LETTER

Hand-Dorsa Vein Recognition Based on Selective Deep Convolutional Feature

Zaiyu PAN[†], ***Student Member and*** **Jun WANG**[†a)], ***Nonmember***

SUMMARY A pre-trained deep convolutional neural network (DCNN) is adopted as a feature extractor to extract the feature representation of vein images for hand-dorsa vein recognition. In specific, a novel selective deep convolutional feature is proposed to obtain more representative and discriminative feature representation. Extensive experiments on the lab-made database obtain the state-of-the-art recognition result, which demonstrates the effectiveness of the proposed model.

key words: *hand-dorsa vein recognition, pre-trained DCNN, selective deep convolutional feature*

1. Introduction

Biometrics is an automatic user authentication technology, which uses human physiological and/or behavioral characteristics with several desirable properties like universality, distinctiveness, permanence and acceptability. In recent years, there has been an increasing interest in vein recognition, which is motivated by the advantages of live identification, non-invasive, and non-contact image capture, and high security over other biometric recognition techniques (e.g., fingerprint, face, iris, voice, gait, etc.). However, traditional vein recognition model is usually designed by the extraction methods of handcrafted feature such as Local Binary Pattern (LBP) and Scale Invariant Feature Transform (SIFT), it is difficult to establish a more robust and discriminative vein recognition model due to the insufficient capacity of handcrafted feature representation.

Recently, deep learning has been successfully applied on large-scale image recognition tasks because of the capacity in learning discriminative and representative features. However, due to the fact that the performance of DCNN models relies seriously on the size of vein databases, it is difficult to build a more effective vein recognition system by DCNN models. Thus, applying DCNN models to vein recognition is still challenging because it is hard to meet the demand for a large number of training samples to train a DCNN. To address this problem, the training strategy of DCNN [1] and the design of task-specific network architecture [2] are proposed in recently years. Although these methods successfully apply DCNN models to vein recognition tasks, some of them need complex training process and some of them obtain the DCNN for vein recognition which need further improve its feature representation ability. Thus, to utilize the discriminative feature representation ability of DCNN for vein recognition, a novel selective deep convolutional feature (SDCF) based on a pre-trained DCNN is proposed. Our proposed SDCF model does not require the complex and long-time training process, which greatly enhances the robustness and feasibility of vein recognition systems. The framework of the proposed model is as shown in Fig. 1.

Manuscript received November 21, 2019.
Manuscript revised February 4, 2020.
Manuscript publicized March 4, 2020.

[†]The authors are with School of Information and Control Engineering, China University of Mining and Technology, Xuzhou 221116, China.
a) E-mail: wj999lx@163.com

DOI: 10.1587/transinf.2019EDL8204

2. Selective Deep Convolutional Feature

Activations of a DCNN pre-trained on a large-scale database, such as ImageNet, can be used as a universal image representation and this method has obtained high performance on some image recognition tasks. However, some researches [3], [4] have indicated that directly utilizing the activations of convolutional layer as image representation produces inferior performance because the convolutional activations contain the background information and noisy information. To better analyze the traits of convolutional features with vein information, we randomly select three vein images on our lab-made database and extract their last convolutional features by a pre-trained DCNN model. The visualization results of feature maps of convolutional layer (pool5 layer) based on vein information are shown in Fig. 2. It can be observed from Fig. 2 that the single cell in the pool5 can correspond to two kinds of vein patches including non-vein regions and vein regions, which causes the low discriminability of convolutional activations. Therefore, if we directly adopt original convolutional features as feature representation, it will affect the performance of vein recognition systems. How to remove the non-vein information in convolutional features to explore its potential feature descriptor ability remains a key issue.

Recently, Wei et al. [5] propose a selective convolutional descriptor aggregation model based on the global mean threshold method to obtain the useful convolutional

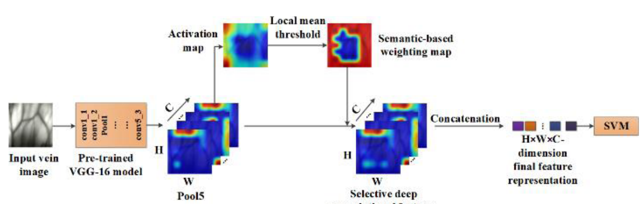

Fig. 1 The framework of the proposed SDCF model for vein recognition

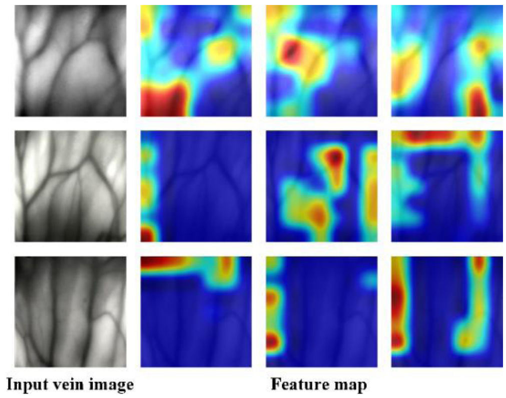

Fig. 2 The visualization results of feature maps of pool5 layer

descriptors for fine-grained image retrieval. However, if we employ this method to the convolutional activations with vein information, it will lose some vein information because the weak responses of feature maps in pool5 layer may also correspond to vein regions. Therefore, this method is not appropriate for hand-dorsa vein recognition tasks. Besides, during the visualization process of feature maps, we also find the fact that the weakest parts of the strong responses in the feature maps generally correspond to non-vein regions, and the strongest parts of the weak responses in the feature maps generally correspond to vein regions. Inspired by this fact, a novel Selective Deep Convolutional Feature (SDCF) model based on local mean threshold method is proposed to obtain more discriminative and useful convolutional features for vein recognition.

Like [5], we do not also employ the semantic information of single channel because it does not reflect the situation of semantic information of local vein regions. Instead, we add up the feature maps of pool5 layer through the depth direction. Thus, we can acquire a convolutional activations of 2-dimension, which is named as activation map. Given the feature maps of pool5 layer $\mathrm{X} \in R^{H\times W\times C}$, the activation map is obtained as:

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

Where A is the activation map and X_n is the n-th feature map of pool5 layer. Each single cell in the activation map corresponds to the vein patch of input vein image, which may be vein region or non-vein region. Based on the above analysis that the weakest parts of the strong responses in the feature maps generally correspond to non-vein regions, and the strongest parts of the weak responses in the feature maps generally correspond to vein regions, it is straightforward to say that the local stronger responses in the activation map A always correspond to the vein region. Therefore, a novel local mean threshold is proposed to decide which cells correspond to vein region of input vein image. The semantic-based weighting map w can be defined as follows:

$$w = \begin{cases} 1 & A(i,j) > Amean \\ 0 & else \end{cases} \tag{2}$$

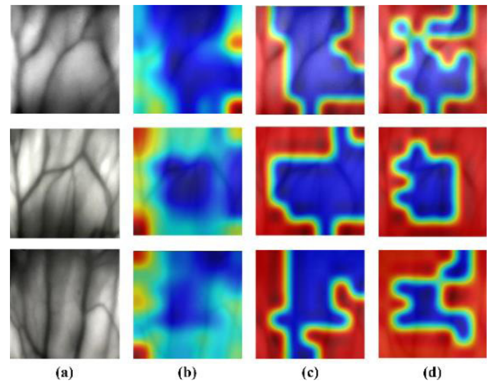

Fig. 3 The visualization results of different semantic-based maps. (a) Input vein image. (b) Activation map. (c) Semantic-based weighting map generated by global mean threshold method. (d) Semantic-based weighting map generated by local mean threshold method.

Where $Amean = \frac{1}{3\times 3}\sum_{i-1}^{i+1}\sum_{j-1}^{j+1} A(i,j),\ i \in [1,H],\ j \in [1,W]$, and it is a local mean operation with 3×3 window conducted on activation map to obtain semantic-based weighting map. To acquire the semantic-based weighting map which is identical in size with activation map, the padding is used in performing the operation.

To better analyze the effect of the proposed selective deep convolutional feature model, we randomly select three vein images on our vein database, and visualize the obtained activation map and semantic-based weighting map. The visualization results of different semantic-based maps are as shown in Fig. 3. As can be seen from Fig. 3, compared with the semantic-based weighting generated by global mean threshold method, the semantic-based weighting map generated by local mean threshold method can effectively preserve the vein information of weaker responses regions of activation map and remove the non-vein information of stronger response regions of activation map. At the same time, it also demonstrates the initial idea that the weakest parts of the strong responses in the feature maps generally correspond to non-vein regions, and the strongest parts of the weak responses in the feature maps generally correspond to vein regions. Therefore, we employ the semantic-based weighting map generated by the local mean threshold method to select discriminative and meaningful deep convolutional features. The selective deep convolutional feature are obtained as follows:

$$\mathrm{f} = \sum_{i}^{H}\sum_{j}^{W} d(i,j)w(i,j) \tag{3}$$

Where f refers to the selected convolutional descriptor set and $d(i,j)$ indicates the local feature vector of 512 dimension in the position (i, j).

3. Lab-Made Hand-Dorsa Vein Database

To obtain persuasive and satisfactory classification result, a comprehensive hand-dorsa vein database containing 200 individuals where male and female are respectively 100 is built, and for each person, 10 right hand-dorsa vein images

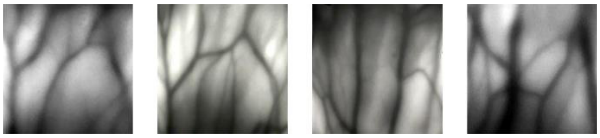

Fig. 4 Samples of lab-made database

are captured. All hand-dorsa vein images in our database are acquired in two specifically sessions separated by a time interval of more than 10 days, and at each time, five samples are acquired from each subject at the wavelength of 850nm. To the fullest of the dorsal vein information, we set the size of the images as 460*680 with extremely high-quality. Figure 4 shows some samples of home-made database. The ROI extraction process [6] specifically designed for this database is conducted followed by the grey and size normalization.

4. Experiments and Analysis

In this part, rigorous comparison experiments on lab-made hand-dorsa vein database are designed to comprehensively evaluate the performance of the proposed SDCF model. In our experiments, we adopt the VGG-16 model [7] trained on ImageNet database as the pre-trained DCNN model to extract deep convolutional features. To speed up the training process of SVM, PCA is utilized to reduce the dimension of final feature vector. In the training process of SVM, the train samples contain 200*5 images, and the test samples contain 200*5 images. Note that the radial basis function is regarded as the kernel function of SVM, the penalty parameter C of SVM is set to 128 and the gamma is set to 0.0078.

4.1 Performance Evaluation of SDCF

In this section, we design two experiments on our vein database to evaluate the effectiveness of SDCF model. It should be noted that the feature maps of pool5 layer are used as convolutional features in our experiments. Different selection deep convolutional features models is conducted to verify the performance of our proposed model in the first experiment. The experimental results are shown in Table 1.

It can be concluded from Table 2 that compared with the first two kinds of aggregation methods, the proposed SDCF based on local mean threshold realizes the highest recognition result, which fully verifies its effectiveness. At the same time, it also illustrates the fact that compared with employing the global mean threshold methods to select deep convolutional features, using the local mean threshold to select the deep convolutional features can better remove the non-vein and noisy information of convolutional activations and preserve the discriminative and useful vein information of convolutional activations.

The design of the second experiment aims to evaluate the performance of the proposed aggregation method for convolutional features. Therefore, to fully verify the advantage of the proposed model, we select the several most

Table 1 The experimental results of different selective deep convolutional feature models

Convolutional activation	Aggregation methods	Recognition rate (%)
Pool5	Directly concatenate	93.96
	Global mean threshold +concatenate	93.19
	Local mean threshold +concatenate	96.53

Table 2 The recognition results of different aggregation methods

Convolutional activation	Aggregation methods	Recognition rate (%)
Pool5	Max-pooling	88.54
	Average-pooing	91.15
	FV	83.97
	VLAD	88.13
	CL	92.35
	Proposed	96.53

commonly used encoding and aggregations approaches for convolutional features such as max-pooling, average-pooling, FV, VLAD and Cross-Convolutioal-Pooling (CL) [8] as comparison algorithms. The recognition results of different aggregation methods for convolutional features are illustrated in Table 2. It can be seen from Table 2 that the proposed SDCF model for vein recognition obtains the best recognition result compared with other aggregation methods, which demonstrates the advantage of our proposed SDCF.

4.2 Comparison with State of the Art Models

In this part, extensive matching experiments are designed on hand-dorsa vein database to verify the advantage of the proposed SDCF model for vein recognition over two representative state-of-the-art feature extraction algorithms including hand-crafted feature-based methods and deep feature-based methods. The genuine matching and imposter matching are conducted with the trained SVM for calculating false rejection rate (FRR) and false acceptance rate (FAR) on our vein database, with which we can obtain the equal error rate (EER) results.

4.2.1 Comparison with Vein Recognition Models Based on Hand-Crafted Feature

Vein recognition methods based on hand-crafted feature consist of two representative algorithms such as LBP and

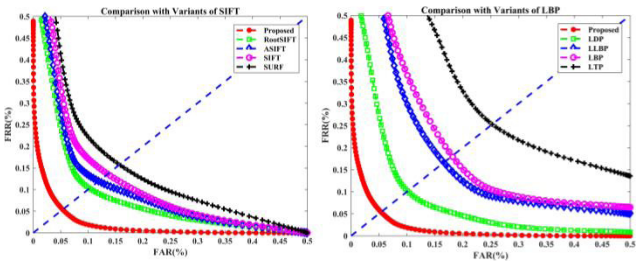

Fig. 5 Comparison of ROC curves between the proposed model and representative state-of-the-art handcrafted methods

Table 3 The EER results of different vein recognition models based on deep feature

Method	Model A [9]	Model B [10]	Model C [11]	Model D [12]	Proposed
EER	0.093	0.075	0.058	0.06	0.054

its variants as well as SIFT and its variants. The former mainly contains LBP, LDP, LTP, and LLBP, and such model is widely employed to vein recognition tasks due to its efficiency, which also provides acceptable recognition rates. The latter mainly includes SIFT, SURF, RootSIFT, ASIFT and has the advantages of being invariant to rotation, translation, scale uncertainty and even uniform illumination, which makes it become the one of the best feature representation algorithms among all the hand-crafted feature extraction models. The ROC curves of two representative vein recognition models based on hand-crafted feature on our vein database are shown in Fig. 5. As can be observed from Fig. 5, the best EER result of SIFT and its variants is 0.105% which is achieved by RootSIFT model as well as the best EER result of LBP and its variants is 0.113% which is generated by LDP model. Our proposed SDCF model respectively achieves the EER result with 0.054%, and the performance is better than that of RootSIFT and LDP, which also verifies the superiority of our proposed SDCF model.

4.2.2 Comparison with Vein Recognition Models Based on Deep Feature

In this experiment, we select four vein recognition models based on deep feature [9]–[12] as comparison experiments to evaluate the advantage of our proposed SDCF. It can be concluded from Table 3 that the EER result achieved by our proposed SDCF is far better than the EER results obtained by other four vein recognition methods based on deep feature, which evidences the superiority of our proposed SDCF model for vein recognition tasks.

5. Conclusions

In this paper, a novel SDCF model based on a pre-trained DCNN is proposed to obtain more representative and discriminative convolutional features for hand-dorsa vein recognition. First, the pre-trained DCNN model such as VGG-16 are used to extract convolutional features of input vein images. Second, the semantic-based weighting map is generated by utilizing the local mean threshold method on activation map which is obtained by adding up all feature maps of pool5 layer. Finally, the semantic-based weighting map is used to select useful and discriminative convolutional features and discard background and noisy information. State-of-the-art vein recognition results demonstrate the effectiveness of our proposed SDCF model. In the future, we will design an end-to end attention model to obtain more discriminative and richer convolutional features for vein recognition.

Acknowledgments

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

References

[1] 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, 2018.
[2] J. Wang and G. Wang, "Hand-dorsa vein recognition with structure growing guided CNN," OPTIK, vol.149, pp.469–477, 2017.
[3] P. Agrawal, R. Girshick, and J. Malik, "Analyzing the performance of multilayer neural networks for object recognition," Proc. Eur. Conf. Comp. Vis., 2014.
[4] M. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," Proc. Eur. Conf. Comp. Vis., 2014.
[5] 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, 2017.
[6] J. Wang, G. Wang, M. Li, K. Wang, and H. Tian, "Hand vein recognition based on improved template matching," Int. J. Bioautomation, vol.18, no.4, pp.337–348, Dec. 2014.
[7] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," arXiv: 1409.1556, 2014.
[8] 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.
[9] H. Hai, L. Chen, H. Song, and J. Yang, "Dorsal hand vein recognition based on convolutional neural networks," 2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 2017.
[10] N. Al-johania and L. Elrefaei, "Dorsa Hand Vein Recognition by Convolutional Neural Networks: Feature Learning and Transfer Learning Approaches," International Journal of Intelligent Engineering and Systems, vol.12, no.3, pp.178–191, 2019.
[11] J. Wang, G. Wang, and M. Zhou, "Bimodal Vein Data Mining via Cross-Selective-Domain Knowledge Transfer," IEEE Trans. Inf. Forensics Security, vol.13, no.3, pp.733–744, March 2018.
[12] 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.