

Received June 17, 2019, accepted July 2, 2019, date of publication July 8, 2019, date of current version July 24, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2927230

Multi-Layer Convolutional Features Concatenation With Semantic Feature Selector for Vein Recognition

ZAIYU PAN¹, JUN WANG^{1,2}, ZHENGWEN SHEN¹, XIAOLING CHEN¹, AND MING LI^{1,2}

¹School of Information and Control Engineering, China University of Mining and Technology, Xuzhou 221116, China

²National Demonstration Center for Experimental Electrotechnics and Electronics Education, China University of Mining and Technology, Xuzhou 221116, China

Corresponding author: Jun Wang (wj9991x@163.com)

This work was supported in part by the Postgraduate Research and Practice Innovation Program of Jiangsu Province under Grant KYCX19_2143, and in part by the Postgraduate Research and Practice Innovation Program of the China University of Mining and Technology under Grant KYCX19_2143.

ABSTRACT Pre-trained DCNN trained on a large-scale image database can be used as a universal feature representation for image classification, which has achieved significant progress in some image recognition tasks. Compared with other image recognition tasks, directly utilizing a single convolutional feature as feature representation for vein recognition task cannot achieve the impressive result due to the sparse distribution of vein information. Therefore, to obtain more representative and discriminative convolutional feature for vein recognition, a novel multi-layer convolutional features' concatenation with semantic feature selector is proposed in this paper. In the pre-trained DCNN, different convolutional layers can encode different-level feature information. High-level convolutional features with vein information cover more semantic information and low-level convolutional features with vein information cover more detail information. However, low-level convolutional features also contain some background information. Therefore, in order to remove the background information of low-level convolutional features, a novel semantic feature selector is presented. First, the proposed local max-pooling of preserving spatial position (LMP-PSP) information is applied on activation map obtained by adding up all feature maps of the high-level convolutional layer to generate the semantic weighting map, which reflects key vein information of high-level convolutional features. Then, semantic weighting map is regarded as a feature selector to discard the background information of the low-level convolutional features and preserve the detail information of low level convolutional features. Finally, low-level convolutional features with vein information are selectively linked to high-level convolutional features with vein information based on the proposed semantic feature selector. A series of rigorous experiments on two lab-made vein databases named CUMT-Hand-Dorsa Vein database and CUMT-Palm Vein database is conducted to verify the effectiveness and feasibility of the proposed model. Besides, additional experiments with PUT Palm Vein database and the subset of PolyU database illustrate its generalization ability and robustness.

INDEX TERMS Pre-trained DCNN, vein recognition, multi-layer convolutional features concatenation, semantic feature selector, LMP-SPS.

I. INTRODUCTION

With the rapid development of society and scientific technology, the protection of personal information security plays a more and more important role in people's life. Identity authentication, one of the most efficient protection measures

The associate editor coordinating the review of this manuscript and approving it for publication was Xiaoyu Zhang.

for personal information security protection, has paid considerable attention in recent years. Compared with traditional identity identification techniques including smart cards, passwords and signature, biometric recognition techniques such as face recognition [1], palmprint recognition [2], fingerprint recognition [3] and vein recognition [4], are developing rapidly in the last few years due to their specific properties. Vein recognition, one of the most popular biometric

identification techniques, is playing more and more essential role in the identity authentication system. Vein recognition system usually consists of vein image preprocessing, feature extraction as well as classifier design and matching. Currently, most of current studies about this field mainly concentrate on generating more robust and discriminative feature representation, which is regarded as the most important and difficult part in vein recognition systems. In traditional vein recognition systems, hand-crafted feature representation algorithms are generally adopted to extract feature information of vein images. However, a more discriminative and robust vein recognition system is difficultly constructed by hand-crafted feature extraction algorithm due to its insufficient feature descriptor ability. Recently, deep convolutional neural network has achieved excellent performance in some image recognition tasks due to its discriminative feature representation capacity. However, because of the lack of sufficient training database like ImageNet, some image recognition tasks such as hand-dorsa vein recognition have great difficulty in training DCNN model. Thus, to utilize the discriminative feature representation ability of DCNN, Fang *et al.* [5] propose the two-stream convolutional network learning model for finger vein recognition system, and this model achieves the impressive performance. Besides, Wang *et al.* [6] present a novel task-specific transfer learning model to address the issues of over-fitting in the training process of DCNN model due to insufficient training vein images. Although acquiring the excellent performance, the above two vein recognition systems based on DCNN model always need the complex training process, leading to the low practicability of vein recognition system in a real environment.

Activations of a DCNN trained on a large-scale database [7], [8] such as ImageNet can be used as a universal image descriptor, and applying this descriptor to various image classification tasks delivers impressive performance. Liu *et al.* [9] present a novel cross-convolutional-layer pooling based on convolutional activation of pre-trained DCNN to better obtain discriminative deep representation for image recognition. Wei *et al.* [10] propose the selective convolutional descriptor aggregation model based global mean threshold method for fine-grained image retrieval. However, compared with other image recognition tasks, directly utilizing single convolutional feature as feature representation for vein recognition task cannot realize the excellent result due to the sparse distribution trait of vein image. Therefore, a novel multi-layer convolutional features concatenation with semantic feature selector is proposed to obtain more discriminative and richer convolutional feature for hand-dorsa vein recognition. In the pre-trained DCNN model, high-level convolutional features can learn more semantic information about samples and low-level convolutional features can learn more detail information about samples. If low-level convolutional features and high-level convolutional features are connected together, richer and more discriminative deep feature representation can be obtained for vein recognition task. However, low-level convolutional features also contain more background information

about samples. Therefore, a novel semantic feature selector is proposed to remove the background information the low-level convolutional features. First, the weighting map based on semantic information, which represents key vein information of high-level convolutional features, is produced by applying the proposed LMP-SPS on activation map generated by adding up all feature maps of the high-level convolutional layer. Second, we adopt the weighting map based on semantic information as a feature selector to select the useful detail information in low-level convolutional features. Finally, low-level convolutional features and high-level convolutional features are connected together as the final feature vector for vein recognition. Besides, PCA is introduced to reduce the dimension of the final feature vector for speeding up the training process of SVM. The outstanding recognition results of series rigorous comparison experiments on four databases including two lab-made databases and two public databases can demonstrate the effectiveness and robustness of the proposed multi-layer convolutional features concatenation with semantic feature selector for vein recognition system.

Overall, the main contributions of this paper are summarized as follows:

- We propose a novel vein recognition model with multi-layer convolutional features concatenation of pre-trained DCNN. Compared with other DCNN-based model, the proposed model not only realizes a simple, efficient, yet discriminative vein recognition system but also is of the high generalization capacity and robustness.
- A novel semantic feature selector is proposed to connect low-level convolutional features and high-level convolutional features for obtaining more discriminative and richer deep convolutional features.
- The LMP-SPS is presented to localize the key vein information such as ending points and cross points in the activation map generated by adding up all feature maps of high-level convolutional feature.

Compared with our previous published work [53], first, the proposed model in this paper fully employ the semantic information of high-level convolutional feature and the detailed information of low-level convolutional feature rather than the semantic information of single convolutional feature. Next, we propose a novel local max-pooling of preserving spatial position information (LMP-SPS) instead of global max-pooling of preserving spatial position information in prior work to better localize the key vein information of convolutional features. Third, the model proposed in this paper obtains more discriminative and richer convolutional features for vein recognition compared with our previous paper. Finally, the proposed model in this paper are evaluated on four databases instead of two databases in previous work to effectively verify its performance for vein recognition.

The remainder of this paper is organized as follows: the related vein recognition methods in recent years are introduced in Section II. In Section III, we propose a novel model for vein recognition with multi-layer convolutional features concatenation of pre-trained DCNN and illustrate the process

of the connection of low-level convolutional features and high-level convolutional features. The experimental results and analyses are presented in Section IV. Finally, we summarize the paper and conclude the future work in Section V.

II. RELATED WORK

Vein recognition, one of the most popular biometric identification technologies, has been received considerable attention due to its higher security. Currently, many studies about this field mainly adopt several vein information such as hand-dorsa vein [11]–[13], palm vein [14], [15] and finger vein [16]–[18] as feature representation for identity verification. The feature extraction methods, whose design is regarded as the most essential and difficult part in vein recognition system, can be summarized into two categories including handcrafted feature extraction methods and deep feature extraction methods.

A. HANDCRAFTED FEATURE EXTRACTION METHODS

Handcrafted feature extraction methods divided into shape feature-based methods and texture feature-based methods have been widely applied to vein recognition tasks, which also achieves excellent performance.

Shape feature-based methods: These approaches usually adopt the shape information of vein image such as the positions and angles of the straight line vectors [19], endpoints and crossing points [20], vein knuckle shapes [21], dominant points [22] as feature representation for later classification. Apart from the above vein recognition model with shape information, Yang *et al.* [23] adopted the tri-branch vein structure as shape information to enhance the recognition performance of the template matching in finger vein recognition system. Huang *et al.* [24] proposed a novel hand-dorsa vein recognition with the fusion of texture information and shape clues generated by the proposed binary coding and graph matching. Yang *et al.* [25] proposed a novel feature representation methods for finger vein recognition based on adaptive vector field estimation by which the finger vein matching accuracy can be effectively improved. However, due to unreliable lighting condition in vein acquisition process, the acquisition of the accurate vein segmentation image for different images of the same sample is extremely difficult, which leads to the insufficient generalization ability of these methods.

Texture feature-based methods: Mainly concentrating on the texture variations within the vein image, these methods can further be divided into two categories: local gray distribution coding and local invariant feature. The former which covers local binary pattern (LBP) [26], local derivative pattern (LDP) [27], local ternary pattern [28] and local line binary pattern (LLBP) [29] is to depict the global gray distribution histogram on the basis of local coding calculation results. Besides, in order to enhance the feature descriptor ability of local gray distribution coding, Wang and Wang [30] proposed the discriminative local binary pattern (DLBP) for vein recognition to decrease the influence of contrast information

during feature coding process. Xi *et al.* [31] presented the discriminative binary codes (DBC) learning method for finger vein recognition, which can obtain more discriminative feature representation and reduce the time cost and storage requirements. Kang and Wu [32] proposed a novel palm vein recognition method with a mutual foreground-based local binary pattern, which can achieve the better matching accuracy. In a word, although achieving the high performance for vein recognition, local gray distribution coding is of the insufficient generalization capacity due to the effect of contrast information of vein image. The latter which contains SIFT [33], SURF [34], RootSIFT [35] and ASIFT [36] generally consists of three main steps including scale-space establishment, extrema detection and descriptors generation. Wang *et al.* [37] adopted SIFT for hand vein describing after simple preprocessing, which refers to the procedure of denoising and contrast enhancement (CE). However, it is argued that the preprocessing would lead to the loss of vein information and the low contrast distribution renders great difficulty in generating enough descriptive keypoints. To remove the negative effect of CE, SIFT or SURF [38], [39] is extracted directly with input without any pre-processing and binarization procedure, resulting in considerable performance improvement. However, the existence of mismatching pairs between unpaired keypoints usually brings up the false acceptance rate (FAR) and equal error rate (EER), which is unacceptable for vein recognition. Inspired by the conclusion in [38]–[40], Kang *et al.* [41] tried to improve the performance of SIFT by way of complementary feature fusion and noisy keypoints removal, and they also import DoG-HE for better CE. In the unpaired matching removal stage, the LBP is adopted for describing the region distribution of the matching keypoints, and then the mismatching points are removed with LBP difference. The analytical disadvantage of the keypoints fusion method lies in the fact that the detected non-vein keypoints describing the palmprint or other skin region is verified to bring up the FAR. In order to address this issue, Huang *et al.* [42] proposed the novel key-point detection and matching for hand-dorsa vein recognition, and the proposed the key-point and detection accounts for the physiological characteristics of dorsal hand as well as the representativeness of the feature points. Thus, this model achieves impressive performance for dorsal hand vein recognition.

B. DEEP FEATURE EXTRACTION METHODS

DCNN has been widely applied to many image recognition tasks, images segmentation tasks and image detection tasks due to its discriminative feature representation ability. Currently, some researchers have been brought DCNN into vein recognition task. Wang *et al.* [6] proposed a novel coarse-to-fine transfer learning model to address the problem of insufficient training database in hand-dorsa vein recognition, which achieves the impressive performance. Fang *et al.* [5] presented a lightweight deep-learning model to obtain representative and discriminative feature descriptor for finger vein recognition. Qin and El-Yacoubi [43] proposed a DCNN

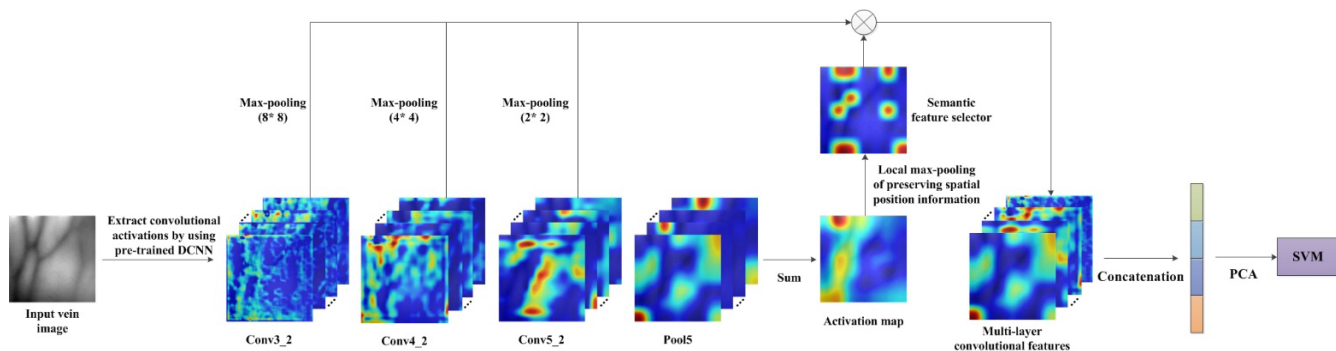


FIGURE 1. The framework of our proposed multi-layer convolutional features concatenation with semantic feature selector for vein recognition.

model to extract and recover vein features based on limited a priori knowledge, which improves the finger vein verification accuracy. Wang and Wang [44] designed a structure growing guided CNN model to learn more discriminative deep feature for vein recognition, which also improves the performance of vein recognitions.

III. METHODOLOGY

To obtain more discriminative and richer convolutional activation of pre-trained DCNN model, we propose a multi-layer convolutional feature concatenation model with semantic feature selector for hand-dorsa vein recognition system which is illustrated in Fig. 1. First, we apply pre-trained DCNN model such as VGG-16 to extract the convolutional features of the input vein image. Next, the semantic feature selector is generated by applying the proposed LMP-SPS on activation map obtained by adding up all feature maps of the high-level convolutional layer. Then, the semantic feature selector is used to remove the background information of low-level convolutional features and acquire more discriminative low-level convolutional features. Finally, low-level convolutional features and high-level convolutional features are connected together as the final feature vector for vein recognition. Besides, PCA is introduced to reduce the dimension of the final feature vector for speeding up the training process of SVM.

A. THE ANALYSIS OF CONVOLUTIONAL FEATURES WITH VEIN INFORMATION

Convolutional activations of a pre-trained DCNN model can be widely utilized as image representation, which also achieves the excellent performance in some image classification and retrieval tasks due to the fact that convolutional activations contain richer spatial and semantic information. However, compared with other images, the vein image is of sparse distribution traits, which results in the fact that adopting single convolutional feature of vein information as image representation cannot achieve the acceptable result.

In the pre-trained DCNN model, high-level convolutional features generally cover richer semantic information of input image and low-level convolutional generally contain richer

detail information of input image. If connecting the high-level convolutional features and low-level convolutional features, we will obtain richer and more discriminative convolutional features for vein recognition. However, the study [45] indicated that directly connecting the high-level convolutional features and low-level convolutional features as images representation cannot obtain the excellent result due to the fact that low-level convolutional contain some background information.

To better analyze the specific property of convolutional activation of vein image, we employ the pre-trained DCNN model to extract the convolutional activations of four vein images randomly selected on CUMT-Hand-Dorsa Vein database, and the visualization results of some convolutional features are shown in Fig. 2. It can be observed from Fig. 2 that the responses of feature maps in high-level convolutional features mainly focus on several vein regions including the end points, cross points and local key vein information and the responses of feature maps in low-level convolutional features mainly concentrate on several regions including edges and background, which effectively demonstrates the fact that high-level convolutional features generally cover richer semantic information of vein images and low-level convolutional features generally contain richer detail information of vein images but also cover some non-vein information.

B. SEMANTIC FEATURE SELECTOR

Through the above analysis for the feature maps of convolutional layer based on vein information, it can be concluded that low-level convolutional features based on vein information not only contain more detail information, but also cover some non-vein information. Therefore, in order to effectively connect low-level convolutional feature with high-level convolutional features to obtain more discriminative and richer convolutional features, it is necessary to remove the non-vein information of low-level convolutional features. However, it can be seen from Fig. 2 that the distribution of background information of low-level convolutional features is scattered, which leads to the fact that it is difficult to directly remove non-vein information by using some measures on low-level convolutional features.

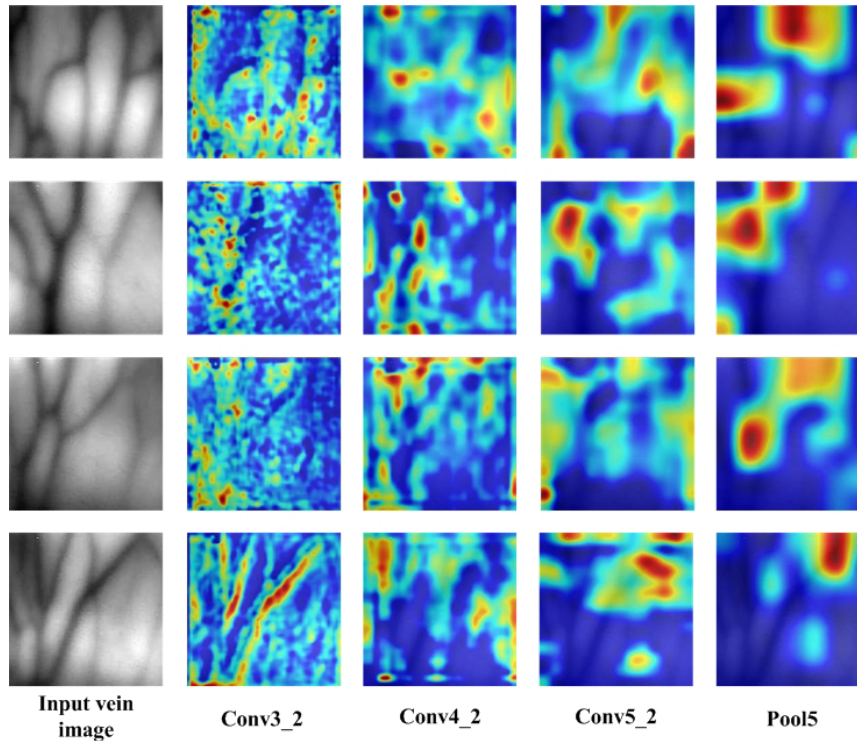


FIGURE 2. The visualization results of feature maps in some convolutional layers.

During the visualization process of feature maps, we find the fact that the strongest parts of the strong responses in the feature maps of high-level convolutional layer generally correspond to some key vein information such as ending point or crossing point, and the strongest parts of the weak responses in the feature maps of high-level convolutional layer generally also correspond to some key vein information such as end point or cross point. Inspired by this idea, a novel LMP-SPS is proposed to obtain key vein information of high-level convolutional features based on vein information. We do not employ the semantic information of single feature map of high-level convolutional layer due to the fact that it does not reflect the situation of semantic information of local vein region. Instead, we add up all feature maps of high-level convolutional layer to obtain activation map A . Given the feature maps of high-level convolutional layer $S \in R^{H_1 \times W_1 \times C_1}$, the activation map is obtained as:

$$A = \sum_{n=1}^C S_n \tag{1}$$

where A is the activation map and S_n is the n -th feature map of high-level convolutional layer. Based on the above analysis, the semantic feature selector W can be obtained by applying LMP-SPS on activation map A . In the neighborhood of 3×3 , the proposed LMP-SPS can be represented as follow:

$$W_{3 \times 3}(i,j) = \begin{cases} 1 & \text{if } A_{3 \times 3}(i,j) == Tmax \\ 0 & \text{else} \end{cases} \tag{2}$$

where $A_{3 \times 3}$ is the neighborhood of 3×3 in the activation map, $W_{3 \times 3}$ is the partial semantic feature selector generated by applying the proposed method on the neighborhood of 3×3 and $Tmax$ is maximum of the neighborhood of 3×3 . It should be noted that the feature maps of pool5 layer in VGG-16 model are adopted as high-level convolutional features in our experiment.

In order to evaluate the performance of the proposed LMP-SPS in obtaining key vein information of high-level convolutional features, we randomly select six vein images on CUMT Hand-Dorsa Vein database and visualize the obtained semantic feature selector which are shown in Fig. 3. It can be seen from Fig. 3 that employing LMP-SPS on activation map can accurately obtain key vein information of high-level convolutional features, which also evidences the effectiveness of the proposed LMP-SPS.

C. MULTI-LAYER CONVOLUTIONAL FEATURES CONCATENATION

After semantic feature selector is generated by utilizing LMP-SPS on activation map, it is used to remove background information of low-level convolutional features. Given the original feature maps of low-level convolutional layer $D \in R^{H_2 \times W_2 \times C_2}$ and then the low-level convolutional features D' which is identical in size with semantic feature selector W can be acquired by conducting a max-pooling operation on the original low-level convolutional features. Low-level convolutional features which is removed non-vein information

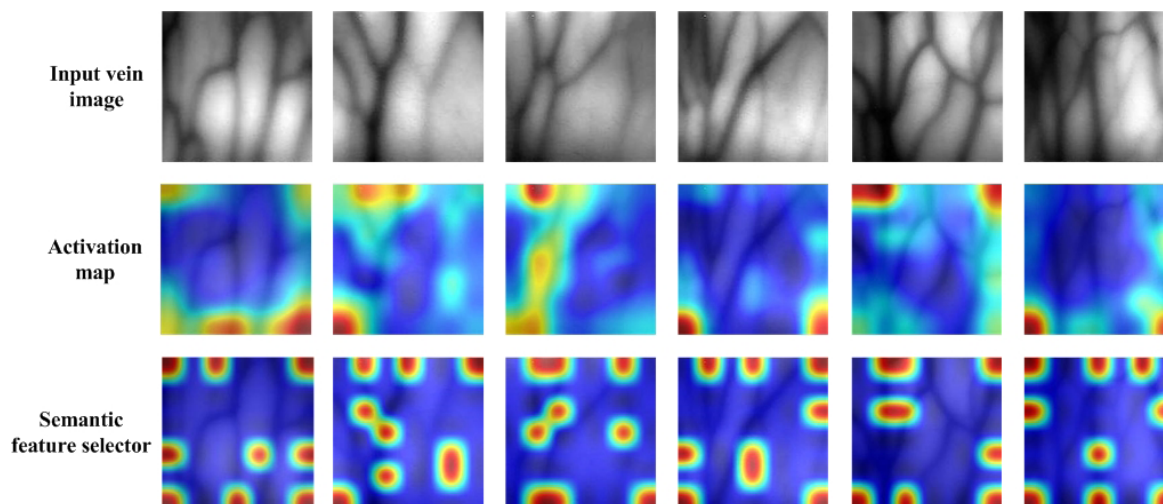


FIGURE 3. The visualization results of semantic feature selectors.

D'' can be obtained as:

$$D'' = \sum_i^{H_1} \sum_j^{W_1} D' * W \quad (3)$$

It should be noted that the Conv3_3, Conv4_2 and Conv5_2 of VGG-16 model are adopted as low-level convolutional features.

The concatenation of high-level convolutional features and low-level convolutional features can be achieved as:

$$f = [D'', S] \quad (4)$$

where f is the final multi-layer convolutional features. The detailed process of selective convolutional descriptor model is illustrated in Fig. 1.

The whole process of proposed multi-layer convolutional features concatenation with semantic feature selector for vein recognition is summarized in Algorithm 1.

IV. EXPERIMENTS AND DISCUSSION

A. DATABASE

In this section, to evaluate the effectiveness of the proposed multi-layer convolutional features with semantic feature selector for vein recognition, a series of experiments is conducted on our two lab-made vein database respectively named as CUMT-Hand-Dorsa Vein database (CUMT-HDV) and CUMT-Palm Vein database (CUMT-PV). Besides, additional experiments are performed on two public database including PUT Palm Vein database [46] and PolyU Palmprint database [47] to verify the generalization ability of the proposed model, and state-of-the-art recognition result fully demonstrates the generalization capacity and robustness of the proposed model.

- Two lab-made vein databases, which contain CUMT-HDV database and CUMT-PV database, are captured by self-designed image capturing device which can be referenced from our previous works [48]. Two lab-made

Algorithm 1 The Whole Process of Proposed Model

Input:

- An input vein ROI image I (224×224)
 - A pre-trained DCNN model (VGG-16)
-

Procedure:

1. Extract low-level convolutional features and high-level convolutional features $D \in \mathbb{R}^{H_2 \times W_2 \times C_2}, S \in \mathbb{R}^{H_1 \times W_1 \times C_1}$ from I using the pre-trained DCNN model.
 2. Obtain semantic feature selector W by using the LMP-SPS on activation map A generated by adding up all feature maps of high-level convolutional features.
 3. Semantic feature selector W is employed to remove the background information of low-level convolutional features by performing multiplication operation $D'' = \sum_i^{H_1} \sum_j^{W_1} D' * W$
 4. Concatenate low-level convolutional features and high-level convolutional features to obtain more discriminative and richer multi-layer convolutional features.
-

vein databases contains $224 \times 2 \times 10$ images which are acquired by capturing the hand-dorsa vein information and palm vein information of 224 individuals. All images on our two vein databases are acquired in two sessions and the time interval of two sessions is more than 10 days. Besides, in the acquisition process of each session, five samples are obtained from each person at the wavelength of 850nm. The ROI extraction [49] process specifically designed for this database is conducted followed by the gray and size normalization. Note that the size of vein ROI image is 161×161 and resized to 224×224 by using “imresize” operation of MATLAB in our experiment.

- PUT Palm Vein database, which consists of 1200 palm vein images, is captured by 100 palms and 12 images are acquired for each palms in three sessions. In our experiment, the palm vein images of the first sessions are used to evaluate the performance of the proposed model.
- PolyU Database, which contains 6000 images, are acquired by 500 palms in two sessions and 6 images are captured in each session. It should be noted that PolyU Multispectral Palmprint database (near-infrared part) is utilized to verify the generation ability of the proposed model.

Some samples of four database including two lab-made databases and two public databases are as shown in Fig. 4.

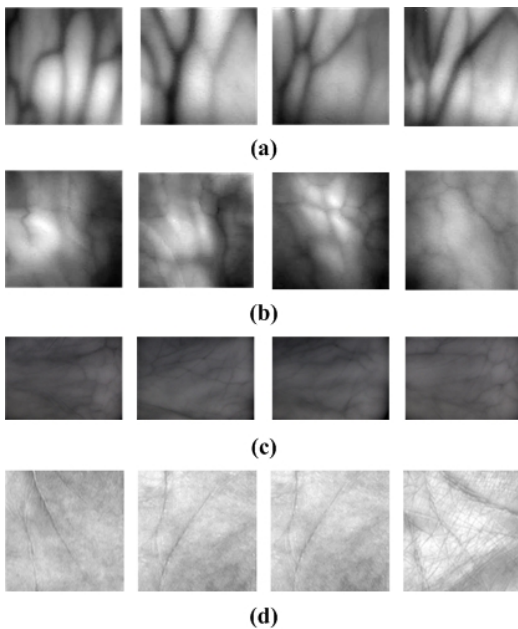


FIGURE 4. Some samples of four databases. (a) CUMT-HDV. (b) CUMT-PV. (c) PUT Palm Vein database. (d) PolyU Palmprint Multispectral database.

B. EXPERIMENT DETAILS

In our experiments, we utilize the pre-trained VGG-16 model [50] which is trained on ImageNet database to extract convolutional activations of input vein images. Besides, the pool5 layer of VGG-16 model is adopted as the high-level convolutional features and the Conv3_2, Conv4_2 and Conv5_2 layers of VGG-16 model are regarded as low-level convolutional features. Thus, S is the $7 \times 7 \times 512$ convolutional features, and D_1, D_2, D_3 are respectively $56 \times 56 \times 256$, $28 \times 28 \times 512$ and $14 \times 14 \times 512$ convolutional features. Due to the fact that low-level convolutional features is not identical in size with high-level convolutional features, the corresponding max-pooling operation is conducted before the connection of high-level convolutional features and low-level convolutional features. The size of multi-layer convolutional features obtained by linking high-level convolutional features and low-level convolutional features is $7 \times 7 \times 1792$. Therefore, the final feature

representation formed by concatenating multi-layer convolutional features is a 87808-dimension feature vector. Due to the fact that the size of obtained feature representation is too large, PCA is used to decrease the dimension of input feature vector of SVM for speeding up its training procedure. In addition, the vein feature representation which is 240 feature vector is regarded as the input data to train SVM. The training samples contain 200×5 images, and the test samples contain 200×5 images. The configuration of SVM in our experiments is that its kernel function adopt the radial basis function and its penalty parameter as well as gamma are respectively set as 128 and 0.0078.

C. PERFORMANCE EVALUATION OF MULTI-LAYER CONVOLUTIONAL FEATURES CONCATENATION WITH SEMANTIC FEATURE SELECTOR

In this section, the effectiveness of multi-layer convolutional features concatenation with semantic feature selector for vein recognition is evaluated on our two lab-made vein databases such as CUMT-HDV database and CUMT-PV database. To verify the advantage of the proposed method in obtaining more discriminative and richer deep convolutional feature, several encoding and pooling methods such as max-pooling, average-pooling, FV, VLAD, CL [9] and SCDA [10] are adopted as comparison experiments. The experimental results as shown in Tables 1. It should be noted that the feature maps of pool5 layer are regarded as convolutional features in comparison experiments. Besides, to fully evaluate the performance of the proposed method, we respectively connect high-level convolutional feature and different low-level convolutional features with or without semantic feature selector. The performance of different parts of the proposed model is illustrated in Table 2.

Judging from the experimental results in Table 1, the proposed multi-layer convolutional features concatenation with semantic feature selector compared with other encoding and pooling methods achieves the best performance. The state-of-the-art recognition rates with 97.66% and 96.82% demonstrate the effectiveness of multi-layer convolutional features concatenation with semantic feature selector for vein recognition. It can be observed from Table 2 that the multi-layer convolutional features with semantic feature selector achieves the highest recognition rate, which demonstrates the fact that convolutional features from a single convolutional feature are not sufficient to obtain excellent performance and also indicates the fact that the proposed model can effectively remove the background information of low-level convolutional features and obtain more discriminative and richer convolutional features for vein recognition.

D. COMPARISON WITH STATE-OF-THE-ART MODEL

In this part, several experiments are designed on our two lab-made vein database including CUMT HDV database and CUMT-PV database to verify the advantage of the proposed multi-layer convolutional features concatenation with semantic feature selector for vein recognition over other

TABLE 1. The recognition results of different aggregation methods on CUMT-HDV database and CUMT-PV database.

Convolutional activations	Methods	Recognition rate (%)	
		CUMT-HDV	CUMT-PV
Pool5	Max-pooling	88.46	85.60
	Average-pooling	91.52	89.21
	FV	81.08	80.73
	VLAD	87.27	83.52
	CL	93.19	91.88
	SCDA	92.84	91.35
Multi-layer convolutional features	Multi-layer convolutional features concatenation with semantic feature selector	97.66	96.82

TABLE 2. The performance of different parts of the proposed model on CUMT-HDV database and CUMT-PV database.

Convolutional activations	Methods	Recognition rate (%)	
		CUMT-HDV	CUMT-PV
Pool5	Direct concatenation	92.86	90.48
Pool5、Conv5_2	Two-layer convolutional features concatenation with semantic feature selector	96.35	95.19
Pool5、Conv5_2、Conv4_2	Three-layer convolutional features concatenation with semantic feature selector	96.95	95.92
Pool5、Conv5_2、Conv4_2、Conv3_2	Multi-layer convolutional features concatenation without semantic feature selector	97.02	96.37
	Multi-layer convolutional features concatenation with semantic feature selector	97.66	96.82

state-of-the-art feature extraction algorithms. Multi-modal experiments in the scenario of verification are designed. In such scenario, the first image is regarded as gallery whereas the remaining images are exploited as probe. Therefore, the number of genuine and imposter scores is 10080 ($224 \times 10 \times 9/2$) and 2497600 ($10 \times 10 \times 224 \times 223/2$).

Three kinds of representative vein recognition algorithms which contain local invariant feature, local gray distribution coding and deep feature extraction model are utilized as comparison model to fully validate the advantage of the proposed multi-layer convolutional features concatenation with semantic feature selector for vein recognition. Local invariant feature models such as SIFT [33], SURT [34], RootSIFT [35] and ASIFT [36] are regarded as one of the most best hand-crafted feature representation algorithms in vein recognition tasks due to their advantage in the invariant of rotation, translation and scale uncertainly. Local gray distribution coding models mainly contain LBP [26], LDP [27], LTP [28] and LLBP [29], which have been widely applied to vein recognition tasks in recently years. The deep feature extraction algorithms including transfer learning model [6], structure

growing guide CNN [44] CL [9] and SCDA [10] achieve outstanding recognition results due to their robustness feature descriptor ability. The ROC curves of three feature extraction models are illustrated in Fig. 5 and Table 3.

It can be concluded from Fig. 6 that the proposed multi-layer convolutional features concatenation with semantic feature selector model performs far better than the LIF (Local Invariant Feature) models with EER as 1.28% and 1.74% whereas the best of LIF is 2.25% and 2.63% with RootSIFT as well as the best of LBPs is 1.96% and 2.38% with LDP, and the state-of-the-art recognition results fully evidence the discriminative and representative feature representation ability of the proposed model for vein recognition task. It can be observed from Table 3 that compared with the coarse-to-fine transfer learning model [6] and structure growing guide CNN [44], our proposed model is simpler. Besides, the EER result as 1.28% and 1.74% achieved by the proposed model is better than the EER results as 1.72% and 2.21% as well as 2.41% and 3.15% respectively obtained by the coarse-to-fine transfer learning model and structure growing guide CNN, which also verifies the advantage of the proposed

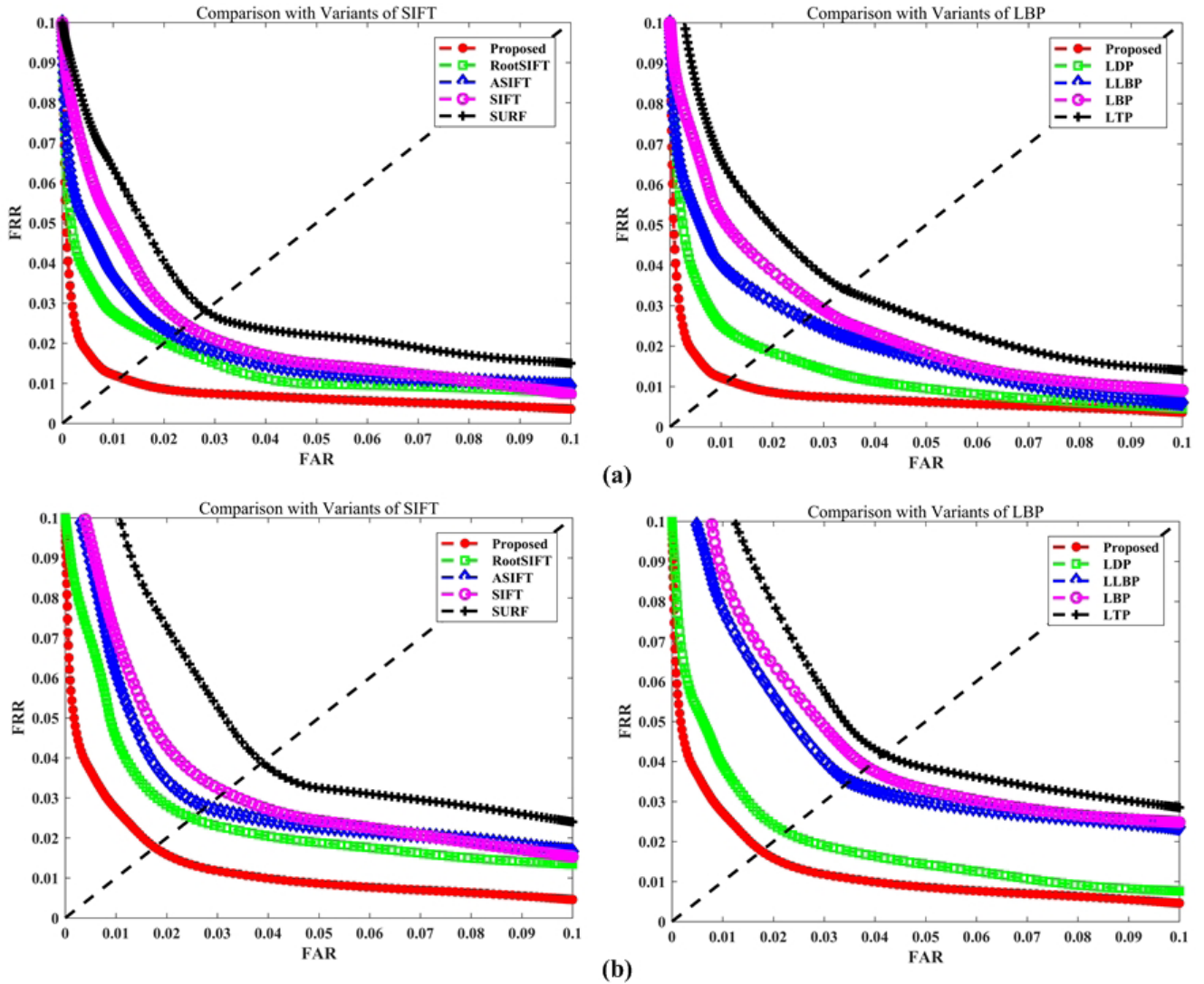


FIGURE 5. Comparison of ROC curves between the proposed feature learning methods and representative handcrafted methods (left: SIFTs, right: LBPs). (a) The ROC curves on CUMT-HDV database. (b) The ROC curves on CUMT-PV database.

TABLE 3. The recognition results of different DCNN-based methods for vein recognition on CUMT-HDV database and CUMT-PV database.

Methods	EER (%)	
	CUMT-HDV	CUMT-PV
Coarse-to-fine transfer learning model [6]	1.72	2.21
Structure growing guide CNN [44]	2.41	3.15
CL [9]	1.78	2.25
SCDA [10]	2.13	2.76
Proposed	1.28	1.74

model. Compared with CL [9] and SCDA [10], our proposed method can achieve higher performance for hand-dorsa vein recognition task. In addition, the EER result as 1.28% and 1.74% achieved by our proposed model is better than the EER results as 1.78% and 2.25% as well as 2.13% and 2.76% respectively obtained by the CL and SCD, which also

illustrates the high performance of the proposed model for hand-dorsa vein recognition.

E. PERFORMANCE EVALUATION OF TIME COST

Recognition rate and time consumption are two important indicators to measure identity recognition systems based on

TABLE 4. Average time consumption of each link of the proposed model.

System Link	Average Time Consumption (S)
Extract the multi-layer convolutional features	0.336
Multi-layer convolutional features concatenation	0.012
Classification	0.0153
Total	0.3633

TABLE 5. The EER results of single series of recently published vein recognition models and different DCNN-based models for vein recognition on PUT Palm Vein database.

Methods	EER (%)
Segmentation with local threshold model [46]	1.1
Coarse-to-fine transfer learning model [6]	0.624
Structure growing guide CNN [44]	0.922
CL [9]	0.644
SCDA [10]	0.714
Proposed	0.582

TABLE 6. Summary of EERs derived from recently published vein recognition models using PolyU Palmprint database.

Reference	Year	Methodology	Performance(EER)
[51]	2013	Combine curvelet transform and Gabor filter for feature extraction and employ score level fusion strategy to obtain matching score for verification	0.1023%
[52]	2014	Employ Gaussian-Random transform to extract the orientation matrix and then compute the principal direction based on the matrix as palm vein feature, and matching between feature bins is used for identification	0.14%
[30]	2017	Extract vein features with quality-specific discriminative LBP algorithm and adopt improved Chi-square for verification	0.079%
[6]	2018	Extract vein features with proposed DNN model fin-tuned under “coarse-to-fine” scheme and adopt LDM for identification	0.058%
[53]	2018	Extract vein features by aggregating deep convolutional features of pre-trained DCNN model based on vein minutiae feature masks and adopt SVM for classification	0.053%
Proposed method	2018	Extract vein features by multi-layer convolutional features concatenation with semantic feature selector and adopt SVM for classification	0.041%

vein information. Identity recognition systems in real environment should not only have higher recognition result but also reduce their time-consuming as much as possible. Sometimes, recognition rate of identity recognition system may be appropriately decreased to optimize the time cost of the system. The key time consumption parts of our proposed multi-layer convolutional features concatenation with semantic feature selector mainly contain the extraction process of multi-layer convolutional features, the process of multi-layer convolutional features concatenation with selector and classification. The evaluation experiment of time cost of our proposed model is conducted by MATLAB 2017 on a PC with CPU 3.30GHz and 8.00GB memory. The average time

consumption of each image is shown in Table 4. It can be observed from Table 4 that the time cost of the proposed method is acceptable on two lab-made vein database, which demonstrates the fact that the proposed model can better meet the demand of identity recognition systems based on vein information for time consumption.

F. GENERALIZATION EVALUATION OF THE PROPOSED MODEL

After the effectiveness of the proposed model are evaluated on our two lab-made vein databases, two experiments are designed on two public database including PUT Palm Vein database (single session) and PloyU Multispectral Palmprint

database (near-infrared part) to verify the generalization ability of the proposed model. In our experiments, the first image is regarded as gallery whereas the remaining images are exploited as probe. The experiments results derived from recently published methods on two public database are shown in Table 5 and Table 6. It should be noted that several methods based on DCNN model for hand-dorsa vein recognition are utilized as comparison experiments with PUT Palm Vein database to better evaluate the generalization capacity of the proposed due to the fact that recently published methods on PUT Palm Vein database are insufficient. It can be observed from Table 5 and Table 6 that the proposed model could achieve state-of-the-art EER level compared with others methods, which effectively verifies the high performance of the proposed model in the generalization ability and robustness.

V. CONCLUSION

In this paper, a novel multi-layer convolutional features concatenation with semantic feature selector is proposed to increase the feature representation ability of a pre-trained DCNN for vein recognition. First, a pre-trained DCNN model is used to extract the multi-layer convolutional features of input vein images. Second, the weighting map based on semantic information, which includes key vein information of high-level convolutional features, is produced by applying LMP-SPS on activation map generated by adding up all feature maps of the high-level convolutional layer. Third, we adopt the weighting map based on semantic information as a feature selector to select the useful detail information in low-level convolutional features. Finally, low-level convolutional features and high-level convolutional features are connected together as final feature vector for vein recognition. Besides, PCA is introduced to reduce the dimension of final feature vector for speeding up the training process of SVM. The outstanding recognition results of series rigorous comparison experiments on four databases including two lab-made databases and two public databases can demonstrate the effectiveness and robustness of proposed model for vein recognition system.

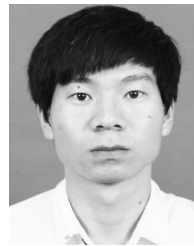
We also argue that the proposed multi-layer convolutional features concatenation with semantic feature selector is also applicable for other computer vision task.

In the future, we can try to design an end-to-end DCNN with a novel attention model for hand-dorsa vein recognition.

REFERENCES

- [1] L. He, H. Li, Q. Zhang, and Z. Sun, "Dynamic feature matching for partial face recognition," *IEEE Trans. Image Process.*, vol. 28, no. 2, pp. 791–802, Feb. 2019.
- [2] S. Zhang, H. Wang, W. Huang, and C. Zhang, "Combining modified LBP and weighted SRC for palmprint recognition," *Signal, Image Video Process.*, vol. 12, no. 6, pp. 1035–1042, 2018.
- [3] C. Lin and A. Kumar, "Contactless and partial 3D fingerprint recognition using multi-view deep representation," *Pattern Recognit.*, vol. 83, pp. 314–327, Nov. 2018.
- [4] C.-L. Lin and K.-C. Fan, "Biometric verification using thermal images of palm-dorsa vein patterns," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 2, pp. 199–213, Feb. 2004.
- [5] Y. Fang, Q. Wu, and W. Kang, "A novel finger vein verification system based on two-stream convolutional network learning," *Neurocomputing*, vol. 290, pp. 100–107, May 2018.
- [6] J. Wang, G. Wang, and M. Zhou, "Bimodal vein data mining via cross-selected-domain knowledge transfer," *IEEE Trans. Inf. Forensics Security*, vol. 13, no. 3, pp. 733–744, Mar. 2018.
- [7] Y. Kalantidis, C. Mellina, and S. Osindero, "Cross-dimensional weighting for aggregated deep convolutional features," 2015, *arXiv:1512.04065*. [Online]. Available: <https://arxiv.org/abs/1512.04065>
- [8] J. Xu, C. Wang, C. Qi, C. Shi, and B. Xiao, "Unsupervised semantic-based aggregation of deep convolutional features," *IEEE Trans. Image Process.*, vol. 28, no. 2, pp. 601–611, Feb. 2019.
- [9] L. Liu, C. Shen, and A. van den Hengel, "Cross-convolutional-layer pooling for image recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 11, pp. 2305–2313, Nov. 2017.
- [10] X.-S. Wei, J.-H. Luo, J. Wu, and Z.-H. Zhou, "Selective convolutional descriptor aggregation for fine-grained image retrieval," *IEEE Trans. Image Process.*, vol. 26, no. 6, pp. 2868–2881, Jun. 2017.
- [11] J. Wang, G. Wang, and L. Li, "Image pattern similarity index and its application to task-specific transfer learning," *IEICE Trans. Inf. Syst.*, vol. E100-D, no. 12, pp. 3032–3035, Dec. 2017.
- [12] Y. Wang, K. Li, J. Cui, L. Shark, and M. Varley, "Study of hand-dorsa vein recognition," in *Proc. Int. Conf. Intell. Comput.*, Changsha, China, 2010, pp. 490–498.
- [13] F. Li, T. Zhang, Y. Liu, and G. Wang, "Hand-dorsa vein recognition based on scale and contrast invariant feature matching," *IEICE Trans. Inf. Syst.*, vol. E100, no. 12, pp. 3054–3058, 2017.
- [14] E. Picuccio, E. Maiorana, and P. Campisi, "Palm vein recognition using a high dynamic range approach," *IET Biometrics*, vol. 7, no. 5, pp. 439–446, Sep. 2018.
- [15] X. Yan, W. Kang, F. Deng, and Q. Wu, "Palm vein recognition based on multi-sampling and feature-level fusion," *Neurocomputing*, vol. 151, pp. 798–807, Mar. 2015.
- [16] L. Yang, G. Yang, Y. Yin, and X. Xi, "Finger vein recognition with anatomy structure analysis," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 28, no. 8, pp. 1892–1905, Aug. 2018.
- [17] W. Yang, Z. Chen, C. Qin, and Q. Liao, " α -trimmed Weber representation and cross section asymmetrical coding for human identification using finger images," *IEEE Trans. Inf. Forensics Security*, vol. 14, no. 1, pp. 90–101, Jan. 2019.
- [18] H. Qin and M. A. El-Yacoubi, "Deep representation for finger-vein image-quality assessment," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 28, no. 8, pp. 1677–1693, Aug. 2018.
- [19] J. M. Cross and C. L. Smith, "Thermographic imaging of the subcutaneous vascular network of the back of the hand for biometric identification," in *Proc. Int. Carnahan Conf. Secur. Technol.*, Sanderstead, U.K., Oct. 1995, pp. 20–35.
- [20] K. Wang, Y. Yuan, Z. Zhang, and D. Zhuang, "Hand vein recognition based on multi supplemental features of multi-classifier fusion decision," in *Proc. Int. Conf. Mechatronics Automat.*, Luoyang, China, Jun. 2006, pp. 1790–1795.
- [21] P. MacGregor and R. Welford, "Imaging for security and personnel identification," *Adv. Imag.*, vol. 6, no. 7, pp. 52–56, 1991.
- [22] K. Mikołajczyk and C. Schmid, "An affine invariant interest point detector," in *Proc. Eur. Conf. Comput. Vis.*, Copenhagen, Denmark, 2002, pp. 128–142.
- [23] L. Yang, G. Yang, X. Xi, X. Meng, C. Zhang, and Y. Yin, "Tri-branch vein structure assisted finger vein recognition," *IEEE Access*, vol. 5, pp. 21020–21028, 2017.
- [24] D. Huang, X. Zhu, Y. Wang, and D. Zhang, "Dorsal hand vein recognition via hierarchical combination of texture and shape clues," *Neurocomputing*, vol. 214, pp. 815–828, Nov. 2016.
- [25] J. Yang, Y. Shi, and G. Jia, "Finger-vein image matching based on adaptive curve transformation," *Pattern Recognit.*, vol. 66, pp. 34–43, Jun. 2017.
- [26] E. C. Lee, H. C. Lee, and K. R. Park, "Finger vein recognition using minutia-based alignment and local binary pattern-based feature extraction," *Int. J. Imag. Syst. Technol.*, vol. 19, no. 3, pp. 179–186, 2009.
- [27] B. J. Kang, K. R. Park, J. H. Yoo, and J. N. Kim, "Multimodal biometric method that combines veins, prints, and shape of a finger," *Proc. SPIE*, vol. 50, no. 1, 2011, Art. no. 017201.
- [28] X. Tan and B. Triggs, "Enhanced local texture feature sets for face recognition under difficult lighting conditions," *IEEE Trans. Image Process.*, vol. 19, no. 6, pp. 168–182, Jun. 2010.

- [29] B. A. Rosdi, W. S. Chai, and S. A. Suandi, "Finger vein recognition using local line binary pattern," *Sensors*, vol. 11, no. 12, pp. 11357–11371, 2011.
- [30] J. Wang and G. Wang, "Quality-specific hand vein recognition system," *IEEE Trans. Inf. Forensics Security*, vol. 12, no. 11, pp. 2599–2610, Nov. 2017.
- [31] X. Xi, L. Yang, and Y. Yin, "Learning discriminative binary codes for finger vein recognition," *Pattern Recognit.*, vol. 66, pp. 26–33, Jun. 2017.
- [32] W. Kang and Q. Wu, "Contactless palm vein recognition using a mutual foreground-based local binary pattern," *IEEE Trans. Inf. Forensics Security*, vol. 9, no. 11, pp. 1974–1985, Nov. 2014.
- [33] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.
- [34] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (SURF)," *Comput. Vis. Image Understand.*, vol. 110, no. 3, pp. 346–359, 2008.
- [35] R. Arandjelović and A. Zisserman, "Three things everyone should know to improve object retrieval," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 2911–2918.
- [36] J.-M. Morel and G. Yu, "ASIFT: A new framework for fully affine invariant image comparison," *SIAM J. Imag. Sci.*, vol. 2, no. 2, pp. 438–469, 2009.
- [37] H. Wang, L. Tao, and X. Hu, "Novel algorithm for hand vein recognition based on retinex method and sift feature analysis," *Elect. Power Syst. Comput.*, vol. 99, pp. 559–566, Jun. 2011.
- [38] L. Xiuyuan, L. Tiegeng, D. Shichao, H. Jin, and W. Yun, "Fast recognition of hand vein with surf descriptors," *Chin. J. Sci. Instrum.*, vol. 32, no. 4, pp. 831–836, 2011.
- [39] H.-G. Kim, E. J. Lee, G.-J. Yoon, S.-D. Yang, E. C. Lee, and M. Y. Sang, "Illumination normalization for sift based finger vein authentication," in *Advances in Visual Computing*. Berlin, Germany: Springer, 2012, pp. 21–30.
- [40] P.-O. Ladoux, C. Rosenberger, and B. Dorizzi, "Palm vein verification system based on sift matching," in *Advances in Biometrics*. Berlin, Germany: Springer, 2009, pp. 1290–1298.
- [41] W. Kang, Y. Liu, Q. Wu, and X. Yue, "Contact-free palm-vein recognition based on local invariant features," *PLoS ONE*, vol. 9, no. 5, 2014, Art. no. e97548.
- [42] D. Huang, R. Zhang, Y. Yin, Y. Wang, and Y. Wang, "Local feature approach to dorsal hand vein recognition by centroid-based circular keypoint grid and fine-grained matching," *Image Vis. Comput.*, vol. 58, pp. 266–277, Feb. 2017.
- [43] H. Qin and M. A. El-Yacoubi, "Deep representation-based feature extraction and recovering for finger-vein verification," *IEEE Trans. Inf. Forensics Security*, vol. 12, no. 8, pp. 1816–1829, Aug. 2017.
- [44] J. Wang and G. Wang, "Hand-dorsa vein recognition with structure growing guided CNN," *Optik-Int. J. Light Electron Opt.*, vol. 149, pp. 469–477, Nov. 2017.
- [45] C. Du, C. Wang, Y. Wang, C. Shi, and B. Xiao, "Selective feature connection mechanism: Concatenating multi-layer CNN features with a feature selector," 2018, *arXiv:1811.06295*. <https://arxiv.org/abs/1811.06295>
- [46] R. Kabacinski and M. Kowalski, "Vein pattern database and benchmark results," *Electron. Lett.*, vol. 47, no. 20, pp. 1127–1128, Sep. 2011.
- [47] D. Zhang, W.-K. Kong, J. You, and M. Wong, "Online palmprint identification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 9, pp. 1041–1050, Sep. 2003.
- [48] J. Wang, G. Wang, M. Li, W. Yu, and H. Tian, "An improved hand vein image acquisition method based on the proposed image quality evaluation system," *Comput. Model. New Technol.*, vol. 18, no. 11, pp. 1204–1208, 2014.
- [49] J. Wang, G. Wang, M. Li, K. Wang, and H. Tian, "Hand vein recognition based on improved template matching," *Int. J. Bioauto.*, vol. 18, no. 4, pp. 337–348, 2014.
- [50] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*. [Online]. Available: <https://arxiv.org/abs/1409.1556>
- [51] J. Sun and W. Abdulla, "Palm vein recognition by combining curvelet transform and Gabor filter," in *Proc. 8th Chin. Conf. Biometric Recognit.*, Jinan, China, 2013, pp. 314–321.
- [52] P. O. Staff, "Palm-vein classification based on principal orientation features," *PLoS ONE*, vol. 9, no. 11, 2014, Art. no. e112429.
- [53] J. Wang, K. Yang, Z. Pan, G. Wang, M. Li, and Y. Li, "Minutiae-based weighting aggregation of deep convolutional features for vein recognition," *IEEE Access*, vol. 6, pp. 61640–61650, 2018.



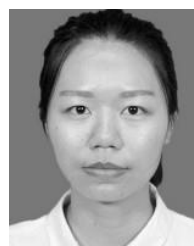
ZAIYU PAN received the B.S. degree in electrical engineering and automation from Chaohu University, Anhui, China, in 2016. He is currently pursuing the Ph.D. degree in control science and engineering with the China University of Mining and Technology. His current research interests include image processing, machine learning, computer vision, and pattern recognition, especially on vein recognition and other biometrical identification methods.



JUN WANG received the B.S., M.S., and Ph.D. degrees in control theory and control engineering from the China University of Mining and Technology, Jiangsu, China, in 2003, 2009, and 2015, respectively, where he has been an Associate Professor with the School of Information and Electrical Engineering. He was also with the Jiangsu Province Laboratory of Electrical and Automation Engineering for Coal Mining. His current research interests include intelligent control system design, computer vision, and pattern recognition, especially on vein recognition and other biometrical identification methods.



ZHENGWEN SHEN received the B.S. degree from the College of Physical and Electronic Information, Luoyang Normal University, Henan, China, in 2017. He is currently pursuing the M.S. degree in information and communication engineering with the China University of Mining and Technology. His current research interests include image processing, machine learning, computer vision, and pattern recognition, especially on vein recognition and other biometrical identification methods.



XIAOLING CHEN received the B.S. degree in communication engineering from Nanchang University, in 2018. She is currently pursuing the M.S. degree in information and communication engineering from the China University of Mining and Technology. Her current research interests include image processing, machine learning, computer vision, and pattern recognition, especially on vein recognition and other biometrical identification methods.



MING LI received the B.S. and M.S. degrees in automation and the Ph.D. degree in control theory and control engineering from the China University of Mining and Technology, Jiangsu, China, in 1983, 1989, and 2002, respectively, where he has been a Professor with the School of Information and Electrical Engineering. He was also with the Jiangsu Province Laboratory of Electrical and Automation Engineering for Coal Mining. His current research interests include intelligent control system design, computer vision, and pattern recognition, especially on vein recognition and other biometrical identification methods.

...