Finger Vein Recognition Based on Multi-Receptive Field Bilinear Convolutional Neural Network

Kaixuan Wang, Guanghua Chen, and Hongjia Chu

***Abstract*—In order to solve the problems of many model parameters during training, long training time and low recognition accuracy, a finger vein recognition method was proposed that uses multi-receptive field bilinear convolutional neural network (MRF-BCNN). The network can obtain the second-order features of finger veins, and better distinguish finger veins with small differences between classifications. Then the lightweight neural network is used to reduce network parameters and computational complexity, so that the algorithm can be easily implemented in the hardware with limited resources. Finally, a dimensional interactive attention mechanism (DIAM) is designed to enhance the correlation between channels and spaces, and further improve the recognition accuracy. Experimental results show that this method not only improves the accuracy of finger vein recognition, but also reduces the training time and model parameters, so it is suitable for the practical application of finger vein recognition.**

***Index Terms*—Multi-receptive field, lightweight, attention mechanism, depthwise separable convolution, model parameters.**

I. Introduction

With the development of society, people's demand for identity information security is rising. Traditional identity recognition technology has been difficult to meet people's needs, so it is necessary to develop biometric identification technology with higher security [1]. Finger veins are hidden under the skin of the finger. Compared with other biological characteristics, the structure is complex and cannot be obtained under visible light. It has higher stability, concealment and anti-counterfeiting, and has a wider application prospect [2].

Finger vein recognition methods are divided into two categories: traditional recognition methods and deep learning [3]–[7]. Although the traditional recognition method has achieved good results [8], but its recognition process is more complicated, and the extracted features are the shallow features of the image, which are easily affected by the image quality. Convolutional neural networks have strong feature expression capabilities and are not easily affected by image quality. In recent years, with the development of deep learning and the improvement of hardware resources, the use of deep learning for finger vein recognition has become the focus of research [9]–[12]. Li *et al*. proposed a feature encoding method based on the local map structure, which has a good performance in improving the accuracy of finger vein recognition [13]. Ahmad Radzi *et al*. used a four-layer CNN to recognize finger veins. The recognition accuracy reached 100%, but the results were obtained in their own database, which is not universal [14]. Das *et al*. proposed a convolutional neural network composed of five convolutional layers, which achieved good recognition accuracy in four public databases, but did not consider factors such as training time and model size [15]. Hong *et al*. used the VGG network to recognize finger veins and found that as the training samples increase, the recognition accuracy will increase. At the same time the computational complexity and parameter amount of the model will increase [16]. Li *et al*. used improved GCNs to recognize finger veins and obtained better recognition accuracy, but image preprocessing was needed to construct a weighted map of finger veins [17]. He *et al*. enhanced the feature extraction ability of the network by adding convolution layer, which improved the recognition accuracy with less training samples, and also increased the complexity of the network [18].

In response to these problems, the multi-receptive field bilinear convolution neural network (MRFBCNN) finger vein recognition is proposed. First, bilinear pooling is used to extract deep digital vein features in the network. Then, reduce the network parameters and training time by using a lightweight method for the network, which is convenient for the implementation of the algorithm in hardware. Finally, a dimensional interactive attention mechanism is designed to strengthen the correlation between channels and space, and improve the network's ability to extract finger vein features in space and channels.

II. Lightweight Multi-Receptive Field Bilinear Convolutional Neural Network

A. Multi-Receptive Field Bilinear Convolutional Neural Network

Due to the great similarity of finger vein images, finger vein image recognition requires high visual feature expression ability. Meanwhile, the difference between the finger vein image may also be affected by posture, angle, position and other factors, thereby increasing the difficulty of identification. Generally, a single-channel neural network can only capture a single-scale receptive field, and it is very difficult to accurately recognize

Manuscript received March 19, 2021; revised May 7, 2021; accepted May 12, 2021. Date of publication July 7, 2021; date of current version August 17, 2021. The associate editor coordinating the review of this manuscript and approving it for publication was Mr. Ville M. Hautamäki. *(Corresponding author: Guanghua Chen.)*

Kaixuan Wang and Hongjia Chu are with the Microelectronics R&D Center, Shanghai University, Shanghai 200444, China (e-mail: 18361258215@163.com; chu_hongjia@163.com).

Guanghua Chen is with the School of Mechatronic Engineering and Autuomation, Shanghai University, Shanghai 200444, China (e-mail: chghua@shu.edu.cn).

Digital Object Identifier 10.1109/LSP.2021.3094998

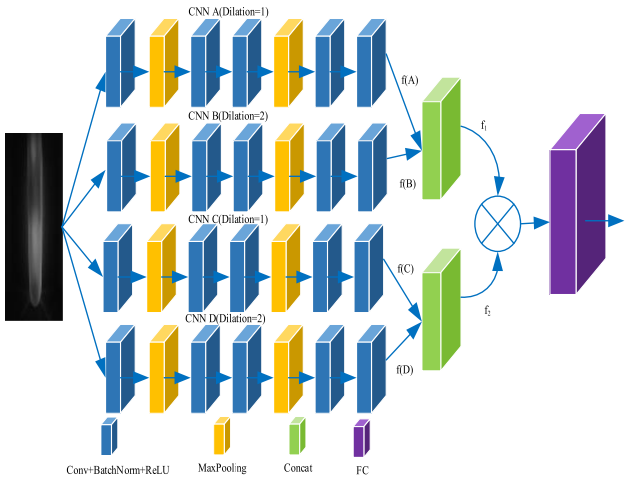


Fig. 1. Multi-receptive field bilinear convolutional neural network.

multi-scale targets such as finger veins. In order to solve these problems, aiming at the feature that the difference between the types of finger veins is too small, this paper improves on the basis of bilinear convolutional neural network [19] and designs MRBCNN. Its structure is shown in Fig. 1.

The network has four identical branches, and the convolution dilated rate of two branches is set to 2 [20]. It can be seen from Fig. 1 that branches A and C, B and D are exactly the same, so in the actual operation of the model, branches C and D can be omitted, and the results obtained by A and B can be used directly. This increases the receptive field on the basis of a small amount of calculation, retains more detailed information, which significantly improves the accuracy of image restoration. Fuse the features extracted from the networks of different receptive fields:

$$f_1 = concat(f(A), f(B)) \tag{1}$$

$$f_2 = concat(f(C), f(D)) \tag{2}$$

where $f_1 \in R^{T\times M}$, $f_2 \in R^{T\times M}$. The merged features are subjected to bilinear pooling to obtain higher-order features. The specific realization formula of multi-receptive field bilinear pooling is:

$$b(I, f_1, f_2) = {f_1}^T(I) f_2(I) \tag{3}$$

$$x = \sum b(I, f_1, f_2) \tag{4}$$

$$y = sign(x)\sqrt{|x|} \tag{5}$$

$$z = y/\|y\|_2 \tag{6}$$

Firstly, the bilinear features of the two path features at the same position I are subjected to outer product. The correlation between the two features is calculated to obtain a second-order feature b. Since the fully connected layer does not accept multi-dimensional tensors, we deal with the second-order feature to get one-dimensional feature tensor x. Then the square root operation and L2 normalization processing are performed on the obtained tensor z to avoid the slow operation speed and over-fitting phenomenon in the fully connected layer. Finally, the processed one-dimensional tensor is sent to the full connective layer for classification and prediction of finger vein images.

The MRFBCNN has achieved excellent results in the task of finger vein recognition. The model completes the recognition of finger vein images through the coordination of four networks. Among them, the second-order feature obtained by performing the outer product of two features by bilinear pooling contains richer detailed information, which is very suitable for classification tasks with great similarity such as finger veins. At the same time, the network also has obvious shortcomings. Bilinear pooling obtains second-order features. Although the feature information is more abundant, but the dimension of the feature vector is very high, the weight parameter of the fully connected layer is increased, and the training time is increased. The following will improve for these shortcomings.

B. Lightweight Network

The MRFBCNN has a large amount of parameters, which is not applicable in some environments with limited computing resources. In practical applications, finger vein recognition algorithms usually need to be deployed in devices with limited resources. Therefore, there are certain requirements for the size and computational complexity of the model. In order to ensure the effectiveness of the algorithm, it is very necessary to carry out a lightweight design for the network. In the convolutional neural network, most of the calculation amount exists in the convolutional layer, and most of the weight parameters exist in the fully connected layer. Therefore, a lightweight idea is proposed, replacing conventional convolution with depthwise separable convolution to reduce the amount of calculation [21]. In the last layer of the neural network, the 1×1 convolution kernel is used to reduce the dimensionality of the features, thereby solving the shortcoming of high vector dimensions caused by bilinear pooling.

If the dimension of the input feature map is $W\times H\times C_I$, the size of the convolution kernel is $D_K\times D_K$, and the number is C_O. For the conventional convolutional layer, the output size of the convolutional layer is $W\times H\times C_O$. At this time, the number of parameters of the convolution kernel is $D_K\times D_K\times C_I\times C_O$, and the calculation amount is $D_K\times D_K\times W\times H\times C_I\times C_O$. The depthwise separable convolution just solves the problem of excessive parameter amount, which is composed of depth convolution and point convolution. For depthwise convolution, the number of parameters of its convolution kernel is $D_K\times D_K\times C_I$, and the amount of calculation is $D_K\times D_K\times C_I\times W\times H$. For pointwise convolution, the number of its parameters is $C_I\times C_O$, and the amount of calculation is $W\times H\times C_I\times C_O$.

Compared with the conventional convolution, the computing resources and the number of parameters of the depthwise separable convolution have been significantly reduced, and the reduction ratio is shown in the following formula. The size of the convolution kernel used in this paper is 3×3, so the number of computing resources and parameters is about 1/9.

$$\frac{D_K \times D_K \times C_I + C_I \times C_O}{D_K \times D_K \times C_I \times C_O} = \frac{1}{C_O} + \frac{1}{{D_K}^2} \tag{7}$$

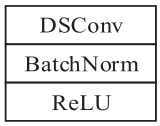


Fig. 2. Lightweight module.

TABLE I
LIGHTWEIGHT NETWORK ARCHITECTURE

Layer number	Layer type	Convolution/Pooling core number	Convolution/Pooling kernel size
1	lightweight module	64	3×3
2	MaxPool	64	2×2
3	lightweight module	128	3×3
4	lightweight module	128	3×3
5	MaxPool	128	2×2
6	lightweight module	256	3×3
7	Conv	128	1×1

$$\frac{D_K \times D_K \times W \times H + C_I \times C_O \times W \times H}{D_K \times D_K \times W \times H \times C_I \times C_O} = \frac{1}{C_O} + \frac{1}{{D_K}^2} \tag{8}$$

After solving the computational complexity, then you need to consider the weight parameters of the fully connected layer. We use a 1×1 convolution kernel to replace the original conventional convolution kernel, which greatly reduces the weight parameters and obtains a more compact network structure. This paper uses depthwise separable convolution to build a lightweight module, and replaces the last layer of convolution with a size of 1×1. The lightweight module is as shown in Fig. 2.

The architecture of the lightweight multi-receptive field bilinear convolutional neural network (L-MRFBCNN) given in Table I.

Using a lightweight network model greatly reduces the amount of parameters and calculations, but depthwise convolution is responsible for filtering, and pointwise convolution is responsible for converting channels, which destroys the correlation between channels and space. When extracting the characteristic information in the channel, the characteristic information between the channels will be lost, which will eventually affect the network performance.

C. Dimensional Interactive Attention Mechanism

The similarities between finger veins are great, classification and recognition of finger vein patterns extracted by the network alone is not enough for accuracy. This requires the vein pattern feature as the main body, and the space and channel features as auxiliary, to jointly complete the classification and prediction of finger veins and improve the recognition accuracy. The DIAM is designed for finger veins, which can not only strengthen the extraction of detailed features on the image space and channel, but also strengthen the relationship between space and channel. The structure is shown in Fig. 3.

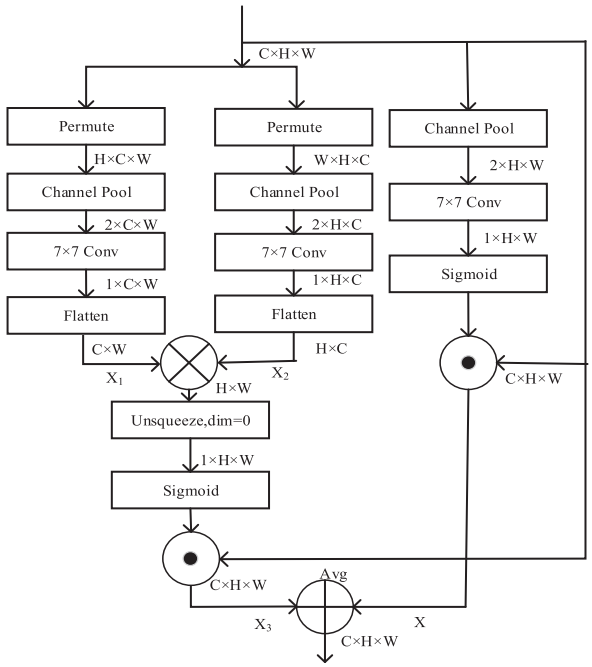


Fig. 3. Dimensional interactive attention mechanism.

The DIAM consists of three parallel branches. Among them, two are responsible for capturing the cross-dimensional interaction of interaction channels C and H, C and W, and the remaining branch is used to construct spatial attention. Finally, the outputs of all three branches are aggregated. The general attention mechanism is to calculate the spatial attention mechanism and the channel attention mechanism separately, so the relationship between the two is not considered. The dimensions of the spatial attention mechanism are interacted, which not only strengthens the feature extraction ability of space and channel, but also captures the dependence between different dimensions. Given an input tensor $X \in R^{C \times H \times W}$, in the first step, rotate the tensor to obtain $X \in R^{C \times H \times W}$, $X_1 \in R^{H \times C \times W}$, $X_2 \in R^{W \times H \times C}$, three Different shapes realize the interaction of dimensions, and then use a channel pool to reduce their shape to $X \in R^{2 \times H \times W}$, $X_1 \in R^{2 \times C \times W}$, $X_2 \in R^{2 \times H \times C}$, channel pool is composed of average pooling and max pooling. Followed by a 7×7 convolution kernel, then flatten X_1 and X_2, and perform outer product on X_1 and ${X_2}^T$ to obtain richer high-order features in the space and channel. Finally, the features obtained by X_3 extension and X are aggregated on an average basis to obtain the features we need:

$$Y = \frac{1}{2}((X_1 \times {X_2}^T) + X) \tag{9}$$

The MRFBCNN is designed for finger vein recognition tasks, which can extract information-rich second-order features of finger veins. Then the network is lightweighted, which greatly reduces the amount of model parameters and computational complexity. Finally, the DIAM is introduced to enhance the dependence of the channel and space, make full use of the information on the channel and the space domain, and improve the accuracy of network prediction and classification.

TABLE II
COMPARISON OF PERFORMANCE INDICATORS OF DIFFERENT MODELS

Method	Accuracy/%		Time/ ms	Parameter	Flops
	FV-USM	SDUMLA			
Hong et al.'s method [16]	95.83	95.77	49	44.0×10^6	32×10^8
AlexNet	92.28	96.32	16	26.7×10^6	1.7×10^8
ResNet-18	96.04	97.79	16	19.7×10^6	4.3×10^8
BCNN	99.39	98.42	15.6	14.7×10^7	31.6×10^8
MRFBCNN	99.80	98.35	12	11×10^6	24×10^8
L-MRFBCNN	99.59	98.90	9	9.1×10^6	2.1×10^8
L-MRFBCNN+DIAM	100	99.82	10	9.1×10^6	2.1×10^8

III. EXPERIMENT AND RESULT ANALYSIS

A. Database and Operating Environment

In order to verify the effectiveness of the method proposed in this paper, experiments were carried out on two public databases. Considering that there may be physical differences between the East and the west, the two databases are FV-USM and SDUMLA [22], [23]. The FV-USM image collection quality is relatively high. 123 volunteers collected six pictures each of the left index finger, left middle finger, right index finger and right middle finger, which were collected twice. Both collection results are used. The quality of SDUMLA was poor, and it was greatly affected by translation and rotation. The index finger, middle finger and ring finger of the left and right hands of 106 volunteers were collected, and each finger was captured six times. The public database was collected in stages from different fingers of different volunteers, the influence of angle, occlusion and other factors were considered to meet the requirements of the experiment. In order to ensure that the training set and test set images do not overlap, the database is randomly assigned as the test set and the training set according to 1:9. In this experiment.

The method of this experiment is implemented under the Pytorch framework. Training and testing are run on NVIDIA GeForce GTX1660 (6GB) GPU. The parameters are set as follows during training: epoch is 200, learning rate is 0.0001, batchsize is 24, and the model is trained using backpropagation algorithm and Adam optimizer.

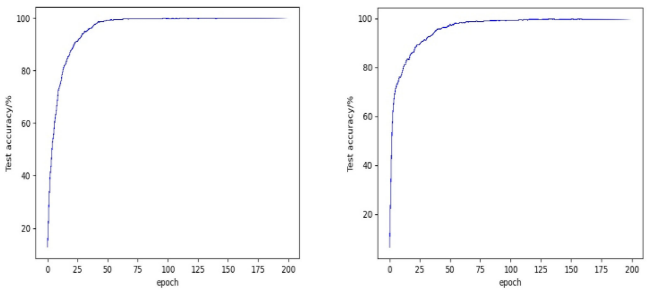
B. Experiment Analysis

The network and DIAM proposed in this paper are respectively tested and compared with the classic network. Set the network parameters to the same and the experimental results are given in Table II.

As can be seen from the table, compared with other networks, the BCNN has obvious advantages. MRFBCNN is improved on the basis of BCNN. It has better performance in test time and parameter amount, but the amount of calculation is still very large. After MRFBCNN has been lightweighted and added to the DIAM module, the amount of calculation and parameters are reduced, and the accuracy is further improved. It can be seen from the experimental results in the above table that each step of the operation is very effective and can achieve the expected effect.

TABLE III
COMPARISON WITH THE RESULTS OF TRADITIONAL RECOGNITION ALGORITHMS

Method	Accuracy/%
LBP	96.10
LLBP	97.78
Yang et al.'s method [2]	99.52
Li et al.'s method [13]	99.87
Our method	100



(a) (b)

Fig. 4. Test accuracy. (a)FV-USM;(b)SDUMLA.

Experiment with the method proposed in this paper and the traditional finger vein recognition algorithm in FV-USM, the results are given in Table III.

Compared with some classic traditional finger-vein recognition algorithms, the method proposed in this paper is almost the same in recognition accuracy. However, our method does not need to preprocess the finger vein image when recognizing finger veins, which greatly reduces the time cost compared with traditional recognition methods. Judging from the experimental data, the ideas and methods in this paper are completely correct and effective. The accuracy curve of the public database is shown in Fig. 4.

IV. CONCLUSION

According to the characteristics of finger vein, the MRF-BCNN is designed in this paper, which has obvious advantages compared with the classic network model. In the model, we used a lightweight method to optimize the model, reducing the amount of parameters and calculations, making the algorithm more convenient for hardware implementation, but the lightweight method destroys the correlation between space and channels. In order to make up for the deficiencies caused by lightweight and improve the network's ability to extract detailed features, the DIAM is designed and introduced into the network to further improve the recognition accuracy. Experimental results show that the accuracy of the proposed method is up to 7.72% higher than that of the classical network. After lightweight processing, the parameter amount is reduced to 1/10 of the original. In the future, we will study the hardware implementation of the algorithm.

REFERENCES

[1] Y. Vasquez, C. Beltran, M. Gomez, M. Florez, and J. L. Vazquez-Gonzalez, "Features extraction in images on finger veins with hybrid curves," in *Proc. IEEE Mex. Humanitarian Technol. Conf.*, Puebla, 2017, pp. 34–38.

[2] J. F. Yang, Y. H. Shi, and G. M. Jia. "Finger-vein image matching based on adaptive curve transformation," *Pattern Recognit.*, vol. 66, pp. 34–43, 2017.

[3] H. Liu, L. Song, G. Yang, L. Yang, and Y. Yin, "Customized local line binary pattern method for finger vein recognition," in *Proc. Chin. Conf. Biometric Recognit.*, 2017, pp. 314–323.

[4] D. Zhou, J. Cao, Y. H. Jiang, Y. C.Feng, and Q. He, "Speckle design method based on principal component analysis," *Laser Optoelectron. Prog.*, vol. 57, pp. 248–253, 2020.

[5] S. Y. Li, B. Zhang, S. P. Zhao, and J. F. Yang, "Local discriminant coding based convolutional feature representation for multimodal finger recognition," *Inf. Sci.*, vol. 547, pp. 1170–1181, 2021.

[6] C. B. Yu and H. F. Qin, *Biometric Identification Technology: Finger Vein Identification Technology*. Beijing, China: Tsinghua Univ. Press, 2009, pp. 81–87.

[7] X. H. Zhao, L. F. Yin, Y. N. Zhu, P. Wang, and X. K. Wang, "Improved image classificational algorithm based on principal component analysis network.," *Laser Optoelectron. Prog.*,vol. 56, pp. 73–82, 2019.

[8] S. Li, B. Zhang, L. Fei, and S. Zhao, "Joint discriminative feature learning for multimodal finger recognition," *Pattern Recognit.*, vol. 111, 2021, Art. no. 107704.

[9] Z. Y. Tao, Y. L. Hu, and S. Lin, "Finger vein recognition based on improved AlexNet," *Laser Optoelectron. Prog.*, vol. 57. pp. 58–66, 2020.

[10] X. L. Zhao, M. Z. Huang, and Y. C. Fu, "Research on recognition method of dorsal hand vein with liveness detection function," *Acta Optica Sinica*, vol. 41, pp. 1–13, 2021.

[11] L. Chen, J. Wang, S. Yang, and H. He, "A finger vein image-based personal identification system with self-adaptive illuminance control," *IEEE Trans. Instrum. Meas.*, vol. 66, no. 2, pp. 294–304, Feb. 2017.

[12] C. Xie and A. Kumar, "Finger vein identification using convolutional neural network and supervised discrete hashing," *Pattern Recognit. Lett.*, vol. 119, pp. 148–156, 2019.

[13] S. Y. Li, H. G. Zhang, G. M. Jia, and J. F. Yang, "Finger vein recognition based on weighted graph structural feature encoding," in *Proc. Chin. Conf. Biometric Recognit.*, 2018, pp. 29–37.

[14] S. A. Radzi, M. Khalil-Hani, and R. Bakhteri, "Finger-vein biometric identification using convolutional neural network," *Turkish J. Elect. Eng. Comput. Sci.*, vol. 24, pp. 1863–1878, 2016.

[15] R. Das, E. Piciucco, E. Maiorana, and P. Campisi, "Convolutional neural network for finger-vein-based biometric identification," *IEEE Trans. Inf. Forensics Secur.*, vol. 14, no. 2, pp. 360–373, Feb. 2019.

[16] H. G. Hong, M. B. Lee, and K. R. Park, "Convolutional neural network-based finger-vein recognition using NIR image sensors," *Sensors*, vol. 17, pp. 1–21, 2017.

[17] R. Li, Z. G. Su, H. G. Zhang, and J. F. Yang, "Application of improved GCNs in feature representation of finger-vein," *J. Signal Process.*, vol. 36, pp. 550–561, 2020.

[18] X. He and X. Chen, "Finger vein recognition based on improved convolution neural network," *Comput. Eng. Design*, vol. 40, pp. 562–566, 2019.

[19] T. Lin, A. Roy Chowdhury, and S. Maji, "Bilinear CNN models for fine-grained visual recognition," in *Proc. IEEE Int. Conf. Comput. Vision*, 2015, pp. 1449–1457.

[20] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions." Accessed: Apr. 2016. [Online] Available: https://arxiv.org/pdf/1511.07122.pdf

[21] F. Mamalet and C. Garcia. "Simplifying convnets for fast learning," in *Proc. Int. Conf. Artif. Neural Netw.*, 2020, pp. 58–65.

[22] S. M. Asaarim, S. A. Suandi, and B. A. Rosdi, "Fusion of band limited phase only correlation and width centroid contour distance for finger based biometrics," *Expert Syst. Appl.*, vol. 41, pp. 3367–3382, 2014.

[23] Y. Yin, L. Liu, and X. Sun, "SDUMLA-HMT: Amultimodal biometric database," in *Proc. Chin. Conf. Biometric Recognit.*, 2011, pp. 260–268.