mechanism is introduced to obtain a feature vector containing more information. Our model has average accuracy values of 0.953, 0.928, and 0.910 for the three difficulty subsets of WiderFace, all of which are higher than the comparison algorithm. Our method also has the highest average accuracy on the MAFA dataset. Experiments have shown that our method can achieve higher accuracy in face detection and outperform the comparison method in performance.

However, this article does not consider that there may be data imbalance in the dataset, which may also affect the detection accuracy of the face detection algorithm. In addition, this article does not consider the influence of the algorithm on shadow, angle, occlusion, and other factors, which will also affect the detection accuracy. The essence of face detection research at this stage is to first extract features related to faces from images and then classify the features to detect faces. Therefore, whether we can design and summarize a more novel network structure different from classification is worth further study. Our face detection method can also be applied to CPUs and embedded devices. However, due to time and device limitations, relevant testing has not yet been conducted on mobile terminals and embedded devices. The next step can be further studied in embedded devices and mobile terminals.

Funding information: This work was supported by the Planning subject for the 14th five-year plan of Shanxi Education Sciences (No. GH21105 and No. GH-220335) and Educational Reform and Innovation Project of Higher Education in Shanxi Province (No. J20221040).

Author contributions: Yue Qi (Writing - original draft, Project administration, Writing - review & editing, Conceptualization, Resources), Yiqin Wang (Writing - original draft, Project administration, Writing - review & editing, Data curation, Software), Yunyun Dong (Formal Analysis, Supervision).

Conflict of interest: Authors state no conflict of interest.

Data availability statement: The data used to support the findings of this study are included within the article. All image data comes from the public data set http://shuoyang1213.me/WIDERFACE/ and https://github.com/cabani/MaskedFace-Net.

References

[1] Ahonen T, Hadid A, Pietikainen M. Face description with local binary patterns: Application to face recognition. IEEE Trans Pattern Anal Mach Intell. 2006 Oct;28(12):2037–41.
[2] Zhao W, Chellappa R, Phillips PJ, Rosenfeld A. Face recognition: A literature survey. ACM Comput Surv (CSUR). 2003 Dec;35(4):399–458.
[3] Phillips PJ, Moon H, Rizvi SA, Rauss PJ. The FERET evaluation methodology for face-recognition algorithms. IEEE Trans Pattern Anal Mach Intell. 2000 Oct;22(10):1090–104.
[4] Wang Y, Zhang C, Lu J, Bai L, Zhao Z, Han J. Weld reinforcement analysis based on long-term prediction of molten pool image in additive manufacturing. IEEE Access. 2020 Apr;8:69908–18.
[5] Phillips PJ, Wechsler H, Huang J, Rauss PJ. The FERET database and evaluation procedure for face-recognition algorithms. Image Vis Comput. 1998 Apr;16(5):295–306.
[6] Yu H, Yang J. A direct LDA algorithm for high-dimensional data – with application to face recognition. Pattern Recognit. 2001 Oct;34(10):2067–70.
[7] Liu C, Wechsler H. Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition. IEEE Trans Image Process. 2002 Apr;11(4):467–76.
[8] Lawrence S, Giles CL, Tsoi AC, Back AD. Face recognition: A convolutional neural-network approach. IEEE Trans Neural Netw. 1997 Jan;8(1):98–113.
[9] Pankaj P, Bharti PK, Kumar B. A new design of occlusion-invariant face recognition using optimal pattern extraction and CNN with GRU-based architecture. Int J Image Graph. 2023 Jul;23(4):2350029.
[10] Qiu H, Chen X, Liu W, Zhou G, Wang Y, Lai J. A fast ℓ1-solver and its applications to robust face recognition. J Ind Manag Optim (JIMO). 2012;8:163–78.
[11] Morton J, Johnson MH. CONSPEC and CONLERN: a two-process theory of infant face recognition. Psychol Rev. 1991 Apr;98(2):164.
[12] Chien JT, Wu CC. Discriminant waveletfaces and nearest feature classifiers for face recognition. IEEE Trans Pattern Anal Mach Intell. 2002 Dec;24(12):1644–9.
[13] Liao S, Zhu X, Lei Z, Zhang L, Li SZ. Learning multi-scale block local binary patterns for face recognition. In Advances in Biometrics: International Conference, ICB 2007, Seoul, Korea, August 27–29, 2007. Proceedings 2007. Springer Berlin Heidelberg; p. 828–37.
[14] Nelson CA. The development and neural bases of face recognition. Infant Child Dev: An Int J Res Pract. 2001 Mar;10(1–2):3–18.

[15] Deng W, Hu J, Guo J. Extended SRC: Undersampled face recognition via intraclass variant dictionary. IEEE Trans Pattern Anal Mach Intell. 2012 Jan;34(9):1864–70.

[16] Schroff F, Kalenichenko D, Philbin J. Facenet: A unified embedding for face recognition and clustering. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2015. p. 815–23.

[17] Chen LF, Liao HY, Ko MT, Lin JC, Yu GJ. A new LDA-based face recognition system which can solve the small sample size problem. Pattern Recognit. 2000 Oct;33(10):1713–26.

[18] Zhang L, Yang M, Feng X. Sparse representation or collaborative representation: Which helps face recognition? In 2011 International Conference on Computer Vision. IEEE; 2011. p. 471–8.

[19] Nguyen HV, Bai L, Shen L. Local gabor binary pattern whitened pca: A novel approach for face recognition from single image per person. In Advances in Biometrics: Third International Conference, ICB 2009, Alghero, Italy, June 2–5, 2009. Proceedings 3 2009. Springer Berlin Heidelberg; p. 269–78.

[20] Pankaj P, Bharti PK, Kumar B. A new design of occlusion invariant face recognition using optimal pattern extraction and CNN with GRU-based architecture. Int J Inf Secur Priv (IJISP). 2022 Jan;16(1):1–25.

[21] Koley S, Roy H, Dhar S, Bhattacharjee D. Illumination invariant face recognition using fused cross lattice pattern of phase congruency (FCLPPC). Inf Sci. 2022 Jan;584:633–48.

[22] Kasongo SM. A deep learning technique for intrusion detection system using a Recurrent Neural Networks based framework. Comput Commun. 2023 Feb;199:113–25.

[23] Krishnaraj M, Raj RJ. Video frame-based deep learning face detection-a review. In 2021 3rd International Conference on Signal Processing and Communication (ICPSC); 2021. IEEE; p. 207–13.

[24] Castelblanco A, Rivera E, Solano J, Tengana L, Lopez C, Ochoa M. Dynamic face authentication systems: Deep learning verification for camera close-Up and head rotation paradigms. Comput Secur. 2022 Apr;115:102629.

[25] Tayyab M, Marjani M, Jhanjhi NZ, Hashem IA, Usmani RS, Qamar F. A comprehensive review on deep learning algorithms: Security and privacy issues. Comput Secur. 2023;131:103297.

[26] Sathyamoorthy B, Snehalatha U, Rajalakshmi T. Facial emotion detection of thermal and digital images based on machine learning techniques. Biomed Eng: Appl Basis Commun. 2023 Feb;35(1):2250052.

[27] Ge H, Zhu Z, Dai Y, Wang B, Wu X. Facial expression recognition based on deep learning. Comput Methods Prog Biomed. 2022 Mar;215:106621.

[28] Farfade SS, Saberian MJ, Li LJ. Multi-view face detection using deep convolutional neural networks. In Proceedings of the 5th ACM on International Conference on Multimedia Retrieval; 2015. p. 643–50.

[29] Li H, Lin Z, Shen X, Brandt J, Hua G. A convolutional neural network cascade for face detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2015. p. 5325–34.

[30] Jiang H, Learned-Miller E. Face detection with the faster R-CNN. In 2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017); 2017. IEEE; p. 650–7.

[31] Yu J, Jiang Y, Wang Z, Cao Z, Huang T. Unitbox: An advanced object detection network. In Proceedings of the 24th ACM International Conference on Multimedia; 2016. p. 516–20.

[32] Woo S, Park J, Lee JY, Kweon IS. Cbam: Convolutional block attention module. In Proceedings of the European Conference on Computer Vision (ECCV); 2018. p. 3–19.

[33] Peng Y, He X, Zhao J. Object-part attention model for fine-grained image classification. IEEE Trans Image Process. 2017 Nov;27(3):1487–500.

[34] Zeng J, Li J, Feng L. Face recognition based on Improved residual network and channel attention. Autom Control Comput Sci. 2022 Oct;56(5):383–92.

[35] Yan W, Liu T, Liu S, Geng Y, Sun Z. A lightweight face recognition method based on depthwise separable convolution and triplet loss. In 2020 39th Chinese Control Conference (CCC); 2020. IEEE; p. 7570–5.

[36] Chollet F. Xception: Deep learning with depthwise separable convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2017. p. 1251–8.

[37] Powers DM. Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. arxiv preprint arxiv:2010.16061; 2020 Oct.