IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

IEEE CONSUMER TECHNOLOGY SOCIETY SECTION

Received October 5, 2021, accepted October 20, 2021, date of publication November 1, 2021, date of current version November 11, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3124631

Contactless Multispectral Palm-Vein Recognition With Lightweight Convolutional Neural Network

YUNG-YAO CHEN[1], (Member, IEEE), CHIH-HSIEN HSIA[2], (Member, IEEE), AND PING-HAN CHEN[3]

[1]Department of Electronic and Computer Engineering, National Taiwan University of Science and Technology, Taipei 106335, Taiwan
[2]Department of Computer Science and Information Engineering, National Ilan University, Yilan 26047, Taiwan
[3]Transportation and Surveillance Laboratory, Chunghwa Telecom Laboratories, Taoyuan 326402, Taiwan

Corresponding author: Chih-Hsien Hsia (chhsia625@gmail.com)

This work was supported by the Ministry of Science and Technology of Taiwan under Grant MOST 109-2221-E-197-029-.

ABSTRACT The development of information technology has made it possible to replace traditional keys and passwords with biometric recognition. Among the various human recognition technologies, contactless palm-vein authentication is becoming increasingly popular because it is hygienic and safe. In the field of deep learning (DL), system security and multispectral compatibility are crucial issues that require outright solutions. One of the most widely investigated DL algorithms is the convolutional neural network (CNN), which has been proven to have strong feature extraction capability. However, the training of CNN requires large samples and thus entails a heavy computational load, resulting in high hardware and software costs. Therefore, this paper proposes an adaptive Gabor filter with enhanced imaging features and triplet loss function that captures sufficient palm-vein data. A multispectral palm database from the CASIA public database was employed in this study to analyze the proposed system. The experimental results show that the proposed method has a low recognition error rate of 0.0556% and uses only a few network parameters in a multispectral environment.

INDEX TERMS Biometric, palm-vein recognition, convolutional neural networks, triplet loss function, handcrafted features.

I. INTRODUCTION

Information security has advanced considerably with the development of information technology. Methods and regulations have thus increasingly been designed to improve the security of personal information. Traditional information security systems include passwords, personal identification cards, and smart cards, which can be forgotten or lost as well as deciphered or stolen by individuals with an ulterior motive. These systems fail to meet modern society's demand for security, reliability, and convenience. Therefore, a new method that makes use of unique biological characteristics would be the best means of protecting personal information. Biological characteristics are suitable for use instead of passwords for identity verification for four main reasons. (1) Universality—everyone has their own unique biometric information such as voice print, facial features, fingerprints, finger vein, palm print, palm vein, and iris. (2) Permanence—biological characteristics do not change greatly over time, with any changes being slight and observed linearly. (3) Distinctiveness— each individual's biological characteristics are recognizable and different from those of others; not only are a person's right and left hands different, but the palm-vein information of twins is also not identical [1]. (4) Collectability—biological information can be obtained quickly and conveniently through special instrument.

Unique biological information on the human body is considerably safer, more reliable, and more convenient than a traditional identification card or password because the information can be used to verify a person's identity quickly and accurately, is less likely to be stolen, and cannot be forgotten. Biological characteristics can be broadly divided into two types: external characteristics, which are found on the surface of the human body and are directly visible such as the face [2], [3], iris [4], and palm print, and internal characteristics, which are inside the body and include the finger vein [5], [6], [22], palm vein [7], [12], hand's dorsal vein [8], and wrist vein [9].

As shown in Table 1, facial features and fingerprints, both external characteristics, can be easily stolen and used by unauthorized individuals. Moreover, fingerprints are not secure because they contain fewer characteristics than the

The associate editor coordinating the review of this manuscript and approving it for publication was Yen-Lin Chen.

TABLE 1. Comparison of commonly used biometric techniques.

Characteristics	Captured	Security	Privacy	Comfort	Hygiene
Face	RGB	Middle	Low	High	Low
Iris	NIR	High	High	Low	High
Finger-print	Pressoptic	Middle	Middle	Middle	Low
Palm-vein	NIR	High	High	High	High

face. Iris recognition becomes uncomfortable over time because near-infrared (NIR) light is needed to capture information. The COVID-19 pandemic has forced people to reduce physical contact to prevent virus transmission. Because fingerprint recognition requires people to touch the surface of a sensor and facial recognition requires people to remove the face mask, these two methods have become inconvenient and unhygienic. The palm vein, an internal characteristic, is more advantageous and relevant given the current global situation. Because the palm 'vein is inside the human body, it is stable and difficult to imitate; moreover, common stains and injuries do not affect recognition. The vein also does not change greatly over time and is highly unique even between twins. A palm-vein image shows a network of veins under the human skin; this network is unidentifiable under visible light and can only be captured from a living person by using NIR light, which is absorbed by hemoglobin in the blood. Reflection of light by muscles results in an NIR spectrum ranging from 750 to 1050 nm [10]. Identification using the palm vein has been proven to work more favorably than that using other biological characteristics because its information cannot be obtained maliciously through improper means [1], [11] and people do not need to touch the sensor during recognition. We refer to the related papers [1], [46]–[50] and summarize the advantages and disadvantages of the various identify methods in Table 1.

Because the palm vein can only be visualized by irradiating the hand with NIR light, the collection of samples for deep learning (DL) model training is difficult. Compared with traditional algorithms, DL finds the best features through superposition, which is more effective when the dataset is large. The computing speed of hardware has improved continually over the past years, and many embedded devices already contain a CUDA parallel computing module, enabling lightweight neural network computing. As the cost of equipment continues to decrease, the popularity of edge computing devices is expected to increase greatly. Traditional recognition systems have mainly been verified using public databases with a simple database background to enable the hand contour to be completely segmented. However, some interference may occur, and hand rotation, shifting, or zooming could influence recognition; evidently, these problems must be resolved (Fig. 1). Finally, different devices are expected to capture different spectra of vein images, and their system adaptability will differ. Thus, a model that is adaptive to different spectra must be developed.

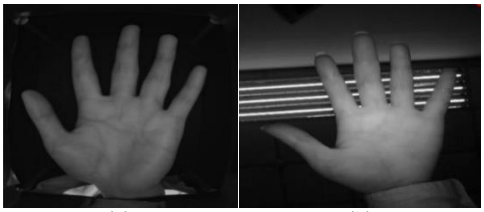

FIGURE 1. Palm image interference: (a) The public dataset without interference. (b) Actual image with interference.

II. LITERATURE REVIEW

In recent years, vein-based biometric methods, including palm-vein [1], [4], [7], finger-vein [5], [6], [12], dorsal hand-vein [8], wrist-vein [9], and forearm vein [14] have attracted a lot of attention. Among these, palm-vein is considered the most convenient as it is easy-to-capture and has a variety of features; thus, palm-vein recognition was chosen as the research object in this paper. In Fig. 2, the standard palm-vein recognition system acquires images mainly through NIR cameras; then, locations and features of the region of interest (ROI) are extracted, and the resulting images are compared to those found in the database. However, the method of feature extraction directly affects the system performance. The current palm-vein recognition system is divided into handcrafted features and CNN-based features. There are three methods to extract handcrafted features:

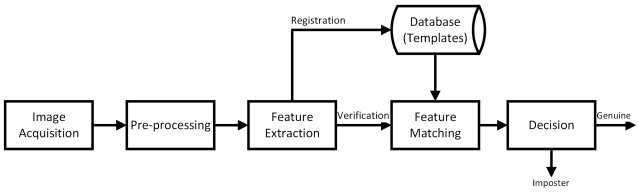

FIGURE 2. The flow of palm-vein recognition system.

1) Geometry-based method [7], [15], which mainly uses vein texture to show a continuous linear structure on the image, captures the information on each line, curve, and point that are close to the vein texture and shape. This method needs to be rectified through ROI first, and the features extracted are often fuzzy or sparse, which makes it difficult to deal with those rotating, scaling, or displaced samples.

2) Statistical-based method [16], uses statistical information to identify image characteristics, such as local binary histogram and image invariant moments. The latter can be classified as local statistics, which includes local derivative patterns (LDP) and local binary patterns (LBP) [17]; or it can also be classified as full-ranged statistics which includes image invariance [18]. These methods are also unstable when dealing with rotating, scaling, and displaced hands of users.

3) Local invariant-based method [19], which is inspired by the classical computer vision (CV) algorithm of scale-invariant feature transform (SIFT), can directly extract local invariant palm-vein characteristics. Although it can counteract the effect of feature displacement and hand

rotation, it is susceptible to changes caused by the light source in the equipment or environment, thus fewer consistent features can be extracted, and it becomes less useful as a recognition system.

With the decrease in hardware cost and the improvement of computing power, DL has been widely used in various fields. Previous studies proposed that the use of PVSNet [20] and handcrafted features was employed as the training target through a well-designed autoencoder CNN, then the encoded characteristics are sent into fully-connected layers and users' authentication are verified through a classification network. With handcrafted features, the ability of DL to find the best solution on its own is removed. Babalola *et al.* [13] proposed a combines binarized statistical image features descriptor method and CNN approaches using a decision-level fusion strategy for palm-vein recognition system that the experimental analysis for identification only. Das *et al.* [21] proposed a CNN framework for finger-vein to obtain images with the same quality and carried out extensive experiments to prove its effectiveness. However, this method directly output classification results, so, whenever a user registered, the network had to be retrained again, requiring a large number of parameters. Fang *et al.* [22] proposed a lightweight network collocation to analyze the entropy and multiple ROI of finger-vein images and they combined it with a two-stream network for classification verification. However, the training and testing data came from images acquired at the same time, which is unacceptable in practice because there is a chance that real data will be slightly changed in an unpredictable manner during each test. It is important to ensure that the training and testing data are collected during different time periods, as a way to demonstrate the robustness of the algorithm.

Most of the current palm-vein recognition systems still have some drawbacks. For instance, traditional algorithms have multiple adjustable parameters, but this takes a lot of time to achieve better accuracy based on the experience of user's [44]. Further, some algorithms use only a part of the data in the experiment, and the error rate tends to increase as the number of users increase [7]. Moreover, many problems can still be encountered by the current DL-based algorithms, such as CNN's large demand for data and the overfitting datasets. In addition, the public palm-vein database is smaller than the finger-vein database, and the number of photos that can be captured for each user is not enough. As a result, the development of DL lags behind the finger-vein database. Nevertheless, this paper focuses on the palm-vein, which has more characteristics and is safer than finger-vein recognition.

III. PROPOSED METHODOLOGY

The palm-vein recognition system shown in Fig. 3 was developed based on previous studies and CNN. It consists of two parts, training, and testing. First, the ROI method was used to locate the area to be identified, and the ROI image and Gabor filter were used for convolution calculation to extract the characteristics. In the deep network training phase, input fusion was performed between the raw ROI image and Gabor features. When the next batch of deep network training was performed, triplet loss and cross entropy will be calculated to optimize the deep network weight through backpropagation (BP).

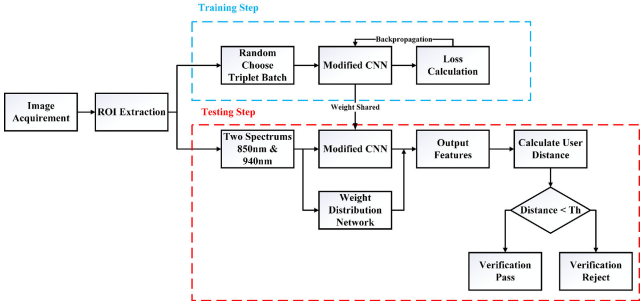

FIGURE 3. Proposed architecture.

A. SYSTEM FLOWCHART

This study proposed a weight selection mechanism to choose the best image in a dual spectrum. The selected sample was given a larger weight in the subsequent validation process where the Euclidean distance was used. When the distance between the testing and registered data was below the threshold, the user's information was considered correct; otherwise, the user's information was rendered invalid.

B. ROI POSITIONING

Except for the selected information (veins), all other information is deemed unimportant; thus, a good background subtraction algorithm is needed to effectively improve the system's performance. Biometric information can be obtained through either contact or contactless devices. Although the displacement of contact devices can be avoided to some extent by physical setting, contactless devices is viewed as the mainstream device in the future since users do not need to touch the devices, which makes it more hygienic [1]. However, contactless devices generate more invalid results because the images captured at different times may have significant changes according to palm proportions, hand displacement, and rotation. Therefore, several algorithms were used to ensure that the captured palm-vein information was not affected by these problems in this study.

In Fig. 4, considering the reality, semantic image segmentation (DeepLab V3+) was used for background subtraction before palm image acquisition. Next, the radial distance function (RDF) was utilized to locate the finger-tip and finger-valley on the basis of the distance between the reference point and the outline point. Then, the foreground information was selected from the captured hand size and rotation correction angle. The detailed steps are as follows:

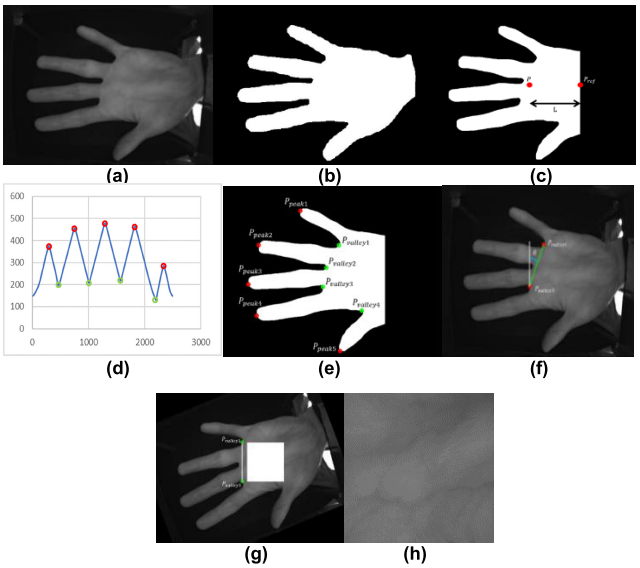

FIGURE 4. ROI position for palm-vein acquisition: (a) Original vein image. (b) Binarized image. (c) Defined reference and barycenter points. (d) RDF. (e) Defined finger-tips and valleys. (f) Calculated rotation angle θ. (g) Normalized image. (h) ROI image.

1) PALM CONTOURING

The traditional method extracts accurate ROI images, and then uses gray scale images to contour the palms. However, it cannot extract valid background and foreground information due to the changeable ambient light. In this research, DeepLab V3+ [23], [49] was used to segment the palm image in front of the background, and the trained network effectively distinguished the foreground palm from the background interference light source, as shown in Fig. 4(b). Based on the actual samples obtained, users often wear accessories such as watches and bracelets which are usually attached to the palm and cannot be removed during the contouring segmentation. Thus, the authors calculated the position of the barycenter in the image, estimated the horizontal distance L (The distance L is a variable parameter, and the L is composed of the center of gravity of the hand image and the position of about 0.8 palm width from left to right), and deleted the part that exceeds the length L, as shown in Fig. 4(c). This method effectively removed the wrist interference and improved the stability of the system.

2) LOCATING THE FINGER-TIPS AND VALLEYS

The position of the fingertips and valleys were acquired next. First, the barycenter (P_{ref}) of hand contour was set as the reference point, and the European distance between the hand contour and mass center was calculated to obtain the RDF(see Fig. 4(c)). Five areas of maximum value and four areas of minimum value were obtained, which corresponded to the position of the finger-tips (Ppeak1 to Ppeak5) and the valleys (Pvalley1 to Pvalley4) in the hand contour image respectively (see Fig. 4(d)).

3) NORMALIZATION AND ROI EXTRACTION

The authors used Pvalley1 and Pvalley3 as reference points, defined d to represent the shortest distance between two reference points and used θ to represent the angle between the straight-line d and the perpendicular line as shown in Fig. 4(e) and in Equations1 and 2 in which xPvalley1 and xPvalley4 represented corresponding coordinates. Further, the palm angle was corrected by using bilinear interpolation. As shown in Fig. 4(f), the square palm-vein image was captured after correcting the two reference points shown in Fig. 4(g).

$$d = \sqrt{(x_{P_{valley1}} - x_{P_{valley3}})^2 - (y_{P_{valley1}} - y_{P_{valley3}})^2} \quad (1)$$

$$\theta = \tan^{-1}(x_{P_{valley1}} - x_{P_{valley3}})/(y_{P_{valley1}} - y_{P_{valley3}}) \quad (2)$$

C. ADAPTIVE GABOR FILTERS

The Gabor filter is a common feature extraction method, which is capable of analyzing specific frequencies and has been applied to various fields because of its excellent performance in frequency analysis and feature extraction which can be shown in Eq. 3. Gabor filters are fine-tuned by using five different parameters which allow a high degree of freedom to adapt to a variety of samples; different arrangements need to be tested in order to determine the best parameters. This research employed a two-dimensional (2-D) Gabor filter with self-adapted parameters to improve the adaptability of the filter.

$$G_{\lambda,\theta,\psi,\sigma,\gamma}(x,y) = \exp\left(\frac{-x'^2 + \gamma^2 y'^2}{2\sigma^2}\right) \cdot \exp\left(i\left(2\pi\frac{x'}{\lambda} + \psi\right)\right) \quad (3)$$

The 2D Gabor filter was simplified by using only the parameters that have significant impacts on the retained vein sample. The improved equation is shown as follows:

$$G_{\sigma,\mu,\theta}(x,y) = g_\sigma(x,y) \cdot \exp\left(2\pi j\mu\left(x\cos\theta + y\sin\theta\right)\right) \quad (4)$$

In Eq. 4, j denotes the imaginary unit and $g_\sigma(x,y)$ is further expanded in Eq. 5.

$$g_\sigma(x,y) = 1/2\pi\sigma^2 \cdot \exp\left(-\left(x^2 + y^2\right)/2\sigma^2\right) \quad (5)$$

where σ denotes the standard deviation (SD), μ denotes the central frequency of the sample, and θ denotes the main angle.

The 2D Gabor filter in Eq. 4 was further divided into real- and imaginary part functions. The real part function was used for saddle detection of vein, while the imaginary part function was used for edge detection. The Euler's formula was used to decompose $G_{\sigma,\mu,\theta}$ into the real part $R_{\sigma,\mu,\theta}$ and the imaginary part $I_{\sigma,\mu,\theta}$.

$$R_{\sigma,\mu,\theta} = g_\sigma(x,y) \cdot \cos\left[2\pi\mu\left(x\cos\theta + y\sin\theta\right)\right] \quad (6)$$

$$I_{\sigma,\mu,\theta} = g_\sigma(x,y) \cdot \sin\left[2\pi\mu\left(x\cos\theta + y\sin\theta\right)\right] \quad (7)$$

The best performance can be obtained when the parameters σ, μ, and θ of the 2D Gabor filter are matched [24]. However,

applying fixed parameters to the palm-vein recognition is difficult because the results are not the same in every acquisition due to the equipment and the complicated human vein structure. To address this situation, this research utilized an adaptive parameter that divided the original palm-vein image into multiple sub regions, and then, the best parameter for each sub region was obtained.

1) FILTER STANDARD DEVIATION

In this study, the SD of Gaussian distribution was represented by σ. This parameter adjusted the width of the envelop in the filter. A large value of σ makes the filter more resilient to interference; on the contrary, a small value of σ acquires more texture information [25]. In this research, we referred to the [27] to find the filter parameters suitable for the palm-vein image in the experiment. The SD of the sub region was obtained via Eqs. 8 and 9, and the best sigma parameter was obtained in Eq. 10.

$$E\left(I\left(i,j\right)\right) = \sum_{i=1}^{m}\sum_{j=1}^{n} I\left(i,j\right)/m \times n \tag{8}$$

$$D\left(I\left(i,j\right)\right) = \sqrt{\frac{\sum_{i=1}^{m}\sum_{j=1}^{n}\left(I\left(i,j\right) - E\left(I\left(i,j\right)\right)\right)^2}{m \times n}} \tag{9}$$

where $E(I_{(i,j)})$ denotes the mean value of the sub-region, and $D(I_{(i,j)})$ denotes the SD of the subregion.

The four variables in Eq. 10 represent the stable zone, slow zone, moderate zone, and rapidly changing zone respectively, which better fit most of the samples.

$$\sigma = \begin{cases} 1, & \text{if D(I(i,j))} \le 1 \\ \sqrt{2}, & \text{if 1<D(I(i,j))} \le 1.4 \\ 2\sqrt{2}, & \text{if 1.4<D(I(i,j))} \le 2.8 \\ 4\sqrt{2}, & \text{if } D(I(i,j)) > 2.8 \end{cases} \tag{10}$$

2) FILTER CENTER FREQUENCY

The gray scale change in the main direction can be treated as a sinusoidal waveform in order to calculate the distance T between the lowest or highest values of the two regions, and the center frequency μ can be calculated by $\mu = 1/\mathrm{T}$. However, the contrast between veins and muscle tissue of the samples in this experiment was not as clear as that in natural image, so it was hard to distinguish the border areas [26]. In order to make a better system performance, other methods were needed to obtain a representative value μ. In this study, we proposed to divide μ into four different ranges, which corresponded to the four different SDs. After experimental observation, it was found that a lower value of T on the vein indicated that the image contained more complex texture features, and the value of SD was higher. Therefore, we can use SD to estimate the required μ.

$$\mu = \begin{cases} 0, & \text{if } \sigma = 1 \\ 0.12, & \text{if } \sigma = \sqrt{2} \\ 0.8, & \text{if } \sigma = 2\sqrt{2} \\ 2, & \text{if } \sigma = 4\sqrt{2} \end{cases} \tag{11}$$

3) FILTER MAIN ANGLE

After dividing the vein image into several sub regions, the texture features of each sub region were analyzed to determine the main direction of this segment. First, the gradient variations of the input sub regions in the vertical and horizontal directions were identified, and the maximum angle of each pixel was obtained in Eq. (12).

$$I_\theta(i,j) = \tan^{-1}\frac{dy(i,j)}{dx(i,j)} \tag{12}$$

In order to reduce the computational effort, the angles were divided into six main angles in determining the angle of each pixel. According to [27], a balance between performance and computation can be achieved by six main angles. A linear difference was used in the process of angular segmentation to maintain the best resolution for each angle. The max major angle in the subregion was obtained in Eq. 13.

$$\theta = \arg_\varphi \max\left(\sum\nolimits_{x_0=1}^{n}\sum\nolimits_{y_0=1}^{m} I_\theta(i,j)\right) \tag{13}$$

By using the angular order $\varphi \in \{0°, 30°, 60°, 90°, 120°, 150°\}$, the optimal filter parameters σ, μ, and θ were obtained for each sub region, which were further decomposed into real part $CR_{\sigma,\mu,\theta}(i,j)$ and imaginary part $CI_{\sigma,\mu,\theta}(i,j)$. Lastly, each sub region was convolved with its corresponding real- and imaginary part of Gabor filters. The image after the convolution was binarized with eigenvalue, as shown in Eq. 14.

$$F_R(i,j) = \begin{cases} 1, & \text{if } \mathrm{CR}_{\sigma,\mu,\theta}(i,j) \ge 0 \\ 0, & \text{if } \mathrm{CR}_{\sigma,\mu,\theta}(i,j) < 0 \end{cases}$$

$$F_I(i,j) = \begin{cases} 1, & \text{if } \mathrm{CI}_{\sigma,\mu,\theta}(i,j) \ge 0 \\ 0, & \text{if } \mathrm{CI}_{\sigma,\mu,\theta}(i,j) < 0 \end{cases} \tag{14}$$

D. MODIFIED CONVOLUTION NEURAL NETWORK

After the palm-vein images were pre-processed and normalized, the results were recorded into the improved CNN. The input data were transformed into an embedding feature via CNN for user validation in the next step. The contributions of this study in comparison with the recent DL-based palm-vein recognition are as follows:

(1) Image feature derecognition: Most of the existing palm-vein recognition systems combined with DL use multiple classifications [21], which can make a quick recognition. To increase the number of users, that the network must be optimized again, which is of little value for practical applications [37]. The proposed method avoided this drawback by transferring the input image into embedding features, when verifying, comparing the feature distance between the test data and the register data. However, it can delay the training cycle, that the threshold value can be set for retained according to practical applications. Therefore, it is necessary to avoid the situation where new users must be trained immediately.

(2) Solving the problem of insufficient data due to small sample sizes: Although DL can be very effective for natural

images recognition, it requires a large number of images to train a robust weight, which is very disadvantageous for small databases. The palm-vein database is so insufficient that it only allows three samples of each type for training. In view of this, this study enhanced the filter features, and added triplet loss and its special data structure to solve this problem.

(3) Network design for lightweight network in vein images: For the DL method, a large network framework requires a lot of computation time to train the network properly. In CNN framework, convolutional layer and fully-connected layer account for a large proportion of the parameters and computation time. For the biometric system commonly seen in low-end embedding devices that cannot handle huge computation volume, Sandler *et al.* [29] proposed a depthwise separable convolution layer to replace the general convolutional layer. The procedure of flattening feature images and of connecting them with neurons were performed by the global average pooling (GAP), as a way to lighten the weight of the network framework.

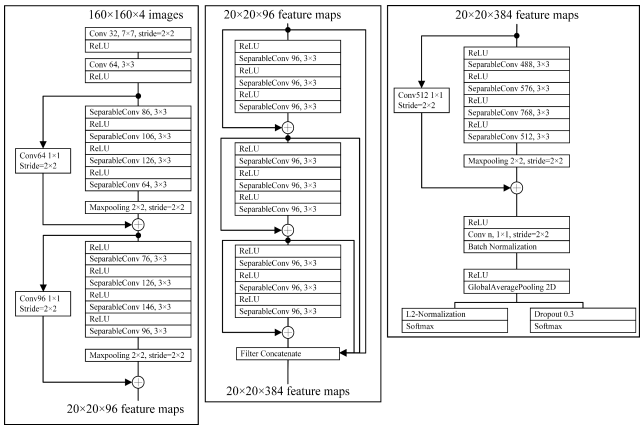

FIGURE 5. **The proposed modified CNN framework.**

1) DESIGN OF NETWORK FRAMEWORK

As shown in Fig. 5, the proposed modified CNN framework consisted of 27 convolutional layers, which were divided into eight modules. Except for the input and output modules, the rest of the modules used residuals to pass information between each other. In this work, a batch of normalization (BN) layer behind the convolutional layer and a dropout layer before the output were added to avoid the common problems of overfitting and internal covariate shift in DL. This yielded three main layers namely, the input layer, intermediate layer, and output layer.

a: INPUT LAYER

The scale conversion could limit the amount of texture information of the palm vein, so the size of the input layer was modified into 160 × 160× 4. This also allowed input fusion, improving the feature fusion effect. The first module of the input layer was a normal convolution layer with a large filter to capture more useful feature information which was then passed to the subsequent network. Finally, every module in the input layer reduced the dimension through a sub-sampling layer, converting the 160 × 160× 4 original image into a 20 × 20× 96 feature map.

b: INTERMEDIATE LAYER

After feature extraction and sub-sampling in the input layer, the resulting size of the input features was 20 × 20× 96. This study used smaller convolutional cores for multiple convolutions since multiple small volume layers are functionally equivalent to one large volume layer but with reduced parameter usage [30]. The intermediate module was subdivided into three small blocks; each block was connecting with a residual. The modules were connected by a dense block to preserve the best feature information and to emphasize feature reuse; in this way, the problem of gradient disappearance was resolved. After module connection, the size of the output feature map was 20 × 20× 384. after the module connection.

c: OUTPUT LAYER

A feature extraction module was first connected to integrate the features from the intermediate process and optimize the resulting output in this layer. This module enlarged the size of the feature map from 20 × 20× 384 to 20 × 20× 512, and then compressed the data via a 1 × 1 volume layer. Among them, the shield number n was the output neural cell and its amount was scaled according to the different sizes of the database, with the smallest amount being larger than the minimum number of users. Next, the GAP layer was used to compute the feature map from a size of 20 × 20× n to neurons of n. Then, L2-Normalization was performed on the n neurons. The BN layer was not used because it can smooth out the differences among the features, producing undesirable image features as the output.

2) TRIPLET LOSS FUNCTION

In the palm-vein recognition system, it is important that the network learning can quantify similarities among the input images on its own. This study employed a triplet loss function that could calculate the inter-sample similarity in real-time during network training and send feedbacks to the neural network for weight update. For deep networks, the main goal is to train models that can distinguish similarities between images. The most important characteristic of the triplet loss function is its ability to shorten the distance between the anchor and the positive sample, and to move the anchor away from the negative sample. In Fig. 6, an image of a user is selected as the anchor and other images of the same user are selected as the positive sample; then, an image from one of the remaining users is selected as the negative sample. The data are generated in this order until every sample in the database is input as an anchor. After passing through the layers, the palm-vein image will transform into embedding features. First, we defined two palm-vein images P, Q and we hire the squared Euclidean distance to confirm the similarity

between two embedded features.

$$\mathrm{D}\left(f\left(P\right),f\left(Q\right)\right)=\left\|f\left(P\right)-f\left(Q\right)\right\|_2^2 \tag{15}$$

where f denotes the function for mapping the image to the embedding layer, and $D\left(.,.\right)$ denotes the square of the Euclidean distance in space.

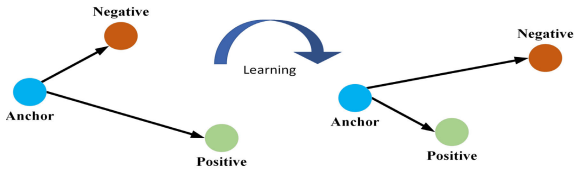

FIGURE 6. The learning process of the triplet loss function.

The closer the distance of $D(P, Q)$ is, the greater the similarity between the original image P and Q will be. Based on the data from the anchor, the positive and negative samples were defined as a set of triplets $b_i = (pi, pi^+, pi^-)$. Where pi, pi+ denote the same user but different pictures, pi- denotes the image of different users and randomly selected until each image has sorted as an anchor. After combining the triplet with the squared Euclidean distance and referring to other types of loss functions [31], the triplet loss function was calculated by Eq. 16 in which g represents the gap parameter that normalizes the two comparison images $(pi,\ pi^+)$ and (pi, pi^-).

$$loss(pi, pi^+, pi^-) = \max\left\{0, g + D\left(f(pi), f(pi^+)\right) - D\left(f(pi), f(pi^-)\right)\right\} \tag{16}$$

Based on the aforesaid loss function, DL was used to learn from the sample itself to determine the important features for similarity judgement, which is different from traditional methods that use handcrafted features. During the training of CNN, a batch of images extracted by the CNN were inputted, as shown in Fig. 7. At the same time, images pi, pi+, pi- were converted from 2D images into embedding features. Equation 16 was used to obtain the average of the loss function after calculation. The average loss value was used to prevent extreme samples from affecting network learning, and the loss value obtained from per batch was used to optimize the weight of the CNN with the method of BP.

E. DUAL IMAGE WEIGHT SELECTION MECHANISM

Previous studies have shown that oxygenated hemoglobin and deoxygenated hemoglobin in human blood absorb light with wavelengths in the range of 750nm to 1050nm, while water absorbs those at 965nm [6], [7], [10]. In actual vein data, not all vein images perform better under the light at 850nm which is an ideal wavelength for absorption rate. Moreover, users may be affected by several factors such as the equipment in use or their physical condition at the time of capturing images, which may result in inconsistent data under the same spectrum. To solve this problem, this paper employed a weight selection mechanism to analyze multispectral images

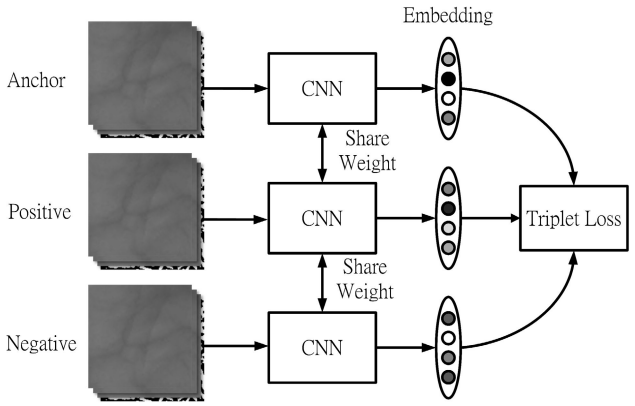

FIGURE 7. Training of triplet loss function network.

taken by users, and to automatically find the best spectrum for each one.

In order to define the distance relation between the images, we cross-check the training data to obtain the relative distance relation between the samples. Compare 850nm and 940nm in pairs and use Eq. 17 to build a train label. Finally, the weight of the neural network is selected through the same neural network structure as the input layer.

$$\begin{cases} \text{distance 850nm} = \text{distance 940nm}, & \alpha = 1, \beta = 1 \\ \text{distance 850nm} > \text{distance 940nm-gap}, & \alpha = 1, \beta = 0 \\ \text{distance 850nm} < \text{distance 940nm-gap}, & \alpha = 0.65, \\ & \beta = 0.35 \\ \text{distance 940nm} > \text{distance 850nm-gap}, & \alpha = 0, \beta = 1 \\ \text{distance 940nm} < \text{distance 850nm-gap}, & \alpha = 0.35, \\ & \beta = 0.65 \end{cases} \tag{17}$$

The parameter of α and β represent the label values assigned to the different wavelength.

IV. RESULTS AND ANALYSIS

A. ENVIRONMENTAL

To evaluate the technical objectivity presented in this work, the multispectral palm database CASIA [32] and PUT [38] were used. First, a CASIA database provided 7,200 images captured by a contactless device. Right- and left-hand images from 100 users in total were acquired. The samples were collected in two sessions, with an interval of one-month. Three images were collected each time, and six samples under different wavelengths (460, 630, 700, 850, 940nm, and white) were simultaneously captured. The authors analyzed the data of the vein captured under the spectrum of 850nm and 940nm. In order to do more comparison between the different samples, the left-hand image and right-hand image of the same person were considered as images from different users, expanding the subjects to 200 users with 12 images per user. The session 1 for testing data and session 2 for training data. Second, a PUT vein database provided 1,200 images

TABLE 2. System specification.

Items	Specifications
CPU	Intel® Core™ i7-6700@3.4GHz
Memory	DDR4 32GB
Graphics	Nvidia Geforce RTX 2080 Ti @1770MHz
Operating System (OS)	Microsoft Windows 10
Programming Language	C++、Python
Software	Tensorflow 1.10.0、Keras 2.2.4、OpenCV 3.4.3

captured by a contact device. Right- and left-hand images from 50 users in total were acquired. The samples were collected in three sessions, with an interval of one-week. Four images were collected each time, and 880nm samples under fixed wavelength were simultaneously captured. In order to do more comparison between the different samples, the left-hand image and right-hand image of the same person were considered as images from different users, expanding the subjects to 100 users with 12 images per user. The sessions 1 and 2 for training and session 3 for testing data.

In this work, Keras and OpenCV were used for algorithm development. The detailed specifications are shown in Table 2. The triple data structure for training was constructed, and then the anchor was taken up by all the samples. A single anchor was paired with 4 positive samples and 18 negative samples, thus a total of 86,400 triplet loss datasets were obtained. These were then fed into the network for an epoch before a new round of triplet loss dataset construction. The optimizer used Adam with its learning rate set at 0.001 and its batch size at six triplet loss datasets.

The security assessment of the biometric system was verified by the equal error rate (EER). The system encounters two possible error patterns: the false reject rate (FRR), which means that the tester should have passed the verification; and the false acceptance rate (FAR), which means that the fake tester should have not passed the test. Adjust the threshold of the system, so that when the security is gradually adjusted from high to low level FAR will gradually approach 1 from 0, and FRR will gradually approaches 0 from 1. In this process, the two data will eventually converge on the same point, which is called EER. This is the moment when optimal balance of the system performance shows. Therefore, the verification system used the EER as the system safety indicator. The definitions of FAR and FRR are shown in Eq. 18. In addition to EER, the receiver operating characteristic curve (ROC) was also used. The FAR was set as the x-axis, and the FRR was reversed to obtain the genuine acceptance rate (GAR) which was set as the y-axis. At the same time, a straight line was set with a slope of 1. When the curve drawn by FAR and FRR intersected on a line segment with a slope of 1, their intersecting point was the EER point of the algorithm.

$$\begin{aligned} FAR &= FP/(FP+TN) \\ FRR &= FN/(TP+FN) \\ GAR &= 1-FRR \end{aligned} \tag{18}$$

B. GABOR FILTERS

In this work, the features captured by an adaptable Gabor filter were recorded and merged with the original vein images. The features, regarded as one of the channels in the image, were inputted into the network for training and verification. Gabor features were added because DL requires a large number of samples to train a robust weight after many repetitions; however, for the CASIA vein database, only 1200 (200 persons × 3 sheets × 2 frequency spectra) images can be used for training under a fair distribution of data. Since the amount of data was inadequate compared to other fields, the Gabor features were utilized to enhance the original image features without disturbing the balance of the data set. ROC curves were then drawn for comparison. In Fig. 9, the red and blue lines represent the recognition rate under 850nm and 940nm light with Gabor features added, while the brown and yellow lines represent the recognition rate under 850nm and 940nm light without Gabor features. It can be observed that the recognition rate under 940nm light was higher than that under 850nm light, and the recognition rate of the two frequency bands was significantly improved after the Gabor feature was added.

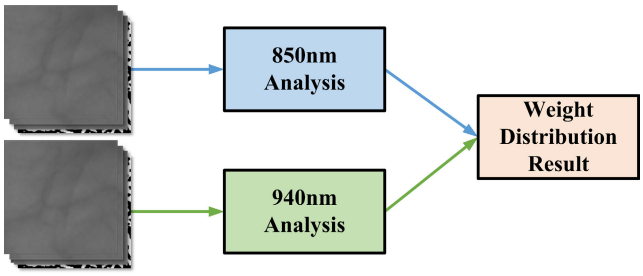

FIGURE 8. Weighting of weight selection mechanism.

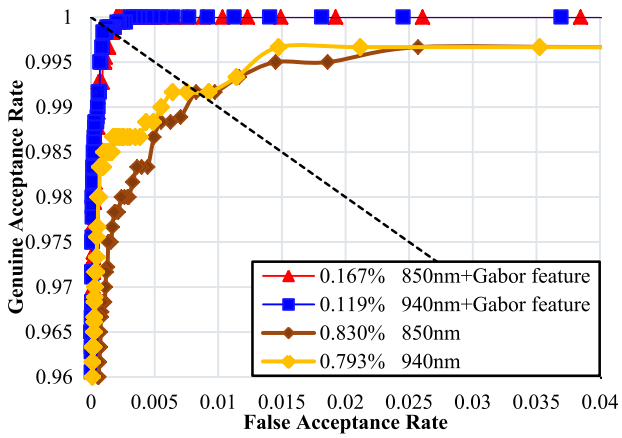

FIGURE 9. ROC curve with Gabor features added.

C. WEIGHT SELECTION MECHANISM

In the CASIA database used in this study, users' vein information captured only under wavelengths of 850nm and 940nm was provided. In order to verify that the proposed

weight selection mechanism can effectively distinguish the most suitable wavelength for users, this study used a score fusion mechanism to analyze the images of the 850nm and 940nm frequency bands, and to give higher weight to the better-quality band. It can be seen in Fig. 10 that the weight selection mechanism effectively integrated the information of the two frequency bands and provided a more robust verification system.

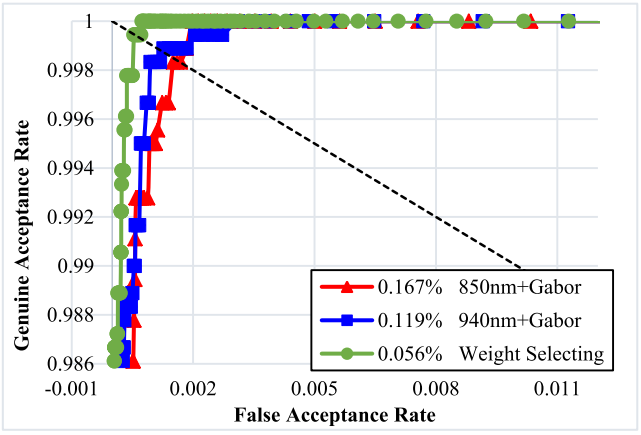

FIGURE 10. ROC curve with weight selection mechanism.

D. COMPARISON WITH RELATED WORKS

This paper compared the proposed techniques with those in recent works under the same test conditions as shown in Table 3. It is divided into two approaches: traditional CV and DL-based methods. The EER of the CV-based algorithm was lower than that of the DL-based algorithm because CV can be fine-tuned with fixed parameters to reduce the EER under specific equipment and spectrum. However, it is more difficult to achieve the same performance on different devices or in different environments. Some of the CV-based algorithms used incomplete databases for comparison, and the error rate may be higher than that of the data tested in this paper because of the increasing number of users. Furthermore, few studies have used DL in the study of palm-vein recognition; thus, studies that used DL in finger-vein recognition with the same sample characteristics were included for comparison.

This work also developed a lightweight DL network framework for palm-vein recognition. Compared with other frameworks used in the vein field, our design has fewer parameters than others, including MobileNetV2 [29], which is mainly used in embedding devices. This also allowed the work to conduct future real-world applications without equipment limitations encountered in handheld devices and edge computing platforms, as shown in Fig. 11.

Table 4 shows the results obtained using the PUT Vein Dataset [9], which contains 1200 images of size 1280 × 960 pixels with a wavelength of 880 nm. The FYO dataset contains only one image for training and one for testing, but the present work proposes a triplet loss method in which at least two images are required for effective training; thus, we use the PUT dataset. The experimental results reveal that the proposed method has advantage compared with the other methods. Compared with the three former methods [40]–[45], the proposed method and wave atom transform [43] perform more strongly. To further examine the practical performance between the method presented in [43] and the proposed method, semantic image segmentation is used to remove the complex background, as shown in Fig. 1(b). In addition, in the PUT Vein Dataset, we adjusted the input image size to 320 × 320 × 4, which considers the GPU's ability to training and the impact of distortion caused by image resizing is reduced. This study adds two convolution layers to the CNN,

TABLE 3. Comparison between EER validation rates of methods.

Methods	Years	Users	EER(%)
CV(Handcrafted) Method -Based			
NMRT [7]	2011	200	0.5100
Hessian phase [7]	2011	200	0.4300
Mutual foreground LBP [1]	2014	100	0.2670
RootSIFT [19]	2014	100	0.9960
Feature-level fusion [33]	2015	100	0.1600
Adaptive 2D Gabor filter [27]	2017	200	0.1200
CS-LBP [34]	2018	200	2.6100
HDPLS [12]	2019	100	0.0292
TERRM [4]	2020	200	4.2800
DL Method-Based			
ResNet101 [36]	2018	200	6.1620
PVSNet [20]	2019	200	3.7100
MobileNetV2[29]	2019	200	5.9970
U-Net[50]	2020	100	0.47
DenseNet161 [35]	2020	200	1.0540
CNN + Bayesian[51]	2021	100	0.0683
GPWLD[52]	2021	100	0.24
Proposed-850 *nm*			0.1670
Proposed-940 *nm*	2020	200	0.1190
Proposed-Weight Selecting			0.0556

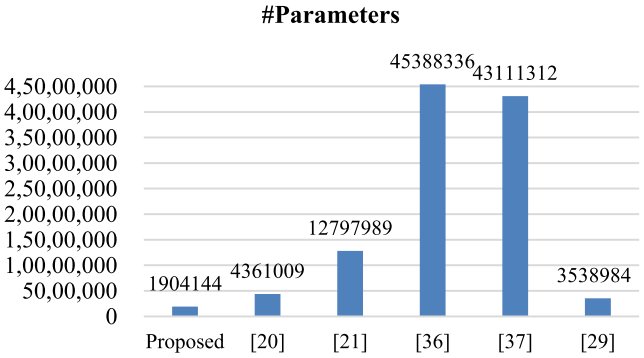

FIGURE 11. Comparison between the number of network parameters.

TABLE 4. Comparison the EER of methods in PUT vein dataset.

Methods	Years	Users	EER(%)
2D Correlation [38]	2011	100	1.10
Improved SIFT [39]	2017	100	1.08
Gradient-based 2D Gabor [40]	2019	100	1.60
Wave Atom Transform [41]	2020	100	0
MWCDE [42]	2020	100	0.62
MSDRA [43]	2021	100	0.791
Proposed-880 *nm*	2020	100	0

with depths 32 and 64 and both with the ReLU activation function. The first convolution layer has a 2×2 stride.

V. CONCLUSION

The Vein-based identity recognition has been actively developed in recent years and has been proven to be an effective and reliable recognition method. However, many problems still need to be overcome in the DL-based algorithm. In view of this, this study proposed a complete lightweight system that effectively solved most of the problems encountered in previously developed systems. The ROI positioning after palm image input was able to resist a certain degree of rotation and displacement, which reduced the system errors caused by the user while maintaining its hygienic contactless acquisition. In addition, a lightweight network for palm-vein was employed with fewer parameters than those used in recent studies. In the network training, a triplet loss function and Gabor features were utilized to fuse the input layer and to allow the CNN network to learn to distinguish the similar features between images. It is able to effectively train the CNN network even when the data in the public database is insufficient. The weight selection mechanism also selected a better sample automatically to improve the system's adaptability for dual spectrum. The results show that the proposed network framework required fewer parameters and had a better error rate of 0.0556%.

REFERENCES

[1] 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.

[2] C.-H. Hsia, J.-S. Chiang, and C.-Y. Lin, "A face detection method for illumination variant condition," *Scientia Iranica*, vol. 22, no. 6, pp. 2081–2091, 2015.

[3] D. Menotti, G. Chiachia, A. Pinto, W. R. Schwartz, H. Pedrini, A. X. Falcao, and A. Rocha, "Deep representations for iris, face, and fingerprint spoofing detection," *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 4, pp. 864–879, Apr. 2015.

[4] S. Cho, B.-S. Oh, K.-A. Toh, and Z. Lin, "Extraction and cross-matching of palm-vein and palmprint from the RGB and the NIR spectrums for identity verification," *IEEE Access*, vol. 8, pp. 4005–4021, 2020.

[5] C.-H. Hsia, "New verification strategy for finger-vein recognition system," *IEEE Sensors J.*, vol. 18, no. 2, pp. 790–797, Jan. 2018.

[6] C.-H. Hsia, "Improved finger-vein pattern method using wavelet-based for real-time personal identification system," *J. Imag. Sci. Technol.*, vol. 26, no. 3, p. 30402, 2018.

[7] Y. Zhou and A. Kumar, "Human identification using palm-vein images," *IEEE Trans. Inf. Forensics Security*, vol. 6, no. 4, pp. 1259–1274, Dec. 2011.

[8] F. Liu, S. Jiang, B. Kang, and T. Hou, "A recognition system for partially occluded dorsal hand vein using improved biometric graph matching," *IEEE Access*, vol. 8, pp. 74525–74534, 2020.

[9] O. Toygar, F. O. Babalola, and Y. Bitirim, "FYO: A novel multimodal vein database with palmar, dorsal and wrist biometrics," *IEEE Access*, vol. 8, pp. 82461–82470, 2020.

[10] M. Zhou, L. Lin, M. Wang, X. Li, and G. Li, "Influence of water on noninvasive hemoglobin measurement by dynamic spectrum," *Anal. Methods*, vol. 5, no. 18, pp. 4660–4665, 2013.

[11] J.-D. Wu and S.-H. Ye, "Driver identification using finger-vein patterns with Radon transform and neural network," *Expert Syst. Appl.*, vol. 36, no. 3, pp. 5793–5799, Apr. 2009.

[12] W. Wu, S. J. Elliott, S. Lin, and W. Yuan, "Low-cost biometric recognition system based on NIR palm vein image," *IET Biometrics*, vol. 8, no. 3, pp. 206–214, May 2019.

[13] F. O. Babalola, Y. Bitirim, and Ö. Toygar, "Palm vein recognition through fusion of texture-based and CNN-based methods," *Signal, Image Video Process.*, vol. 15, pp. 1–8, Apr. 2020.

[14] H. Zhang, C. Tang, A. W.-K. Kong, and N. Craft, "Matching vein patterns from color images for forensic investigation," in *Proc. IEEE 5th Int. Conf. Biometrics, Theory, Appl. Syst. (BTAS)*, Sep. 2012, pp. 77–84.

[15] N. Miura, A. Nagasaka, and T. Miyatake, "Extraction of finger-vein patterns using maximum curvature points in image profiles," *IEICE Trans. Inf. Syst.*, vol. 90, no. 8, pp. 1185–1194, Aug. 2007.

[16] L. Mirmohamadsadeghi and A. Drygajlo, "Palm vein recognition with local binary patterns and local derivative patterns," in *Proc. Int. Joint Conf. Biometrics (IJCB)*, Washington, DC, USA, Oct. 2011, pp. 1–6.

[17] 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.

[18] X. Li, S. Guo, F. Gao, and Y. Li, "Vein pattern recognitions by moment invariants," in *Proc. 1st Int. Conf. Bioinf. Biomed. Eng.*, Wuhan, China, Jul. 2007, pp. 612–615.

[19] 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, pp. 1–12, 2014.

[20] D. Thapar, G. Jaswal, A. Nigam, and V. Kanhangad, "PVSNet: Palm vein authentication Siamese network trained using triplet loss and adaptive hard mining by learning enforced domain specific features," in *Proc. IEEE 5th Int. Conf. Identity, Secur., Behav. Anal. (ISBA)*, Hyderabad, India, Jan. 2019, pp. 1–8.

[21] R. Das, E. Piciucco, E. Maiorana, and P. Campisi, "Convolutional neural network for finger-vein-based biometric identification," *IEEE Trans. Inf. Forensics Security*, vol. 14, no. 2, pp. 360–373, 2019.

[22] 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.

[23] C.-H. Hsia and C.-F. Lai, "Embedded vein recognition system with wavelet domain," *Sensors Mater.*, vol. 32, no. 10, pp. 3221–3234, 2020.

[24] R. Mehrotra, K. R. Namuduri, and N. Ranganathan, "Gabor filter-based edge detection," *Pattern Recognit.*, vol. 25, no. 12, pp. 1479–1494, Dec. 1992.

[25] J.-C. Lee, C.-H. Lee, C.-B. Hsu, P.-Y. Kuei, and K.-C. Chang, "Dorsal hand vein recognition based on 2D Gabor filters," *Imag. Sci. J.*, vol. 62, no. 3, pp. 127–138, 2014.

[26] W.-Y. Han and J.-C. Lee, "Palm vein recognition using adaptive Gabor filter," *Expert Syst. Appl.*, vol. 39, no. 18, pp. 13225–13234, Dec. 2012.

[27] X. Ma, X. Jing, H. Huang, Y. Cui, and J. Mu, "Palm vein recognition scheme based on an adaptive Gabor filter," *IET Biometrics*, vol. 6, no. 5, pp. 325–333, 2017.

[28] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "ImageNet large scale visual recognition challenge," 2014, *arXiv:1409.0575*.

[29] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted residuals and linear bottlenecks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 4510–4520.

[30] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.

[31] S. Tang, S. Zhou, W. Kang, Q. Wu, and F. Deng, "Finger vein verification using a Siamese CNN," *IET Biometrics*, vol. 8, no. 5, pp. 306–315, Sep. 2019.

[32] *CASIA Palm Image Database*. Accessed: Jul. 15, 2018. [Online]. Available: https://biometrics.idealtest.org/dbDetailForUser.do?id=5

[33] 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.

[34] S. Bhilare, G. Jaswal, V. Kanhangad, and A. Nigam, "Single-sensor hand-vein multimodal biometric recognition using multiscale deep pyramidal approach," *Mach. Vis. Appl.*, vol. 29, no. 8, pp. 1269–1286, Nov. 2018.

[35] K. J. Noh, J. Choi, J. S. Hong, and K. R. Park, "Finger-vein recognition based on densely connected convolutional network using score-level fusion with shape and texture images," *IEEE Access*, vol. 8, pp. 96748–96766, 2020.

[36] W. Kim, J. M. Song, and K. R. Park, "Multimodal biometric recognition based on convolutional neural network by the fusion of finger-vein and finger shape using near-infrared (NIR) camera sensor," *Sensors*, vol. 18, no. 7, p. 2296, 2018.

[37] E. Jalilian and A. Uhl, "Finger-vein recognition using deep fully convolutional neural semantic segmentation networks: The impact of training data," in *Proc. IEEE Int. Workshop Inf. Forensics Secur.*, Dec. 2018, pp. 1–8.
[38] R. Kabacinski and K. Kowalski, "Vein pattern database and benchmark results," *Electron. Lett.*, vol. 47, no. 20, pp. 1127–1128, 2011.
[39] G. Wang and J. Wang, "SIFT based vein recognition models: Analysis and improvement," *Comput. Math. Methods Med.*, vol. 2017, pp. 1–14, Jun. 2017.
[40] S. Bharathi and R. Sudhakar, "Biometric recognition using finger and palm vein images," *Soft Comput.*, vol. 23, no. 6, pp. 1843–1855, Mar. 2019.
[41] F. Ahmad, L.-M. Cheng, and A. Khan, "Lightweight and privacy-preserving template generation for palm-vein-based human recognition," *IEEE Trans. Inf. Forensics Security*, vol. 15, pp. 184–194, 2020.
[42] G. Wang, C. Sun, and A. Sowmya, "Multi-weighted co-occurrence descriptor encoding for vein recognition," *IEEE Trans. Inf. Forensics Security*, vol. 15, pp. 375–390, 2020.
[43] Z. Pan, J. Wang, G. Wang, and J. Zhu, "Multi-scale deep representation aggregation for vein recognition," *IEEE Trans. Inf. Forensics Security*, vol. 16, pp. 1–15, 2021.
[44] H. Wan, L. Chen, H. Song, and J. Yang, "Dorsal hand vein recognition based on convolutional neural networks," in *Proc. IEEE Int. Conf. Bioinf. Biomed. (BIBM)*, Nov. 2017, pp. 1215–1221.
[45] C.-H. Hsia, J.-M. Guo, and C.-S. Wu, "Finger-vein recognition based on parametric-oriented corrections," *Multimedia Tools Appl.*, vol. 76, no. 23, pp. 25179–25196, Dec. 2017.
[46] T. E. Boult, "PICO: Privacy through invertible cryptographic obscuration," in *Proc. Comput. Vis. Interact. Intell. Environ. (CVIIE)*, Nov. 2005, pp. 27–38.
[47] V. G. Moshnyaga, J. Shioyama, and K. Hashimoto, "A camera-based approach to prevent fingerprint hacking," in *Proc. IEEE Int. Workshop Signal Process. Syst.*, Oct. 2018, pp. 235–240.
[48] K. Shaheed, H. Liu, G. Yang, I. Qureshi, J. Gou, and Y. Yin, "A systematic review of finger vein recognition techniques," *Information*, vol. 9, no. 9, p. 213, Aug. 2018.
[49] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proc. Eur. Conf. Comput. Vis.*, Munich, Germany, 2018, pp. 833–851.
[50] P. Wang and H. Qin, "Palm-vein verification based on U-Net," *IOP Conf. Mater. Sci. Eng.*, vol. 806, Apr. 2020, Art. no. 012043.
[51] M. I. Obayya, M. El-Ghandour, and F. Alrowais, "Contactless palm vein authentication using deep learning with Bayesian optimization," *IEEE Access*, vol. 9, pp. 1940–1957, 2021.
[52] M. El-Ghandour, M. I. Obayya, B. Yousef, and N. F. Areed, "Palmvein recognition using block-based WLD histogram of Gabor feature maps and deep neural network with Bayesian optimization," *IEEE Access*, vol. 9, pp. 97337–97353, 2021.

YUNG-YAO CHEN (Member, IEEE) received the B.S. and M.S. degrees in electrical and control engineering from the National Chiao Tung University, Hsinchu, Taiwan, in 2004 and 2006, respectively, and the Ph.D. degree in electrical engineering from Purdue University, USA, in 2013. Before being a Faculty, he has worked at HP Labs-Printing and Content Delivery Lab about one year. He is currently an Associate Professor with the Department of Electronic and Computer Engineering, National Taiwan University of Science and Technology, Taiwan. His current research interests include vision-based automation, automated/wisdom factory, self-driving car, and human–computer interaction. He is a member of Golden Key International Honor Society and Phi Tau Phi. He was a recipient of the First Prize Paper Award of the 2015 International Conference on Advanced Robotics and Intelligent Systems, the Rotary Foundational Scholarship, and the Ta-Yu Wu Memorial Award from Taiwan's Ministry of Science and Technology.

CHIH-HSIEN HSIA (Member, IEEE) was born in Taipei, Taiwan, in 1979. He received the Ph.D. degree from Tamkang University, New Taipei, Taiwan, in 2010.

In 2007, he was a Visiting Scholar with Iowa State University, Ames, IA, USA. From 2010 to 2013, he was a Postdoctoral Research Fellow with the Department of Electrical Engineering, National Taiwan University of Science and Technology, Taipei. From 2013 to 2015, he was an Assistant Professor with the Department of Electrical Engineering, Chinese Culture University, Taiwan. He was an Associate Professor with Chinese Culture University and the National Ilan University, Taiwan, from 2015 to 2017. From 2019 to 2020, he was the Director of the Research Planning Division, Research and Development, National Ilan University. He is currently a Professor and a Chairman with the Department of Computer Science and Information Engineering, National Ilan University, where he is the Director of the Multimedia and Intelligent Technical Laboratory. His research interests include DSP IC design, AI in computer vision, and cognitive learning. He received the Outstanding Young Scholar Award of the Taiwan Association of Systems Science and Engineering, in 2020, and the Outstanding Young Scholar Award of the Computer Society of the Republic of China, in 2018. He is the Chapter Chair of IEEE Young Professionals Group, Taipei Section, and the Director of the IET Taipei Local Network. He has served as an Associate Editor for the *Journal of Imaging Science and Technology*, the *Journal of Imaging*, and the *Journal of Computers*.

PING-HAN CHEN was born in Taoyuan, Taiwan, in 1996. He received the B.S. degree from the Department of Electrical Engineering, Chinese Culture University, Taipei, Taiwan, in 2018, and the M.S. degree from the Institute of Automation Technology, National Taipei University of Technology, Taiwan, in 2021. He is currently an Associate Researcher at Chunghwa Telecom Laboratories. His research interests include biometrics, autonomous cars, computer vision, and deep learning.