

# Palm Vein Recognition Based on Convolutional Neural Network

Yong-Yi FANJIANG<sup>1</sup>, Cheng-Chi LEE<sup>2,3,\*</sup>, Yan-Ta DU<sup>1</sup>,  
Shi-Jinn HORNG<sup>4</sup>

<sup>1</sup> Department of Computer Science and Information Engineering, Fu Jen Catholic University, No. 510, Zhongzheng Rd., Xinzhuang Dist., New Taipei City 24205, Taiwan, R.O.C.

<sup>2</sup> Department of Library and Information Science, Research and Development Center for Physical Education, Health, and Information Technology, Fu Jen Catholic University, No. 510, Zhongzheng Rd., Xinzhuang Dist., New Taipei City 24205, Taiwan, R.O.C.

<sup>3</sup> Department of Computer Science and Information Engineering, Asia University, Wufeng Shiang, Taichung 41349, Taiwan, R.O.C.

<sup>4</sup> Department of Computer Science and Information Engineering, National Taiwan University of Science and Technology, No. 43, Sec. 4, Keelung Rd., Taipei 10617, Taiwan, R.O.C.

e-mail: [yfyanj@csie.fju.edu.tw](mailto:yfyanj@csie.fju.edu.tw), [clee@mail.fju.edu.tw](mailto:clee@mail.fju.edu.tw), [horngsj@yahoo.com.tw](mailto:horngsj@yahoo.com.tw)

Received: April 2020; accepted: September 2021

**Abstract.** Convolutional neural networks (CNNs) were popular in ImageNet large scale visual recognition competition (ILSVRC 2012) because of their identification ability and computational efficiency. This paper proposes a palm vein recognition method based on CNN. The four main steps of palm vein recognition are image acquisition, image preprocessing, feature extraction, and matching. To reduce the processing steps in the recognition of palm vein images, a palm vein recognition method using a CNN is proposed. CNN is a deep learning network. Palm vein images are acquired using near-infrared light, under which the veins in the palm of the hand are relatively prominent. To obtain a good vein image, many previous methods used preprocessing to further enhance the image before using feature extraction to find feature matches for further comparison. In recent years, CNNs have been shown to have great advantages and have performed well in image classification. To reduce early-stage image processing, a CNN is used to classify and recognize palm vein images. The networks AlexNet and VGG depth CNN were trained to extract image features. The palm vein recognition rates by VGG-19, VGG-16, and AlexNet were 98.5%, 97.5%, and 96%, respectively.

**Key words:** palm vein recognition, CNN, AlexNet, VGG, biometrics.

## 1. Introduction

User authentication identification methods include passwords, cards, and biometrics (O’Gorman, 2003). With the increase in financial activities and safety awareness, followed by the development of science, technology, and societal progress, traditional identity authentication, such as passwords, personal identification numbers, and smart cards

---

\*Corresponding author.

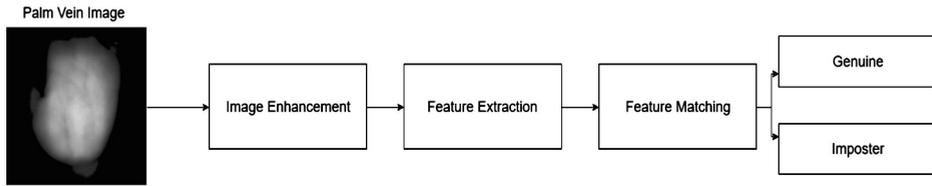


Fig. 1. Palm vein recognition process.

are largely unable to meet today's convenience and reliability needs. Biometrics technologies classify authentication identification into physiological features, such as the user's face, fingerprint, iris, and veins, or personal behaviour features, such as signature, gait, and voice (Jain *et al.*, 2006). Among these features, internal biometrics (e.g. veins) are more secure than external biometrics (e.g. fingerprint, face). Facial recognition is very popular and identifies individuals by their unique facial features. However, facial recognition may be degraded due to external factors, such as hair, wearable objects, and injuries. Fingerprint recognition is also a widely used identification technology, but it may be considered unsanitary due to the need for contact and result in user resistance. In addition, under environmental conditions, such as dirt and injuries, fingerprint recognition may be insufficient and performance may be reduced. Moreover, handwriting or signatures can be easily forged. In contrast, internal biometrics are difficult to counterfeit, but acquisition is more challenging, less environmentally friendly, and must effectively prevent external damage. Vein recognition stands out as a promising recognition technology. In 1992, Shimizu first considered the application of the hand blood vessel method of authentication (Shimizu, 1992).

The process of recognizing palm veins is complicated by image preprocessing, feature extraction, and matching methods. The process is shown in Fig. 1. First, a palm vein image is preprocessed to make the image features more obvious. Second, a method must be found to extract the features of the image. Finally, feature matching must be performed via distance measures to identify the user. To reduce the manual selection process, deep learning is proposed, and target features can be automatically selected through model training. Deep learning excels in issues related to visual recognition, voice recognition, natural language, and time series. Convolutional neural networks (CNNs) were further studied by LeCun *et al.* in the 1990s, and their structure has been proposed (Lécun *et al.*, 1998).

The major contributions of this paper are as follows:

- We propose a palm vein recognition method based on a CNN that uses palm vein features to identify users.
- CNNs reduce the labourious manual selection process and simplifies authentication.
- Finally, we test the performance metrics of CNN, such as the False Acceptance Rate (FAR) and the False Rejection Rate (FRR).

### 1.1. Related Work

Vein recognition includes palm veins (Kang and Wu, 2014; Lu and Yuan, 2017; Raut and Humbe, 2015), finger veins (Huang *et al.*, 2017), hand dorsal veins (Khan *et al.*, 2009; Raghavendra *et al.*, 2015), wrist veins (Mohamed *et al.*, 2017; Pascual *et al.*, 2010), forearm veins (Choras, 2017), and sclera veins (Lin *et al.*, 2014; Suganya and Sivitha, 2014). Of these, we use palm vein images for recognition. The literature includes many relevant palm vein recognition algorithms (Kong *et al.*, 2009), such as:

- Structure-based methods use line features and point features. This information is highly dependent on the selected coordinates and very sensitive to spatial occlusion; thus, the region of interest (ROI) must be carefully selected. Usually, when poorly performing images and feature points are used in feature extraction, difficulty in matching occurs. This method is also relatively sensitive to scaling, rotation, and displacement (Akin-sowon and Alese, 2013; Xu, 2015).
- Appearance-based (subspace-based) methods use features to reduce dimensionality from high-dimensional space to low-dimensional space and retain the required information. Common methods include PCA, LDA, ICA, and subspace clustering (SSC) (Liu and Zhang, 2011; Raut and Humbe, 2015).
- Statistical-based methods use local statistics, such as the mean and variance of each small area, which are calculated and considered as characteristics. Gabor, wavelet, and Fourier transforms have been applied, including local binary patterns (LBP) or local derivative patterns (LDP) (Aglio-Caballero *et al.*, 2017; Kang and Wu, 2014).
- Local invariant-based methods use an image pyramid that is constructed to form a three-dimensional image space. The local maxima of each layer are obtained using a Hessian matrix, and a neighbourhood corresponding to the scale is selected at the feature point to find the main direction. With the main direction as the axis, coordinates can be established at each feature point with methods including scale invariant feature transform (SIFT) and speeded up robust features (SURF) (Gurunathan *et al.*, 2016).
- Fusion, also a very popular method, can improve the accuracy and security of the system, but the relative amount of data required is large. There are many fusion methods, including the use of palm, finger, vein, and face and many fusion rules, including algorithms, SVM, and neural networks (Garg *et al.*, 2016; Kim *et al.*, 2010).
- Deep learning. In recent years, the prevalence of neural networks has made CNNs the most popular image classification method; the recognition rate is not inferior to previous methods (Huang *et al.*, 2017).

### 1.2. Our Work

With the many research topics discussed above, because of the convenience and practicality of CNNs, we propose a palm vein recognition method based on CNN. The main contributions of our work are as follows. The vein pattern of the palm of the hand was photographed with a near-infrared camera using a pretrained CNN model called AlexNet (Krizhevsky *et al.*, 2012). VGG (Simonyan and Zisserman, 2014) was trained and the

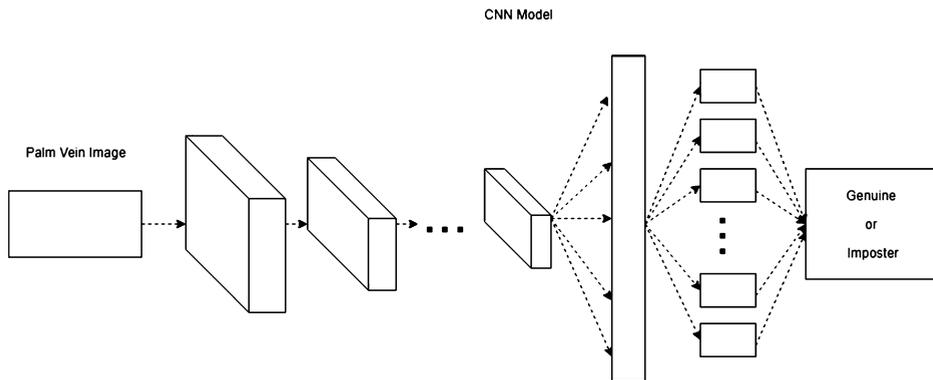


Fig. 2. Using the CNN framework for palm vein recognition.

result was identified. Although the size of the training set is very different from that of the competition, good performance is still achieved. The simplified steps required for the overall palm vein recognition are shown in Fig. 2. The first layer is the input layer, followed by the interaction of the convolution layer, pooling layer, and activation function, and finally, classification by the fully connected layer and the classifier.

## 2. Preliminaries

In this section, we introduce the CNN model and the network architecture. In the proposed scheme, the model is trained and used for identification. A CNN has advantages in image recognition. According to the pioneering work of Lenet (LéCun *et al.*, 1998), CNNs have the following three characteristics: local perception, downsampling, and weight sharing. Since 2010, ImageNet has held the Large-Scale Visual Recognition (ILSVR) Competition annually. In ILSVR 2012, the appearance of AlexNet stood out, and it became the future of CNNs. It was followed by the appearance of famous models, such as ZFNet, VGG, GoogLeNet, ResNet, and SENet. Here, we use AlexNet and VGG.

### 2.1. Convolution Neural Network Structure

In this paper, a CNN is used to analyse the important parts of the network and the overall architecture. A CNN is a type of deep neural network. It has important advantages in image recognition processing. CNN feature extraction is different in each layer, and as the number of layers increases, feature classification improves.

Previous neural networks were still very limited when dealing with problems, such as computer vision, natural language processing and voice recognition. Due to the appearance of the convolution kernel, CNN has advantages for processing multidimensional data. The most important concepts of the CNN are the convolution layer, activation function, pooling layer, and fully connected layer.

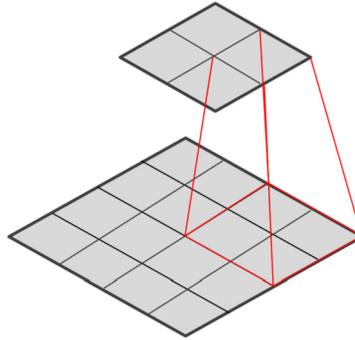


Fig. 3. Convolution operation of  $4 \times 4$  image with  $2 \times 2$  filters.

## 2.2. Convolution Layer

The processing performed by the convolutional layer is called the “convolutional operation”. It uses filters to process and extract features and reduce noise. As the local links and weights are shared, the training parameters are significantly reduced. In high-dimensional images, all connected neurons will need to train a large number of parameters, which causes the calculation to be time consuming and leads to severe overfitting. The characteristics of local links and weight sharing simplifies parameter training. As shown in Fig. 3, the image is  $4 \times 4$  in size and convoluted with a  $2 \times 2$  filters.

## 2.3. Activation Function

The activation function is used to add nonlinear factors. If the activation function is not used in the neural network, the output and input cannot be separated from each other using a linear relationship. Therefore, it is meaningless to make a deep neural network without the activation function. The Sigmoid and Tanh functions were previously used as the activation function, but gradient disappearance easily occurs when using these functions. When the neural network is deepened, training obstacles often occur. The ReLU function can effectively overcome gradient disappearance and requires less computation. The ReLU function gradually replaced the Sigmoid and Tanh functions.

$$\text{Sigmoid: } \varphi(X) = \frac{1}{1 + e^{-x}}, \quad (1)$$

$$\text{Tanh: } \varphi(X) = \frac{e^x - e^{-x}}{e^x + e^{-x}}, \quad (2)$$

$$\text{ReLU: } \varphi(X) = \max(0, x). \quad (3)$$

Figure 4 shows the three common activation functions.

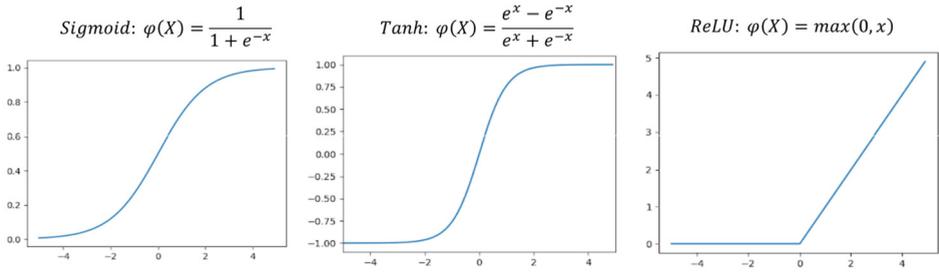


Fig. 4. The activation functions.

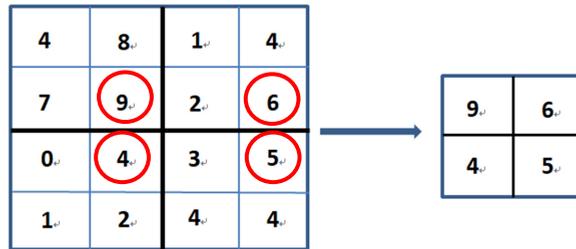


Fig. 5. The max pooling processing.

### 2.4. Pooling Layer

The data extracted by the convolutional layer have a high data dimension. The primary role of the pooling layer is to reduce the data dimension to avoid overfitting. The commonly used methods are max pooling and average pooling. In addition to reducing data dimensionality, pooling is resistant to image translation and slight deformation. Figure 5 shows the max pooling processing. For reducing the data dimensionality, pooling extracts the largest features with a specific window size to avoid overfitting problems.

### 2.5. Fully Connected Layer

The entire CNN acts as a classifier. If the operations of the convolutional layer, pooling layer, and activation function map the original data to the hidden layer feature space, the fully connected layer will learn the “distributed feature representation” maps to the role of the sample tag space. Among them, the result is classified by Softmax and normalized to obtain the probability value, as shown in (4).

$$S_j = \frac{e^{a_j}}{\sum_{k=1}^T e^{a_k}}, \tag{4}$$

where  $a_j$  denotes the  $j$ th value in the vector;  $a_k$  in the denominator denotes the value in the input vector, so it will have the summation symbol.

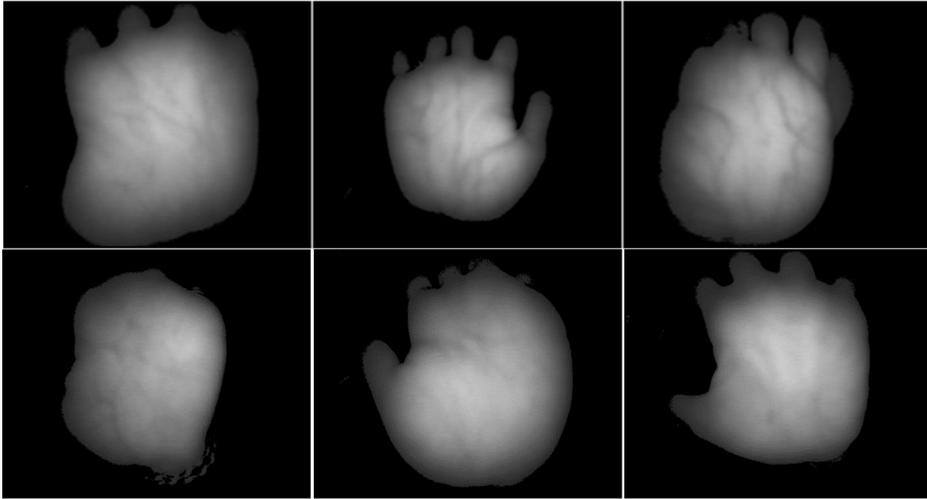


Fig. 6. Original palm vein images.

### 3. Methodology

#### 3.1. Palm Vein Database

The two palm vein databases currently published have a multispectral palm vein database. The database equipment used by the Hong Kong Polytechnic University has a fixed contact type acquisition device (Zhang *et al.*, 2010). The Chinese Academy of Sciences Institute of Automation uses five wavelengths for spectral acquisition (Hao *et al.*, 2008), in which images with wavelengths of 850 to 940 nm are clear. The equipment used is a noncontact acquisition device. We established a database of palm veins. Twenty images from each of 50 individuals were collected, for a total of 1 000 experimental images. The size of the captured image is  $640 \times 480$  pixels. For AlexNet and VGGNet, the size must be changed to  $227 \times 227$  and  $224 \times 224$ , respectively. Figure 6 shows the palm vein images; each image is from a different person.

#### 3.2. AlexNet

Krizhevsky *et al.* (2012) proposed a CNN in the ImageNet contest and earned first place in image classification. Subsequently, this convolutional neural network became popular. The network architecture of AlexNet included five convolution layers and three fully connected layers; the network structure is shown in Fig. 7. The five convolutional layers contain the processes of the convolutional layers, pooling layers, activation function, and local normalization. Because the network uses two GTX580 GPUs, it is divided into upper and lower parts, which can be used for interaction. The fully connected layers have 4,096 neurons each. The image is classified by Softmax.

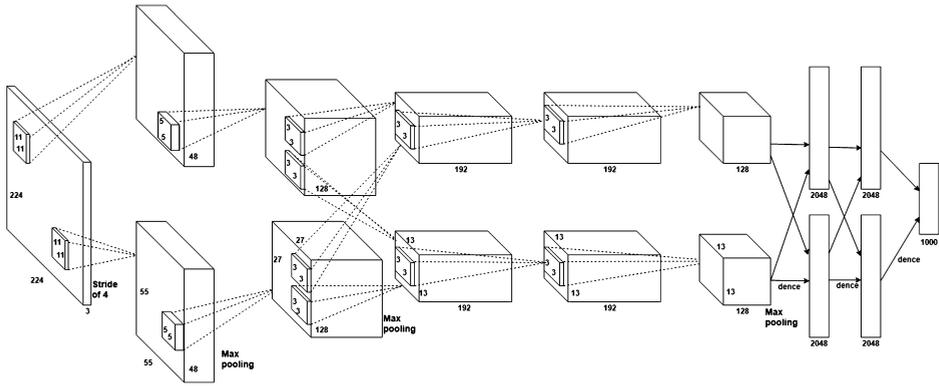


Fig. 7. Structure of AlexNet.

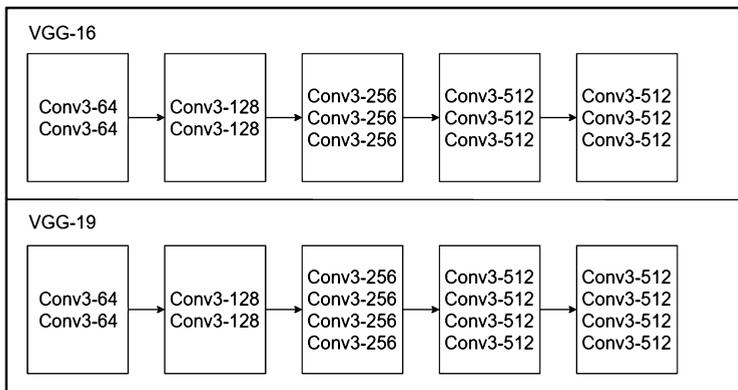


Fig. 8. Structure of VGG16 and VGG19.

### 3.3. VGG-16 and VGG-19

In 2014, an image recognition method based on a VGG network was developed by the University of Oxford. VGG was the first to use small  $3 \times 3$  filters to check the image in each convolutional layer. VGG uses multiple  $3 \times 3$  filters to extract high-dimensional features; it is not designed to use larger convolutional kernels, as in previous network architectures, and it avoids using  $1 \times 1$  convolutional kernels, as in shallow networks. If a larger convolution kernel is used, the parameters will increase significantly and the computation time will increase. For example, a  $3 \times 3$  convolution kernel has only 9 weight parameters; in a  $7 \times 7$  convolution kernel, the number of weight parameters increases to 49. The appearance of VGG demonstrates that the depth of a CNN upgrade can achieve better recognition results. VGG has a variety of structures, with the configuration depth increasing from (A) to (E). We choose structures D and E. The VGG structures used have 13 and 16 convolution layers and three fully connected layers. The structure is shown in Fig. 8.

### 3.4. Transfer Learning

In most machine learning, deep learning, and data mining tasks, we assume the training and inference use data following the same distribution and coming from the same feature space. However, in practical applications, this assumption is difficult to establish and often encounters some problems:

- The number of marked training samples is limited. For example, when dealing with the classification of the target domain, there is a lack of sufficient training samples. Additionally, there is a large number of training samples in the B (source) domain related to the A field, but the B field and the A field are in different feature spaces or their samples obey different distributions.
- The data distribution will change. The data distribution can be related to time, location or other dynamic factors. With a change in dynamic factors, the data distribution will change. Thus, the previously collected data are outdated, and so it must be recollected and the model must be rebuilt.

At this time, knowledge transfer, i.e. transferring the knowledge in the B field to the A field and improving the classification in the A field, whose data do not require much time to mark, is a good choice. Migration learning, a new method of transfer learning, was proposed to solve this problem (Pan and Yang, 2010).

Transfer learning is a research issue in machine learning. The model will be trained by transferring the trained model parameters to new problems. Considering that most of the data or tasks are relevant, we can learn the model parameters (this can also be understood as model learned knowledge) through transfer learning in a way that optimizes and improves the speed of the learning efficiency of the model and does not require learning from zero, as with most networks.

We use pretrained AlexNet and VGG to move from the 1 000 classes of the original model to our palm vein categories.

Table 1 shows the network architecture of AlexNet, including the use of each layer, stride and padding in the convolution layer, following the rules to adjust the image, and the dropout in the fully connected layer. We used transfer learning to convert the last three layers into our classification, as in the AlexNet example. VGG16 and VGG19 use the same method to replace the last three layers.

Table 1 is the initial network architecture. The last three layers are initially set to 1 000 categories, and we adjust for the new categories in Table 2.

In Table 2, we adjust the final fully connected layer, Softmax layer, and output layer to our experimental classification.

## 4. Experiment and Result Analysis

In this section we discuss the results of the CNN palm vein recognition experiments and analysis. We constructed AlexNet and VGG using MATLAB and the deep learning convolutional neural network framework. The main process includes the following: image acquisition, model training, parameter adjustment, and result identification.

Table 1  
Architecture network of AlexNet.

1	'data'	Image Input	227x227x3 images with 'zerocenter' normalization
2	'conv1'	Convolution	96 11x11x3 convolutions with stride [4 4] and padding [0 0 0 0]
3	'relu1'	ReLU	ReLU
4	'norm1'	Cross Channel Normalization	cross channel normalization with 5 channels per element
5	'pool1'	Max Pooling	3x3 max pooling with stride [2 2] and padding [0 0 0 0]
6	'conv2'	Convolution	256 5x5x48 convolutions with stride [1 1] and padding [2 2 2 2]
7	'relu2'	ReLU	ReLU
8	'norm2'	Cross Channel Normalization	cross channel normalization with 5 channels per element
9	'pool2'	Max Pooling	3x3 max pooling with stride [2 2] and padding [0 0 0 0]
10	'conv3'	Convolution	384 3x3x256 convolutions with stride [1 1] and padding [1 1 1 1]
11	'relu3'	ReLU	ReLU
12	'conv4'	Convolution	384 3x3x192 convolutions with stride [1 1] and padding [1 1 1 1]
13	'relu4'	ReLU	ReLU
14	'conv5'	Convolution	256 3x3x192 convolutions with stride [1 1] and padding [1 1 1 1]
15	'relu5'	ReLU	ReLU
16	'pool5'	Max Pooling	3x3 max pooling with stride [2 2] and padding [0 0 0 0]
17	'fc6'	Fully Connected	4096 fully connected layer
18	'relu6'	ReLU	ReLU
19	'drop6'	Dropout	50% dropout
20	'fc7'	Fully Connected	4096 fully connected layer
21	'relu7'	ReLU	ReLU
22	'drop7'	Dropout	50% dropout
23	'fc8'	Fully Connected	1000 fully connected layer
24	'prob'	Softmax	softmax
25	'output'	Classification Output	crossentropyex with 'tench', 'goldfish', and 998 other classes

Table 2  
Fully connected layer and output layer adjustment.

23	'fc8'	Fully Connected	1000 fully connected layer
24	'prob'	Softmax	softmax
25	'output'	Classification Output	crossentropyex with 'tench', 'goldfish', and 998 other classes
23	'fc'	Fully Connected	50 fully connected layer
24	'softmax'	Softmax	softmax
25	'classoutput'	Classification Output	crossentropyex with 'test (1)', 'test (10)', and 48 other classes

#### 4.1. Dataset

We established two set of datasets of palm veins. The first dataset contains twenty images from each of 50 individuals, for a total of 1 000 experimental images. The second dataset contains twenty images from each of 63 individuals, for a total of 1 260 experimental images. The size of a captured image is  $640 \times 480$  pixels. For AlexNet and VGGNet, the size must be changed to  $227 \times 227$  and  $224 \times 224$ , respectively. The image acquisition uses a near-infrared camera to acquire palm vein images, and two sets of image sets, i.e. the training set and validation set, were established in the experimental process. Of these images, 80% were used for training, and 20% were used for validation.

Table 3  
Equipment of the first dataset.

Equipment	Device configuration
System	Windows 10 64-bit
CPU	Intel(R)Core(TM)i7-6500 CPU @ 2.50 GHz
RAM	8 GB
GPU	GTX940MX
Framework	MATLAB

Table 4  
Equipment of the second dataset.

Equipment	Device configuration
System	Windows 10 64-bit
CPU	Intel(R)Core(TM)i7-8700 CPU @ 3.20 GHz
RAM	32 GB
GPU	GTX1080
Framework	MATLAB

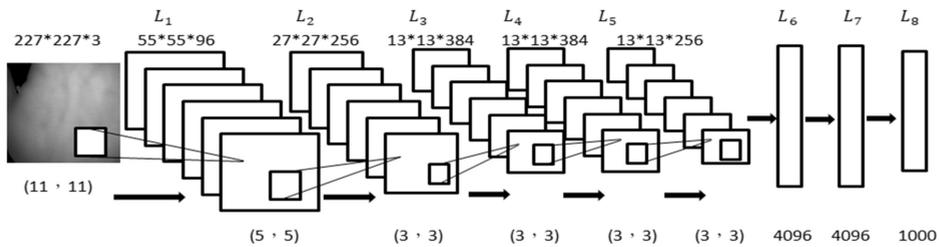


Fig. 9. Training process of AlexNet.

#### 4.2. Equipment

The experimental environment for our equipment of the first dataset is shown in Table 3. We use a Windows 10 64-bit system, and 8 G memory, NVIDIA GTX940MX; the framework is MATLAB. The equipment of the second dataset is shown in Table 4. We use a Windows 10 64-bit system, and 32 G memory, NVIDIA GTX1080; the framework is MATLAB.

#### 4.3. Training Process

The training of CNN is to optimize various parameters. Different parameters, as well as the network structure and usage, will affect the effectiveness and training speed of CNN recognition. The AlexNet neural network model is used as an example to illustrate the training process of palm vein recognition, whose network structure is shown in Fig. 9. The input image size of the first layer is 227\*227\*3, where 3 represents the colour of the

Table 5  
Training iterations results of the first dataset.

Models	Iterations	Result
AlexNet	800	96%
VGG-16	800	97.5%
VGG-19	1000	98.5%

image (RGB). The equation for the convolution and pooling layers is

$$\begin{aligned} OH &= \frac{H + 2P - FH}{S + 1}, \\ OW &= \frac{W + 2P - FW}{S + 1} + 1, \end{aligned} \quad (5)$$

where  $OH$  and  $OW$  depict the output size;  $H$  and  $W$  denote the height and width of the input;  $FH$  and  $FW$  are the size of the filter;  $P$  denotes Padding; and  $S$  is the Stride. The input and output size can be calculated using the above equation.

The first convolution layer ( $L_1$ ), where the Kernel\_size is 11 and the Stride is 4. After this layer, the image size becomes  $55*55*96$ . After the ReLU excitation function, the output image size remains unchanged and the output of the previous layer is the input of the next layer. Then Maxpooling is done, where Kernel\_size is 3, Stride is 2, and the output image size is  $27*27*96$ . The convolution process of the second layer ( $L_2$ ) is the same as the first layer. After 256 convolutions with Kernel\_size of 5 and Stride of 1, the parameters remain the same. However, in the second layer, the Padding is 2, and the output image size is  $27*27*256$  after doing the same Maxpooling operation. The third ( $L_3$ ) and fourth layers ( $L_4$ ) only do convolution and ReLU, the fifth layer ( $L_5$ ) is similar to the first layer, and the final output becomes a long vector. Here, three fully connected layers are used, the first two layers have 4096 Kernels and the last layer has 1000 nodes. The final output is classified using the Softmax layer.

#### 4.4. Experimental Results

##### 4.4.1. Experimental Results of the First Dataset

In this paper, the experimental process includes the following: randomly selecting the training image and validation sets, adjusting the batch value according to the performance of the hardware, and adjusting the learning rate. The hyperparameter adjustment on the model will affect the experimental results. In this process, the final numbers of iterations were 800, 800, and 1000 in AlexNet, VGG-16 and VGG-19, respectively, as shown in Table 5, and the accuracy rates were 96%, 97.5%, and 98.5%, respectively. Experimental process diagrams are shown in Figs. 10, 11 and 12. The loss function was used to evaluate the degree of inconsistency between the predicted value and the true value of the model. It is a nonnegative real-valued function; the smaller the loss function, the better

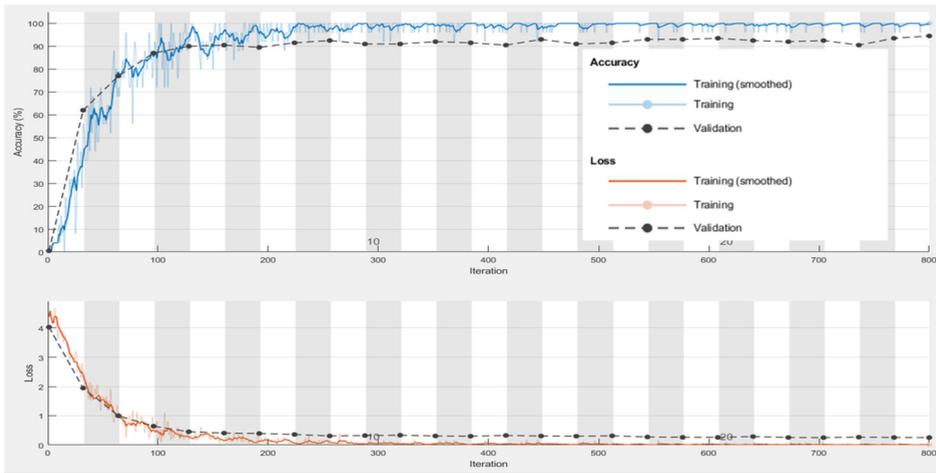


Fig. 10. Accuracy, iterations and loss for iterations of AlexNet of the first dataset.

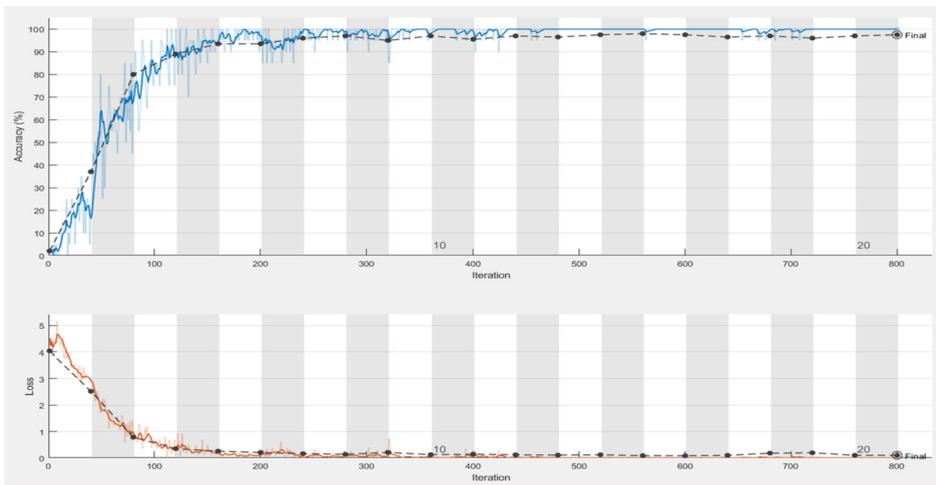


Fig. 11. Accuracy, iterations and loss for iterations of VGG-16 of the first dataset.

the robustness of the model. The loss function is as follows:

$$E(t, y) = - \sum_j t_j \log y_j, \tag{6}$$

where  $t$  and  $y$  are neural network target tags and outputs, respectively.

The number of iterations in Table 5 is either 800 or 1000, and the recognition rate is as high as 98.5%.

The comparison with other authors' research results are shown in Table 6. Frontasari and Gunawan (2017) proposed using DCLBP, which is a LBP, average LBP, median LBP

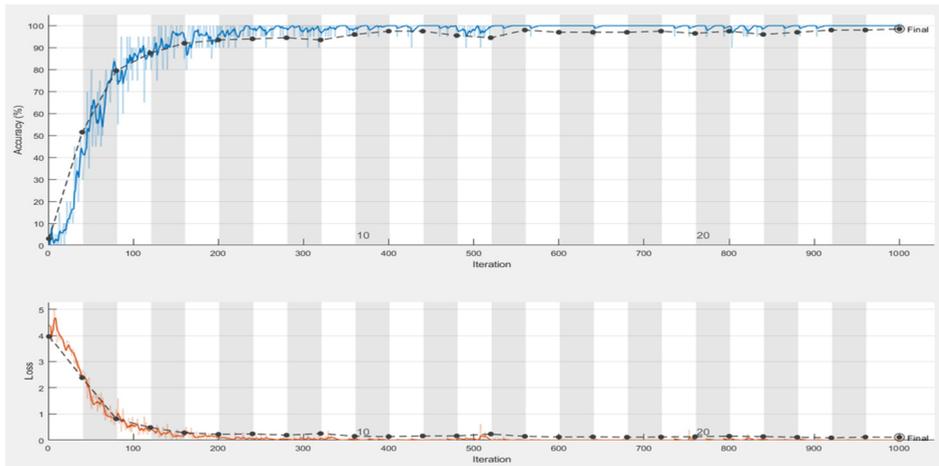


Fig. 12. Accuracy, iterations and loss for iterations of VGG-19 of the first dataset.

Table 6  
Test results for various algorithms of the first dataset.

Algorithm	Result
LBP+PNN (Fronitasari and Gunawan, 2017)	98%
SURF (Badrinath <i>et al.</i> , 2013)	96.7%
Gabor (Lee, 2012)	96.37%
LBPROT (Pratiwi <i>et al.</i> , 2016)	96%
AlexNet (ours)	96%
VGG-16 (ours)	97.5%
VGG-19 (ours)	98.5%

and LBP original corrections, and feature matching process using PNN. (Badrinath *et al.*, 2013) proposed using the geometric hash of the SURF keypoint to index and perform a fractional fusion of the voting strategy. Lee (2012) proposed using the 2-D Gabor filter to represent a palm vein image using its texture features and the hamming distance. In addition, the author proposed a direction coding to code the palm vein features in two-bit representation. (Pratiwi *et al.*, 2016) proposed using the LBPROT feature extraction as a palm vein image matching process by using the Cosine Distance. Among the three models we used, the highest accuracy rate was 98.5%. Compared with other papers, we simplified the entire palm vein recognition process, and the preprocessing part only adjusted the image size. The experimental results in Table 5 show that, in the previous research methods, the steps taken are too complex. Compared with the previous CNN-based palm vein recognition methods, the only pretreatment we have done is image size reduction and step reduction; the entire process obtained excellent results.

Based on the performance of biometrics-based verification systems (Seshikala *et al.*, 2012; Vaid and Mishra, 2015), the false acceptance rate (FAR) and the false rejection rate (FRR) are the primary metrics used to evaluate the performance of the model in this paper. Among them, there are the four categories of True Positive (*TP*), False Positive (*FP*), False

Table 7  
Performance metrics of the first dataset.

Performance metrics			
No.	Models	FAR (%)	FRR (%)
1	AlexNet	0	0.77
2	VGG-16	0	0.65
3	VGG-19	0	0.6

Negative ( $FN$ ) and True Negative ( $TN$ ) used in these two metrics. If the test is positive and classified as positive, it is  $TP$ . If the test is positive and classified as negative, it is  $FP$ . If the test is negative and classified as positive, it is  $FN$ . If the test is negative and classified as negative, it is  $TN$ . Our performance metrics used in this paper are FAR and FRR. FAR and FRR can be measured with the following formulas:

$$FAR = \frac{FP}{FP + TN}, \quad (7)$$

$$FRR = \frac{FN}{TP + FN}. \quad (8)$$

The statistics of our performance metrics for the three models are shown in Table 7. In the table, our error rates are compared using the performance metrics rules. We also have a fairly low error rate, and with the three models, FRR is also a fairly low.

From Table 7, it can be seen that performance of CNN is quite remarkable among the unprocessed images.

#### 4.4.2. Experimental Results of the Second Dataset

In this study, we divided the image dataset into two sets, namely a training set and a validation set. For the training and validation of the three models AlexNet, VGG-16, and VGG-19, we set the number of iterations to be 1000 times.

The palm vein images in the database were collected by using a near-infrared light camera. Since these were contactless shots, each image had a different resolution and needed to be preprocessed. To obtain the best result of image contrast enhancement while avoiding noise amplification, we used a new technique for contactless palm vein recognition. To preprocess the input image, our design uses contrast limited adaptive histogram equalization (CLAHE) to enhance the image quality and thus the feature distinguishability. Figure 13(a) shows a raw image, and Fig. 12(b) shows the enhanced image.

The training processes of the three models are shown in Fig. 14. In Table 8, we present the final results of accuracy of the three models. All three models gave excellent results with the accuracy rate of AlexNet being 99.35%, the accuracy rate of VGG-16 being 99.45%, and the accuracy rate of VGG-19 being 99.5%.

The smaller the loss function, the better the stability. The loss function is the core part of the empirical risk function and an important part of the structural risk function. The common loss function includes the following factors:

- Hinge Loss: Mainly used in support vector machines (SVM).

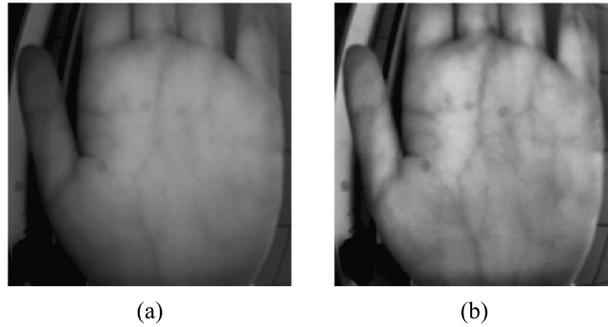


Fig. 13. (a) Original image; (b) Enhanced image.

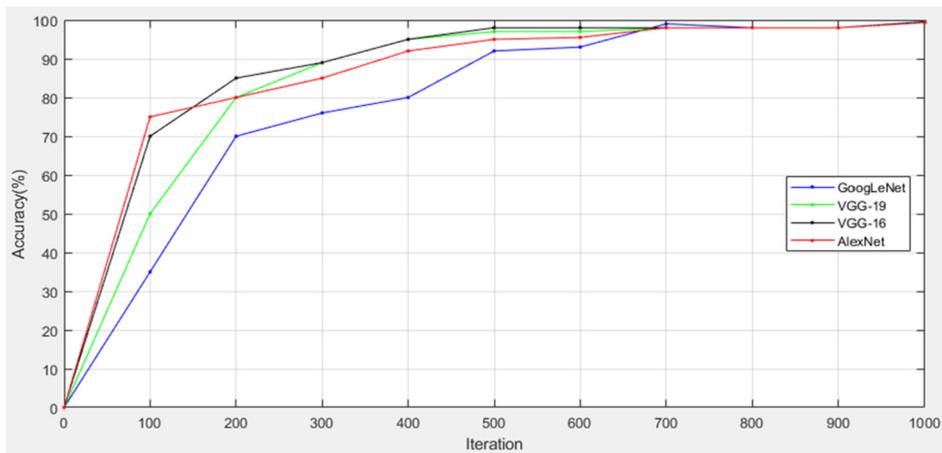


Fig. 14. Training processes and accuracy of the models of the second dataset.

Table 8  
Training iteration results of the second dataset.

Models	Iterations	Result
AlexNet	1000	99.35%
VGG-16	1000	99.45%
VGG-19	1000	99.5%

- Cross Entropy Loss; Softmax Loss: Used in Logistic Regression and Softmax Classification.
- Square Loss: Mainly in Ordinary Least Squares (OLS).
- Exponential Loss: Mainly used in the Adaboost integrated learning algorithm.
- Other loss (such as 0–1 loss, absolute value loss)

Here we mainly use Cross Entropy Loss, or Softmax Loss. Figure 15 shows the loss function of the training processes of the four models. It can be seen that all four models were

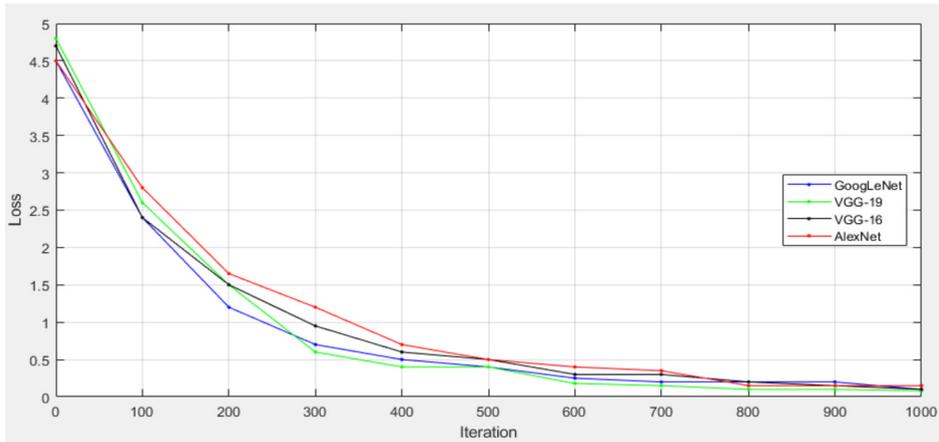


Fig. 15. Training process and loss of the models of the second dataset.

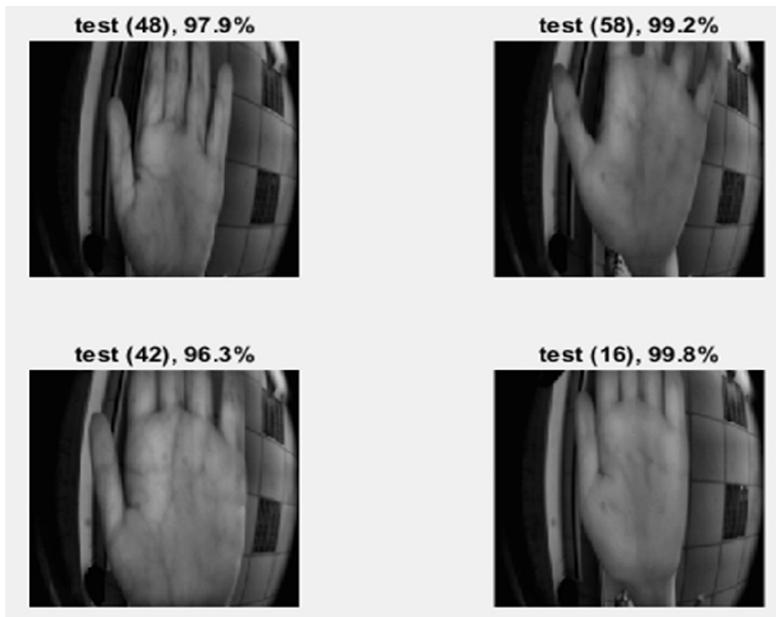


Fig. 16. Training results of four random users.

very stable after training. Figure 16 shows the training and verification results for four different users randomly selected after the final validation.

In addition, we also tested the models with two different graphics cards, and the training time and accuracy differed accordingly. The results are shown in Table 9. Please note that in this part of our experiment, the test image had not been preprocessed.

The statistics of our performance metrics for the three models of the second dataset are shown in Table 10.

Table 9  
Performance comparison among models when using different graphics cards.

Models	GTX940		GTX1080	
	Training time	Accuracy	Training time	Accuracy
AlexNet	695 s	96%	83 s	97%
VGG-16	757 s	97.5%	515 s	98%
VGG-19	832 s	98.5%	560 s	98.5%

Table 10  
Performance metrics of the second dataset.

Performance metrics			
No.	Models	FAR (%)	FRR (%)
1	AlexNet	0	0.5
2	VGG-16	0	0.35
3	VGG-19	0	0.3

Table 11  
Performance comparison among similar methods.

Methods	FAR (%)	FRR (%)	Accuracy (%)
Gabor (Cancian <i>et al.</i> , 2017)	0.32	1.75	99.6
LeNet (Zheng <i>et al.</i> , 2017)	0	0.86	99.1
AlexNet (Zheng and Han, 2018)	0	0.7	99.2
AlexNet (ours)	0	0.5	99.35
VGG-16 (ours)	0	0.35	99.45
VGG-19 (ours)	0	0.3	99.5

The performance comparison between our approaches with others are shown in Table 11. Cancian *et al.* (2017) proposed a method that was the combination of the Gabor filter and histogram calculations to create a biometric template. Zheng *et al.* (2017) proposed a LeNet structure to process palm veins, and the method achieved an accuracy rate of 99.1%. Zheng and Han (2018) proposed an AlexNet design to do palm vein image ROI extraction. We compared our approaches with the three similar methods above in terms of FAR, FRR, as well as accuracy rate, and the results are shown in Table 11. As the results show, all the models of ours had lower error rates and higher accuracy rates, demonstrating the superiority of our approached over the others.

## 5. Conclusion

This paper proposed a palm vein recognition method based on a convolutional neural network. Through AlexNet and VGG, the entire palm vein recognition process is simplified, and image processing, feature extraction, and matching methods are no longer as complicated. The unique advantages of CNNs compared to other methods include greater convenience and excellent accuracy rates; there are also good results for the performance

metrics. In our first palm vein dataset, 1000 images from 50 individuals were adopted for testing, and an FRR of 0.6% was achieved. The above three models provide a new research method for palm vein recognition and prove the advantages of deep learning in the image field. The second dataset, including 1260 images from 63 individuals, was adopted for testing, and an FRR of 0.3% was achieved. We tried to preprocess the palm vein image, use the CLAHE method to increase the contrast of the image, highlight its feature, and improve the accuracy of the three types of CNN to 99%. When using different graphics cards, the training time will have an enormous impact, and the accuracy will be slightly affected.

In the future, this study can be improved in two directions. First, the palm vein images used in our method need to be clear and uncontaminated files, and the effect on shadows is not very good. In the future, we can further improve the pre-processing method of CLAHE proposed in this paper to remove some noise and shadows, which should improve the recognition rate. The second one is towards the finger vein recognition so that it can be applied to the security system control of handheld mobile devices.

## Funding

This research was partially supported by the Ministry of Science and Technology (MOST), Taiwan, R.O.C., under contract no.: MOST 108-2410-H-030-074 and 110-2410-H-030-032. This work was also partially supported by the “Center for Cyber-Physical System Innovation” from the Featured Areas Research Center Program within the framework of the Higher Education Sprout Project by the Ministry of Education (MOE) in Taiwan and MOST under 109-2221-E-011-115, 110-2221-E-011-125, 110-2218-E-011-006-MBK.

## References

- Aglío-Caballero, A., Ríos-Sánchez, B., Sánchez-Ávila, C., de Giles, M.J.M. (2017). Analysis of local binary patterns and uniform local binary patterns for palm vein biometric recognition. In: *International Carnahan Conference on Security Technology (ICCST)*, pp. 1–6.
- Akinsowon, O.A., Alese, B.K. (2013). Edge detection methods in palm-print identification. In: *8th International Conference for Internet Technology and Secured Transactions (ICITST-2013)*, pp. 422–426.
- Badrinath, G.S., Gupta, P., Mehrotra, H. (2013). Score level fusion of voting strategy of geometric hashing and SURF for an efficient palmprint-based identification. *Journal of Real-Time Image Processing*, 8(3), 265–284.
- Cancian, P., Donato, G.W.D., Rana, V., Santambrogio, M.D. (2017). An embedded Gabor-based palm vein recognition system. In: *IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)*, pp. 405–408.
- Choras, R.S. (2017). Personal identification using forearm vein patterns. In: *International Conference and Workshop on Bioinspired Intelligence (IWOBI)*, pp. 1–5.
- Fronitasari, D., Gunawan, D. (2017). Palm vein recognition by using modified of local binary pattern (LBP) for extraction feature. In: *15th International Conference on Quality in Research (QIR): International Symposium on Electrical and Computer Engineering*, pp. 18–22.
- Garg, S.N., Vig, R., Gupta, S. (2016). Multimodal biometric system based on decision level fusion. In: *International Conference on Signal Processing, Communication, Power and Embedded System (SCOPE5)*, pp. 753–758.
- Gurunathan, V., Sathiyapriya, T., Sudhakar, R. (2016). Multimodal biometric recognition system using SURF algorithm. In: *10th International Conference on Intelligent Systems and Control (ISCO)*, pp. 1–5.

- Hao, Y., Sun, Z., Tan, T., Ren, C. (2008). Multispectral palm image fusion for accurate contact-free palmprint recognition. In: *15th IEEE International Conference on Image Processing*, pp. 281–284.
- Huang, H., Liu, S., Zheng, H., Ni, L., Zhang, Y., Li, W. (2017). DeepVein: novel finger vein verification methods based on deep convolutional neural networks. In: *IEEE International Conference on Identity, Security and Behavior Analysis (ISBA)*, pp. 1–8.
- Jain, A.K., Bolle, R., Pankanti, S. (2006). *Biometrics: Personal Identification in networked Society*. Springer Science & Business Media, Vol. 479.
- Kang, W., Wu, Q. (2014). Contactless palm vein recognition using a mutual foreground-based local binary pattern. *IEEE Transactions on Information Forensics and Security*, 9(11), 1974–1985.
- Khan, M.H.-M., Subramanian, R.K., Khan, Ali Mamode, N. (2009). Representation of hand dorsal vein features using a low dimensional representation integrating cholesky decomposition. In: *2nd International Congress on Image and Signal Processing*, pp. 1–6.
- Kim, Y., Toh, K.A., Teoh, A.B.J. (2010). An online learning algorithm for biometric scores fusion. In: *Fourth IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, pp. 1–6.
- Kong, A., Zhang, D., Kamel, M. (2009). A survey of palmprint recognition. *Pattern Recognition*, 42(7), 1408–1418.
- Krizhevsky, A., Sutskever, I., Hinton, G.E. (2012). ImageNet classification with deep convolutional neural networks. In: *International Conference on Neural Information Processing Systems*, pp. 1097–1105.
- Lécun, Y., Bottou, L., Bengio, Y., Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11), 2278–2324.
- Lee, J.C. (2012). A novel biometric system based on palm vein image. *Pattern Recognition Letters*, 33(12), 1520–1528.
- Lin, Y., Du, E.Y., Zhou, Z., Thomas, N.L. (2014). An efficient parallel approach for sclera vein recognition. *IEEE Transactions on Information Forensics and Security*, 9(2), 147–157.
- Liu, J., Zhang, Y. (2011). Palm-dorsa vein recognition based on two-dimensional fisher linear discriminant. In: *International Conference on Image Analysis and Signal Processing*, pp. 550–552.
- Lu, W., Yuan, W.Q. (2017). Comparison of four local invariant characteristics based on palm vein. In: *IEEE International Conference on Computational Science and Engineering (CSE) and Embedded and Ubiquitous Computing (EUC)*, pp. 850–853.
- Mohamed, C., Akhtar, Z., Eddine, B.N., Falk, T.H. (2017). Combining left and right wrist vein images for personal verification. In: *Seventh International Conference on Image Processing Theory, Tools and Applications (IPTA)*, pp. 1–6.
- O’Gorman, L. (2003). Comparing passwords, tokens, and biometrics for user authentication. *Proceedings of the IEEE* 91(12), 2021–2040.
- Pan, S.J., Yang, Q. (2010). A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, 22(10), 1345–1359.
- Pascual, J.E.S., Uriarte-Antonio, J., Sanchez-Reillo, R., Lorenz, M.G. (2010). Capturing hand or wrist vein images for biometric authentication using low-cost devices. In: *Sixth International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, pp. 318–322.
- Pratiwi, A.Y., Budi, W.T.A., Ramadhani, K.N. (2016). Identity recognition with palm vein feature using local binary pattern rotation Invariant. In: *4th International Conference on Information and Communication Technology (ICoICT)*, pp. 1–6.
- Raghavendra, R., Surbiryala, J., Busch, C. (2015). Hand dorsal vein recognition: sensor, algorithms and evaluation. In: *IEEE International Conference on Imaging Systems and Techniques (IST)*, pp. 1–6.
- Raut, S.D., Humbe, V.T. (2015). Palm vein recognition system based on corner point detection. In: *IEEE International WIE Conference on Electrical and Computer Engineering (WIECON-ECE)*, pp. 499–502.
- Seshikala, G., Kulkarni, U., Giriprasad, M.N. (2012). Biometric parameters & palm print recognition. *International Journal of Computer Applications*, 46(21), 31–34.
- Shimizu, K. (1992). Optical trans-body imaging feasibility of optical CT and functional imaging of living body. *Medicina Philosophica*, 11, 620–629.
- Suganya, A., Sivitha, M. (2014). A new biometric using sclera vein recognition for human identification. In: *IEEE International Conference on Computational Intelligence and Computing Research*, pp. 1–4.
- Simonyan, K., Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. [arXiv:1409.1556](https://arxiv.org/abs/1409.1556).
- Vaid, S., Mishra, D. (2015). Comparative analysis of palm-vein recognition system using basic transforms. In: *IEEE International Advance Computing Conference (IACC)*, pp. 1105–1110.

- Xu, J. (2015). Palm vein identification based on partial least square. In: *2015 8th International Congress on Image and Signal Processing (CISP)*, pp. 670–674.
- Zhang, D., Guo, Z., Lu, G., Zhang, L., Zuo, W. (2010). An online system of multispectral palmprint verification. *IEEE Transactions on Instrumentation and Measurement*, 59(2), 480–490.
- Zheng, S.W., Han, J.G. (2018). Deep learning for contact-less palmprint recognition. *Microelectronics & Computer*, 35(4), 98–102.
- Zheng, S.W., Han, J.G., Wang, Y.F., Jing, J.D. (2017). Convolutional neural network for palmprint recognition. *Science Technology and Engineering*, 17(35), 1–7.

**Y.-Y. Fanjiang** received his PhD degrees in computer science and information engineering from National Central University, Taiwan, in 2004. Currently, he is a professor in the Department of Computer Science and Information Engineering, and the director of Information Technology Center, Fu Jen Catholic University. His research interests include service-oriented computing, software engineering, semantic web, and artificial intelligence in internet of things. He is a member of IEEE.

**C.-C. Lee** received the PhD degree in computer science from National Chung Hsing University (NCHU), Taiwan, in 2007. He is currently a distinguished professor with the Department of Library and Information Science at Fu Jen Catholic University. Dr. Lee is currently an editorial board member of *Mathematics*, *Electronics*, *Future Internet*, *International Journal of Network Security*, *Journal of Computer Science*, *Cryptography*, *International Journal of Internet Technology and Secured Transactions*, *Journal of Library and Information Studies*, *Journal of InfoLib and Archives*, and guest editor of *Sensors*. He also served as a reviewer of many SCI-index journals, other journals and conferences. His current research interests include data security, cryptography, network security, mobile communications and computing, wireless communications. Dr. Lee had published over 200 scientific articles on the above research fields in international journals and conferences. He is a member of IEEE, the Chinese Cryptology and Information Security Association (CCISA), the Library Association of The Republic of China, and the ROC Phi Tau Phi Scholastic Honor Society.

**Y.-T. Du** received the BS and MS in computer science and information engineering, Fu Jen Catholic University, New Taipei City 24205, Taiwan, R.O.C., in 2016 and 2018. His current research interests include information security, cryptography, and mobile communications.

**S.-J. Horng** (also known as Hsi-Chin Hung) received the BS degree in electronics engineering from National Taiwan Institute of Technology, Taipei, the MS degree in information engineering from National Central University, Taiwan, and the PhD degree in computer science from National Tsing Hua University, Taiwan, in 1980, 1984, and 1989, respectively. He was the Dean of the College of Electrical Engineering and Computer Science, National United University, Taiwan. Currently, he is a chair professor in the Department of Computer Science and Information Engineering, NTUST. He was a visiting professor at Tokyo Institute of Technology, in 2008; Georgia State University, in 2007; University of Dayton, Ohio, in 2000; National Mongolia University, in 2004; Southwest Jiaotong University, in 2004. He also worked as a PMTS at AT&T Bell Laboratories from 1990 to 1991. His research interests include deep learning, biometric recognition, multi-medium, image processing, and information security. He has published more than 200 research papers and received many awards; especially, the Distinguished Research Award between 2004 and 2006 from the National Science Council in Taiwan.