



OPEN

Finger vein recognition based on bilinear fusion of multiscale features

Bin Ma¹, Kaixuan Wang² & Yueli Hu¹✉

Finger veins are widely used in various fields due to their high safety. Existing finger vein recognition methods have some shortcomings, such as low recognition accuracy and large model size. To address these shortcomings, a multi-scale feature bilinear fusion network (MSFBF-Net) was designed. First, the network model extracts the global features and local detail features of the finger veins and performs linear fusion to obtain second-order features with richer information. Then, the mixed depthwise separable convolution replaces the ordinary convolution, which greatly reduces the computational complexity of the network model. Finally, a multiple attention mechanism (MAM) suitable for finger veins was designed, which can simultaneously extract the channel, spatial, directional, and positional information. The experimental results show that the method is very effective, and the accuracy of the two public finger vein databases is 99.90% and 99.82%, respectively.

With the development of science and technology and the continuous improvement of safety regulations, more and more people pay attention to information security. Biometric identification technology can guarantee people's information security and gradually replace the status of traditional identity identification technology in daily life. Compared with other biometric identification technologies, finger veins have the advantages of high security, privacy, non-contact identification, and low cost due to their hidden under the skin surface. Therefore, finger vein recognition technology has better application prospects and has become a research hotspot in the field of biometric recognition.

Medical research shows that everyone's finger veins are unique, and even twins can't have exactly the same vein texture characteristics¹. With the increase of age, the pattern of the finger vein will not change. Even if the fingers are worn, the veins of the fingers are not affected. Finger veins can only be collected on a living body using specialized equipment. In recent years, a large number of researchers have carried out research on finger vein recognition technology and have made good progress. Finger vein recognition technology can be divided into two categories: traditional methods of finger vein recognition and deep learning-based finger vein recognition methods²⁻⁴. The traditional finger vein recognition method preprocesses the finger vein images and then uses traditional algorithms such as PCA, LBP and LDP to extract and classify the texture features of the finger vein images^{5,6}. Hu et al. improved the maximum curvature method and proposed that the curvature value of the best direction not only extracted the vein network, but also mined the vein trunk, and verified the effectiveness of the method on two public databases⁷. Yang et al. Proposed a block-based (2D) 2PCA method based on LBP and PCA, which preserves the local information of vein images. The experiment was carried out on FV-USM, and the recognition accuracy was 99.32%⁸. Meng et al. proposed a region-based detail recognition technology to solve the problem of finger vein image distortion. Combining the method based on the region of interest with the method of detail matching, the experiment was carried out on SDUMLA, and excellent results were achieved⁹. Although the traditional methods have a very good performance in the task of finger vein recognition, the recognition results are very dependent on the quality of the image. Before recognition, the finger vein image needs to be preprocessed, and the recognition steps are tedious. With the development of deep learning, more and more neural networks are applied to finger vein recognition tasks. The recognition method based on deep learning has strong feature expression ability. Compared with the traditional finger vein recognition algorithm, it can obtain better results under the same image quality without preprocessing. Kamaruddin et al. proposed a new PCANet to extract semantic information from finger vein images, which has better performance compared with other state-of-the-art methods. This method needs to consider both the original grayscale image and the extracted vein line image¹⁰. Ren et al. proposed a novel lightweight convolutional neural network that can filter out low-quality images and mine different attributes hidden in low-quality images to improve network performance.

¹School of Mechatronic Engineering and Automation, Shanghai University, Shanghai 200444, China. ²Microelectronics R&D Center, Shanghai University, Shanghai 200444, China. ✉email: huyueli@shu.edu.cn

The method can be combined with existing databases and recognition algorithms, which guarantees excellent performance of the algorithm, but the recognition time is too long, the recognition time of a single image is 9.2ms¹¹. Zhang et al. proposed a joint Bayesian framework based on partial least squares discriminant analysis, which can effectively extract the texture and orientation features of finger veins. Experiments on public databases obtained high recognition accuracy and fast recognition speed. However, when converting the original features into a low-dimensional form, there may be a loss of features¹². Noh et al. proposed a deep CNN method for finger vein recognition, which fused the texture features and shape features of finger veins, effectively reducing the impact of image noise on the recognition results. This method requires two input images which results in excessive processing time¹³. Hong et al. proposed a novel finger vein recognition network based on VGG16 and achieved excellent results. But the input to the network model needs to use the augmented image, the original unprocessed image, and the difference image obtained from the two finger vein images¹⁴. Yang et al. proposed the FV-GAN network model based on a generative adversarial network, which is the first time to apply a generative adversarial network to the field of finger veins and has strong robustness against outliers and blood vessels ruptures. But requires a large number of training samples¹⁵. Tang et al. proposed a pre-trained weighted convolutional neural network based on Siamese, which effectively improved the network performance. But the extraction of ROI is not considered¹⁶. Ren et al. proposed to apply the encryption algorithm to the finger vein recognition task, which greatly improves the security, but due to the need for encryption, it leads to additional time overhead³¹. It can be seen from the above analysis that the existing identification methods have achieved good results in recognition performance, but there are still some shortcomings in recognition time and computational complexity. The method in this paper comprehensively considers the four aspects of recognition accuracy, model parameters, computational complexity, and recognition time, and obtains good performance.

The methods based on deep learning have achieved better results without image preprocessing. However, the quality of the image is still an important factor affecting the performance of the network model. When the image quality is too poor or the database samples are too few, it is difficult for the network to extract the effective features of finger veins, so there is a disadvantage that the recognition accuracy is not high enough. Convolutional neural networks also have shortcomings such as many model parameters, and long training and recognition time. To solve these problems, we propose a multi-scale feature bilinear fusion network combined with the multiple attention mechanism. Firstly, a multi-scale feature bilinear fusion network is designed. The shallow structure of the network model extracts the detailed features of finger veins, and the deep structure extracts the overall contour of finger veins. Then the shallow features and deep features are bilinear fused to obtain the second-order features with richer information. Secondly, the lightweight method based on mixed depthwise separable convolution is adopted to reduce the parameters and computational complexity of the network, and speed up network model training and recognition. Finally, to further improve the recognition accuracy, a multiple attention mechanism module is designed and embedded into the network model to simultaneously extract the space, channel, direction, and position information of the finger vein images.

Proposed method

In this section, a novel end-to-end finger vein recognition network framework is proposed to extract finger vein features.

Multi-scale feature bilinear fusion network. Finger veins are composed of vein patterns and have relatively simple features, which results in a certain similarity between different finger veins. Recognition only by extracting vein pattern features will leads to recognition errors. To solve these problems, a multi-scale feature bilinear fusion network was designed. The network structure is shown in Fig. 1.

Take the Conv0 layer as an example, where 3×3 represents the size of the convolution kernel, and 64 represents the number of convolution kernels. Firstly, the shallow features of the finger vein image are extracted by two convolution layers. The shallow features include the grain details of the finger vein, and the shallow features are divided into two paths, one of which is reserved for classification. The other way continues to use two convolution layers to extract the deep features including outline of the finger vein from the shallow features. The retained shallow features are dimensionally increased by using a 1×1 convolution kernel, and bilinear fusion is performed with deep features to obtain second-order features containing rich information. The formula is as follows:

$$b(p, I, f_1, f_2) = f_1^T(p, i) f_2(p, i) \quad (1)$$

where $f_1(p, i) \in \mathbb{R}^{T \times M}$ and $f_2(p, i) \in \mathbb{R}^{T \times M}$ is the deep feature and shallow feature of image i at position p . The deep feature and shallow feature are bilinearly fused to obtain matrix b , which contains the correlation of these two features.

$$z = \text{vec} \left(\sum_p b(p, i, f_1, f_2) \right) \quad (2)$$

Equation (2) uses sum pooling to aggregate bilinear features on b of all positions in image i , and the feature dimension is $M \times M$. In the convolutional neural network, the fully connected layer is responsible for classifying the final features, while the fully connected layer can only accept one-dimensional feature vectors, so the features need to be flattened to obtain $z \in \mathbb{R}^{(M \times M) \times 1}$.

$$x = \text{sign}(z) \sqrt{|z|} \quad (3)$$

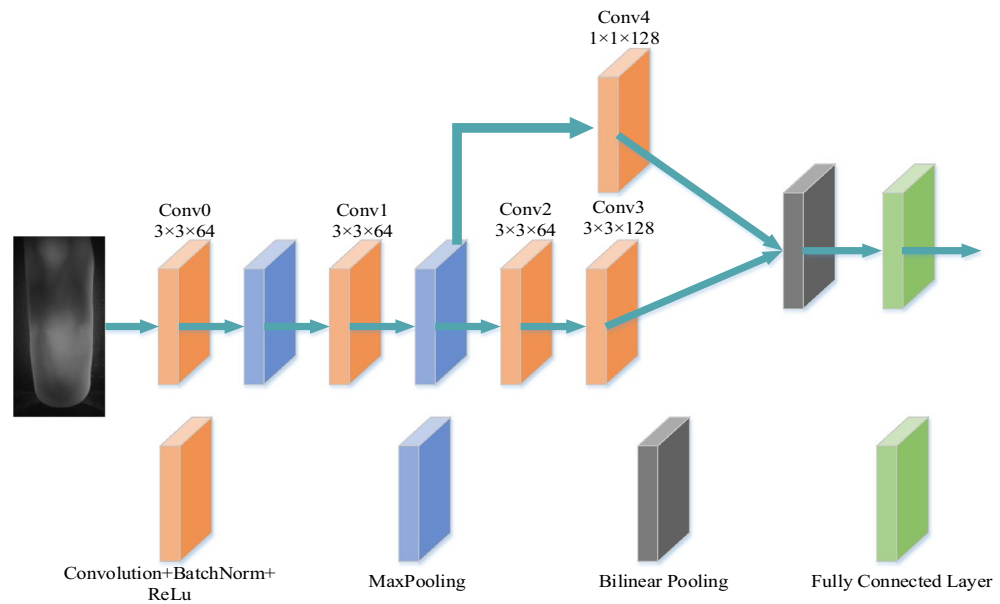


Figure 1. Multi-scale feature bilinear fusion network structure.

$$y = \frac{x}{\|x\|_2} \quad (4)$$

Equation (3) normalizes the obtained z vector. Equation (4) is to perform L2 normalization on the vector x , where $\|x\|_2$ represents the norm of the vector x . The signed square root and L2 normalization operations are performed on the obtained one-dimensional features, which can improve the operation speed of the fully connected layer, avoid overfitting, and increase the network performance. Finally, the fully connected layer processes the feature y to complete the classification of finger vein images.

Since the finger vein features are relatively simple, the network model is not very deep, resulting in high efficiency. The multi-scale feature bilinear fusion network is divided into two branches. One branch locates the finger veins and extracts the shallow features (local features) of the finger veins. The other branch continues to process the shallow features and extracts the deep features (vein contours). The two-way features are bilinear fused to obtain second-order features, which contain more useful information. The global and local features of finger vein are fully mined, which makes the prediction and classification of finger vein recognition task more accurate.

Network lightweight. The convolutional neural network has a large number of parameters and computations, which ensures the powerful visual feature expression ability of the network model, but pays the cost of parameter redundancy, which will reduce the training and recognition speed. Because the finger vein recognition algorithm must operate in real-time on hardware with constrained resources, there are somewhat strict limitations on the size of the network model and the number of model parameters. It is required to create a lightweight network model in order to condense the model's size and satisfy the neural network's low-latency operating criteria. We adopted a shallower bilinear fusion network with deep multiscale features, allowing for a smaller number of parameters in the fully connected layer. The lightweight method in this paper is to directly optimize the convolutional layer and replace the conventional convolution with a mixed depthwise separable convolution to reduce the computational complexity of the network model.

The depthwise separable convolution decomposes the convolution process and divides the ordinary convolution into depthwise convolution and pointwise convolution, which greatly reduces the number of parameters and computations. Generally, the depthwise separable convolution uses a 3×3 convolution kernel, which is not suitable for finger vein recognition tasks. Due to the particularity of the structure of finger veins, the combination of large receptive field features and small receptive field features can obtain better results. The mixed depthwise separable convolution divides the feature map channels into groups, and each group uses depthwise separable convolutions with different kernel sizes, which can extract receptive field features of different sizes, so it is more suitable for finger vein recognition tasks.

Figure 2 shows the convolution process of the mixed depthwise separable convolution. As can be seen from the figure, the mixed depthwise separable convolution receives input tensors, groups the channels of the tensors, and each group uses convolution kernels of different sizes. The features of different receptive fields can be extracted in the same convolution layer, and then different groups of features are combined together. The depthwise separable convolution is divided into two steps, which just separate feature extraction and feature combination. The tensors output in the feature extraction stage will be discharged according to the size of the receptive field, and the output tensors will be mixed in the feature combination stage, so there is no need to worry about the transmission of different receptive fields characteristics. The multi-scale feature bilinear network

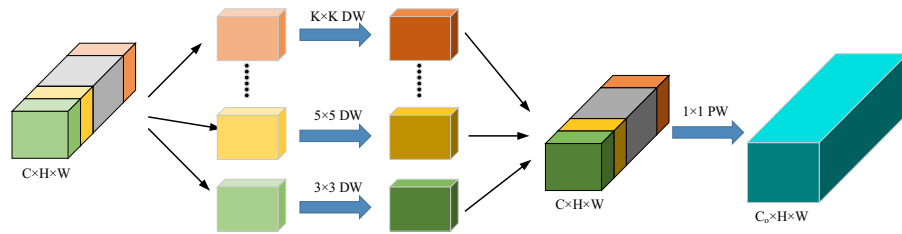


Figure 2. Mixed depthwise separable convolution.

extracts the shallow features and deep features of the finger vein images. Therefore, the lightweight idea is to use a single set of 3×3 mixed depth separable convolutions in the shallow network to extract the contour features of the finger veins, and use 3×3 and 5×5 mixed depthwise separable convolutions in the deep network to extract the detailed features of different receptive fields of finger vein. To further reduce the computational complexity of the network model, the 5×5 convolution kernel is replaced with a 3×3 (dilation = 2) dilated convolution. The network structure is shown in Table 1.

Compared with ordinary convolution, the mixed depthwise separable convolution has significantly reduced the computational cost and the number of parameters, resulting in a better performance in terms of lightweight. However, depthwise separable convolution also has some shortcomings that cannot be ignored. The depthwise convolution performs independent convolution on each channel of the input tensor, ignoring the spatial location information. Pointwise convolution performs a weighted combination of input tensors in depth, which does not make efficient use of channel information. The combination of them separates the channel and space of the convolution, which affects the transfer of feature information. In response to this problem, a multiple attention mechanism module suitable for finger veins is designed, which will be described in detail in the next section.

Multiple attention mechanism. For the finger vein recognition task, a network that can extract key features is very important. However, only relying on the basic network to extract finger vein features for prediction and classification, the results often fail to meet their expectations. Firstly, finger vein recognition demands high quality images, while collecting clear and noiseless high-quality finger vein image is always challenging. Secondly, A convolutional neural network will not locate the region of interest of the finger vein map like the human brain but process the whole image, which will extract many useless features. The process of feature classification is a waste of time and will be misled by these useless features, leading to classification errors. To overcome above two issues, The attention mechanism is adopted to save the energy wasted on useless information, and invest more in the key area in order to obtain detailed information in the area. Existing attention mechanisms are roughly divided into spatial attention mechanisms and channel attention mechanisms. If you want to use these two modules at the same time, it is also realized through a series connection, and the information may be lost in the transmission between modules. In view of the characteristics of finger veins and the shortcomings caused by lightweight processing, a multiple receptive field module is designed. The structure is shown in Fig. 3.

The multiple attention mechanism receives an input tensor $X \in R^{C \times H \times W}$ and outputs a tensor $Y \in R^{C \times H \times W}$ with feature enhancement capability. The multiple attention mechanism module is divided into multiple branches for parallel processing. First, the input tensor enters the module, and the first two branches are pooled by $(H, 1)$ and $(1, W)$ through X Avg Pool and Y Avg Pool respectively. Each channel encoding is checked so that features are aggregated along two spatial directions, capturing spatial interactions with location information. The output of channel c with height h and width w can be obtained as follows:

$$s_c^h = \frac{1}{W} \sum_{1 \leq i \leq H} X \tag{5}$$

| Layer number | Layer name | Kernel number | Kernel size |
|--------------|---------------------------------------|---------------|---------------------|
| Conv0 | Mixed depthwise separable convolution | 64 | [3] |
| Max0 | Max pooling | 64 | 2 |
| Conv1 | Mixed depthwise separable convolution | 64 | [3] |
| Max1 | Max pooling | 64 | 2 |
| Conv2 | Mixed depthwise separable convolution | 64 | [3,3(dilation = 2)] |
| Conv3 | Mixed depthwise separable convolution | 128 | [3,3(dilation = 2)] |
| Conv4 | Mixed depthwise separable convolution | 128 | [1] |

Table 1. Multi-scale feature bilinear fusion network lightweight structure.

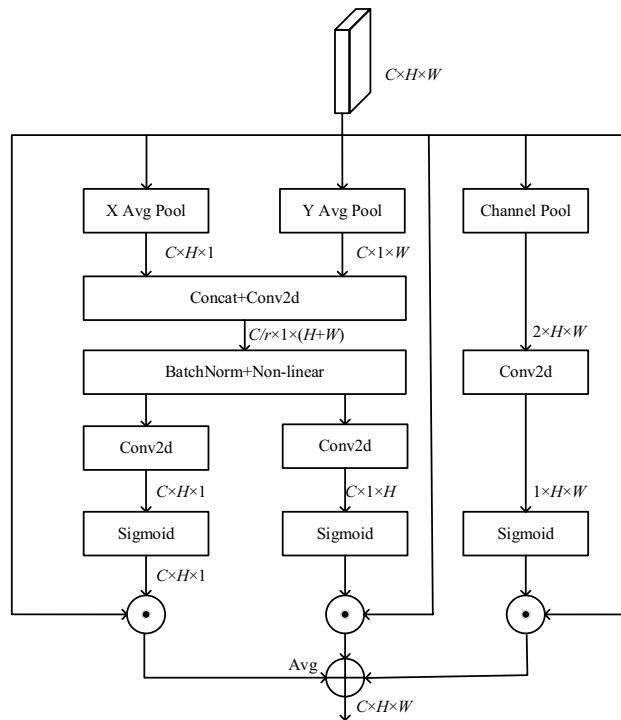


Figure 3. Multiple attention mechanism.

$$s_c^w = \frac{1}{H} \sum_{1 \leq i \leq W} X \tag{6}$$

After helping the network model to find the region of interest of the finger vein, the two-way features are cascaded, and then the 1×1 size convolution is used to reduce the dimension of the features, which reduces the computational complexity. The formula is as follows:

$$f = \delta(F_1([s^h, s^w])) \tag{7}$$

where F_1 represents a 1×1 size convolution operation, and δ represents non-linear activation function. Finally, f is divided into two independent tensors along the spatial dimension, and the sigmoid function is used to limit the features to 0–1 after using 1×1 convolution to increase the dimension. The final result is used as a weight parameter to enhance important features and weaken unimportant feature information:

$$g^h = \sigma(F_1(f^h)) \tag{8}$$

$$g^w = \sigma(F_1(f^w)) \tag{9}$$

The first two branches complete the horizontal and vertical directions of channel attention and then introduce the spatial attention branch. After the input feature map is pooled, the channel dimension becomes 1, and the spatial information is aggregated to obtain average pooling feature $F_{avg} \in \mathbb{R}^{1 \times H \times W}$ and max pooling features $F_{max} \in \mathbb{R}^{1 \times H \times W}$ respectively. The two feature maps are then concatenated together, and a conventional convolution is used to generate an attention image with spatial information. The calculation method is as follows:

$$g = \sigma(F_7([AvgPool(X); MaxPool(X)])) = \sigma(F_7([f_{avg}^S; f_{max}^S])) \tag{10}$$

where σ represents the sigmoid function and F_7 represents a 7×7 size convolution operation. The sigmoid function can assign corresponding weights according to the different degrees of information contribution at different positions. Equation (11) multiplies and sums multiple input features and the weights obtained by the corresponding attention mechanism, and then calculates the average value to obtain the final feature

$$Y = \frac{(X \times g^h + X \times g^w + X \times g)}{3} \tag{11}$$

The multiple attention mechanism embeds the position information into the channel attention and combines with the spatial attention mechanism, which not only extracts the channel domain and spatial domain information, but also captures the directional and positional information of the finger veins, and performs attention operations in large areas. Improved network performance at a lower computational cost. The multiple attention mechanism assigns weights to different information on the finger vein image according to the degree of contribution, which is very suitable for finger vein recognition tasks and makes up for the deficiencies caused by depthwise separable convolution. The combination of a multi-scale feature bilinear fusion network with mixed depthwise separable convolution and multiple attention mechanism is the final model proposed for the finger vein recognition task in this paper. The network structure is shown in Fig. 4.

Experiments and analysis

Finger vein database. To evaluate the performance of the proposed network model from multiple aspects, we conduct experiments on two public finger vein databases, SDUMLA¹⁷ and FV-USM¹⁸.

SDUMLA. SDUMLA considers the influence of translation, rotation, and other factors during acquisition, so the image quality is slightly worse. A total of 106 volunteers participated in the collection work. The index finger, middle finger, and ring finger of both hands were collected. Each finger captured six images. Finally, 636 finger vein categories and 3816 finger vein images were obtained, and the image size was 320×240 .

FV-USM. The quality of vein image collection in FV-USM is high. 123 volunteers participated in the collection work. Six images of their left and right index fingers and middle fingers were collected respectively. The images were collected in two times, and six finger vein images were captured in each collection. Finally, 492 finger vein categories and 5904 finger vein images are obtained, and the image size is 640×480 .

The public finger vein database takes into account differences in age, gender, and region of volunteers when collecting, and also considers the influence of non-ideal images such as angle, occlusion, and dark light. In order to facilitate the analysis and comparison of the experimental results, the database image size is adjusted to 60×180 . The pictures in the public database are randomly divided according to the ratio of 1:5, one is the test database, and the five are the training database, to ensure that they are independent of each other.

Evaluation indicators. This paper uses the recognition accuracy, recognition time, Parameter and Flops as the evaluation indicators of the model. The recognition accuracy represents the number of correct classifications by the network model, and the recognition time is the time it takes the network to recognize a picture. Parameter(M) stands for Mega-parameter, Flops(GFlops) stands for Giga float point operations, which can be used as a measure of whether it is suitable for hardware deployment.

Training configuration. We implement our proposed method using PyTorch and trained the model on the NVIDIA GeForce GTX1660 (6 GB) GPU. Adam optimizer is employed during training, and the batch size is set at 8. We trained the model using dataset FV-USM and SDUMLA respectively. When trained on dataset FV-USM, the duration of a training epoch is 5.86 s. The final trained model is obtained after 300 epochs. Thus,

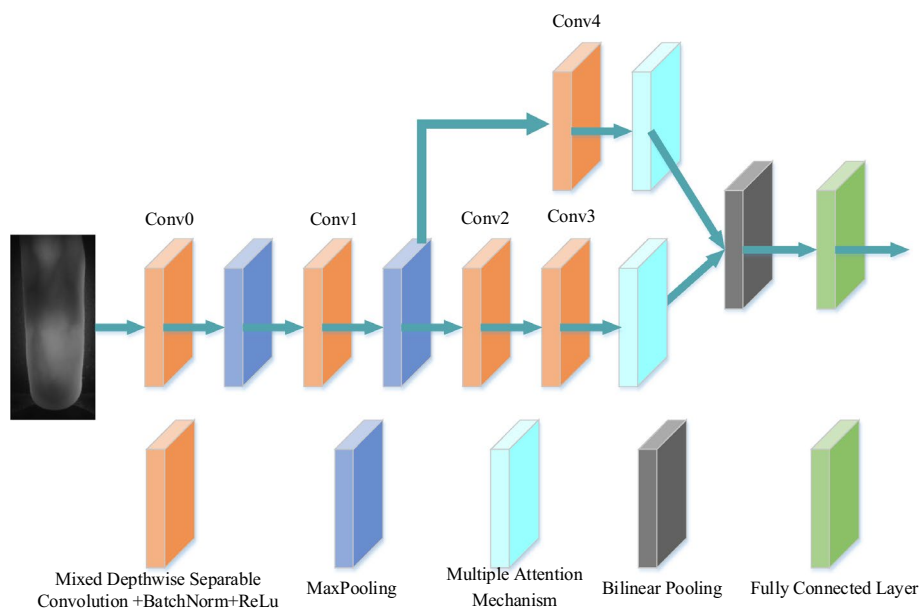


Figure 4. Proposed lightweight MSFBB-Net with attention mechanism.

the total training time is 1758s. When trained on dataset SDUMLA, Each epoch takes 4.06 s to train. The final trained model is also obtained after 300 epochs. Thus the total training time is 1218 s.

Comparison with classical network. As can be seen from the data from Table 2, compared with the classical network model, the multi-scale feature bilinear fusion network performs very well. This paper proposes MSFBB-Net, and then obtains two variant networks based on this network model, which are the lightweight MSFBB-Net(Lightweight MSFBB-Net) and the lightweight MSFBB-Net combined with multiple attention mechanism(Lightweight MSFBB-Net+MAM). The recognition accuracy of MSFBB-Net on two public databases is as high as 99.59% and 99.53%. Single image recognition time is only about 3 ms. Both model parameters and computational complexity have certain advantages. After a comprehensive comparison, it is concluded that the multi-scale feature bilinear fusion network is the most suitable for the finger vein recognition task.

Ablation experiment. In order to verify the correctness of ideas and methods and the effectiveness of each step, it is necessary to conduct ablation experiments. The algorithm is evaluated from the following four aspects: recognition accuracy, model parameters, computational complexity, and single image recognition time. Next, FV-USM data are used for analysis, and the experimental results are shown in Table 3. Although the multi-scale feature bilinear fusion network model has obtained a good recognition accuracy, the parameters and computational complexity are very large. After the network model is lightweight, a lot of computational complexity is reduced, and the recognition accuracy decreased from 99.59 to 99.29%, which is consistent with our expectation. The decrease in recognition accuracy is due to the use of mixed depthwise separable convolutions in the network. The depthwise separable convolution destroys the correlation between channels and spaces, resulting in a decrease in recognition accuracy. Finally, the multiple attention mechanism is embedded into the lightweight network model. It can be seen that the accuracy rate has increased from 99.29% to 99.90%, and the amount of computation has hardly increased, still 0.043GFlops. It is very convenient for the model to be deployed on hardware devices later. Comparing the lightweight model with the novel lightweight network model, we find that the method in this paper maintains high recognition accuracy, low computational complexity and less recognition time, while the amount of model parameters is acceptable. This is because the model parameters are composed of convolution kernel parameters and weight parameters. There are few convolution layers in our network model, so it is not obvious to reduce the parameters of the lightweight model. But most of the computation is in convolutional layers, which reduces Flops drastically from 0.21 to 0.043 GFlops. The recognition time of a single image of the final model is only 4 ms. This is consistent with the theoretically expected results, demonstrating the effectiveness of the algorithm at each step.

Compared with the traditional algorithm. Table 4 shows the experimental results on FV-USM. It can be seen that the recognition accuracy of the lightweight multi-scale feature bilinear fusion network combined with multiple attention mechanism is better than that of the excellent traditional finger vein recognition algorithm. In addition, when the network model is used for finger vein recognition, the finger vein database image is directly used without preprocessing, which greatly reduces the time cost.

| Method | Accuracy/% | | Time/ms | Parameter/M | Flops/GFlops |
|-------------------------|------------|--------|---------|-------------|--------------|
| | FV-USM | SDUMLA | | | |
| VGG-16 ¹⁹ | 95.83 | 95.71 | 49 | 44.0 | 3.20 |
| DesNet ²⁰ | 96.95 | 98.16 | 34 | 26.7 | 0.17 |
| AlexNet ²¹ | 92.28 | 96.32 | 16 | 19.7 | 0.43 |
| ResNet-18 ²² | 96.03 | 97.79 | 16 | 11.0 | 2.40 |
| MSFBB-Net | 99.59 | 99.53 | 3 | 8.2 | 0.21 |

Table 2. Comparison of classic network performance.

| Method | Accuracy/% | | Time/ms | Parameter/M | Flops/GFlops |
|-------------------------------|------------|--------|---------|-------------|--------------|
| | FV-USM | SDUMLA | | | |
| EfficientNet-B0 ²³ | 99.09 | 98.58 | 12 | 5.3 | 0.069 |
| MobileNet-V1 ²⁴ | 94.72 | 94.18 | 4 | 3.7 | 0.14 |
| GhostNet ²⁵ | 95.83 | 95.60 | 11 | 4.7 | 0.039 |
| MSFBB-Net | 99.59 | 99.53 | 3 | 8.2 | 0.21 |
| Lightweight MSFBB-Net | 99.29 | 99.21 | 2 | 8.1 | 0.043 |
| Lightweight MSFBB-Net + MAM | 99.90 | 99.82 | 4 | 8.1 | 0.043 |

Table 3. Ablation experiment results.

| Method | Accuracy/% |
|------------------------------------|------------|
| LBP | 96.10 |
| LBP + SVM | 97.86 |
| SURF ²⁶ | 97.17 |
| Yang et al.'s method ²⁷ | 99.52 |
| Ours | 99.90 |

Table 4. Experimental results of traditional algorithms.

Compared with existing deep learning methods. The lightweight multi-scale feature bilinear fusion network combined with the multiple attention mechanism and the existing convolutional neural network are tested on two public databases, and the experimental results are shown in the Table 5. Among the excellent finger vein recognition algorithms in recent years, Ren et al.'s method has the best comprehensive performance, but our method has better performance in recognition accuracy and recognition time. Figure 5 shows the recognition curves of our method and some other network models on two public databases. It can be seen that our method has advantages in recognition accuracy. Figure 6 is the training loss curve and valid loss curve obtained by using the lightweight MSFBF-Net architecture with attention mechanism proposed in this paper to experiment with FV-USM and SDUMLA. It can be seen from the figure that the model gradually converges after 100 epochs.

Figure 7 shows some poor-quality images with obvious noise and unclear vein lines in public databases. Other methods of identification always identify errors. The lightweight multi-scale feature bilinear fusion network combined with multiple attention mechanism is designed for the characteristics of finger veins, which can extract the multi-scale features with rich information, and simultaneously extract the channel, spatial, directional, and positional information to assist recognition.

Conclusion

Aiming at the problems of large models, many parameters, and long recognition time in existing finger vein recognition methods, a lightweight multi-scale feature bilinear fusion network combined with multiple attention mechanism was proposed. Firstly, the multi-scale feature bilinear fusion network extracts the global features and

| Method | Accuracy/% | | Time/ms |
|---------------------------------------|------------|--------|---------|
| | FV-USM | SDUMLA | |
| Zhao et al.'s method ²⁸ | 97.95 | 98.41 | 9 |
| Shen et al.'s method ²⁹ | 99.59 | 99.30 | 14 |
| Shaheed et al.'s method ³⁰ | 99.08 | 98.50 | 45 |
| Ren et al.'s method ³¹ | 99.59 | 96.70 | 5 |
| Ours | 99.90 | 99.82 | 4 |

Table 5. Experimental results of deep learning models.

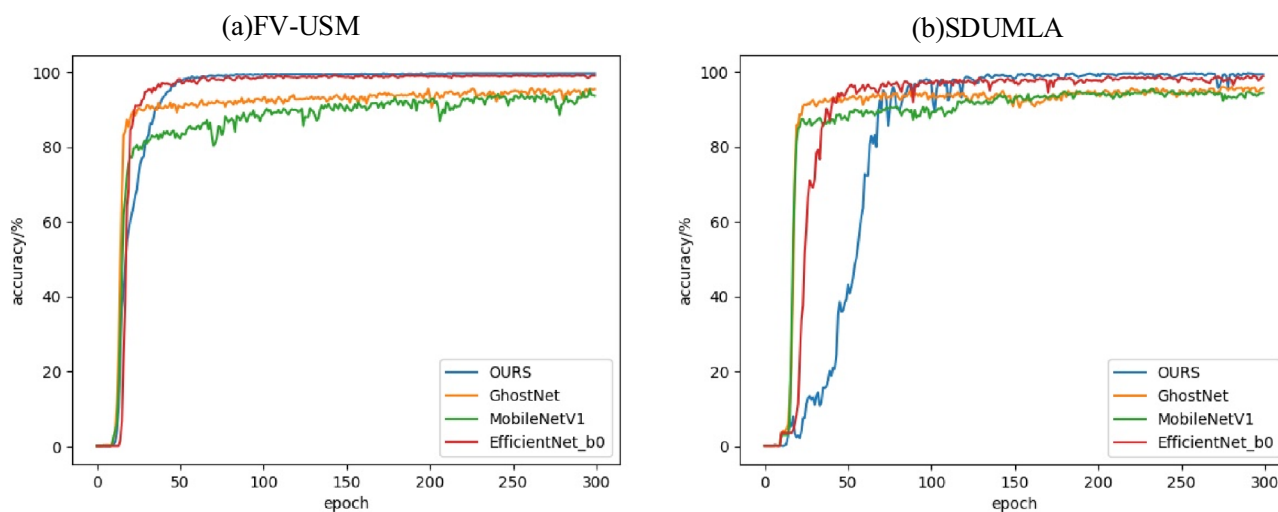


Figure 5. Recognition accuracy curve.

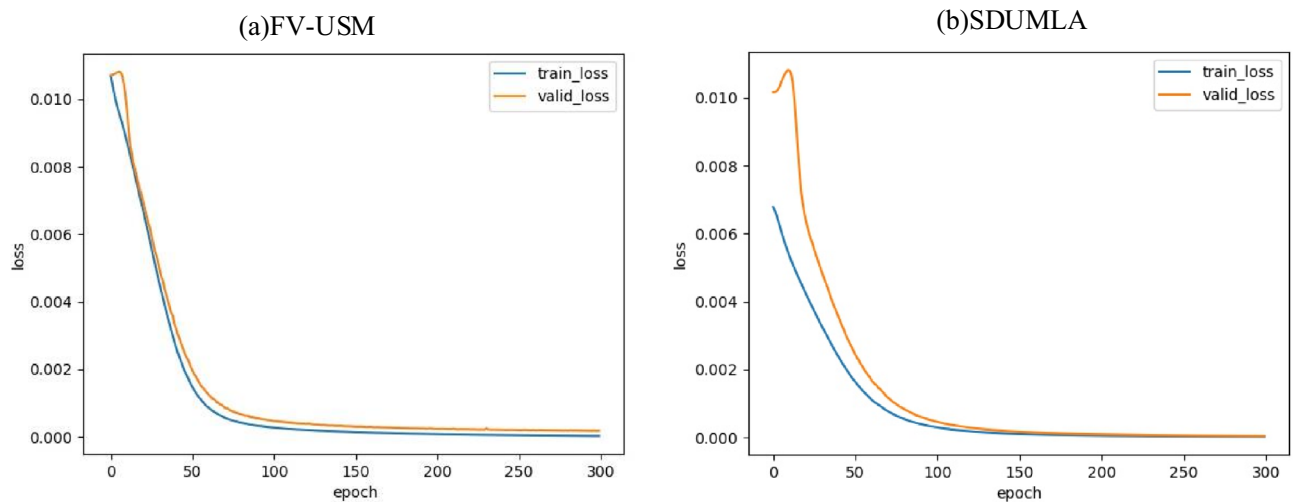


Figure 6. Loss curve.

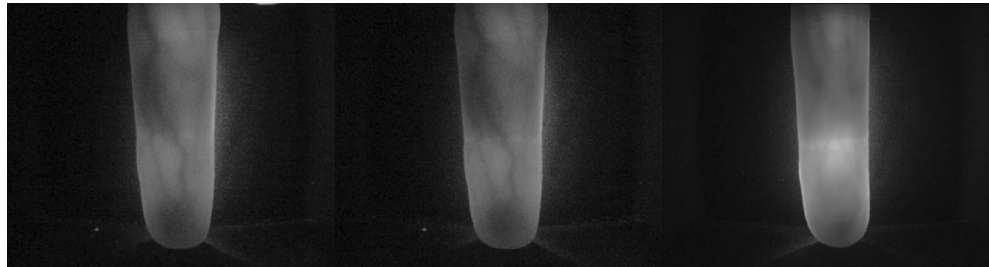


Figure 7. Poor quality images in public databases.

local detail features of the finger vein image. After bilinear pooling, a second-order feature containing richer information is obtained. Then, the mixed depthwise separable convolution is used to lightweight the network, which greatly reduces the computational complexity of the network. Finally, a multiple attention mechanism is designed to make up for the shortcomings of the mixed depthwise separable convolution and strengthen the network model's ability to extract subtle features. The proposed method is tested on two public databases, and the recognition accuracy is up to 99.90% and 99.82%, which is 7.62% higher than other methods. After the lightweight processing of the network, the computational complexity is only one-fifth of the original, and the recognition time of a single image is only 4 ms. Comprehensive comparison, the effect of the method in this paper is significantly improved, which is very suitable for finger vein recognition tasks.

Data availability

The data that support the findings of this study are available from SDUMLA (<https://time.sdu.edu.cn/kycg/gksjk.htm>) and FV-USM (http://drfendi.com/fv_usm_database/) but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data are however available from the corresponding author upon reasonable request and with permission of official.

Received: 5 September 2022; Accepted: 3 January 2023

Published online: 05 January 2023

References

- Shahin, M., Badawi, A. & Kamel, M. Biometric authentication using fast correlation of near infrared hand vein patterns. *Int. J. Biol. Med. Sci.* **2**(3), 141–148 (2017).
- Chen, L., Wang, J., Yang, S. & He, H. A finger vein image-based personal identification system with self-adaptive illuminance control. *IEEE Trans. Instrum. Meas.* **66**(2), 294–304 (2016).
- Das, R., Piciuccio, E., Maiorana, E. & Campisiet, P. Convolutional neural network for finger-vein-based biometric identification. *IEEE Trans. Inf. Forensics Secur.* **14**(2), 360–373 (2018).
- Wang, J., Pan, Z., Wang, G., Li, M. & Li, Y. Spatial pyramid pooling of selective convolutional features for vein recognition. *IEEE Access.* **6**, 28563–28572 (2018).
- Wu, J. D. & Liu, C. T. Finger-vein pattern identification using SVM and neural network technique. *Expert Syst. Appl.* **38**(11), 14284–14289 (2011).
- Raghavendra, R., Surbiryala, J., & Busch, C. An efficient finger vein indexing scheme based on unsupervised clustering. In *IEEE International Conference on Identity, Security and Behavior Analysis*. 1–8(2015).

7. Hu, N., Ma, H. & Zhan, T. Finger vein biometric verification using block multi-scale uniform local binary pattern features and block two-directional two-dimension principal component analysis. *Optik* **208**, 163664 (2020).
8. Yang, L., Yang, G., Yin, Y. & Xi, X. Finger vein recognition with anatomy structure analysis. *IEEE Trans. Circuits Syst. Video Technol.* **28**(8), 1892–1905 (2017).
9. Meng, X., Zheng, J., Xi, X., Zhang, Q. & Yin, Y. Finger vein recognition based on zone-based minutia matching. *Neurocomputing* **423**, 110–123 (2021).
10. Kamaruddin, N. M. & Rosdi, B. A. A new filter generation method in PCANet for finger vein recognition. *IEEE Access.* **7**, 132966–132978 (2019).
11. Ren, H., Sun, L., Guo, J., Han, C. & Cao, Y. A high compatibility finger vein image quality assessment system based on deep learning. *Expert Syst. Appl.* **196**, 116603 (2022).
12. Zhang, L. *et al.* A joint bayesian framework based on partial least squares discriminant analysis for finger vein recognition. *IEEE Sens. J.* **22**(1), 785–794 (2021).
13. Noh, K. J., Choi, J., Hong, J. S. & Park, K. Finger-vein recognition based on densely connected convolutional network using score-level fusion with shape and texture images. *IEEE Acces.* **8**, 96748–96766 (2020).
14. Hong, H. G., Lee, M. B. & Park, K. R. Convolutional neural network-based finger-vein recognition using NIR image sensors. *Sensors.* **17**(6), 1297 (2017).
15. Yang, W., Hui, C., Chen, Z., Xue, J. & Liao, Q. FV-GAN: Finger vein representation using generative adversarial networks. *IEEE Trans. Inf. Forensics Secur.* **14**(9), 2512–2524 (2019).
16. Tang, S., Zhou, S., Kang, W., Wu, Q. & Deng, F. Finger vein verification using a Siamese CNN. *IET Biometrics.* **8**(5), 306–315 (2019).
17. Yin, Y., Liu, L., & Sun, X. SDUMLA-HMT: A multimodal biometric database. In *Chinese Conference on Biometric Recognition*, 260–268. Springer, Berlin, Heidelberg (2011).
18. Asaari, M. S. M., Suandi, S. A. & Rosdi, B. A. Fusion of band limited phase only correlation and width centroid contour distance for finger based biometrics. *Expert Syst. Appl.* **41**(7), 3367–3382 (2014).
19. Simonyan, K., & Zisserman, A. Very deep convolutional networks for large-scale image recognition. <https://arxiv.org/pdf/1409.1556.pdf>(2014).
20. Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 4700–4708 (2017).
21. Krizhevsky, A., Sutskever, I. & Hinton, G. E. Imagenet classification with deep convolutional neural networks. *Commun. ACM* **60**(6), 84–90 (2017).
22. He, K., Zhang, X., Ren, S., & Sun, J. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 770–778 (2016).
23. Tan, M., & Le, Q. Efficientnet: Rethinking model scaling for convolutional neural networks. In *International conference on machine learning*. 6105–6114 (2019).
24. Howard, A. G., Zhu, M., & Chen, B. Mobilenets: Efficient convolutional neural networks for mobile vision applications. <https://arxiv.org/pdf/1704.04861.pdf>(2017).
25. Han, K., Wang, Y., Tian, Q., Guo, J., Xu, C., & Xu, C. Ghostnet: More features from cheap operations. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 1580–1589 (2020).
26. Bay, H., Ess, A., Tuytelaars, T. & Van Gool, L. Speeded-up robust features (SURF). *Comput. Vis. Image Underst.* **110**(3), 346–359 (2008).
27. Yang, J., Shi, Y. & Jia, G. Finger-vein image matching based on adaptive curve transformation. *Pattern Recogn.* **66**, 34–43 (2017).
28. Zhao, D., Ma, H., Yang, Z., Li, J. & Tian, W. Finger vein recognition based on lightweight CNN combining center loss and dynamic regularization. *Infrared Phys. Technol.* **105**, 103221 (2020).
29. Shen, J. *et al.* Finger vein recognition algorithm based on lightweight deep convolutional neural network. *IEEE Trans. Instrum. Meas.* **71**, 1–13 (2021).
30. Shaheed, K. *et al.* DS-CNN: A pre-trained Xception model based on depth-wise separable convolutional neural network for finger vein recognition. *Expert Syst. Appl.* **191**, 116288 (2022).
31. Ren, H., Sun, L., Guo, J., Han, C. & Wu, F. Finger vein recognition system with template protection based on convolutional neural network. *Knowl.-Based Syst.* **227**, 107159 (2021).

Acknowledgements

The authors received no specific funding for this work.

Author contributions

B.M. conducts experiments and writes papers. K.X.W. data collection and discussed the results. Y.L.H. guides and supervises the work. All authors reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to Y.H.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023