Research Article Vol. 15, No. 12 / 1 Dec 2024 / Biomedical Optics Express 6845
Biomedical Optics EXPRESS

Mask-guided network for finger vein feature extraction and biometric identification

HAOHAN BAI,[1,2,†] YUBO TAN,[2,†] AND YONG-JIE LI[1,2,*]

[1] *Yangtze Delta Region Institute (Huzhou), University of Electronic Science and Technology of China (UESTC), Huzhou 313001, China*
[2] *School of Life Science and Technology, UESTC, Chengdu 610054, China*
[†] The authors contributed equally to this work.
[*] *liyj@uestc.edu.cn*

Abstract: The problems of complex background, low quality of finger vein images, and poor discriminative features have been the bottleneck of feature extraction and finger vein recognition. To this end, we propose a feature extraction algorithm based on the open-set testing protocol. In order to eliminate the interference of irrelevant areas, this paper proposes the idea of segmentation-assisted classification, that is, using the rough mask of the finger vein to constrain the feature learning process so that the network can focus on the vein area and learn greater weight for the vein. Specifically, the feature maps of the shallow layers of the network are first sent to the feature pyramid module to fuse the primary features of different scales, which are then sent to the spatial attention module to obtain the spatial weight map of the image. Based on the results of several classical vein skeleton extraction algorithms, a weighting method is used to obtain a more accurate mask to constrain the learning of the spatial weight map. Finally, a hybrid loss function combining triplet loss and cross-entropy loss is used to reduce the distance between feature vectors of the same categories and increase the distance between feature vectors of different categories in the Euclidean space, thereby improving feature discriminability. Good recognition results were achieved on the three public data sets of SDUMLA, MMCBNU, and FVUSM, and the values of equal error rate (EER) on them are as low as 2.50%, 0.20%, and 0.14%, respectively.

1. Introduction

Biometric identification is a type of authentication method that aims to implement identification by using people's natural physiological characteristics or behavioral characteristics [1], which has better stability, security and uniqueness. At present, commonly used identification methods based on physiological characteristics include face [2], fingerprint [3], finger vein [4], palm vein [5], iris [6], etc. Among them, face, fingerprint and iris are external characteristics of people, and their security is not as good as that of internal characteristics such as finger vein. Kono *et al.* [7] first introduced finger vein to the biometric identification field in 2002, and since then, finger vein recognition has become a research hotspot because of its good security, inexpensive acquisition equipment, and potential for wide range of applications. For example, Zhang *et al.* introduced a three-dimensional (3D) finger vein biometric authentication approach and system, leveraging the innovative technology of photoacoustic tomography [8].

Due to the acquisition principle, finger vein images have the following characteristics:

- Poor contrast. Because of the possible scattering of near-infrared light inside tissues such as muscles or the influence of ambient light, the acquired images are easily confused between the venous and non-venous areas.
- Rotation or translation. One person may have different finger placement during multiple acquisitions, which requires the feature extraction algorithms to have a certain degree of rotation and translation invariance.

#535390 https://doi.org/10.1364/BOE.535390
Journal © 2024 Received 12 Jul 2024; revised 14 Sep 2024; accepted 25 Sep 2024; published 15 Nov 2024

Biomedical Optics EXPRESS

- Small sample size. The number of finger vein images for each individual in existing public datasets is relatively small.

Over the years, research in this field has progressed significantly, transitioning from conventional methodologies to deep neural networks. These traditional approaches encompass vein skeleton-based [9,10], texture-based [11,12], minutia-based [13,14], and subspace-based methods [15,16], all of which heavily rely on high-resolution digital vein images and necessitate extensive professional expertise for design and fine-tuning. Consequently, these methods struggle to generalize effectively to novel images. In contrast, deep learning have emerged as a promising alternative, incorporating transfer learning based classification models operating in closed or open set environments [17–20], as well as networks augmented with attention modules [21,22]. However, it is noteworthy that the performances of transfer learning based models may falter facing the relatively scarce medical data in the field of finger vein recognition. Furthermore, many existing attention strategies are inadequately tailored to address the unique challenges posed by this task.

Due to the above characteristics, although a large number of excellent finger vein recognition algorithms have emerged in recent years, it is still a challenge to extract vein features with high robustness, good discrimination, and generalization abilities from the vein images acquired by different devices. To this end, this paper proposes a deep learning based feature extraction algorithm that uses the mask of finger vein as a priori information to guide the feature extraction process. The main contributions of this work are summarized as follows.

- To exclude the interference of irrelevant regions, this paper proposes a segmentation-assisted recognition idea that uses the mask of finger veins to constrain the feature learning process so that the feature extraction network can focus on learning the features of the vein region.
- A mask guided module (MGM) is proposed, which serves to constrain the weight distribution of the shallow feature maps. To this end, the mask of finger vein is obtained by fusing the results of several traditional efficient vein skeleton extraction algorithms, which is then used as priori information to guide the feature extraction network to learn the vein features better.
- A deep learning based finger vein feature extraction framework is proposed, which is based on the open-set testing protocol and combined with efficient image pre-processing, resulting in better inter-class dispersion and intra-class aggregation for the extracted vein features.
- Extensive experiments are conducted to evaluate the proposed method from three aspects, including the performance comparison of the method before and after improvement, the ablation study of the designed modules, and the comparison with other mainstream algorithms. In short, the results demonstrate the effectiveness of the proposed method.

2. Related work

2.1. Traditional finger vein recognition methods

Traditional methods for finger vein recognition can be divided into the following four categories:

2.1.1. Vein skeleton based methods

The general idea of this type of methods is to apply threshold segmentation on the whole image to get the skeleton structure of finger veins, and then use the geometry or topology of the segmented image to perform template matching to get a matching score, and finally determine whether they

Biomedical Optics EXPRESS

are from the same persons according to the score. The representative methods include RLT [9], MC [10], WLD [23], etc. Such methods require high quality finger vein images and are susceptible to shadows and noises.

2.1.2. Texture-based methods

They mainly compare the gray values of the central element and the neighborhood elements and then perform binary encoding, and finally realize image matching and recognition based on certain distance, including Local Binary Patterns (LBP) and a series of variants [11,12,24]. These methods are characterized by the requirement of designing features manually according to specific needs and only the target elements and elements in the neighborhood are considered in the coding process, ignoring the elements in other regions, resulting in weak feature representation.

2.1.3. Minutia-based approaches

This type of methods mainly extracts different types of point features in the image, for example, bifurcation points and ending points, which are normally the intersections of multiple veins. Yang *et al.* [25] used the intersection and its surrounding venous branches (called tri-branch) as minutiae for template matching. Liu *et al.* [26] proposed a minutiae matching algorithm based on singular value decomposition. Yu *et al.* [27] first extracted the vein skeleton, then extracted the ending points and bifurcation points from the vein skeleton, and finally used the improved Hausdorff distance for matching feature points for recognition. In addition, a series of methods based on SIFT (Scale Invariant Feature Transform) [13,14] can also automatically extract feature points from finger vein images for matching. Among others, Qin *et al.* [28] combined finger shape, orientation and SIFT feature for finger vein recognition.

2.1.4. Subspace-based methods

In these methods, the finger vein images are transformed into a lower dimensional space for recognition by learning the transformation matrix from the training data. For example, Wu *et al.* [15] used Principal Components Analysis (PCA) to extract features for identity recognition. PCA requires reducing a two-dimensional(2D) image to a one-dimensional vector, which is computationally tedious and time-consuming. 2DPCA directly projects the image to reduce the time overhead, and Ma *et al.* [16] then used 2DPCA to implement feature extraction. Wu *et al.* [29] implemented finger vein recognition using a combination of PCA and Linear Discriminate Analysis (LDA). However, when recognition is performed in an open-set environment, such methods require to update the transformation matrix, which is not very practical.

2.2. Deep learning based finger vein recognition

Compared with the traditional algorithms that require manual design of feature extractors, deep learning-based methods use deep neural networks, e.g., CNN (Convolutional Neural Networks), to automatically extract high-level abstract features from vein images, so they have stronger feature representation ability. Based on the testing protocol, when evaluating the performance, these methods can be classified as close-set and open-set testing protocol based ones.

2.2.1. Closed-set protocol

The closed-set protocol means that all categories of finger vein images will appear in the training set, that is, some samples are selected from each category to compose the training set, and the remaining untrained images in each category are used for testing. This protocol is suitable for situations where the number of categories is fixed, and the disadvantage is that the recognition models need to be retrained once a new category is added.

Among this group, Radzi *et al.* [17] implemented finger vein recognition using a simple four-layer CNN model, which was claimed to be the first attempt to apply CNN to the finger vein

Biomedical Optics EXPRESS

recognition. Hou *et al.* [30] used Auto-Encoder for feature extraction and then CNN for finger vein recognition, which achieved a large improvement compared with traditional algorithms. Jalilian *et al.* [31] used a semantic segmentation network directly to extract finger vein patterns without using pre-processing and post-processing and achieved better results. Yang *et al.* [32] used a multi-task learning approach to integrate the recognition task and the aliveness detection task into a unified lightweight CNN, which ensures the real-time performance. Zhang *et al.* [18] combined CNN with Gabor filters using a method that adaptively learns the parameters of Gabor filters, solving the difficult problem of parameter selection. Lu *et al.* [33] designed a CNN-based local descriptor (CNN-CO) using a pre-trained network, selecting certain convolutional kernels of the first layer of CNN to act as filters, whereby different CNN-CO images were obtained. Zhang *et al.* [34] proposed a finger vein recognition pipeline by combining the multi-directional local information and the features automatically learned by a CNN network. This pipeline has the advantage of compensating for some finger vein features that may be overlooked.

2.2.2. Open-set protocol

In the open-set protocol, the finger vein images of some categories are used for training, and the vein images of the other categories are used for testing, with no intersection between the two category groups at all. In this open-set situation, finger vein recognition is no longer a multi-classification task. Usually, for a category to be recognized, its feature vectors are first extracted using a feature extraction model, and this category is then identified according to the similarity between its feature vectors and the feature vectors of existing categories in the database. For a new category, its feature vector can be stored into the database by a registration operation. It is clear that such open-set protocol is more suitable for actual usage scenarios.

Based on this type of testing protocol, Song *et al.* [19] investigated a method that utilizes the DenseNet and is robust to noise by feeding a composite image of two finger vein images to a CNN. Hou *et al.* [35] used metric learning loss function and cosine distance to enhance the representation ability of the feature extraction model. Ou *et al.* [36] used vertical flipping to increase the number of categories and employed with intra-class data augmentation techniques to solve the problem of insufficient training data, then combined classification loss and metric learning loss to design a fused loss function. Shen *et al.* [37] proposed to build a lightweight CNN to ensure the real-time performance of feature extraction and recognition. To deal with the lack of large-scale finger vein dataset, Hou *et al.* citehou2022triplet designed a generative adversarial network and combined it with a CNN based on triplet loss to extend the training data and improve the discriminative ability of the algorithm, which was finally tested in both open-set and closed-set scenarios. Lin *et al.* [20] proposed to improve the recognition performance of intra-class matching and inter-class matching using a self-encoder-based internal feature module and an external feature module using siamese network, respectively. Song *et al.* [38] first trained a vein segmentation network using their own labeled dataset, then extracted features from the segmented and the original images and fused them. Unlike other methods that extract features only in the spatial domain, Huang *et al.* [39] proposed to use deep networks for feature extraction in both frequency and spatial domains.

Most finger vein recognition methods mentioned above use images of single view, whereas slight difference of the poses for the same fingers may increase the difficulty of intra-class recognition. For this reason, Zhao *et al.* [40] designed a multi-view low-cost vein image acquisition device and extracted hidden 3D features from each viewpoint through a global backbone network and local features through a local perception module, thus improving the feature recognizability and robustness. Zhang *et al.* [41] designed a reflectance imaging-based acquisition device and constructed a large-scale dataset containing five different light intensities.

Biomedical Optics EXPRESS

To reduce the differences between different illumination data domains, they proposed a domain-adaptive finger vein network to extract light invariance features to improve the robustness to illumination variations.

2.3. Visual attention mechanism

When we pay attention to the image of finger vein, we should focus on information such as the vein skeleton and finger outline in the image, and we can selectively ignore irrelevant information such as the background. In this way, we can distinguish different categories of finger vein images faster and more accurately. In fact, the attention mechanism has been introduced successfully in various tasks of computer vision, such as classification, segmentation, image detection, etc. Classical approaches employing visual attention mechanisms include SENet [21], ECANet [42], and CBAM [22], etc. These attention modules are simple to compute and do not add too much computation, while being plug-and-play and easy to use. For example, Huang *et al.* [43] proposed a joint attention module using an attention mechanism that allows dynamic adjustment and information aggregation of feature maps in spatial and channel dimensions, enabling the network to focus on fine vein patterns rather than non-vein regions. Zhang *et al.* [44] proposed a lightweight deep network to achieve accurate extraction of finger vein features by introducing a convolutional attention module to adaptively assign feature weights. With the emergence of transformer [45], several researchers began to introduce it into the task of finger vein recognition. For example, literatures [46–48] used transformer-based methods to extract the global features of vein images to make up for the insufficiency of convolutional networks.

3. Materials and methods

3.1. Finger vein dataset

Compared with the datasets in other fields, finger vein datasets are generally less in sample number (only a few images for each category) and more difficult to obtain (requiring a specific acquisition device). The datasets used in this paper include SDUMLA-HMT [49], MMCBNU [50], and FVUSM [51], which are briefly described below.

SDUMLA: The dataset is a multimodal biometric database made public by Shandong University [49], which contains 106 volunteers (61 males and 45 females). Each subject was collected the images of six fingers, i.e., the index, middle and ring fingers of the left and right hands, respectively, and each finger is regarded as a category, resulting in a total of 636 categories. For each finger, six images were collected during the same session, obtaining a total of 3816 images in this dataset with a resolution of 320×240 pixels.

MMCBNU: This dataset was collected and released by Chonbuk National University in South Korea in 2013 [50]. 100 volunteers (83 males and 17 females) were selected from more than 20 countries. Each volunteer was collected 6 images of fingers, namely the index, middle and ring fingers of the left and right hands, resulting in a total of 600 categories. They were all acquired in the same session and each finger was collected 10 times, obtaining a total of 6000 images with a resolution of 640×480 pixels.

FVUSM: It is a near-infrared light finger image dataset made public by the Universiti Sains Malaysia in 2014 [51], which provides the processed ROI images. Specifically, this dataset was collected from 123 subjects, 83 males and 40 females. Each subject was collected images of four fingers, i.e., the index and middle fingers of the left and right hands, obtaining a total of 492 categories. The dataset was collected in two sessions, separated by more than 2 weeks, with 6 images for each finger in each session, resulting in a total number of 5904 images with a resolution of 640×480 pixels and large variation in image appearance.

Biomedical Optics EXPRESS

3.2. Evaluation metric

For the open-set testing, the input is an image pair, one is from the dataset and the other is the one to be identified. According to whether the two images are of the same category, they can be divided into genuine pair and impostor pair. To ascertain whether two images form a genuine or impostor pair utilizing a pre-trained feature extractor, the cosine distance between the respective feature vectors extracted by the model is computed. If this distance is less than a predefined threshold, the model classifies the pair as genuine; conversely, if it exceeds the threshold, the pair is deemed an impostor. The evaluation metrics we choose include the widely used false acceptance rate (FAR), false rejection rate (FRR), equal error rate (EER), and receiver operator characteristic (ROC) curves. Specifically, FAR denotes the probability of being incorrectly considered as a genuine pair among all impostor pairs, and FRR denotes the probability of being incorrectly considered as an impostor pair among all genuine pairs, EER represents the error rate at which FAR and FRR are equal.

3.3. Overall architecture

As shown in Fig. 1, the overall framework proposed in this paper includes a backbone network, a feature pyramid module, a mask-guided attention module, and a joint loss function composed of triplet loss and softmax loss. During training, the ROI (Region of Interest) images are sent to the backbone network RestNet18, and then the shallow feature map of the network is extracted and sent to the feature pyramid module to obtain the fused features. These features are sent to the mask guidance module to obtain the corresponding weight map, where the mask is used to constrain the distribution of the weight map so that the network can give more attention to the venous regions and weaken the contribution of the non-venous regions during the learning process. The weight map is then multiplied and summed with the output of the backbone network and the global average pooling operation is used to obtain the feature vector, which is then passed through the classifier to obtain the predicted category. During the test, the mask guidance module will not be used, and the trained network model can be used to directly obtain the feature vector of the input vein image, which is applicable to the open-set scenarios.

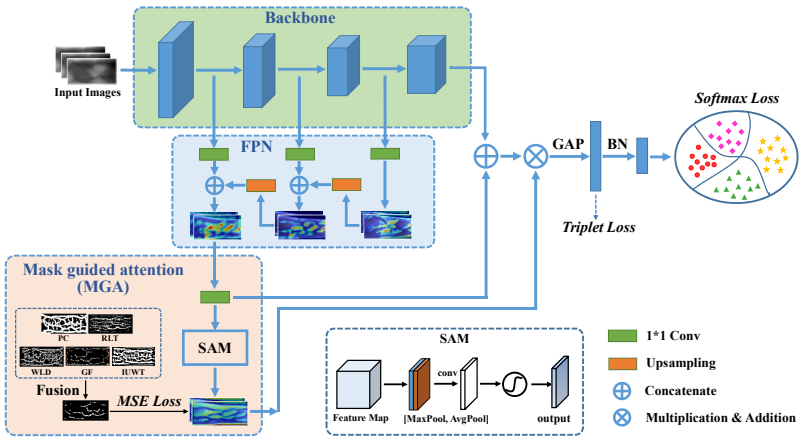


Fig. 1. Overall structure of the proposed method. The backbone network is used to extract the finger vein feature and subsequently generates the feature vectors. The modified Feature Pyramid Network (FPN) [52] is used to fuse the shallow features of the backbone network to facilitate the generation of the weight map. The MGA module guides the network learning by constraining the distribution of the weight map through SAM [22].

Biomedical Optics EXPRESS

3.4. Image preprocessing

The large difference in the finger vein images obtained by the acquisition devices and the interference of irrelevant backgrounds undoubtedly increase the difficulty of subsequent identification. Therefore, we extract a region of interest (ROI) in the finger vein image first and then perform image enhancement [53]. The image preprocessing steps in this paper include finger contour detection, midline fitting, rotation correction, ROI extraction and image enhancement, which are shown in the Fig. 2.

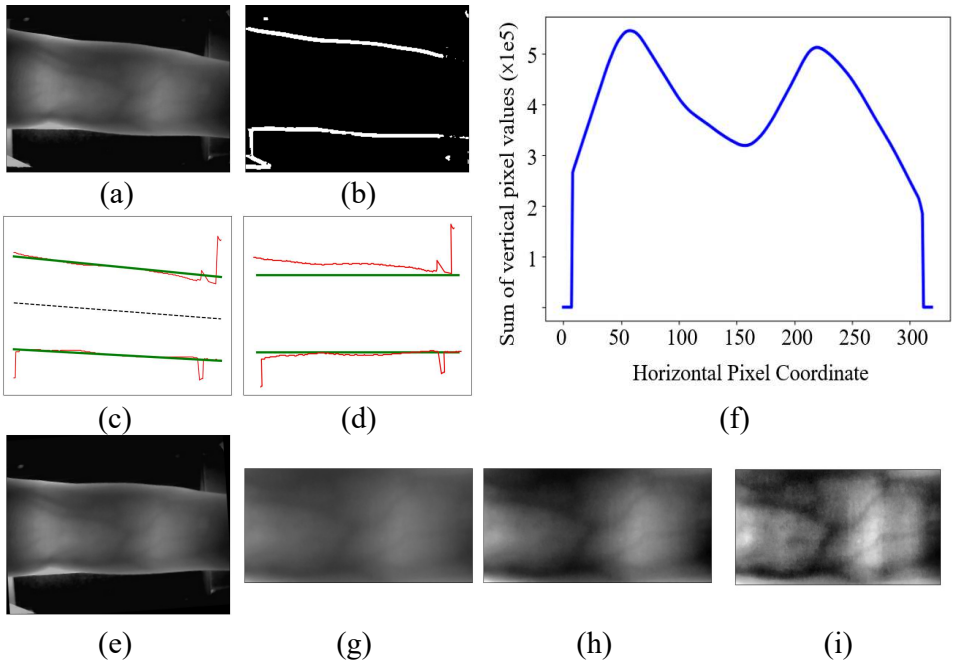


Fig. 2. Finger vein image preprocessing steps. (a). Original image. (b). Finger contour. (c). Midline fitting. (d). Rotation correction. (e). Rotated finger vein image. (f). Gray distribution of finger vein along the horizontal direction. (g). Extracted ROI. (h). Image normalization. (i). Image enhancement.

(1) Contour detection and midline fitting. Finger contour detection is the first step of preprocessing, and the main purpose is to detect the finger edges and exclude the non-finger regions. Here the simple Laplacian operator is chosen to extract the finger contours (Fig. 2(b)). According to the obtained upper and lower edges of the finger (red curves) and the fitted edges (green curves) in Fig. 2(c), the middle line (the dashed line) of the finger is fitted by using the least square method.

(2) Rotation correction. Using the fitted midline, we obtain its slope, calculate the horizontal deflection angle according to the arctangent function, and then correct the rotation of the entire image with the opposite angle, as shown in Eq. (1), where θ represents the angle between the midline and the horizontal direction, and A and B are two points on the midline. x and y represent the horizontal and vertical coordinates of the element, respectively.

$$\theta = arctan(\frac{y_B - y_A}{x_B - x_A}) \tag{1}$$

(3) ROI extraction. The extraction of ROI is carried out in two steps, along the vertical and horizontal directions. In the vertical direction, if we take the external rectangle of the finger edge,

Biomedical Optics EXPRESS

it will introduce some background information, so here we choose the internal rectangle of the finger edge. In the horizontal direction, because the finger does not absorb infrared light well at the joints, their gray values have two peaks, as shown in Fig. 2(f), based on which we choose the content between the two peaks as ROI, so that we can fix the extraction range of each finger.

(4) Image enhancement. To make the grayscale distribution of images more comparable, we first normalize each ROI, as shown in Fig. 2(h), and the normalization is given by

$$Img_{(i,j)} = \frac{I(i,j) - min}{max - min} \times 255 \tag{2}$$

where $I(i,j)$ represents the gray value of the pixel at (i,j) of the image before processing, *min* and *max* are respectively the minimum and maximum gray values. Finally, to improve the contrast between the vein and the non-vein regions, we use the CLAHE (Contrast Limited Adaptive Histogram Equalization) algorithm [54] to enhance the ROI images.

Based on the above steps, we obtain the ROI images for SDUMLA [49], FVUSM [51] and MMCBNU-6000 [50] datasets, which will be used in our experiments. Note that we don't use the provided ROI images in the FVUSM dataset. Generally, there are inconsistencies in the obtained ROI sizes, and the convolutional neural network requires images of same size as input. For this reason, we resize all the extracted ROI images from the three datasets to 128×320 pixels, and Fig. 3 shows the pre-processed ROI images of the three datasets.

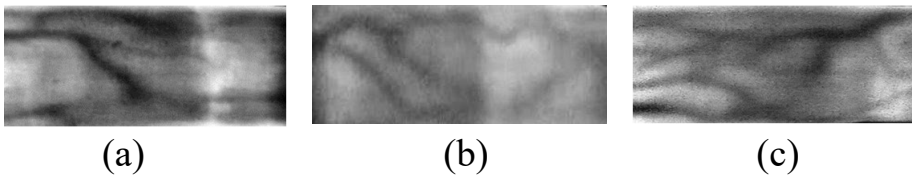


Fig. 3. ROIs on three datasets. (a). SDUMLA. (b). FVUSM (c). MMCBNU.

3.5. Backbone network

The backbone network of our model is the ResNet [55], which is used as a feature extractor. In this paper, the convolution operation at the last stage of the ResNet18 is modified so that the size of the feature map is not reduced at this stage. On this basis, we feed the shallow features of this network (the first three stages) into the feature pyramid module to upsample the small-size feature maps, and then fuse them with the feature maps of the corresponding sizes to obtain larger-size, more semantic features. Finally, to allow the output of the our modified Feature Pyramid Network (FPN) to be stitched together with the features of the last stage of the backbone network in the channel dimension, we connect two convolutional layers to the output of the FPN, each using a 3×3 convolutional kernel followed by a batch normalization operation and a ReLU activation function.

The shallow features of the convolutional neural network are mainly information such as contours and edges, and the deep features are mainly abstract high-level semantic features. Therefore, we expect to use the feature maps of the first few layers of the backbone network to generate the weight maps. Specifically, we use the FPN structure and directly take the output feature maps of the first three stages of the backbone network as input, which can include shallow features at different scales and avoid the above-mentioned selection dilemma. Figure 4 shows the modified FPN structure used in this paper.

Biomedical Optics EXPRESS

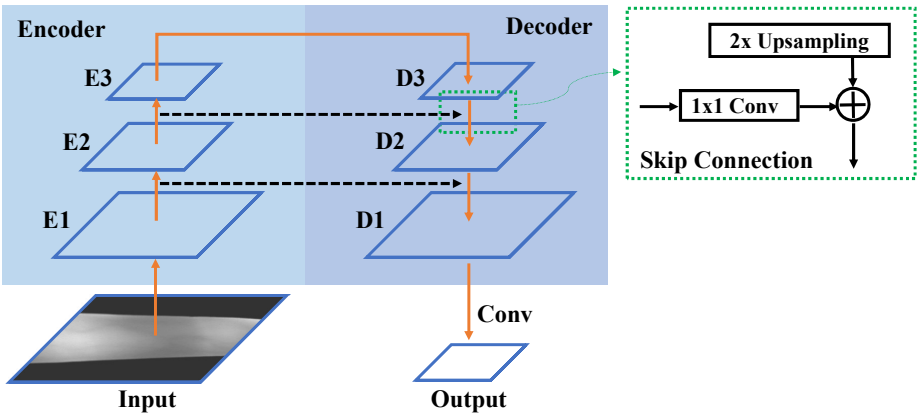


Fig. 4. The modified FPN. We add a convolution operation to the original one to make the output conform to the requirements of the proposed model.

3.6. *Mask guidance module*

The mask guidedance module is designed to pay more attention and assign higher weights to the vein regions in the learning process while reducing the attention to the non-vein regions. The primary features in the shallow feature map of the convolutional network are very useful for vein recognition, so we convert the shallow feature map into a spatial weight map by spatial attention, and use the mask obtained in advance to constrain the learning process of the network. The specific process is as follows. The output of FPN is sent to the spatial attention module, and the new feature map is generated by retaining key information through maximum and average pooling operations. Then the spatial attention map is generated using the convolutional layer and Sigmoid function.

Currently, there are relatively few studies using deep learning to explicitly extract finger vein patterns, and a key reason is the lack of publicly available datasets with manual annotation, which is time-consuming and very prone to over-segmentation or under-segmentation. Inspired by the literature [56], we employ five traditional vein skeleton segmentation algorithms including PC [57], RLT, WLD, GF [58], and IUWT [59], and fuse their results to generate the final mask.

The fusion is conducted as follows. If the gray value at a pixel of the image produced by a vein segmentation model is 255 then it is recorded as 1. Then the above binarized images are summed and if the summed value is greater than or equal to 3, i.e., three or more models consider the point as a vein point, this point is voted as a vein point and vice versa. These operations can be described as

$$N(x,y) = \sum_{i=1}^{5} f_i(x,y) \tag{3}$$

$$P(x,y) = \begin{cases} 255, & if\ N(x,y) \geq 3 \\ 0, & otherwise \end{cases} \tag{4}$$

where $f_i(x,y)$ is the binarized images of the skeleton map extracted by the i_{th} vein segmentation model, and $P(x,y)$ is the mask.

The segmentation results of the five models and the fused result are shown in Fig. 5. This fusion strategy can improve the accuracy of vein segmentation to a certain extent and make up for the shortcomings of individual segmentation models. The mask $P(x,y)$ obtained above is

Biomedical Optics EXPRESS

used to guide the learning of the weight map through the mean square error loss, written as

$$L_{mse} = \frac{1}{M \times N} \sum_{i=1}^{H} \sum_{j=1}^{W} \|P_{i,j} - Q_{i,j}\| \tag{5}$$

where H and W represent respectively the height and width of the weight map, and P and Q represent the mask and the spatial weight map, respectively.

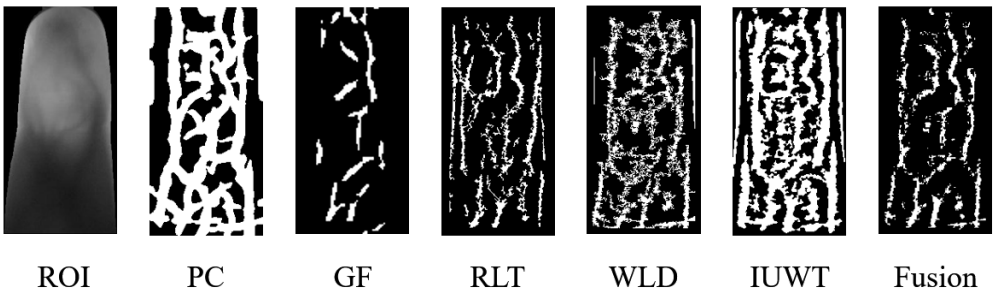


Fig. 5. Segmentation results of five traditional finger vein skeleton extraction algorithms and their fused result.

It should be emphasized that the proposed mask-guided attention mechanism is specifically tailored to the prior knowledge of finger vein recognition, which differs from both the fully self-learned attention and the hard attention that directly applies manually segmented masks. Our approach leverages the mask as a supervisory signal to autonomously identify veins of interest. This, in turn, directs the model's focus towards these critical regions in the finger vein recognition task, thereby imparting novelty to the proposed method. Although inaccurately segmented masks may impact finger vein recognition, the soft attention approach adopted in this work is more robust in learning from masks of different accuracies compared to hard attention commonly adopted by others.

3.7. Fusion loss function

Our model uses a weighted combination of triple loss and cross-entropy loss to train the network, aiming to not only correctly predict the finger category during training, but also strengthen the intra-class aggregation and inter-class separation of feature vectors in the Euclidean space.

Cross-entropy loss is a commonly used loss function in classification tasks. The vector corresponding to the number of category dimensions is obtained through the last fully connected layer of the network, and then normalized to the probability value P_{ic} of the corresponding category by the Softmax function, written as

$$L_{ce} = -\frac{1}{N} \sum_{i} \sum_{c=1}^{M} y_{ic} log(p_{ic}) \tag{6}$$

where N is the number of samples, M is the total number of categories, y_{ic} represents 0 or 1, y_{ic}=1 when the sample belongs to the c_{th} category and 0 if it does not, and p_{ic} represents the probability value that the given finger belongs to the c_{th} category for the i_{th} sample.

The triplet loss function constructs a triplet so that the distance between the anchor sample and the positive sample is smaller than the distance between the anchor sample and the negative

Biomedical Optics EXPRESS

sample, given by

$$L_{tri} = \sum_{i}^{N} [\|f(x_i^a) - f(x_i^a)\|_2^2 - \|f(x_i^a) - f(x_i^n)\|_2^2 + \theta]_+ \tag{7}$$

where x_i^a is the anchor sample, x_i^p represents the finger vein image belonging to the same category as the anchor sample, x_i^n represents an image belonging to a category other than the category of the anchor sample, θ represents the margin, and + means if the value is less than or equal to 0, it is set to 0, otherwise it keeps the original value.

The fusion loss function is calculated as

$$L_{all} = \alpha L_{ce} + \beta L_{tri} + \gamma L_{mse} \tag{8}$$

where α, β, and γ are the weights of the three loss functions, respectively. Specifically, α is the coefficient of the cross-entropy loss, β is utilized to modulate the weight of the triplet loss, while γ serves to adjust the segmentation loss. Based on our empirical observations, when α is set to 1, β is set to 4, and γ is set to 1, the model exhibits optimal performance.

4. Results

This section will demonstrate the effectiveness of the proposed method in terms of both qualitative and quantitative analysis, mainly including the comparison with baseline, ablation analysis, and comparison with other methods. It is important to note that the reported metric values of other methods in the comparison are cited directly from their original papers.

4.1. Implementation details

In this paper, we use the Pytorch deep learning framework to train our model on a Linux operating system with a NVIDIA TITAN RTX GPU with 24 GB memory. Stochastic gradient descent (SGD) is used to update the network parameters with an initial learning rate of 0.01, which becomes 0.001 at epoch 60, and the total epoch is set to 100. When using triplet loss, the batch size is 32 (four images are selected for each category, and a total of 8 categories are selected), when only using the cross entropy loss function, the batch size is 64. The margin is set to 0.2, and the image size is uniformly set to 128×320 pixels.

The three datasets used in this paper are first divided into training set and test set based on the number of categories in a ratio of 1:1, with no intersection between the two sets at all, and randomly selecting one image from each category in the training set form the validation set. Table 1 shows the detailed division of the three datasets. Among them, FVUSM is a two-stage data set, and we only selected the data of the first session. The test set is generated in the form of image pairs, containing genuine pairs and impostor pairs. Obviously, genuine pairs are much less in number than impostor pairs, and the situation of extreme imbalance in the number of positive and negative sample pairs will easily occur. Therefore, when generating impostor pairs, we first randomly select an image in each category, and then randomly select another image from each of other categories, which not only reduces the number of impostor pairs, but also makes each finger category compose an impostor pair with any other finger categories.

Table 1. Data division information for the employed three datasets.

Dataset	Total	Categories	Training	Valid	Genuine pairs	Impostor pairs
SDUMLA	3816	318	1590	318	4770	50403
MMCBNU	6000	300	2700	300	13500	44850
FVUSM	2952	246	1230	246	3690	30135

Biomedical Optics EXPRESS

4.2. Performance comparison before and after model improvement

In order to clearly understand the difference between the proposed method and the baseline, we obtained the confusion matrixes and the cosine similarity distribution maps between the feature vectors on the test sets of the three public datasets. Figure 6 and Fig. 7 show the confusion matrixes of the baseline and the proposed method on the SDUMLA, MMCBNU and FVUSM datasets. Among them, the results of the baseline were obtained using ROI, ResNet18 and cross-entropy loss function. As for the baseline, the accuracy rates of genuine pairs and impostor pairs on the SDUMLA, MMCBNU, and FVUSM datasets are 92.35%, 92.35%, 99.41%, 99.41%, 99.11%, and 99.14%, respectively. We can see that the accuracy rates of both intra-class matching and inter-class matching on the SDUMLA dataset are relatively low compared with the other two datasets, while the MMCBNU dataset has achieved promising results and the inter-class matching on the FVUSM dataset has higher accuracy rate. As for the results of the proposed algorithm, the accuracy rates of genuine pairs and impostor pairs on the three datasets reach 96.94%, 96.94%, 99.79%, 99.80%, 99.78%, and 99.93%, respectively, clearly higher than that achieved by the baseline. Specifically, the SDUMLA dataset has the largest improvement, which is due to the poor image quality of this dataset and the shallow structure of the baseline network, which cannot encode the vein features well. In contrast, the feature representation ability of the proposed network is largely enhanced, and thus can extract more distinguishable features. The best performance is achieved on the FVUSM dataset, where the inter-class matching accuracy reaches 99.93%, with a absolute improvement of 0.79%, and the intra-class matching accuracy is improved by 0.67%, demonstrating that the proposed network can further increase the distances between feature vectors of the same categories and decrease the distances between feature vectors of different categories. In addition, the accuracy of inter-class matching is higher than (or equal to) that of intra-class matching, proving that intra-class matching is more difficult in terms of recognition difficulty. This is due to the fact that the finger images are affected by some subjective or non-subjective factors among different image acquisitions, which may result in unstable feature vectors and false rejection. This suggests a valuable future direction to design new strategy to further compensate for these unwanted variations induced in the stage of image acquisition.

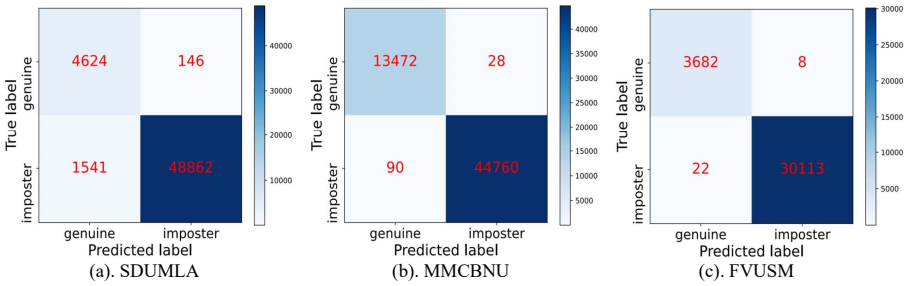
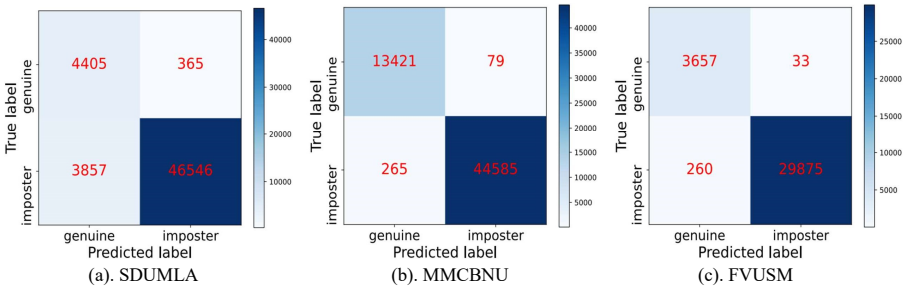


Fig. 6. Confusion matrix of the baseline model on the test set of different datasets, where genuine represents that the two images are of the same category and impostor represents that the two images are of different categories. (a). SDUMLA. (b). MMCBNU. (c). FVUSM.

Figure 8 shows the cosine distance distribution maps of the feature vectors, where the first row lists the results on the baseline, while the second row represents the results of the proposed method. The horizontal coordinate is the cosine distance between two feature vectors, the closer they are to 1, the more similar they are, and the closer they are to -1, the less similar they are. The vertical coordinate is the proportion of the corresponding distance. The distributions of genuine and impostor pairs are plotted in red and blue, respectively. We can see that when using the algorithm proposed in this paper, the overlapping area of intra-class matching and inter-class matching on the SDUMLA dataset is further reduced, which means that the feature discrimination

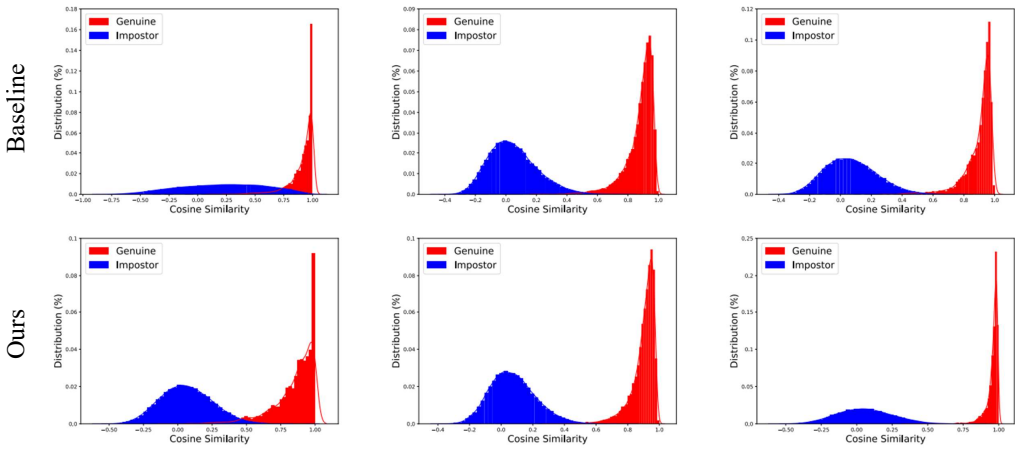
Biomedical Optics EXPRESS



(a). SDUMLA (b). MMCBNU (c). FVUSM

Fig. 7. The confusion matrix of the proposed method on the test set of different datasets, where genuine represents that the two images are of the same category and impostor represents that the two images are of different categories. (a). SDUMLA. (b). MMCBNU. (c). FVUSM.

becomes better and errors are less prone to occur. Especially for the distribution of inter-class matching, it is obvious that most of the blue areas are concentrated around 0, which means that the similarity between different classes of finger feature vectors has been largely reduced. On the MMCBNU dataset, the distances of intra-class matching are more concentrated and close to 1 for our proposed model. On the FVUSM dataset, our method produces almost no overlap between the distributions of intra-class matching and inter-class matching, which indicates that a quite low ERR can robustly obtained by easily selecting a similarity threshold been 0.6 and 0.7.



(a). SDUMLA (b). MMCBNU (c). FVUSM

Fig. 8. The distribution of cosine distances between feature vectors on the baseline and our method on the test sets of the three datasets, (a). SDUMLA. (b). MMCBNU. (c). FVUSM. The first row is for the baseline method and the second row for our method. Genuine represents that the two feature vectors are of the same category and impostor represents that they are of different categories.

4.3. Ablation experiments

To understand the contribution of each component of the proposed algorithm, we conducted ablation experiments on the SDUMLA, MMCBNU, and FVUSM datasets as shown in Table 2.

In each of the three tables, the first column lists six different types of masks (i.e., five masks extracted by five specific vein segmentation methods and one mask by fusing the five ones), the

Biomedical Optics EXPRESS

Table 2. Results of different masks on SDUMLA, MMCBNU, and FVUSM (EER:%). BL is ResNet18.

	SDUMLA				MMCBNU				FVUSM			
Mask type	BL	+SAM	+Mask	+Fusion	BL	+SAM	+Mask	+Fusion	BL	+SAM	+Mask	+Fusion
PC [57]	7.58	6.79	17.87	16.60	0.59	0.34	0.32	0.26	0.78	0.42	0.70	0.16
RLT [9]			7.70	4.46			2.22	0.21			0.41	0.19
WLD [23]			20.98	3.08			0.44	0.36			1.12	0.40
GF [58]			11.88	3.32			0.31	0.33			0.65	0.30
IUWT [59]			5.66	3.21			0.48	0.37			0.37	0.20
Fusion			6.06	**2.50**			0.28	**0.20**			0.35	**0.14**

second column represents the results obtained on the ROI images using the ResNet18 network without using any masks, the third column "+SAM" represents the results of adding FPN and SAM to the original network without using any masks, and the fourth column "+Mask" represents the results of adding various masks as guidance on the basis of the third column, and the fifth column "+Fusion" represents the results of adding a fusion loss function on the basis of the fourth column. The data in the three tables are the EER values, with the best value being bolded. On the whole, the EER values are relatively high on the SDUMLA dataset and low on the MMCBNU and FVUSM datasets, because the SDUMLA dataset has poor image quality due to rotation, overexposure, and low contrast, while the other two datasets have relatively good image quality, with more visible vein and non-vein areas and less complex background.

Although ResNet18 solves the gradient disappearance problem in the form of residual structure, for the finger vein extraction task, ResNet18 may be too shallow to focus on the vein features, so there is much room for improvement. After adding FPN and SAM, large improvements are achieved on the three datasets. This is because the input of FPN is the shallow feature map of the network, and the spatial weight map is obtained by fusing information such as primary features (texture, edge), and then superimposed with the deep features to improve the attention of the network to the vein area, which helps realize the effective extraction of vein features.

For the results guided by six different types of masks, we found that the results of combining the masks of five traditional methods are more stable, and using one mask alone may decrease rather than increase the performance of the model (e.g., PC, WLD, etc. on SDUMLA). This is because the traditional vein skeleton extraction methods require high image quality and are prone to mis-segmentation or over-segmentation, and using such masks may mislead the learning of the network. Furthermore, using the fusion loss function on the basis of mask guidance can increase the inter-class separation and intra-class aggregation of feature vectors on the Euclidean space, making the features more discriminative, thus improving the network performance.

The ROC curves for the three datasets are shown in Fig. 9. The AUC (Area Under Curve) values listed in the lower right corner of each figure also indicate that the introduced strategies are effective in improving the network performance.

4.4. Comparison with other methods

In order to verify the performance of the method proposed in this paper, we select some traditional methods and deep learning based state-of-the-art ones to compare on the three datasets, and the results are shown in Table 3. As mentioned before, all the ERR values of the existing methods are directly cited from their original papers. It is obvious that on the datasets of MMCBNU and FVUSM, our method proposed in this work outperforms all the compared methods in terms of ERR, as shown in Table 3. While on the SDUMLA dataset (Table 3), our model performs better than others, only slightly worse than the method of Lu *et al.* [33] (2.50 vs. 2.37). We noticed

Biomedical Optics EXPRESS

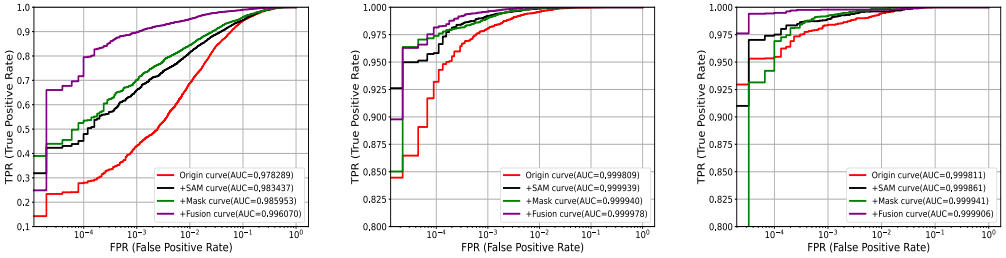


Fig. 9. ROC curves on the three datasets. From left to right, SDUMLA, MMCBNU, and FVUSM.

that the results of all the methods on the SDUMLA dataset are at an unsatisfactory level, which is mainly due to the image quality problem of this dataset.

Table 3. Comparison with other methods on SDUMLA, MMCBNU, and FVUSM (Unit:%).

SDUMLA			MMCBNU			FVUSM		
Method	Year	EER	Method	Year	EER	Method	Year	EER
MC [10]	2005	3.65	MC [10]	2005	1.12	MC [10]	2005	2.63
ITQM [60]	2017	2.78	ITQM [60]	2017	1.33	ITQM [60]	2017	1.05
Yang *et al.* [25]	2017	3.46	Kang *et al.* [61]	2018	0.827	Kang *et al.* [61]	2018	0.216
Lu *et al.* [33]	2019	**2.37**	Yang *et al.* [62]	2019	0.42	Zhang *et al.* [18]	2019	0.57
Choi *et al.* [63]	2020	3.93	Yang *et al.* [32]	2020	1.11	Yang *et al.* [32]	2020	0.95
SRLRR [64]	2021	3.75	Ou *et al.* [36]	2021	1.21	Ou *et al.* [36]	2021	0.48
Zhao *et al.* [65]	2023	3.61	Yang *et al.* [66]	2023	0.42	Zhao *et al.* [65]	2023	2.55
Hong *et al.* [67]	2024	4.72	Qin *et al.* [68]	2024	1.00	Yang *et al.* [66]	2023	0.61
Ours	2024	2.50	Ours	2024	**0.20**	Ours	2024	**0.14**

4.4.1. Analysis of performance

As the compared methods have neither reported their parameter counts nor computational complexity, nor provided accessible code for reproducibility, it is challenging to directly compare the computational complexity of the proposed approach with that of those methods. Consequently, we solely compared our model against the baseline to observe the impact of the introduced modules on performance, as shown in Table 4. In terms of parameter counts, the proposed model contains only 13.36M parameters, representing a mere 0.1M increase over the baseline. Regarding computational complexity, the difference in FLOPs between our model and the baseline is negligible, with only a slight increase in execution time on both CPU and GPU. In summary, our model achieves substantial performance gains while introducing minimal additional parameters and computational complexity, rendering the benefits highly significant and desirable.

4.5. Failure cases

We select two images from the SDUMLA dataset that were incorrectly identified on the validation set, as shown in Fig. 10. Among them, '024/left/ring_6.bmp' represents the name of the original image, and '10' represents the output category of the proposed method. We checked the ROI of the corresponding category to be recognized and the original image, as indicated by the arrow, and observed that this original image has a large axial rotation compared with other five fingers of this category, which might be the main reason leading to the image incorrectly recognized. To further improve the performance of the model in the future, we can start from minimizing the

Biomedical Optics EXPRESS

Table 4. Illustration of the computational complexity and parameters, as well as the inference time for 32 tensors with shape of $64 \times 3 \times 120 \times 320$ on CPU/GPU. (CPU: Intel Xeon CPU E5-2620 v4 @ 2.10GHz. GPU: 24 GB TITAN RTX. ITC: Inference Time on CPU. ITG: Inference Time on GPU.)

Method	FLOPs(G)	Param(M)	ITC (s)	ITG (s)
ResNet18	170.16	12.76	11.05	0.85
Ours	170.26	13.36	11.51	0.93

intra-class variation and solve the problems of rotation translation (as indicated in Fig. 10(a)) and uneven illumination (as indicated in Fig. 10(b)) among fingers of the same categories due to acquisition and other factors. Achieving robust finger vein image enhancement serves as an ideal approach, exemplified by the success of 3DFD U-net in producing clear 3D vascular images of the palm, arms, breasts, and feet of human subjects [69]. This is of great significance for the recognition of finger vein images that are often blurred and contaminated with noise.

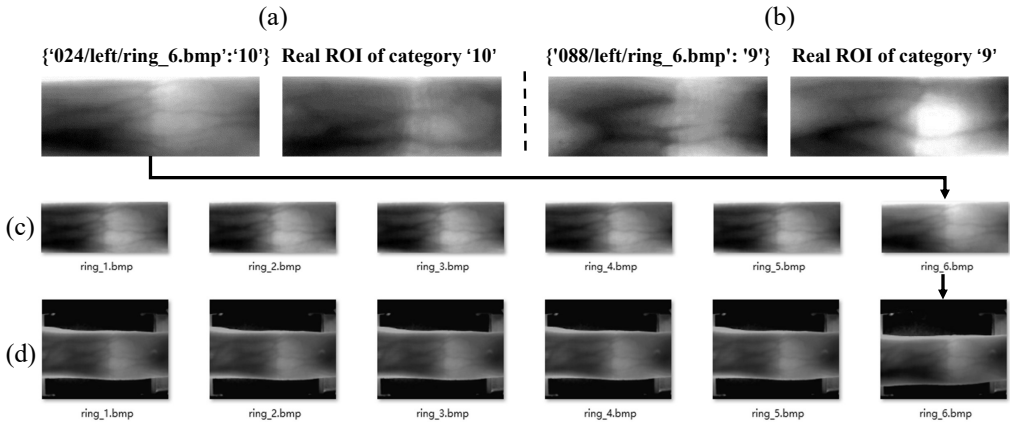


Fig. 10. Two failure cases of the proposed model on SDUMLA. (a) The ROI of the input image 024/*left*/*ring*_6.*bmp* (left) is wrongly recognized as the category 10, whose real ROI is shown on its right side. (b) The ROI of the input image 088/*left*/*ring*_6.*bmp* (left) is wrongly recognized as the category 9, whose real ROI is shown on the right side. (c) All the six ROIs corresponding to the category of 024/*left*/*ring*_6.*bmp* (left of (a)). (d) All the six original images corresponding to the six ROIs listed in (c). The red arrows indicate that the image *ring*_6.*bmp* in (d) is the original finger vein image of the ROI in (a) that is wrongly recognized as the category 10 by the proposed model.

5. Conclusion

In this paper, our primary contributions encompass the design of a novel feature extraction framework specifically tailored for finger vein images, as well as the introduction of a mask attention guidance module grounded on finger vein priors. Within the proposed model, our work, on the one hand, validates the varying degrees of enhancement provided by diverse manual mask extraction methods for finger vein recognition, and meticulously selects the optimal mask fusion strategy to effectively guide the model's feature attention. On the other hand, our work demonstrates the superiority of the multi-faceted approach over other state-of-the-art algorithms.

A notable drawback of the proposed model lies in its reliance on conventional finger vein segmentation algorithms. The finger veins extracted by various segmentation algorithms can vary

Biomedical Optics EXPRESS

significantly. In scenarios where the traditional finger vein segmentation algorithms exhibit poor generalization, the proposed model may experience performance degradation or even failure.

As future work, we aim to introduce a deep model for finger vein segmentation and further incorporate the structural information of them into the classification network, with the goal of achieving more interpretable finger vein recognition.

Funding. Huzhou Science and Technology Program (#2023GZ13).

Disclosures. The authors declare no conflicts of interest.

Data availability. Data underlying the results presented in this paper are available in Ref. [49], Ref. [50], and Ref. [51].

References

1. A. K. Jain, A. Ross, and S. Prabhakar, "An introduction to biometric recognition," IEEE Trans. Circuits Syst. Video Technol. **14**(1), 4–20 (2004).
2. W. Zhao, R. Chellappa, P. J. Phillips, *et al.*, "Face recognition: A literature survey," ACM Comput. Surv. **35**(4), 399–458 (2003).
3. T.-Y. Jea and V. Govindaraju, "A minutia-based partial fingerprint recognition system," Pattern Recognition **38**(10), 1672–1684 (2005).
4. J. Hashimoto, "Finger vein authentication technology and its future," in *2006 Symposium on VLSI Circuits, 2006. Digest of Technical Papers.*, (IEEE, 2006), pp. 5–8.
5. Y. Zhou and A. Kumar, "Human identification using palm-vein images," IEEE Trans.Inform.Forensic Secur. **6**(4), 1259–1274 (2011).
6. R. P. Wildes, "Iris recognition: an emerging biometric technology," Proc. IEEE **85**(9), 1348–1363 (1997).
7. M. Kono, "A new method for the identification of individuals by using of vein pattern matching of a finger," in *Proc. Fifth Symposium on Pattern Measurement, Yamaguchi, Japan, 2000*, (2000), pp. 9–12.
8. Y. Zhan, A. Singh Rathore, G. Milione, *et al.*, "3d finger vein biometric authentication with photoacoustic tomography," Appl. Opt. **59**(28), 8751–8758 (2020).
9. N. Miura, A. Nagasaka, and T. Miyatake, "Feature extraction of finger-vein patterns based on repeated line tracking and its application to personal identification," Machine vision and applications **15**(4), 194–203 (2004).
10. N. Miura, A. Nagasaka, and T. Miyatake, "Extraction of finger-vein patterns using maximum curvature points in image profiles," IEICE Transactions on Information and Systems **E90-D**(8), 1185–1194 (2007).
11. H. C. Lee, B. J. Kang, E. C. Lee, *et al.*, "Finger vein recognition using weighted local binary pattern code based on a support vector machine," J. Zhejiang Univ. - Sci. C **11**(7), 514–524 (2010).
12. B. A. Rosdi, C. W. Shing, and S. A. Suandi, "Finger vein recognition using local line binary pattern," Sensors **11**(12), 11357–11371 (2011).
13. H.-G. Kim, E. J. Lee, G.-J. Yoon, *et al.*, "Illumination normalization for sift based finger vein authentication," in *Advances in Visual Computing: 8th International Symposium, ISVC 2012, Rethymnon, Crete, Greece, July 16-18, 2012, Revised Selected Papers, Part II 8*, (Springer, 2012), pp. 21–30.
14. J. Peng, N. Wang, A. A. Abd El-Latif, *et al.*, "Finger-vein verification using gabor filter and sift feature matching," in *2012 eighth international conference on intelligent information hiding and multimedia signal processing*, (IEEE, 2012), pp. 45–48.
15. J.-D. Wu and C.-T. Liu, "Finger-vein pattern identification using principal component analysis and the neural network technique," Expert Syst. Appl. **38**(5), 5423–5427 (2011).
16. H. Ma, "Finger vein identification based on 2dpca," in *Advanced Materials Research*, vol. 988 (Trans Tech Publ, 2014), pp. 548–551.
17. S. A. Radzi, M. K. Hani, and R. Bakhteri, "Finger-vein biometric identification using convolutional neural network," Turkish Journal of Electrical Engineering and Computer Sciences **24**, 1863–1878 (2016).
18. Y. Zhang, W. Li, L. Zhang, *et al.*, "Adaptive learning gabor filter for finger-vein recognition," IEEE Access **7**, 159821–159830 (2019).
19. J. M. Song, W. Kim, and K. R. Park, "Finger-vein recognition based on deep densenet using composite image," IEEE Access **7**, 66845–66863 (2019).
20. L. Lin, H. Liu, W. Zhang, *et al.*, "Finger vein verification using intrinsic and extrinsic features," in *2021 IEEE International Joint Conference on Biometrics (IJCB)*, (IEEE, 2021), pp. 1–7.
21. J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *CVPR Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, (2018), pp. 7132–7141.
22. S. Woo, J. Park, J.-Y. Lee, *et al.*, "Cbam: Convolutional block attention module," in *ECCV Proc. Eur. Conf. Comput. Vis.*, (2018), pp. 3–19.
23. B. Huang, Y. Dai, R. Li, *et al.*, "Finger-vein authentication based on wide line detector and pattern normalization," in *2010 20th international conference on pattern recognition*, (IEEE, 2010), pp. 1269–1272.
24. Y. Lu, S. Yoon, S. J. Xie, *et al.*, "Finger vein recognition using generalized local line binary pattern," KSII Transactions on Internet Information Systems **8**, 1 (2014).

Biomedical Optics EXPRESS

25. L. Yang, G. Yang, X. Xi, *et al.*, "Tri-branch vein structure assisted finger vein recognition," IEEE Access **5**, 21020–21028 (2017).
26. F. Liu, G. Yang, Y. Yin, *et al.*, "Singular value decomposition based minutiae matching method for finger vein recognition," Neurocomputing **145**, 75–89 (2014).
27. C.-B. Yu, H.-F. Qin, Y.-Z. Cui, *et al.*, "Finger-vein image recognition combining modified hausdorff distance with minutiae feature matching," Interdisciplinary Sciences: Computational Life Sciences **1**, 280–289 (2009).
28. H. Qin, L. Qin, L. Xue, *et al.*, "Finger-vein verification based on multi-features fusion," Sensors **13**(11), 15048–15067 (2013).
29. J.-D. Wu and C.-T. Liu, "Finger-vein pattern identification using svm and neural network technique," Expert Syst. Appl. **38**, 14284–14289 (2011).
30. B. Hou and R. Yan, "Convolutional auto-encoder based deep feature learning for finger-vein verification," in *2018 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, (IEEE, 2018), pp. 1–5.
31. E. Jalilian and A. Uhl, "Finger-vein recognition using deep fully convolutional neural semantic segmentation networks: the impact of training data," in *2018 IEEE International Workshop on Information Forensics and Security (WIFS)*, (IEEE, 2018), pp. 1–8.
32. W. Yang, W. Luo, W. Kang, *et al.*, "Fvras-net: an embedded finger-vein recognition and antispoofing system using a unified cnn," IEEE Trans. Instrum. Meas. **69**(11), 8690–8701 (2020).
33. Y. Lu, S. Xie, and S. Wu, "Exploring competitive features using deep convolutional neural network for finger vein recognition," IEEE Access **7**, 35113–35123 (2019).
34. Z. Zhang, Z. Zhou, X. Yang, *et al.*, "Convolutional neural network based on multi-directional local coding for finger vein recognition," Inf. Sci. **623**, 633–647 (2023).
35. B. Hou and R. Yan, "Arcvein-arccosine center loss for finger vein verification," IEEE Trans. Instrum. Meas. **70**, 1–11 (2021).
36. W.-F. Ou, L.-M. Po, C. Zhou, *et al.*, "Fusion loss and inter-class data augmentation for deep finger vein feature learning," Expert. Syst. Appl. **171**, 114584 (2021).
37. J. Shen, N. Liu, C. Xu, *et al.*, "Finger vein recognition algorithm based on lightweight deep convolutional neural network," IEEE Trans. Instrum. Meas. **71**, 1–13 (2021).
38. Y. Song, P. Zhao, W. Yang, *et al.*, "Eifnet: an explicit and implicit feature fusion network for finger vein verification," IEEE Trans. Circuits Syst. Video Technol. **33**(5), 2520–2532 (2022).
39. J. Huang, A. Zheng, M. S. Shakeel, *et al.*, "Fvfsnet: Frequency-spatial coupling network for finger vein authentication," IEEE Trans.Inform.Forensic Secur. **18**, 1322–1334 (2023).
40. P. Zhao, S. Zhao, L. Chen, *et al.*, "Exploiting multiperspective driven hierarchical content-aware network for finger vein verification," IEEE Trans. Circuits Syst. Video Technol. **32**(11), 7938–7950 (2022).
41. Z. Zhang, F. Zhong, and W. Kang, "Study on reflection-based imaging finger vein recognition," IEEE Trans.Inform.Forensic Secur. **17**, 2298–2310 (2021).
42. Q. Wang, B. Wu, P. Zhu, *et al.*, "Eca-net: efficient channel attention for deep convolutional neural networks," in *CVPR Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, (2020), pp. 11534–11542.
43. J. Huang, M. Tu, W. Yang, *et al.*, "Joint attention network for finger vein authentication," IEEE Trans. Instrum. Meas. **70**, 1–11 (2021).
44. Z. Zhang and M. Wang, "Convolutional neural network with convolutional block attention module for finger vein recognition," arXiv, (2022).
45. A. Dosovitskiy, L. Beyer, A. Kolesnikov, *et al.*, "An image is worth 16x16 words: transformers for image recognition at scale," arXiv, (2020).
46. J. Huang, W. Luo, W. Yang, *et al.*, "Fvt: finger vein transformer for authentication," IEEE Trans. Instrum. Meas. **71**, 1–13 (2022).
47. H. Qin, R. Hu, M. A. El-Yacoubi, *et al.*, "Local attention transformer-based full-view finger-vein identification," IEEE Trans. Circuits Syst. Video Technol. **33**(6), 2767–2782 (2022).
48. Z. Zhao, H. Zhang, Z. Chen, *et al.*, "Transfinger: Transformer based finger tri-modal biometrics," in *Biometric Recognition: 16th Chinese Conference, CCBR 2022, Beijing, China, November 11–13, 2022, Proceedings*, (Springer, 2022), pp. 114–124.
49. Y. Yin, L. Liu, and X. Sun, "Sdumla-hmt: a multimodal biometric database," in *Biometric Recognition: 6th Chinese Conference, CCBR 2011, Beijing, China, December 3-4, 2011. Proceedings 6*, (Springer, 2011), pp. 260–268.
50. Y. Lu, S. J. Xie, S. Yoon, *et al.*, "An available database for the research of finger vein recognition," in *2013 6th International congress on image and signal processing (CISP)*, vol. 1 (IEEE, 2013), pp. 410–415.
51. M. S. M. Asaari, 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. **41**(7), 3367–3382 (2014).
52. T.-Y. Lin, P. Dollár, R. Girshick, *et al.*, "Feature pyramid networks for object detection," in *CVPR Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, (2017), pp. 2117–2125.
53. L. Yang, G. Yang, Y. Yin, *et al.*, "Sliding window-based region of interest extraction for finger vein images," Sensors **13**(3), 3799–3815 (2013).
54. K. Zuiderveld, "Contrast limited adaptive histogram equalization," Graphics Gems (1994), pp. 474–485.
55. K. He, X. Zhang, S. Ren, *et al.*, "Deep residual learning for image recognition," in *CVPR Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, (2016), pp. 770–778.

Biomedical Optics EXPRESS

56. H. Qin and M. A. El-Yacoubi, "Deep representation-based feature extraction and recovering for finger-vein verification," IEEE Trans.Inform.Forensic Secur. **12**(8), 1816–1829 (2017).
57. J. H. Choi, W. Song, T. Kim, *et al.*, "Finger vein extraction using gradient normalization and principal curvature," in *Image Processing: Machine Vision Applications II*, vol. 7251 (SPIE, 2009), pp. 359–367.
58. A. Kumar and Y. Zhou, "Human identification using finger images," IEEE Trans.Inform.Forensic Secur. **21**(4), 2228–2244 (2011).
59. J.-L. Starck, J. Fadili, and F. Murtagh, "The undecimated wavelet decomposition and its reconstruction," IEEE Trans. on Image Process. **16**(2), 297–309 (2007).
60. K. Wang, L. Yang, G. Yang, *et al.*, "Integration of discriminative features and similarity-preserving encoding for finger vein image retrieval," in *Proc. Int. Conf. Image Process*, (IEEE, 2017), pp. 3525–3529.
61. W. Kang, Y. Lu, D. Li, *et al.*, "From noise to feature: Exploiting intensity distribution as a novel soft biometric trait for finger vein recognition," IEEE Trans. Inform. Forensic Secur. **14**(4), 858–869 (2018).
62. L. Yang, G. Yang, K. Wang, *et al.*, "Point grouping method for finger vein recognition," IEEE Access **7**, 28185–28195 (2019).
63. J. Choi, K. J. Noh, S. W. Cho, *et al.*, "Modified conditional generative adversarial network-based optical blur restoration for finger-vein recognition," IEEE Access **8**, 16281–16301 (2020).
64. L. Yang, G. Yang, K. Wang, *et al.*, "Finger vein recognition via sparse reconstruction error constrained low-rank representation," IEEE Trans. Inform. Forensic. Secur. **16**, 4869–4881 (2021).
65. P. Zhao, Z. Chen, J.-H. Xue, *et al.*, "Single-sample finger vein recognition via competitive and progressive sparse representation," IEEE Trans. Biom. Behav. Identity Sci. **5**(2), 209–220 (2022).
66. L. Yang, X. Xu, and Q. Yao, "Finger vein recognition based on unsupervised spiking neural network," in *Chinese Conference on Biometric Recognition*, (Springer, 2023), pp. 55–64.
67. J. S. Hong, S. G. Kim, J. S. Kim, *et al.*, "Deep learning-based restoration of multi-degraded finger-vein image by non-uniform illumination and noise," Eng. Appl. Artif. Intel. **133**, 108036 (2024).
68. H. Qin, C. Fan, S. Deng, *et al.*, "Ag-nas: An attention GRU-based neural architecture search for finger-vein recognition," IEEE Trans. Inform. Forensic Secur. **19**, 1699–1713 (2023).
69. W. Zheng, H. Zhang, C. Huang, *et al.*, "Deep learning enhanced volumetric photoacoustic imaging of vasculature in human," Adv. Sci. **10**(29), 2301277 (2023).