GANet: gabor attention aggregation network for palmvein identification

Hongchao Liao, Xin Jin, Hongyu Zhu, Yuming Fu, Mounim El Yacoubi, Huafeng Qin

GANet: Gabor Attention Aggregation Network for Palmvein Identification

1st Hongchao Liao
School of Artificial Intelligence
CTBU
Chongqing, China
lhc980503@163.com

2nd Xin Jin
School of Artificial Intelligence
CTBU
Chongqing, China
jinxin20001118@163.com

3rd Hongyu Zhu
School of Artificial Intelligence
CTBU
Chongqing, China
zhuhongyu@ctbu.edu.cn

4th Yuming Fu
School of Artificial Intelligence
CTBU
Chongqing, China
1057726877@qq.com

5th Mounim A. El Yacoubi
School of Artificial Intelligence
Telecom SudParis, Institut Polytechnique de Paris
Paris, France
mounim.el_yacoubi@telecom-sudparis.eu

6th Huafeng Qin
School of Artificial Intelligence
CTBU
Chongqing, China
qinhuafengfeng@163.com

***Abstract*—Palm vein recognition has attracted recently wide attention thanks to its robust feature representation and high accuracy. Despite advancements in the literature, however, existing solutions suffer from the following issues: 1) Insufficient large-scale data for deep learning-based recognition of vein biometrics, resulting in decreased generalization performance and model accuracy. 2) Lack of methods based on machine learning convolutional neural networks capable of capturing the global receptive field for vein biometric recognition. In addressing these issues, this paper proposes a method to acquire the global receptive field, termed GANet, which extracts vein features using Gabor filters and computes an attention mechanism to capture the global receptive field for downstream palm vein recognition models. Initially, vein features are extracted using multi-scale fixed Gabor filters and multi-scale adaptive Gabor filters. Subsequently, self-attention mechanisms are employed to compute relationships between blocks to obtain the global receptive field. To perform recognition, the Euclidean distance between feature vectors is then computed. Our experiments on three datasets show that our approach outperforms existing palm vein recognition methods.**

***Index Terms*—palm vein recognition, Gabor filters, global receptive field**

I. Introduction

Traditional personal identification techniques for information security, *e.g.* keys, and passwords, are too simple for everyday use. Keys can be easily copied and lost, while passwords are frequently forgotten. Biometric identification can be divided into two classes: 1) external, e.g., Face [1], AND Finger print [2]; 2) internal, e.g., Palm-vein [3], Finger-vein [4], and Hand-vein [6]. External biometric identification is widely used in various settings, *e.g.* Control systems, Financial transactions, and electronic payment. However, external biometric identification can be easily duplicated without the user's consent [7] [8], leading to privacy concerns. In contrast, internal biometric features, hidden within the human body, are less prone to duplication and offer inherent liveness detection, providing a high level of security and privacy [6] [7].

Currently, vein identification still with some challenging issues due to a variety of external factors that affect image acquisition, *e.g.* Light [9], Temperature [10], Light scattering [11], identification of some user behaviors [10], identification of dryness or wetness of objects. These factors can add some noise or irregular shadow areas in the images, leading to lower image identification rates. In order to address this issue, researchers have proposed machine learning (ML) and deep learning (DL) methods to improve model robustness and identification rates. DL methods, including *Convolutional Neural Networks* (CNNs), have recently demonstrated significant robustness in computer vision tasks, *e.g.* Classification [12] [21], and Object Detection, *etc.* In recent years, a large number of research works proposed DL methods for vein identification [3] [4] [5]. However, Deep Neural Networks require a large number of training parameters, which can lead to reduced generalization performance and lower model accuracy when dealing with small-scale vein datasets. Additionally, the existing ML methods for vein identification cannot capture a global receptive field of vein images, resulting in inadequate performance and generalization capabilities in vein identification.

To enhance model performance, we propose **GANet** for palm vein identification. We summarize our contributions as follows:

- We propose **GANet**, a Gabor Attention Aggregation Network, for palm vein identification. Specifically, we extract vein texture features by incorporating fixed Multi-scale Gabor filters and Adaptive-scale Gabor filters into CNN model.
- By introducing an attention module into the CNN, we enable the latter to learn long-range dependencies within images and acquire a global receptive field, improving thereby the network's performance and generalization

CTBU denoted Chongqing Technology and Business University.

ability in vein recognition tasks.
- We report several experiments on three palm vein datasets to assess **GANet**'s performance. Experimental results show that **GANet**, a ML-based palm vein identification network, outperforms mainstream palm vein identification networks both on DL and ML in terms of *Equal Error Rate* (EER) and Accuracy.

II. Related Work

Palm vein identification methods usually identify individuals by assessing vein feature representations extracted by network models. Thanks to the robustness of vein texture information, and more importantly, the liveness detection capability of veins, vein identification offers high accuracy, security, and robustness. Consequently, vein identification has been widely applied across various industries. To enhance vein identification accuracy, researchers have proposed various methods for extracting vein features. These methods can be divided into ML-based methods and DL-based methods.

a) Machine Learning-based Methods: ML-based methods mainly rely on some hand-crafted algorithms and shallow learning techniques designed by researchers based on their prior knowledge. Hand-crafted methods depend on manually designing algorithm models for vein feature extraction and classification, including curvature-based methods [22], Gabor filter-based methods [10] [11] [28], and local binary descriptor-based methods [23]. Shallow learning methods employ traditional ML techniques to extract vein features for classification tasks. Representative methods, *e.g.* k-means clustering [24], Support Vector Machine (SVM) [25], Principal Component Analysis (PCA) [26], Two-Dimensional Principal Component Analysis (2D-PCA) [27], and Sparse Representation (SR) [29] have been proposed for vein identification tasks. To improve the performance, Low-Rank Representation (LRR) [30] has been used to reduce the rank of extracted vein features, while adding regularization terms to constrain the low-rank coefficients to improve discriminative power.

b) Deep Learning-based Methods: DL-based methods use DNNs, e;g., CNNs and Transformers, as they were shown to extract robust feature representations with proven effectiveness in various computer vision tasks [14] [39]. This led researchers to apply DL to vein image quality assessment, vein texture segmentation [3], and vein identification [20] [13]. For example, FV-CNN [16] is a convolutional-based model specifically designed for vein identification. To improve the efficiency of models, LightweightCNN [18] proposed a lightweight version of the vein identification model based on a triplet loss function. To extract some robust features, Arcvein [20] introduced a method that utilizes the cosine center loss function to simultaneously learn inter-class and intra-class information, enhancing the network's recognition ability in finger vein recognition. The FVRASNet [19] literature proposes a lightweight CNN model that can be used for recognition and anti-spoofing tasks. Recently, methods based on Transformer have been applied to vein recognition to obtain global feature information of vein texture.Some noteworthy studies [17] are very important in the task of vein recognition.

III. The Proposed Approach

In this section, we present our methodologies **GANet**, a CNN model aggregate with Gabor filters and a Self-attention module, as illustrated in Figure.1.

A. CNN Training

In this section, we provide a detailed explanation of the complete training process of **GANet**. Initially, we introduce the overall framework of the network, followed by detailing the selection process based on Gabor filters. Lastly, we present the parameters of the proposed **GANet** framework.

1) Network Architecture: The GANet proposed in this paper comprises a network framework with two convolutional layers (*L1* and *L2*) and one binarization layer (*L3*), with the network's input layer defined as *L0*. Figure Figure.2 illustrates the overall architecture of the CNN. Details of each layer are explained as follows:

- *L0*: The dimensions of the input layer are $u \times v$, corresponding to the size of the vein region in the palm vein image.
- *L1*: The initial Gabor layer consists of k_n Gabor filters, and each filter is applied to the vein image at the input layer, resulting in k_n venous feature maps at the *L1* layer.
- *L2*: The second layer of the CNN comprises a total of k_m Gabor filters, with each filter applied to one of the k_n vein texture feature images from the first layer. Consequently, the output of the *L2* layer consists of k_n sets, with each set containing k_m vein texture feature maps.
- *L3*: The third layer of the network is the image binarization layer, which applies the function $f(x) = bin(x)$ to every pixel in each of the $k_n \times k_m$ vein feature maps output from the second layer, to binarize the images. Here, $bin(x)$ is defined by equation 1.

$$bin(x) = \begin{cases} 1, & x > 0, \\ 0, & otherwise. \end{cases} \tag{1}$$

2) The filters in both layers *L1* and *L2* are of varying scales and orientations, and are configured as convolutional kernel filters during the training process of vein images. In our proposed method, we utilize filters of different orientations and scales. Gabor filter, which is currently employed in various state-of-the-art vein identification methods due to the ability to effectively extract texture features of veins, thereby improving vein classification accuracy. The Gabor filter is specifically designed for vein features, with its shape defined by the product of a sine wave and a Gaussian function [31].

$$G_f(x, y, \theta, \mu, \sigma) = \frac{1}{2\pi\sigma^2} \cdot e^{-\frac{x^2+y^2}{2\sigma^2}} \cdot e^{2\pi j(\mu x \cos\theta_f + \mu y \sin\theta_f)}, \tag{2}$$

where θ_f represents the direction of the function, σ is the standard deviation, μ is the frequency of the sine wave, and $j = \sqrt{-1}$. $G_f(x, y, \theta, \mu, \sigma)$. $G_f(x, y, \theta, \mu, \sigma)$ contains both

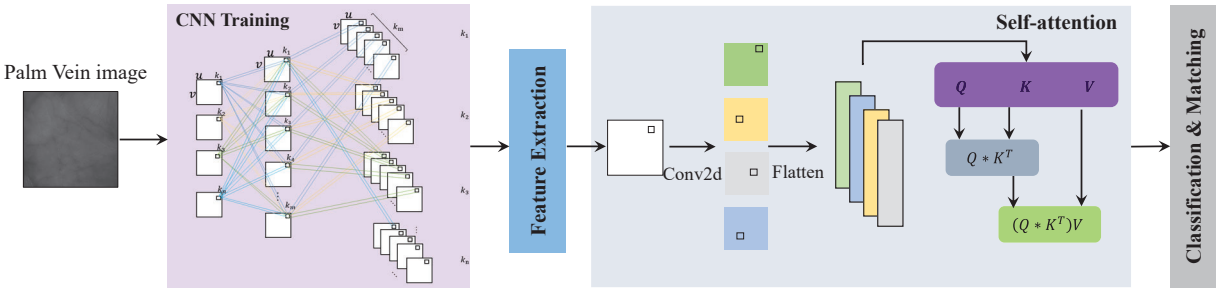


Fig. 1. Palm Vein Recognition Framework Based on GANet.

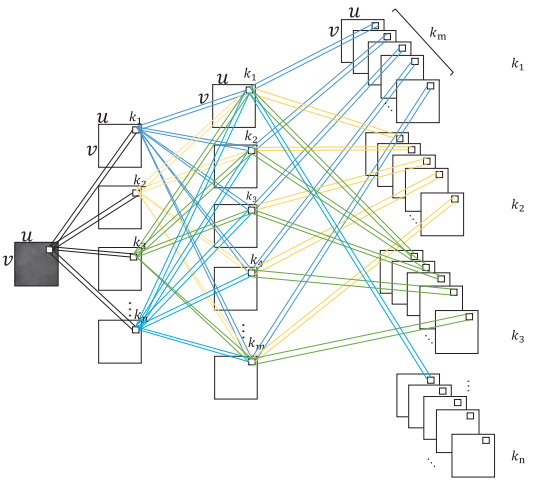


Fig. 2. Topology of the proposed GANet.

real and imaginary parts, but in our specific experiments, only the real part was utilized.

B. Feature Extraction

The feature extraction process can be divided into two steps: I) Loading the pre-trained Gabor filters; II) Encoding the palm vein images.

I) We loading a pre-trained CNN to the **GANet**. Thus, we obtain $k_n k_m$ binary images. Specifically, each set of k_m binary images output from *L3* layers corresponds to one of the k_n images output from *L1* layer, as illustrated in Figure.2. In this way, obtained a total of k_n sets, each containing k_m images.

II) Initially, we consider a set of k_m binary images denoted as B_i, where $i = 1, 2, ..., c$. Each image within this set has dimensions of $u \times v$. For each position (x, y) in the images, a binary vector b is constructed by concatenating the binary values from all k_m images as follows: $b = [B_1(x, y), B_2(x, y), ..., B_{k_m}(x, y)]$. Subsequently, the binary vector b is encoded into a decimal number d using the formula $d = \sum_{j=1}^{k_m} 2^{j-1} b(j)$. This process encodes each position (x, y) to produce the matrix $D(x, y)$. Similarly, the decimal matrices D_l are computed for all k_m sets of images, where $l = 1, 2, ..., k_m$.

C. Slef-Attention Module

The self-attention module takes the query $Q \in \mathbb{R}^{N_q \times C}$, key $K \in \mathbb{R}^{N_k \times C}$, and value $V \in \mathbb{R}^{N_v \times C}$ as inputs. It computes a weighted sum of values for each query, where the weights are calculated as the normalized dot products between the query and the corresponding keys. It is defined in a concise matrix form according to Eq.3:

$$Attn\,(Q, K, V) = Softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V, \tag{3}$$

where the scalar factor $\sqrt{d_k}$ is introduced to avoid weight concentration and gradient vanishing [32]. In Transformers, the actual building block used is the Multi-Head Self-Attention ($MHSA$).Through this, queries Q, keys K, and values V are derived as linear projections of the same input $X \in \mathbb{R}^{N \times C}$. As for *multi-head*, it means splitting the output along the channel dimension into h blocks (*i.e.* heads), each using a set of independent projection weights as according to Eq.4:

$$\begin{aligned} MHSA(X) &= Concat(head_0, head_1, ..., head_h)W^0, \\ head_i &= Attn(XW_i^q, XW_i^k, XW_i^v), \end{aligned} \tag{4}$$

where by $head_i \in \mathbb{R}^{N \times C/h}$ is the output of the i-th attention head. $XW_i^q, XW_i^k, XW_i^v \in \mathbb{R}^{N \times C/h}$ are the corresponding input projection weights. An additional linear transformation, represented by the weight matrix $W^0 \in \mathbb{R}^{C \times C}$, aggregates all the heads.

The complexity of $MHSA(\cdot)$ is O(N^2) because there are N queries, each involving N key-value pairs. Such high complexity poses a serious scalability issue in terms of input space resolution. So we propose a new attention function.

Given a two-dimensional input feature map $X \in R^{(H \times W \times C)}$, we first partition it into $S \times S$ non-overlapping regions, each containing $\frac{HW}{S^2}$ feature vectors. This is achieved by reshaping X into $X^r \in R^{(S^2 \times \frac{HW}{S^2} \times C)}$. Then, linear projections are used to derive query, key, and value tensors $Q, K, V \in R^{(S^2 \times \frac{HW}{S^2} \times C)}$, and self-attention computation is directly performed on the feature map after the CNN feature extraction output $f(x)$, as this process does not require updating the Q,K,V projection weights, so the values of Q, K, and V as Eq.5:

$$Q = X^r, K = X^r, V = X^r. \tag{5}$$

Then, by constructing a directed graph, we found the adjacency relationship (*i.e.*, which regions each given region should attend to). Specifically, we first derive region level queries and keys, $Q^r, X^r \in R^{(S^2 \times \frac{HW}{S^2} \times C)}$, by taking the region-wise averages of Q and K, respectively. Through the matrix multiplication between Q^r and the transpose of K^r,

TABLE I
RESULTS OF THE ABLATION EXPERIMENTS ON THE DATASET C. 2,4,6,8,10 DENOTED THE TRAINING SAMPLES OF EACH CLASS, RESPECTIVELY.

Classifer	2		4		6		8		10	
	Acc(%)	EER(%)	Acc(%)	EER(%)	Acc(%)	EER(%)	Acc(%)	EER(%)	Acc(%)	EER(%)
ResNet18	56.80	8.42	65.20	7.01	76.23	6.10	78.67	5.43	89.87	2.53
FVCNN	61.63	7.95	68.30	6.32	78.83	5.46	79.67	2.73	88.87	2.43
FVRASNet	60.97	8.65	68.93	5.03	77.43	2.94	80.77	3.31	86.00	2.52
LightWeightCNN	62.57	7.33	69.25	5.36	76.13	3.64	82.17	3.42	89.00	2.62
PCANet	94.33	5.45	97.10	4.98	97.67	5.05	98.00	4.88	98.10	4.83
HOG	83.63	9.10	87.80	9.22	91.03	9.09	91.57	9.28	92.77	9.24
LDP	77.27	11.02	81.10	10.88	85.23	10.87	86.43	11.13	88.60	11.08
LLDP	80.90	14.05	85.33	13.89	87.33	14.12	88.73	14.03	90.53	13.97
PalmNet	96.93	2.77	98.53	2.70	98.53	2.71	98.70	2.66	98.87	2.56
Ours	**97.07**	**2.68**	**98.67**	**2.68**	**98.83**	**2.64**	**99.00**	**2.55**	**99.17**	**2.47**

we obtain the adjacency matrix $A^r \in R^{(S^2 \times \frac{HW}{S^2} \times C)}$ for the region affinity graph:

$$A = Q^r (K^r)^T. \tag{6}$$

The entries in the adjacency matrix A can be used to measure the correlation between two regions. Our next step is to appropriately optimize the adjacency matrix, thereby preserving regions with strong correlations and weakening those with poorer correlations.Specifically, we select the *topk* values by comparing the numerical values in the adjacency matrix to obtain the final A.

$$A = topkIndex(A^r). \tag{7}$$

After obtaining the final adjacency matrix A, we utilize A to multiply with V to obtain the final feature map. Then, we add the input feature map to the final feature map as the overall output O:

$$O = A * V + f(x). \tag{8}$$

D. Classification and Matching

GANet can be applied to both image identification and verification [37]. In the image identification stage, for image classification, the feature vectors H extracted from the palm vein are used, and the Euclidean distance between each set of feature vectors is calculated. After obtaining the Euclidean distance matrix, a kNN classifier is used for vein image classification, and this classifier requires no training. We set the hyperparameter k with 1.

In the image verification stage, we perform vein feature matching by comparing the extracted vein feature vectors H_1 and H_2 for feature matching verification.

IV. EXPERIMENTS

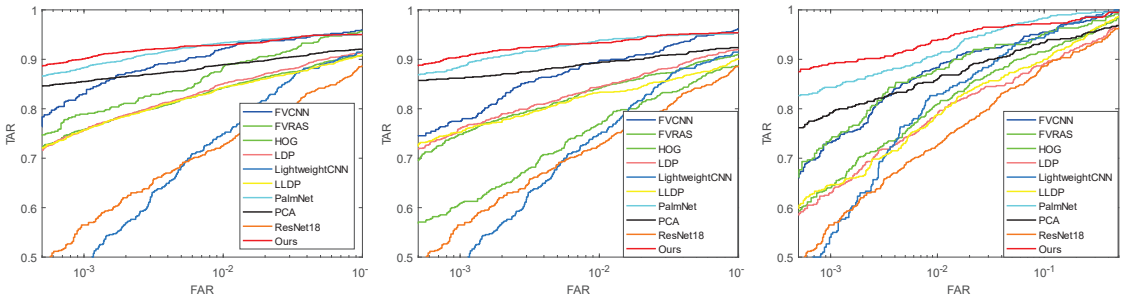
To verify the performance of our approach, we report numerous experiments on three palm vein datasets. To evaluate GANet's effectiveness in terms of improving accuracy, we compare the proposed approach with mainstream machine learning models and deep learning models. *e.g.* ResNet [15], FVCNN [16], FVRASNet [19], LightWeightCNN [18], PCANet [33], HOG [36], LDP [35], LLDP [34], and PalmNet [37].

A. Dataset Information

Dataset A [42]: The CASIA Palm-vein dataset contains 1200 images of palm vein. Those images come from 100 volunteers, each of whom provided images of both their left and right hands. Specifically, each hand of every volunteer was photographed 3 times during two different sessions. Therefore, each volunteer provided a total of 6 images (3 left-hand images and 3 right-hand images) in each session, making a total of 12 images across the two sessions. Dataset A has a total number of images calculated as follows: 100 volunteers $\times$ 2 hands $\times$ 3 images per hand per session $\times$ 2 sessions = 1200 images.

Dataset B [41]: The VERA Palm-vein dataset contains 2200 palm vein images. These images are sourced from 110 volunteers, each contributing palm vein images of their left and right hands. Specifically, each volunteer's hand was captured in two separate acquisition sessions, resulting in a total of 20 images per volunteer (5 images for the left hand + 5 images for the right hand $\times$ 2 acquisition sessions). Dataset B has a total number of images calculated as follows: 110 volunteers $\times$ 2 hands $\times$ 5 images per hand $\times$ 2 acquisition sessions = 2200 images. The resolution of each image is 200 $\times$ 200 pixels.

Dataset C [40]: The Tongji University Palm-vein dataset comprises 12,000 palm vein images from 300 volunteers, collected over two acquisition sessions with an interval of approximately 60 days. Each volunteer's hand was photographed with 10 images in each acquisition session, resulting in a total of 12,000 images (300 subjects $\times$ 2 hands $\times$ 10 images $\times$ 2 sessions). The resolution of each image is 128 $\times$ 128 pixels.



(a). ROC Curve on TJU600 (b). ROC Curve on VERA220 (c). ROC Curve on CASIA200

Fig. 3. ROC results of different methods.

TABLE II
RESULTS OF THE ABLATION EXPERIMENTS ON THE DATASET A. 2,3 DENOTED THE TRAINING SAMPLES OF EACH CLASS, RESPECTIVELY

Classifer	2		3	
	Acc(%)	EER(%)	Acc(%)	EER(%)
ResNet18	74.0	10.98	88.67	3.49
FVCNN	79.67	6.14	89.00	3.47
FVRASNet	78.33	6.82	88.33	4.21
LightWeightCNN	77.00	6.33	89.00	4.53
PCANet	93.33	7.50	89.00	7.46
HOG	88.73	8.67	95.67	8.78
LDP	86.50	10.67	88.73	10.65
LLDP	88.33	10.00	86.67	10.00
PalmNet	95.00	3.97	88.67	4.10
Ours	**97.00**	**3.50**	**97.00**	**3.44**

TABLE III
RESULTS OF THE ABLATION EXPERIMENTS ON THE DATASET B. 2,4 DENOTED THE TRAINING SAMPLES OF EACH CLASS, RESPECTIVELY

Classifer	2		4	
	Acc(%)	EER(%)	Acc(%)	EER(%)
ResNet18	78.55	8.36	87.36	6.00
FVCNN	80.91	5.27	89.45	5.19
FVRASNet	75.27	8.52	89.00	5.35
LightWeightCNN	75.45	8.90	88.56	6.60
PCANet	94.91	7.82	96.18	8.14
HOG	94.91	9.29	92.27	9.36
LDP	90.91	8.36	90.23	8.86
LLDP	88.00	9.93	88.41	9.36
PalmNet	96.91	5.00	97.09	5.18
Ours	**97.64**	**4.82**	**97.82**	**5.14**

B. Experiments Setting

To assess the effectiveness of our proposed method, we divided each dataset into two separate sets. For dataset A, we treated each palm as an individual class, resulting in a total of 200 classes (each volunteer's left and right palms counted as one class, with 100 volunteers in total). We employed two different training-testing approaches: one where we randomly selected two images for model training in the first stage and used the remaining images for testing in the second stage; the other approach involved training with three images in the first stage and testing with the remaining three images in the second stage. For dataset B, we followed a similar methodology: randomly selecting two and four images for training in the first stage, and then using the remaining five images for testing in the second stage. Dataset C was subjected to a different grouping strategy: we randomly selected two, four, six, eight, and ten images for training in the first stage, and used the ten images from the second stage for testing. Considering that original images might contain background regions irrelevant for classification, we adopted the method described in [21] to extract Regions of Interest *Regions of Interest* (ROI) from these images. All ROI images were standardized to a size of 32×32.

C. Classification

We validated the effectiveness of our method on three publicly available palm vein image databases and conducted extensive experiments. Specifically, we have implemented nine networks, which are ResNet [15], FVCNN [16], FVRASNet [19], LightWeightCNN [18], PCANet [33], HOG [36], LDP [35], LLDP [34] and PalmNet [37], and the recognition results and equal error rate are displayed on three datasets, to calculate the recognition results and equal error rate through a unified method, we use the deep learning method are calculated by extracting the feature vector of the trained model, calculating the Euclidean distance and the equal error rate, we also plotting the ROC curve, because each dataset has more than one training subset in our experiment, we only show one ROC curve in each database separately, and Figure.3 shows the ROC curves of different methods on different datasets.

The experimental results (Table I, Table II, Table III) show that compared with other methods, the proposed method improves the recognition rate and obtains the lowest equal error rate. For example, 99.17% accuracy and 2.47% equal error rate (EER) were obtained on Dataset A using our method in Table I.

V. CONCLUSION

In this paper, we present a method to obtain the global receptive field **GANet**, which extracts the vein features through the gabor filter and then calculates the attention mechanism to obtain the global receptive field for downstream palmar vein recognition. Firstly, the multi-scale fixed filter and the multi-scale adaptive Gabor filter were used to extract the venous features, then the self-attention mechanism was used to calculate the relationship between the blocks to obtain the global receptive field, and finally the Euclidean distance between the feature vectors was calculated for vein recognition. Experimental results on three databases show that our method

is superior to the existing palmar vein recognition methods in improving the accuracy of palm vein recognition, and the best recognition results are obtained.

ACKNOWLEDGMENTS

This work was supported in part by the National Natural Science Foundation of China (Grant No.61976030), the Scientific Innovation 2030 Major Project for New Generation of AI (Grant No.2020AAA0107300), the Fellowship of China Post-Doctoral Science Foundation (Grant No.59676651E), the Science Fund for Creative Research Groups of Chongqing Universities (Grant No.CXQT21034, Grant Nos. KJQN201900848 and KJQN201500814).and the Research on Key Technologies of Three-dimensional Finger Vein Authentication System (Grant No.222102210301)

REFERENCES

[1] Kshirsagar V P, Baviskar M R, Gaikwad M E. Face recognition using Eigenfaces[C]//2011 3rd International Conference on Computer Research and Development. IEEE, 2011, 2: 302-306.

[2] Jain A, Hong L, Bolle R. On-line fingerprint verification[J]. IEEE transactions on pattern analysis and machine intelligence, 1997, 19(4): 302-314.

[3] Qin H, El-Yacoubi M A, Li Y, et al. Multi-scale and multi-direction GAN for CNN-based single palm-vein identification[J]. IEEE Transactions on Information Forensics and Security, 2021, 16: 2652-2666.

[4] Qin H, El-Yacoubi M A. Deep representation-based feature extraction and recovering for finger-vein verification[J]. IEEE Transactions on Information Forensics and Security, 2017, 12(8): 1816-1829.

[5] Wang J, Wang G, Zhou M. Bimodal vein data mining via cross-selected-domain knowledge transfer[J]. IEEE Transactions on Information Forensics and Security, 2017, 13(3): 733-744.

[6] Kumar A, Hanmandlu M, Gupta H M. Online biometric authentication using hand vein patterns[C]//2009 IEEE Symposium on Computational Intelligence for Security and Defense Applications. IEEE, 2009: 1-7.

[7] Yu Z, Qin Y, Li X, et al. Deep learning for face anti-spoofing: A survey[J]. IEEE transactions on pattern analysis and machine intelligence, 2022, 45(5): 5609-5631.

[8] Menotti D, Chiachia G, Pinto A, et al. Deep representations for iris, face, and fingerprint spoofing detection[J]. IEEE Transactions on Information Forensics and Security, 2015, 10(4): 864-879.

[9] Huang B, Dai Y, Li R, et al. Finger-vein authentication based on wide line detector and pattern normalization[C]//2010 20th international conference on pattern recognition. IEEE, 2010: 1269-1272.

[10] Kumar A, Zhou Y. Human identification using finger images[J]. IEEE Transactions on image processing, 2011, 21(4): 2228-2244.

[11] Lee E C, Park K R. Image restoration of skin scattering and optical blurring for finger vein recognition[J]. Optics and Lasers in Engineering, 2011, 49(7): 816-828.

[12] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60(6): 84-90.

[13] Jin X, Zhu H, Yacoubi M AE, et al. StarLKNet: Star Mixup with Large Kernel Networks for Palm Vein Identification[J]. arXiv preprint arXiv:2405.12721, 2024.

[14] Liu S, Deng W. Very deep convolutional neural network based image classification using small training sample size[C]//2015 3rd IAPR Asian conference on pattern recognition (ACPR). IEEE, 2015: 730-734.

[15] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.

[16] Das R, Piciucco E, Maiorana E, et al. Convolutional neural network for finger-vein-based biometric identification[J]. IEEE Transactions on Information Forensics and Security, 2018, 14(2): 360-373.

[17] Qin H, Hu R, El-Yacoubi M A, et al. Local attention transformer-based full-view finger-vein identification[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2022.

[18] Shen J, Liu N, Xu C, et al. Finger vein recognition algorithm based on lightweight deep convolutional neural network[J]. IEEE Transactions on Instrumentation and Measurement, 2021, 71: 1-13.

[19] Yang W, Luo W, Kang W, et al. Fvras-net: An embedded finger-vein recognition and antispoofing system using a unified cnn[J]. IEEE Transactions on Instrumentation and Measurement, 2020, 69(11): 8690-8701.

[20] Hou B, Yan R. ArcVein-arccosine center loss for finger vein verification[J]. IEEE Transactions on Instrumentation and Measurement, 2021, 70: 1-11.

[21] Qin H, Zhu H, Jin X, et al. EmMixformer: Mix transformer for eye movement recognition[J]. arXiv preprint arXiv:2401.04956, 2024.

[22] Miura N, Nagasaka A, Miyatake T. Feature extraction of finger-vein patterns based on repeated line tracking and its application to personal identification[J]. Machine vision and applications, 2004, 15: 194-203.

[23] Jun B, Kim D. Robust face detection using local gradient patterns and evidence accumulation[J]. Pattern Recognition, 2012, 45(9): 3304-3316.

[24] Sulaiman D M, Abdulazeez A M, Haron H, et al. Unsupervised learning approach-based new optimization K-means clustering for finger vein image localization[C]//2019 international conference on advanced science and engineering (ICOASE). IEEE, 2019: 82-87.

[25] Kapoor K, Rani S, Kumar M, et al. Hybrid local phase quantization and grey wolf optimization based SVM for finger vein recognition[J]. Multimedia Tools and Applications, 2021, 80(10): 15233-15271.

[26] Wu J D, Liu C T. Finger-vein pattern identification using principal component analysis and the neural network technique[J]. Expert Systems with Applications, 2011, 38(5): 5423-5427.

[27] Yang G, Xi X, Yin Y. Finger vein recognition based on (2D) 2 PCA and metric learning[J]. BioMed Research International, 2012, 2012.

[28] Wang H, Du M, Zhou J, et al. Weber local descriptors with variable curvature gabor filter for finger vein recognition[J]. IEEE Access, 2019, 7: 108261-108277.

[29] Wright J, Ma Y, Mairal J, et al. Sparse representation for computer vision and pattern recognition[J]. Proceedings of the IEEE, 2010, 98(6): 1031-1044.

[30] Yang L, Yang G, Wang K, et al. Finger vein recognition via sparse reconstruction error constrained low-rank representation[J]. IEEE Transactions on Information Forensics and Security, 2021, 16: 4869-4881.

[31] Li G, Kim J. Palmprint recognition with local micro-structure tetra pattern[J]. Pattern Recognition, 2017, 61: 29-46.

[32] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[J]. Advances in neural information processing systems, 2017, 30.

[33] Chan T H, Jia K, Gao S, et al. PCANet: A simple deep learning baseline for image classification?[J]. IEEE transactions on image processing, 2015, 24(12): 5017-5032.

[34] Luo Y T, Zhao L Y, Zhang B, et al. Local line directional pattern for palmprint recognition[J]. Pattern Recognition, 2016, 50: 26-44.

[35] Jabid T, Kabir M H, Chae O. Robust facial expression recognition based on local directional pattern[J]. ETRI journal, 2010, 32(5): 784-794.

[36] Dalal N, Triggs B. Histograms of oriented gradients for human detection[C]//2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05). Ieee, 2005, 1: 886-893.

[37] Genovese A, Piuri V, Plataniotis K N, et al. PalmNet: Gabor-PCA convolutional networks for touchless palmprint recognition[J]. IEEE Transactions on Information Forensics and Security, 2019, 14(12): 3160-3174.

[38] Advances in biometrics for secure human authentication and recognition[M]. CRC Press, 2014.

[39] Qin H, Jin X, Jiang Y, et al. Adversarial AutoMixup[J]. arXiv preprint arXiv:2312.11954, 2023.

[40] Zhang L, Cheng Z, Shen Y, et al. Palmprint and palmvein recognition based on DCNN and a new large-scale contactless palmvein dataset[J]. Symmetry, 2018, 10(4): 78.

[41] Tome P, Marcel S. On the vulnerability of palm vein recognition to spoofing attacks[C]//2015 International Conference on Biometrics (ICB). IEEE, 2015: 319-325.

[42] Liu C L, Yin F, Wang D H, et al. CASIA online and offline Chinese handwriting databases[C]//2011 international conference on document analysis and recognition. IEEE, 2011: 37-41.