Journal of Image Processing Theory and Applications (2024)
Clausius Scientific Press, Canada

DOI: 10.23977/jipta.2024.070117
ISSN 2560-6239 Vol. 7 Num. 1

Lightweight Face Anti-spoofing for Improved MobileNetV3

Sun Zhenlin[1,*], Yan Hao[1], Guo Mengyu[1], Hao Zhiqiang[1]

[1]School of Science, Tianjin University of Commerce, Tianjin, 300134, China
[*]Corresponding author: sunzhenlin2022@163.com

Keywords: Face anti-spoofing, Convolutional block attention mechanism, Central difference convolution, Lightweight neural network

Abstract: Aiming at the security challenges associated with facial authentication in biometric technology, we propose an improved MobileNetV3 model that reduces the complexity and cost of existing face anti-deception methods. This model integrates the Convolutional Block Attention Module (CBAM) and Central Difference Convolution (CDC) techniques. CBAM enhances feature representation, while CDC captures fine-grained information by aggregating intensity and gradient data. Experimental results from the NUAA and Replay-Attack datasets indicate that the improved model achieves a recognition accuracy exceeding 97% without a significant increase in computational requirements, highlighting its potential for mobile applications.

1. Introduction

With the rapid development of information technology, biometrics has become one of the key technologies in the field of identity verification and access control. As an intuitive and easy to obtain biometric feature, face has been widely used in biometric systems. However, as technology continues to evolve, traditional static face authentication methods are encountering increasingly severe security challenges. Attack methods such as photo and video replay can easily bypass authentication systems that rely solely on image comparison[1]. In order to distinguish the real face from the fake face to ensure the security and reliability of the authentication process, the face anti-spoofing technology comes into being. This technology can not only improve the security of biometric systems, but also prevent unauthorized access and identity theft, and has important application value in many fields such as finance, security, and mobile payment. With the rise of deep learning technology, face anti-deception technology has made remarkable progress, and the accuracy, authenticity and robustness of detection have been significantly improved, but the complexity of network structure, high computing cost and large volume are problems. It is still an

important challenge for the practical application of this technology. Therefore, research on how to build a lightweight deep network model by optimizing the network structure and computing efficiency, while maintaining high accuracy, reduce the dependence on computing resources, to meet the real-time and resource-limited environment application requirements has become a hot spot of current research.

In the realm of face anti-deception, traditional methods primarily rely on manual feature extraction techniques, such as SIFT, SURF, HOG, and LBP[2-5], as well as methods based on color local binary pattern descriptors[6]. However, traditional methods often struggle to extract meaningful features, resulting in a focus on surface characteristics. In contrast, deep neural networks can extract deeper image features. For instance, relevant literature introduces CNNs[7] into face anti-deception technology for the first time, framing the face anti-deception challenge as a binary classification problem. A framework based on 3D CNNs[8] has been proposed for video datasets related to face anti-deception, enabling the joint capture of spatial and temporal information while optimizing the original data set. In view of the limitations of fully connected layer neural network in true and false face identification[9], CNN was used to extract deep features, and then dimension was reduced by principal component analysis to avoid overfitting. Finally, the support vector machine was used to effectively distinguish true and false faces. There is also a combination of CNN and Recurrent Neural Network[10], assisted by rPPG (remote Photo Plethysmo Graphy). The CNN-RNN model is used to estimate the face depth, and the estimated depth and rPPG are fused together to distinguish the true and false faces. Zhou et al. [11]proposed the method of progressive principal component analysis to compress CNN, which verified the effectiveness of the compressed network on multiple classical network architectures and also verified the remarkable effect of the compressed network in face anti-deception. With the deepening of research, although traditional deep neural networks show significant advantages in feature extraction compared with manual technology, the problems of high network complexity and huge computing requirements also gradually emerge. Therefore, the application of lightweight network in the field of face detection comes into being. For example, literature[12]proposes an improved YOLOv7 face recognition algorithm, which reduces the complexity and computational burden of the model while maintaining the recognition performance by integrating multi-scale information input, global adaptive feature extraction and attention mechanism, transfer learning and polynomial loss strategies. Li et al.[13]proposed an improvement on MoblieNetV3 and used it as the backbone network. They introduced a face anti-spoofing method that combines near-infrared and visible light with MobileNetV3 to form a dual-mode lightweight network. Zhang et al.[14]introduced an ultra-lightweight network architecture (FeatherNet A/B), which optimizes the defects of global average pooling and reduces the number of parameters through the flow module. Finally, it has good performance and generalization on face anti-deception data. Hu et al.[15]proposed a face recognition method based on improved MobileFaceNet to reduce information loss during feature compression by optimizing patterns and verifying the effectiveness of the method. Since CNN emphasizes deep semantic features while ignoring local details of images, Yu et al.[16]proposed Central Difference Convolution (CDC) networks to enhance the ability to describe fine-grained features under varying lighting conditions. This approach has been applied

to multi-modal face anti-deception technology. In a subsequent study, Yu et al.[17]utilized neural structure search to identify a more powerful CDC network architecture, termed Central Difference Convolution Net++(CDCN++). This architecture incorporates a multi-scale attention fusion module to further enhance the performance of face anti-deception systems. Yang et al.[18]introduced an improvement to the CDC framework by proposing directional differential convolution, which was applied to both the CDC network and the CDCN++ network to enhance feature extraction capabilities and detection robustness. Li et al.[19] demonstrated that by replacing standard convolution with CDC on the MobileNetV3 architecture, and integrating the ECA-Net attention mechanism along with the ACON activation function, the application accuracy in face anti-spoofing reached 96.94%. Bu et al.[20]introduced the Convolutional Block Attention Module (CBAM) based on MobileNetV3 to improve feature extraction capabilities; however, it was not fully optimized regarding computational efficiency.

Although the research has made significant strides in the field of face anti-deception, there remains a need to enhance efforts in reducing network complexity and computational costs while maintaining or improving recognition accuracy. To address this issue, this paper introduces CBAM and CDC technologies based on MobileNetV3:

(1) CBAM enables the network to focus more on the key information within an image, thereby enhancing the model's performance. Its design is both simple and effective, improving the representational power of convolutional neural networks by emphasizing features along both the channel and spatial axes.

(2) The primary role of the CDC in face detection is to enhance the ability to differentiate between live and deceptive samples. This is achieved by capturing more detailed information and improving the model's modeling capabilities, all while ensuring robustness against environmental changes.

2. Network Architecture and Optimization

2.1 An Overview of MobileNet

The MobileNet family is a lightweight deep neural network architecture designed for mobile and embedded vision applications. First proposed by Google researchers in 2017, it aims to optimize deep learning performance on resource-constrained devices. The core concept involves using depthwise separable convolution, which breaks the traditional convolution operation into two distinct parts: the first is depthwise convolution, where the convolution operation is performed independently for each input channel; the second is pointwise convolution, which employs a 1x1 convolution kernel. This approach significantly reduces the computational complexity and the number of parameters in the model, making it suitable for mobile devices. Its exceptional generalization ability allows the model to perform reliably in diverse environments and facilitates easy deployment on various edge devices.

MobileNetV1, the inaugural version of the MobileNet series, introduced deep separable convolution and proposed a method known as "network architecture search" to optimize the

network structure[21]. The subsequent MobileNetV2 enhanced the architecture, improving both the accuracy and efficiency of the model by incorporating the reciprocal residual structure and linear bottleneck[22]. MobileNetV3 employs an innovative network architecture search technique that integrates traditional network architecture search with hardware-aware neural architecture search, achieving a superior balance between efficiency and accuracy[23].

2.2 MobileNetV3 Network

Figure 1 illustrates the schematic diagram of the MobileNetV3 network architecture. Compared to other deep neural networks, MobileNetV3 is more streamlined and lightweight. It integrates depthwise separable convolution and Squeeze-and-Excitation (SE) attention mechanisms within some of its building blocks, enhancing overall network performance. MobileNetV3 is available in both large and small versions. The Large version is designed for scenarios with ample computing resources, while the small version is tailored for resource-constrained environments. In this paper, the large version of the MobileNetV3 network is selected as the backbone network.

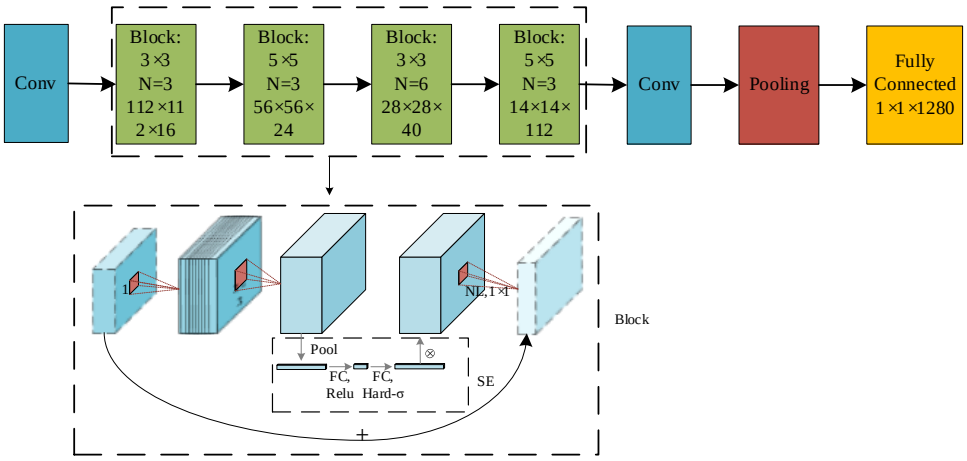

Figure 1: Schematic diagram of Mobilenetv3 network structure.

2.3 Network Improvement

This study integrates CBAM and CDC technologies with MobileNetV3 to develop a deep learning model that is both efficient and accurate.

CDC is a variant of convolutional operations in deep learning that emphasizes capturing locally detailed features in images, particularly high-frequency information. While traditional convolution typically involves a weighted summation of local regions using filters, the core concept of CDC is to extract features by comparing the differences between the central pixel and its surrounding pixels. CDC demonstrates a remarkable ability to describe invariant fine-grained features across various environments, making it particularly suitable for face anti-deception tasks under different lighting conditions. Notably, CDC can replace the conventional convolutional layer in existing neural networks without introducing additional parameters, achieving a plug-and-play effect [16]. Formula (1) illustrates the relationship between CDC and standard convolution.

$$y=(p_0)=\theta\cdot\sum_{p_n\in R}w(p_n)\cdot(x(p_0+p_n)-x(p_0))+(1-\theta)\cdot\sum_{p_n\in R}w(p_n)\cdot x(p_0+p_n) \quad (1)$$

y is the output feature map, x is the input feature map, p_0 is the current position of the input and output feature maps, and p_n is the position of local R. w is the convolution kernel. For the $\theta\in[0,1]$, when the value of θ is higher, it means that the gradient semantic information occupies a significant proportion; Otherwise, the depth-level semantic information occupies a significant proportion.

CBAM is a dual attention mechanism that enhances feature representation through two phases: channel attention and spatial attention. In the context of face anti-deception, CBAM improves the model's ability to recognize subtle differences between real and fake faces. The strength of CBAM lies in its adaptive feature recalibration capability, which dynamically adjusts the importance of each channel and spatial position in the feature map based on task requirements. This adaptability enhances the model's recognition performance. By emphasizing important features, CBAM enables the network to concentrate on the most informative aspects of the input data[24].

The improvement of the MobileNetV3 network in this study is shown in Figure 2. CDC is used instead of the common convolution kernel to enhance the recognition ability of fine-grained features. CBAM is used in each block module to replace the SE attention mechanism. Finally, the channel attention mechanism and spatial attention mechanism of CBAM are analyzed in turn. CBAM provides a more comprehensive feature enhancement approach to further improve model performance.

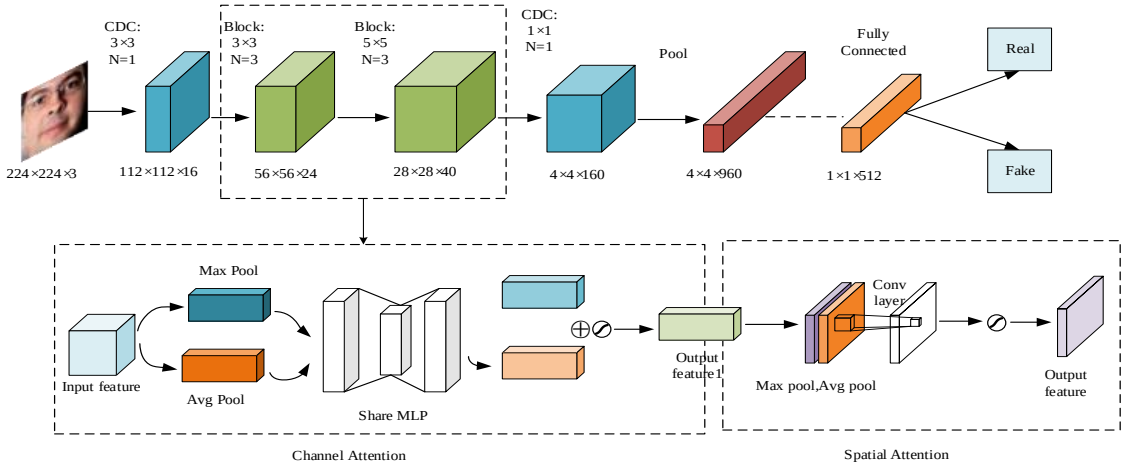

Figure 2: Improved network structure.

3. Experimental Analysis

3.1 Data Sets and Data Set Preprocessing

The datasets selected for this study are the Nanjing University of Aeronautics and Astronautics (NUAA) data set and the Replay-Attack data set. The NUAA data set comprises photographs of 15 individuals' faces captured at a rate of 20 frames per second, specifically sampling each person's frontal facial posture and expressionless state. Everyone contributed 500 photographs with a

resolution of 640 by 480 pixels. The image acquisition for this data set utilized a Canon camera, and the forged face images were created by printing the photographs[25]. The Replay-Attack data set consists of 1,300 videos featuring 50 participants in various lighting environments, with a resolution of 320 by 240 pixels. This data set provides videos of both real and attack scenarios under two different lighting conditions for replay attack detection[26].

The experimental platform utilizes the Windows 11 operating system and features a 12 vCPU Intel(R) Xeon(R) Platinum 8255C CPU. At a clock speed of 2.50 GHz, the GPU is an RTX 2080 Ti, the development environment is PyCharm, and the deep learning framework is TensorFlow 2.3

Replay-attack data set is video data. To facilitate experimental analysis, this paper captures an image every 15 frames. Additionally, the two datasets were divided in a ratio of 3:1:1, as shown in Table 1.

Table 1: Utilizes datasets.

datasets	train	test	verification	total
NUAA	7460	2493	2472	12425
Replay-attack	6724	2671	2631	12026

To remove the redundant information such as background, Open CV is used to detect the face and extract the face image, and then the face key point detection in Dlib library is used to locate 68 key points of the face. According to these key points, all faces are aligned by rotating the operation, and then they are cropped out and uniformly adjusted to the size of 224×224 pixels, which can facilitate subsequent processing work.

3.2 Evaluation Index

The improved network was utilized to conduct experiments on the datasets, primarily using Accuracy (ACC), Loss, Recall, and Area under the Curve (AUC) as evaluation metrics.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{2}$$

$$\mathrm{Re}\,call = \frac{TP}{TP + FN} \tag{3}$$

True Positives (TP): The number of samples that were correctly predicted as positive.

A true negative example (TN): The number of samples that are accurately predicted as belonging to the negative class.

False Positive Cases (FP): The number of samples that are incorrectly predicted to be positive (Type I errors).

False Negative Cases (FN): The number of samples incorrectly predicted to be negative (Type II errors).

3.3 Experimental Result

Table 2 summarizes the performance of the two datasets based on key performance indicators. The analysis reveals that the accuracy of both the NUAA and Replay-attack datasets exceeds 97%, the loss value is reduced to below 0.1, and the recall rate has increased to over 0.96. These indicators collectively demonstrate the model's high efficiency in recognition tasks. Specifically, the AUC value for the NUAA data set is as high as 0.996, which is close to the ideal value of 1, indicating perfect recognition. Meanwhile, the AUC value for the Replay-attack data set remains at a high level, reflecting the proposed model's strong generalization performance across different datasets. Taken together, these results indicate that the model not only offers an effective technical solution for enhancing the security of face recognition systems but also demonstrates its potential and reliability in practical applications.

Table 2: Results under different indicators.

Index	NUAA	Replay-attack
ACC	98% ± 0.42%	97 ± 0.68%
Loss	0.0652	0.0679
Recall	0.9752	0.9663
AUC	0.996	0.9896

3.4 Comparison of Experimental Results

In order to highlight the results of this experiment, we utilized the Replay-attack database to compare various networks and visually present the experimental findings. Table 3 demonstrates that, compared to ResNet18, the proposed algorithm improves accuracy by 17.45% over MobileNetV2 and by 4.65% over MobileNetV3, while significantly reducing loss. The algorithm's Flops are only 450M, which is substantially lower than ResNet18's 7760M. Although the Flops are slightly higher than those of MobileNetV3, the proposed algorithm remains competitive in overall performance. In summary, the model presented in this paper not only achieves a significant improvement in accuracy but also exhibits clear advantages in computational efficiency, providing an effective technical solution for enhancing the security of face anti-deception technology systems.

Table 3: Comparison of Classification Algorithms for Replay Attacks.

Index	RestNet18	MobileNetV2	MobileNetV3	Ours
ACC	80.23%	93.03%	96.74%	97.68%
Loss	0.4241	0.2687	0.1384	0.0679
AUC	0.8803	0.9865	0.9987	0.9896
Flops	7760M	620M	380M	440M

4. Conclusions

To tackle security issues in facial biometrics, we've developed an advanced MobileNetV3 model

that simplifies and reduces the cost of face anti-deception. This model uses CBAM for better feature representation and CDC for detailed information capture. It shows over 97% accuracy on NUAA and Replay-Attack datasets without a big jump in computation, making it suitable for mobile use. Further research is needed to optimize lightweight networks for face anti-deception, and using multiple datasets could boost model robustness and mobile performance, enhancing the security of face recognition systems.

References

[1] Yang X, Luo W, Bao L, et al. Face anti-spoofing: Model matters, so does data[C]//Proceedingsof the IEEE/CVF conference on computervision and pattern recognition. 2019: 3507-3516.

[2] PATEL K, HAN H, JAIN A K. Secure Face Unlock: Spoof Detection on Smartphones [J].IeeeTransactionson Information Forensics and Security,2016, 11(10): 2268-2283.

[3] Boulkenafet Z, Komulainen J, Hadid A. Face antispoofing using speeded-up robust features and fisher vector encoding[J]. IEEE Signal Processing Letters, 2016, 24(2): 141-145.

[4] Komulainen J, Hadid A, Pietikäinen M. Context based face anti-spoofing[C]//2013 IEEESixth International Conference on Biometrics: Theory, Applications and Systems(BTAS). IEEE, 2013.

[5] Boulkenafet Z, Komulainen J, Hadid A. Faceanti-spoofing based on color texture analysis[C]//Proceedings of 2015 IEEE International Conference on Image Processing (ICIP). IEEE, 2015.

[6] de Freitas Pereira T, Anjos A, De Martino J M. LBP-TOP based countermeasure against face spoofing attacks[C]// Computer Vision-ACCV 2012 Workshops: ACCV 2012 International Workshops, Daejeon, Korea, November 5-6, 2012, Revised Selected Papers, Part I 11. Springer Berlin Heidelberg, 2013: 121-132.

[7] Wang Z, Wang Q, Deng W, et al. Learning multi-granularity temporal characteristics forface anti-spoofing[J]. IEEE Transactions on Information Forensics and Security, 2022, 17: 1254-1269.

[8] Gan J, Li S, Zhai Y, et al. 3d convolutional neural network based on face anti-spoofing[C]/2017 2nd international conference on multimedia and image processing (ICMIP). IEEE, 2017: 1-5.

[9] Li L, Feng X, Boulkenafet Z, et al. An original face anti-spoofing approach using partial convolutional neural network[C]/ 2016 sixth international conference on image processing theory, tools and applications (IPTA). IEEE, 2016: 1-6.

[10] Liu Y, Jourabloo A, Liu X. Learning deep models for face anti-spoofing: Binary or auxiliary supervision[C]/ Proceedings of the IEEE conferenceon computer vision and pattern recognition. 2018:389-398.

[11] ZHOU J, QI H B, CHEN Y, et al. Progressive principal component analysis for compressing deep convolutional neural networks [J]. Neurocomputing, 2021, 440: 197-206.

[12] Dai Ying, Ye GUI. Improved YOLOv7 face recognition algorithm for Intelligent elderlycare [J]. Journal of Information Engineering University, 2024, 25 (02):175-180+226.

[13] Li L Z, Gao Z B, Huang L F, Zhang H, LinM J. A dual-modal face anti-spoofing method via light-weight networks[C]//Proceedings of 2019 IEEE 13th International Conference on Anti-counterfeiting, Security, and Identification (ASID). IEEE, 2019: 70-74.

[14] Zhang P, Zou F H, Wu Z W, Dai N L, Mark S, Fu M, Zhao J, Li K. FeatherNets: Convolutional neural networks as light as feather for face anti-spoofing[C]/Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. 2019: 1574-1583.

[15] Hu Jiarong, Meng Wen, ZHAO Jingjing. Face recognition method based on improved MobileFaceNet [J].Semiconductor Optoelectronics, 2022, 43 (01): 164-168.

[16] Yu Z, Qin Y, Li X. Multi-modal face anti-spoofing based on central difference networks[C]/Proceedings ofthe IEEE/CVF Conference on Computer Vision and Pattern RecognitionWorkshops. 2020: 650-651.

[17] Yu Z, Zhao C, Wang Z. Searching central difference convolutional networks for face anti-spoofing[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020:5295-5305.

[18] Yang, Mingye. Research on Face detection in vivo based on Deep learning [D].Qingdao University, 2022.2.
[19] Li Yutong, Lu Wenli, Song Wei, et al. Face detection in vivo based on central difference convolution and frequency domain assistance [J]. Sensors and Microsystems, 2023, 42 (05):117-120+125.
[20] Bu, Chenyu, Shi, Zeyu. Multi-modal face detection in vivo based on lightweight network [J]. Journal of Information Recording Materials, 2023, 24 (12): 1-3+6.
[21] Howard A G. Mobilenets: Efficient convolutional neural networks for mobile vision applications[J]. arXivpreprint arXiv:1704.04861, 2017.
[22] Sandler M, Howard A, Zhu M, et al. Mobilenetv2: Inverted residuals and linear bottlenecks[C]//Proceedings of the IEEE conference on computervision and pattern recognition. 2018: 4510-4520.
[23] Howard A, Sandler M, Chu G, Chen L C,Chen B, Tan M, Wang W, Zhu Y, Pang R, Vasudevan V, Le QV. Searching for mobilenetv3[C]//Proceedings of the IEEE/CVF international conference on computer vision. 2019:1314-1324.
[24] Woo S, Park J, Lee J Y, Kweon I S. Cbam: Convolutional block attention module[C]/Proceedings of the European conference on computer vision (ECCV). 2018: 3-19.
[25] Peixoto B, Michelassi C, Rocha A. Face liveness detection under bad illumination conditions[C]/2011 18th IEEE International Conference on Image Processing. New York: IEEE, 2011:611.
[26] Chingovska I, Anjos A, Marcel S. On the effectiveness of local binary patterns in faceanti-spoofing[C]/2012 BIOSIG-proceedings of the international conference of biometrics special interest group (BIOSIG).2012:1-7.