IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received May 4, 2021, accepted May 27, 2021, date of publication June 7, 2021, date of current version June 15, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3086811

Outside Box and Contactless Palm Vein Recognition Based on a Wavelet Denoising ResNet

WEI WU[1,2], QIANG WANG[1], SIQUAN YU[2,3], QIONG LUO[2,3], SEN LIN[4], ZHI HAN[2], (Member, IEEE), AND YANDONG TANG[2], (Member, IEEE)

[1]Information Engineering Department, Shenyang University, Shenyang 110044, China
[2]State Key Laboratory of Robotics, Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang 110169, China
[3]Institutes for Robotics and Intelligent Manufacturing, Chinese Academy of Sciences, Shenyang 110169, China
[4]School of Automation and Electrical Engineering, Shenyang Ligong University, Shenyang 110159, China

Corresponding author: Wei Wu (wuwei@sia.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61991413, Grant 61821005, Grant 61703285, and Grant 62073205; in part by the China Postdoctoral Science Foundation under Grant 2019M651142; and in part by the Natural Science Foundation of Liaoning Province under Grant 2019-DZ-0554 and Grant 2019-MS-237.

ABSTRACT Palm vein recognition is a high-security biometric. Outside the NIR capture box and contactless palm vein recognition are more popular but challenging. The users feel comfortable outside the NIR capture box but face more optical blurring brought by visible light. Contactless capture gestures solve the hygienic problem but face the image rotation, position translation, and scale transformation which makes classification difficult especially in large-scale databases. To address these problems, we develop a wavelet denoising ResNet, which consists of two models: the wavelet denoising (WD) model and the squeeze-and-excitation ResNet18 (SER) model. The WD model focuses on removing noise from skin scattering and optical blurring from palm vein images. The WD model enhances the low-frequency feature into a deep learning feature by residual learning technology. This strategy increases the weight of an effective handcrafted feature in the deep learning network. The SER model overcomes rotation, position translation, and scale transformation by selectively emphasizing classification features and weakening less useful features. To train and verify the network, an inside box palm vein image database and an outside box palm vein image database are set up. The Tongji contactless palm vein image database was also employed in the experiments. The validity and superiority of our network are verified in a series of experiments.

INDEX TERMS Deep learning, biometrics, palm vein recognition, Resnet, wavelet decomposition, denoise.

I. INTRODUCTION

Palm vein recognition identifies a person by the structure of vessels underneath palm skin [1]. The vessels absorb more NIR (Near-Infrared Ray) light than the surrounding tissues [2]. The vessels reflect less NIR light to the camera. The shadow of vessels represents the structure of people's palm vein images. Palm vein recognition is high security. First, the vascular structures disappear without blood flowing. Second, as interior biological information and captured under NIR, vein patterns are difficult copy. Third, palm veins are always concealed naturally, because people usually half clench their fists.

Outside box and contactless palm vein recognition in large-scale databases is the direction of this burgeoning research field [3]. Outside the NIR capture box makes the user feel comfortable but is affected by unstable and uneven visible light. Visible light and NIR light scatter strongly inside palm tissues [4], and a majority of palm vein images inevitably contain noise and optical blurring, especially with a non-NIR CCD camera. The vessels are obscure in these images. It is difficult to simultaneously remove optical blurring and preserve effective information. Most of the research captures palm vein images in NIR capture boxes or lightproof environments. This capture acquisition prevents the progress

The associate editor coordinating the review of this manuscript and approving it for publication was Varuna De Silva.

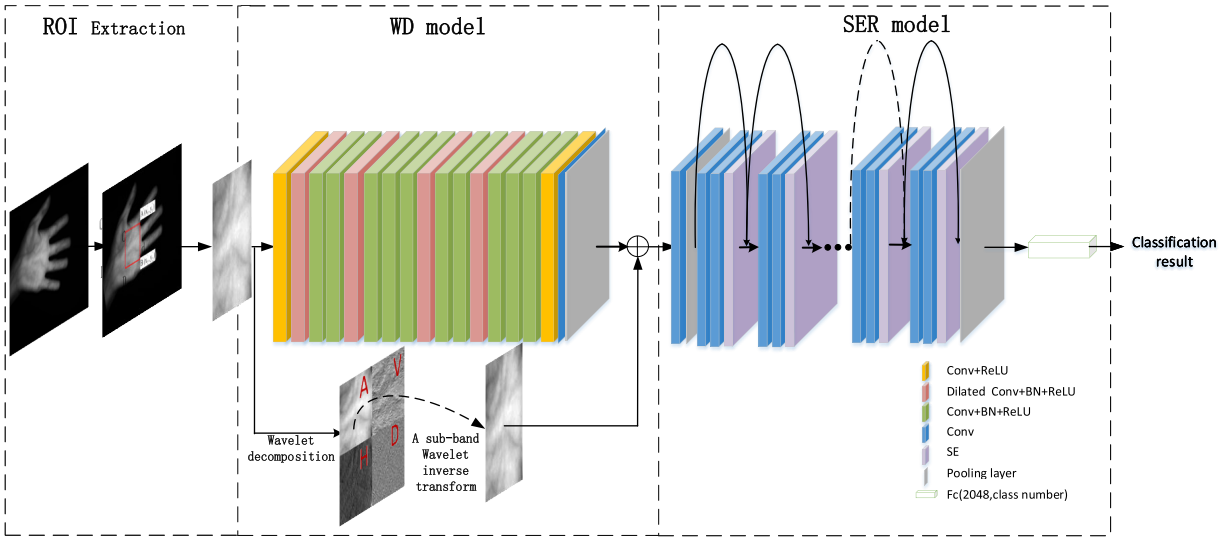


FIGURE 1. Schematic of the proposed method.

of palm vein recognition in practical scenarios. Contactless image solve the hygienic problems but face image rotation, position translation, and scale transformation, which makes classification difficult, especially in large-scale databases. The similarity in different classes and diversity in the same class in large databases is the second difficulty. Motivated by its practical value and the lack of good solutions, this paper develops a deep neural network: wavelet denoising ResNet. The architecture of the proposed network is shown in Fig. 1, which is composed of ROI (region of interest) extraction, the WD model and the SER model.

The WD model addresses the first problem. The noise and optical blurring are mainly concentrated in the high-frequency part. At the same time, some useful information is preserved in the high-frequency part. The main texture information of the palm vein image is concentrated in the low-frequency part. We want to keep other information without noise at high frequency. Traditional methods such as the 2-dimensional discrete wavelet transform (2D-DWT) can extract low-frequency information without other useful information. The WD model uses a deep convolutional neural network method and introduces residual learning into the network to enhance the effective feature representation. Residual learning imports the A sub-band wavelet inverse transform image as a short connection into the denoising deep learning network. This design filters out some ambiguous representations for recognition. The result of the palm vein image ROI after the WD model is shown in Fig. 2.

SER model addresses the second problem. It overcomes rotation, position translation, and scale transformation by selectively emphasizing information features and weakening less useful features. The ResNet component overcomes the vanishing or exploding gradients in a large-scale database. The squeeze-and-excitation component enhances the channel

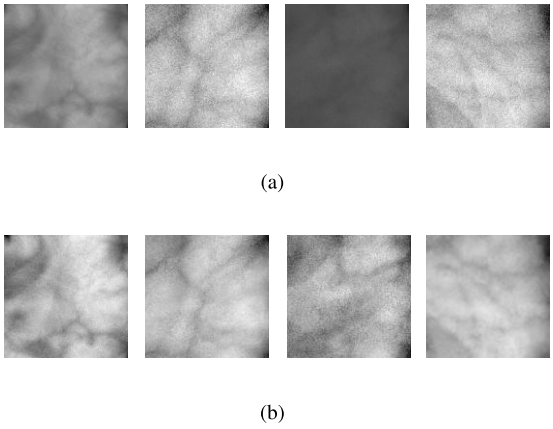


FIGURE 2. Examples of palm vein ROI images processed by our method. (a) Original ROI images. (b) Corresponding results by our method.

eigenvalues that are useful for the recognition task and suppresses the useless channel eigenvalues. It can distinguish the diversity in the different classes.

The main contributions can be included in four aspects:

First, we propose an end-to-end deep network architecture for outside box and contactless palm vein recognition. This network architecture integrates feature extraction and classification into one deep learning network. It improves the recognition performance of outside box and contactless palm vein recognition.

Second, the WD model fuses a sub-band wavelet inverse transform image into the deep learning network by residual learning technology. Through this strategy, we improve the weight of meaningful features and denoise the palm vein image simultaneously. It enhances the effective feature representation for recognition.

Thirdly, the SER model emphasizes the features which are helpful for recognition and settle the problem of degradation,

vanishing, or exploding gradients in a multilayer network. The model provides a solution to the high inter-class similarity and intra-class variation faced by contactless palm vein recognition in large databases.

This paper is summarized as follows. Section II describes related works. Section III explains ROI extraction and the network structure, including the WD model and SER model. Abundant comparison experiments on three databases are reported in Section IV. Section V concludes.

II. RELATED WORK

Palm vein recognition has become popular due to the inherent advantages of palm veins. It provides a variety of palm vein feature description methods. They can be classified into two groups: handcrafted methods and deep learning methods.

A. HANDCRAFTED METHODS

Handcrafted methods can be classified into four categories: geometry-based methods, texture-based methods, local invariant-based methods, and subspace-based methods.

Geometry-based methods [5], [6] use the line feature or point feature of the vasculature to describe the palm vein. These approaches use vein information directly and response structure information. However, when skin scattering and optical blurring obscuring, some of the vessels or parts of the vessels cannot be imaged. In addition, these approaches are not invariant to rotation, scaling, or translation of the palm vein images.

Texture-based methods [7]–[15] use statistical texture features, such as Local Derivative Pattern (LDP), Local Binary Pattern (LBP) and it's variant or structure texture features. Some approaches represent a grey histogram distribution of the palm vein image but suppress positional information on the palm vein texture. Compensation strategies are adopted for the lost information. Sometimes the performance of these approaches is also affected by weak texture palms.

Local invariant-based methods, [6], [16] such as Scale Invariant Feature Transform (SIFT), ASIFT and RootSIFT, are not sensitive to axis changes, scale, or rotation. These approaches are mostly used for contactless recognition systems. While these approaches are not fast in speed.

Subspace-based methods, [17]–[20] project palm vein images into subspaces. Subspace approaches, including Principal Component Analysis (PCA), Linear Discriminative Analysis (LDA), Independent Component Analysis (ICA), take the identified object as a whole. Subspace coefficients are taken as eigenvalues without prior knowledge.

These handcrafted approaches show excellent performance in small-scale databases. In the large-scale database, they suffer from the high inter-class similarity and intra-class variation. To address these problems, deep learning techniques help considerably.

B. DEEP LEARNING METHODS

We classify deep learning methods used in palm vein recognition into two groups: based on palm vein representation and based on network design. The palm vein representation category focuses on making discriminative and robust vein representations by the deep learning method. The network design category focuses on a good network structure to obtain a good recognition performance.

Some literature has explored the method of palm vein representation by deep learning methods. Pan *et al.* [21] represented the palm vein feature by a multi-layer convolutional feature concatenation with a semantic feature selector. They kept the semantic information with high-level convolution, detailed information with low-level convolution, and removed the background information. Wang *et al.* [22] designed a model with multi-weighted co-occurrence representation model. A convolution between an indicator filter in a higher layer with a to-be-reweighted filter in a lower layer and a redundancy-driven indicator filter selection algorithm is designed. Wang *et al.* [23] adopted convolutional activations as the regional representation of palm vein images. To obtain more discriminative and robust feature representation, they proposed a spatial weighting convolutional feature model. QIN *et al.* [24] extracted vein features by an iterative deep belief network (DBN). Zhang *et al.* [25] proposed the Palm-RCNN scheme based on CNN. It used 6 convolutions and a 1 pooling operation to extract shallow layer features.

Other literature focuses on the network structure. Lefkovits *et al.* [26] applied transfer learning and fine-tuning techniques to AlexNet, VGG-16, SqueezeNet, and
ResNet-50. They proposed conv4-fc2 and conv6-fc2 CNN architectures as well. Chantaf *et al.* [27] use Inception V3 and smaller VGGNet architectures in palm vein recognition. After comparing the convolutional layers from 2-5, Trabelsi *et al.* [28] proposed the architecture of a 3-layer CNN network. Wulandari *et al.* [29] implemented the VGG16 and VGG19 for palm vein recognition. Wan *et al.* [30] tested VGG19 for hand vein recognition in practice. Bhilare *et al.* [31] used palm veins as a multimodal model for biometric recognition. During the matching stage, the proposed method adopted the convolutional layer along with subsampling, max-pooling, and rectification layer. Thapar *et al.* [32] proposed PVSNet. Generative domain-specific features were learned by an encoder-decoder structure. After that, a Siamese network was taken as an autoencoder. Kuzu [33] proposed a densely connected convolution autoencoder way to increase the discriminative capability. It connected a supervised CNN with an unsupervised autoencoder.

All these works improved the performance of palm vein recognition. However, the study of outside boxes and contactless palm vein recognition is still an area to research. Capturing palm vein image under an unconstrained environment is sanitary and comfortable. Among the above literature, only Zhang *et al.* [25] and Chantaf *et al.* [27] did studies of palm vein recognition under outside boxes and contactless image environments. There are still many problems to research. Intending to address two challenging issues (i. more

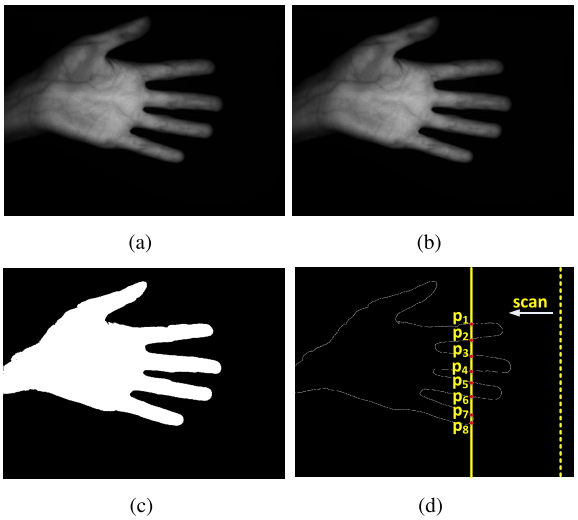


FIGURE 3. Prepare for ROI location. (a) The original palm vein image. (b) The filtering image from (a). (c) Binary image segmented from (b). (d) Eight cross points on the hand contour.

optical brought by visible light, ii. image rotation,position translation, and scale transformation which makes classification difficult, especially in large databases) of outside boxes and contactless palm vein recognition, we proposed a wavelet denoising ResNet for outside box and contactless palm vein recognition.

III. METHOD

We explain our method from ROI extraction to the WD model and SER model.

A. ROI EXTRACTION

Fig. 3, Fig. 4, and Fig. 5 show the extraction procedure of the ROI. Fig. 3 (a) shows an original palm image from an inside box database. It is denoised by the low-pass filter as shown in Fig. 3 (b), and then the image is binarized as shown in Fig. 3 (c). The palm contour is extracted by the expansion method of binocular morphology. The contour image in Fig.3 (d) is scanned by a vertical line from right to left. As shown in Fig. 3 (d), when the vertical line crosses the hand contour with 8 points, we record the position of these points from p_1 to p_8.

We locate two key points A and B in the finger valley. As shown in Fig. 4, Point A is in the valley of the index finger and middle finger. Point B is in the valley of the ring finger and pinkie. First, we fill in the palm contour with white pixels. Point A is searched from point p_2 to point p_3 along with the contour of the index finger and middle finger; there is a blue arc. We take every point on the blue arc as the centre of a circle. The radius of the circle is denoted as r. The length of r equals one-third of $\overline{p_2p_3}$. Comparing the number of white pixels in these circles, we choose the maximum one. The centre of that circle is point A, the valley of the index finger and middle finger. If the number of maximum points is not one, the first point we find is chosen. By this method, we can find B.

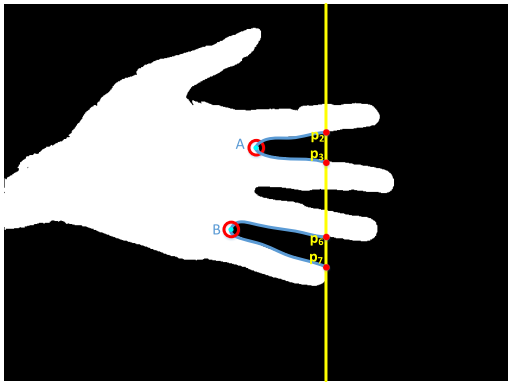


FIGURE 4. Search the key points.

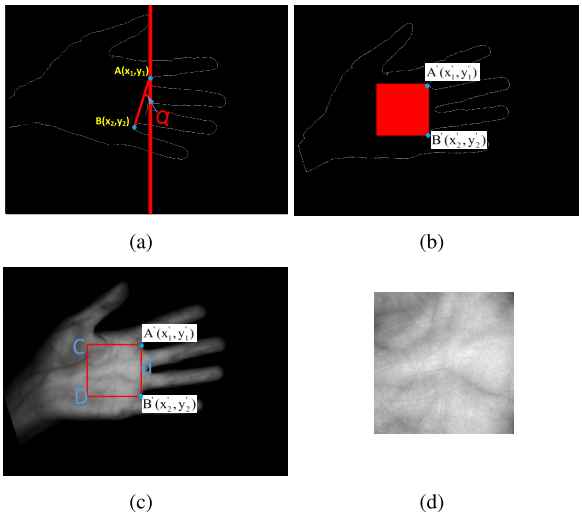


FIGURE 5. Locate and extract the ROI. (a) Determine the rotation angle. (b) Rotate (a) with the angle α and determine the square. (c) Extract the square from the rotated original image (Fig. 3(a)). (d) ROI.

Keypoint A(x_1, y_1) and keypoint B(x_2, y_2) are employed for orientation normalization (Fig.5(a)). Connecting point A and point B, we obtain line AB. The angle between line AB and the vertical line is denoted as α.

We rotate the hand image by the angle α to normalize the orientation variation, as shown in Fig. 5(b). The centre point of rotation is the centre of Fig. 5 (a). We rotate the image by the same angle for Fig. 3 (a). The centre of rotation is still the centre of Fig. 3 (a). As shown in Fig. 5(b), we correspond points A and B in Fig. 5(a) to points A' and B'. The distance from A' to B' is denoted as d. We make a square A'CDB' with one side A'B' and rotate Fig. 3 (a) with the same angle. In the same position, we extract A'CDB' from Fig. 5(c) and resize A'CDB' to $128pixels \times 128pixels$ as the ROI, as shown in Fig. 5 (d).

B. WD MODEL

Compared to inside box capture, outside box capture will be affected by visible light. Under the spectral window (700 nm-900 nm), the NIR absorption rate of haemoglobin (haemoglobin including oxygen-containing haemoglobin and

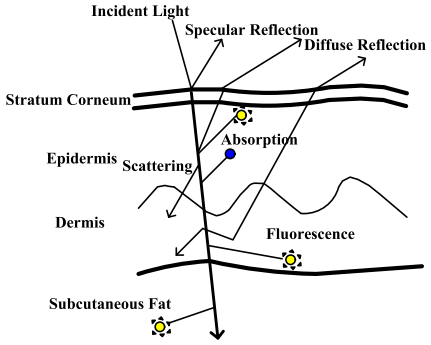


FIGURE 6. Schematic diagram of the incident light in skin tissue [34].

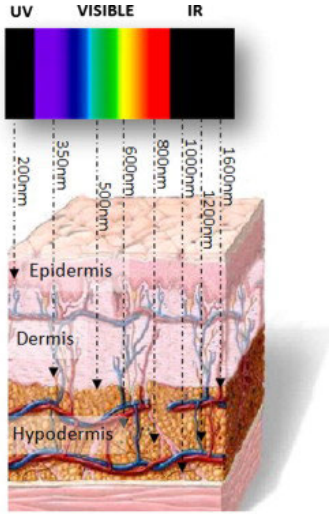


FIGURE 7. Light of different wavelengths penetration into skin.

deoxidizing haemoglobin) in the surrounding tissues is weaker than the rate of the vein vessels [2]. Vessels absorb more NIR light, so the vein structure forms a shadow in the camera. The NIR light must penetrate the palm skin deep enough to reach the blood vessel. There are four layers of palm skin: stratum corneum layer, epidermis layer, dermis layer, and subcutaneous layer [4], as illustrated in Fig. 6. Incident light is reduced by specular reflection, scattering, and diffuse reflection. Specular reflection, scattering, and diffuse reflection come the into the skin scattering and optical blurring. Fluorescence also reduces the fidelity of vein images. The penetration ability of visible light is weaker than that of NIR light [34], as shown in Fig. 7. Visible light forms more skin scattering and optical blurring. During outside box palm vein image capture, we intensified the NIR light to decrease the effect of visible light.

Skin scattering and optical blurring are mainly concentrated in the high-frequency part. We need to keep and strengthen the texture information and remove the noise factor. The traditional method applies 2D-DWT on ROI. We adopt Haar as the kernel of 2D-DWT. As shown in Fig. 8, the 2D-DWT decomposes image X into four sub-bands (Band A, Band V, Band H, and Band D) [35]. The band A

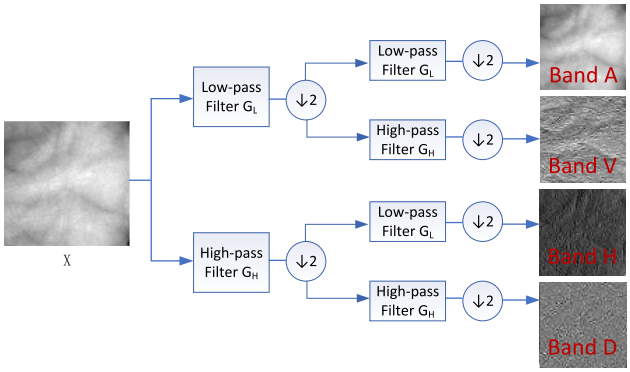


FIGURE 8. Schematic diagram of the 2D-DWT. X is decomposed into four sub-bands: average (Band A); horizontal (Band H); vertical (Band V); and diagonal (Band D).

contains texture features for palm vein recognition. band D represents the high-frequency information of the palm vein image. It shows most of the noises. The classification experiment on the band A did not enhanced the recognition accuracy because some useful information was ignored by the band A. We need a network that cannot only remove skin scattering and optical blurring but also strengthen A channel information.

We proposed a WD model. This model adds handcrafted information into the deep learning network. The handcrafted information is an A sub-band wavelet inverse transform image, which has low-frequency features. The deep learning model has 17 layers, which learn how to denoise. It enhances low-frequency information which benefits recognition by a residual learning technique. The residual learning technique also prevents vanishing and exploding gradient problems. We also use batch normalization (BN) [36] to accelerate the convergence of the trained model. Dilated convolution [37] is adopted to capture more contextual information with a lower computational burden.

Denoising network estimates the latent clean image x from its noisy observation y. The denoising formula can be expressed as $y = x + \mu$. Here, μ is white Gaussian noise added to the original image. Eq. (1) is the loss function to train the WD model. The network parameters p can be learned as follows:

$$l(p) = \frac{1}{N}\sum_{i=1}^{N} \|f(y_i, p) - (y_i - x_i)\|^2. \quad (1)$$

Here y_i denotes the ith noisy observation and x_i denotes the ith original image. x_i is the A sub-band wavelet inverse transform image. Adaptive moment estimation (Adam) is adopted as an optimizer in the WD model.

The structure of the WD model is shown in Fig. 1. The kernel size of all the convolution layers is 3×3. The 1st and 16th layers are composed of convolution and a rectified linear unit (ReLU) activation function. The padding size of these two convolution layers is 1, and the stride is 1. The 2nd, 5th, 9th, and 12th layers have the same structure. They are composed of dilated convolution, batch normalization,

and ReLU. The function of dilated convolution is to increase the texture information and reduce the computational cost. The padding is 2, the stride is 1, and the dilation is 2. The last layer is the convolution layer. The padding size is 1, and the stride is 1. The other layers are composed of convolution, batch normalization, and ReLU. The padding is 1, the stride is 1. There is a shortcut connection in the WD model. This shortcut connection enhances texture features from the previous layer. It retains the low-frequency information. Through this shortcut connection, we input the handcrafted information into the deep learning network.

C. SER MODEL

Outside box and contactless palm vein recognition in large-scale databases suffer from high inter-class similarity and intra-class variation. Too simple a neural network (e.g., 3 layers) cannot achieve good performance. Increasing the depth of the network can improve the accuracy of the results. However, with the deepening of the depth, there will be two problems [38]. The deeper the depth is, the gradient will decay, and the error intensity will be propagated back by multiplying the weight, and the weight will be smaller and smaller. Resnet gives a good solution to these two problems. Also, this recognition model suffers from high inter-class similarity and intra-class variation. The proposed model must address three problems. The first one is the vanishing or exploding gradients. The second problem is degradation. The third problem is to enhance the classification feature.

Therefore, we propose the SER model. The model takes ResNet18 as the backbone plus the squeeze-and-excitation(SE) model. ResNet can settle the problem of vanishing or exploding gradients by normalization [39]. The degradation is addressed by the residual learning framework [40]. ResNet can also retain more of the low-level features (such as texture features, structure features and object edges feature) [41], [42]. SENet [43] addresses the third problem by considering the eigenvalues channel relationship. Without introducing new spatial dimensions, we adopt the ''eigenvalues recalibration'' method to explain the interdependence relationship among the eigenvalue channels. Through deep learning, the model can obtain the degree of importance of the different channels. The eigenvalues useful for classification are enhanced, and useless eigenvalue are suppressed.

We apply the SE model after two convolutions on the residual block of the ResNet18 network.

The squeeze operation is:

$$z_c = F_{sq}(u_c) = \frac{1}{W \times H} \sum_{i=1}^{W} \sum_{j=1}^{H} u_c(i, j). \tag{2}$$

Here, z_c demonstrates the statistical value of the c-th channel of feature u. C is the channel number of feature u. The dimension of each channel is the same, and we represent it as $W \times H$. u_c is the c-th feature map after the 2nd Conv layer of the residual block.

TABLE 1. The network architecture of the SER model.

Output size	SER model		
[1,64,45,45]	Block1:Conv. $[7 \times 7]$, stride=2, max pool, $[3 \times 3]$, stride 2		
[1, 64, 22, 22]	Block2:	$Conv., 3 \times 3, 64$ $Conv., 3 \times 3, 64$ $fc, [4, 64]$	$\times 2$
[1, 128, 11, 11]	Block3:	$Conv., 3 \times 3, 128$ $Conv., 3 \times 3, 128$ $fc, [8, 128]$	$\times 2$
[1, 256, 6, 6]	Block4:	$Conv., 3 \times 3, 256$ $Conv., 3 \times 3, 256$ $fc, [16, 256]$	$\times 2$
[1, 2048, 1, 1]	Block5:	$Conv., 3 \times 3, 2048$ $Conv., 3 \times 3, 2048$ $fc, [128, 2048]$	$\times 2$
[1,64,45,45]	Average pool,FC(2048,530)		

The excitation operation is:

$$s = F_{ex}(z, W) = \sigma(W_2 \delta(W_1 z)). \tag{3}$$

where δ is a ReLU activation function, σ is a sigmoid function, $W_1 \in R^{\frac{C}{r} \times C}$ is the fully connected layer for dimensionality reduction, $W_2 \in R^{C \times \frac{C}{r}}$ is the fully connected layer for dimensionality increase, and r = 16.

The output of the residual block layer is:

$$\widetilde{x_c} = F_{scale}(u_c, s_c) = s_c u_c. \tag{4}$$

where x_c is the final output of the SE model, and $F_{scale}(u_c, s_c)$ refers to channel-wise multiplication between the scalar s_c and the feature map u_c.

The network structure is described in Table 1.

The input image size is 128 *pixels* × 128 *pixels*. The kernel size of the first convolution layer is 7 × 7. The convolution kernel size of other layers is 3 × 3. There are 5 blocks in the SER model. Except for block1, the other blocks include two residual modules. We implement average pooling on the feature map after the last SE layer. The output dimension of the SE module is expressed in the brackets following *fc*. An eigenvector of 2,048 dimensions results from full connection. The classification probability is represented by softmax.

IV. EXPERIMENTS AND RESULTS

In this section, we introduce the experiments and results from the following aspects: A. Databases and experimental environments, B. WD model, C. SER model, and D. ablation experiment and running time.

The general idea is described in the following. First, we train the WD model on a self-built box palm vein image database (marked as Database I). We use images from Database I as clean images and synthesize noisy images by adding Gaussian noise. To evaluate the denoising effect of the WD model, we conduct a quantitative experiment on Database I and a qualitative experiment on a self-built outside box palm vein image database (marked as Database II).

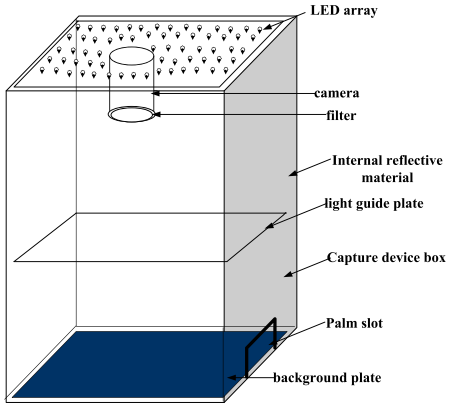


FIGURE 9. Structure diagram of the inside box palm vein image acquisition prototype.

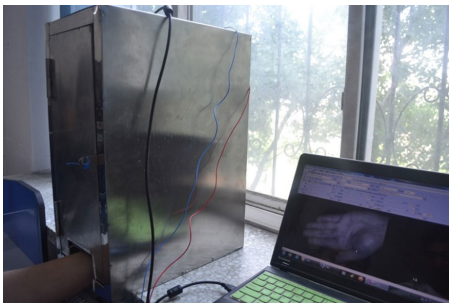


FIGURE 10. Inside box palm vein image capture device.

Second, we train the SER model on Database II. To demonstrate the performance of the WD model, the paper compares it with some other classical palm vein recognition methods on Database II and a Tongji contactless palm vein database (marked as Database III) [25]. Finally, an ablation experiment and time cost experiment on Database II are reported.

A. DATABASES AND EXPERIMENTAL ENVIRONMENTS

We use three databases in this paper: Database I, Database II, and Database III.

1) DATABASE I

We need a benchmark database with clean/noisy palm vein image pairs to train WD. Since there is not a public database that satisfies our requirements, we make one. The clean palm vein images come from our previous work [44]. In [44] we set up an inside box palm vein image database with 1,500 images from 250 volunteers. Fig. 9 shows the structure diagram of the inside box palm vein image acquisition device. In the image capture system, an 850 nm LED array was chosen as the active light source, and an OV5116 CMOS camera was chosen as the camera. Fig. 10 shows the demonstration of the inside box palm vein image capture prototype. The corresponding noisy palm vein images were synthesized with white Gaussian noise. We describe this in the following experiments. Fig. 11 shows the clean palm vein image from Database I.

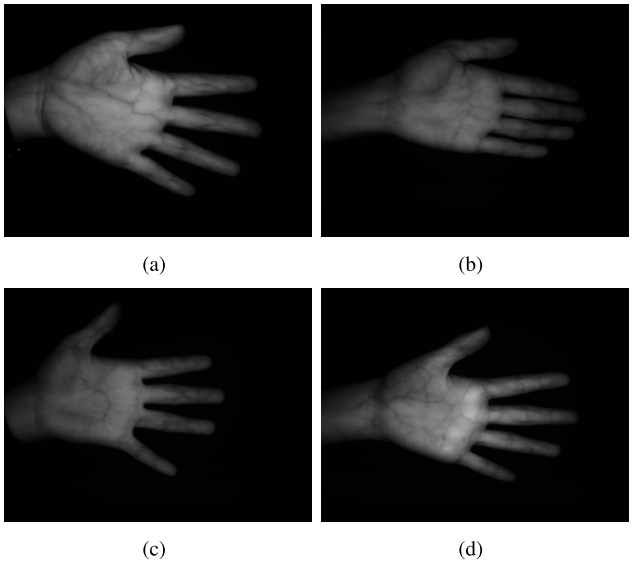


FIGURE 11. Clean palm vein image from database I.

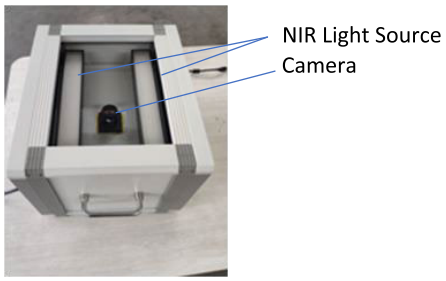


FIGURE 12. Outside box and contactless palm vein image acquisition prototype.

2) DATABASE II

To show the performance of the proposed approach for outside box and contactless palm vein recognition, we set up an outside box and contactless palm vein image database: Database II.

We adopted an MV-VD120SM 1/2" CCD camera as the camera of this self-developed palm vein image acquisition prototype. The resolution of an image was $1,280$ *pixels* $\times$ 960 *pixels* pixels. The active light source part was a two-strip light source of 850 nm LED sets. The device is shown in Fig. 12.

During palm vein image capture, the volunteers held their hands on the capture device, palm facing the camera with approximately 20 cm on the device. Different from other databases, our capture environment was indoors and did not avoid visible light. We did not collect images in a darkroom or inside a box. Our research team conducted contactless palm vein image capture, which included 5,300 images from the hands of 265 individuals who participated in both visits. All the images were collected in both sessions. The average interval between the 1st and 2nd image acquisitions was approximately one week. Each volunteer provided 20 images. Two hands were captured. Five images of each palm vein

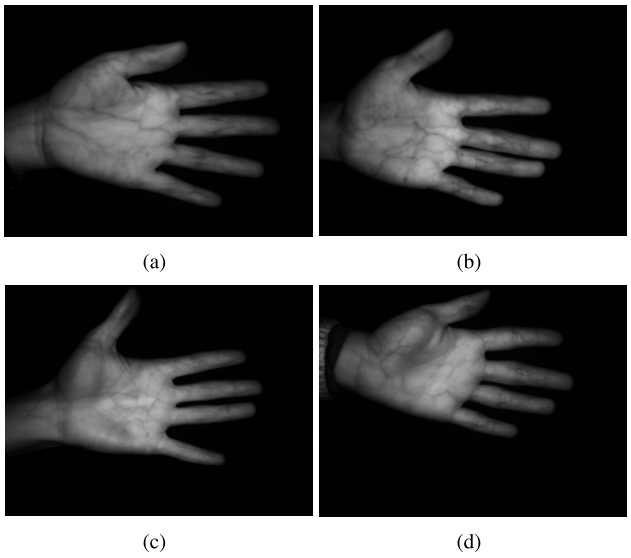


FIGURE 13. Palm vein images from dataset II.

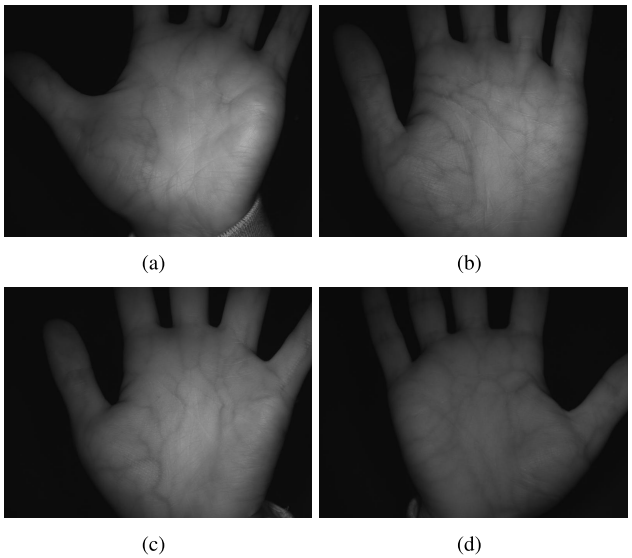


FIGURE 14. Palm vein images from dataset III.

were acquired in each stage. The age distribution of volunteers was from 18 to 56 years old. The image samples of Database II are shown in Fig. 13.

We can see in Fig. 13 that the image quality is not as good as Fig. 11 because it is affected by the visible light.

3) DATASET III

To obtain an objective evaluation, we consider five open access palm vein databases: the Chinese Academy of Sciences' Institute of Automation (CASIA) [45], the PolyU Multispectral Palmprint Database (PolyU) [46], the VERA Palmvein Database (VERA) [47], PUT database (PUT) [26] and the Tongji contactless palm vein image database. The CASIA, VERA, and Tongji databases are contactless palm vein databases. CASIA is an inside box database. The VERA palm vein database only provides the first 50 individuals. Therefore, we evaluated our method on the Tongji contactless palm vein image database [48] (Database III).

Database III contains 12,000 palm vein images from 600 palms. All the images were captured with 940 nm light. The image quality in the Tongji dataset is much better than that in Database II because they adopted an NIR CCD camera (JAI AD-80 GE). The image samples of Dataset III are shown in Fig. 14.

4) EXPERIMENTAL ENVIRONMENTS

Our experimental environment is described in Table 2.

TABLE 2. Experimental environment.

Hardware Condition	Software Condition
CPU: Intel(R)_Core(TM)_i9-9980HK_CPU_@_2.40GHz	Operating System: Ubuntu(Version 16.04)
GPU: NVIDIA Quadro RTX 3000	Run environment: Python 3.5.2+Cuda10.2+CUDNN7.6.5
RAM:64G	IDE:Pycharm2019.3.4

B. WD MODEL

First, we train the parameters of the WD model on Database I. Second, we compare quantitative results with other denoising methods on Database I. Finally, we compare the visually qualitative results with other denoising methods on Dataset II. Several representative traditional denoising methods and deep learning denoising methods with published codes are selected for comparison: BM3D [49], LSM_NLR [50], NL_means [51], DIP [52], ECND [53], and DNCNN [54].

1) TRAINING THE WD MODEL ON DATABASE I

The clean palm vein image is obtained from Database I. This kind of image method avoids visible light and the image quality is good. We extract ROIs from these clean palm vein images as the ground truth (GT). The noisy image is synthesized with white Gaussian noise [53], [54] of different levels ($\sigma = 10,15,25$). In our experiment, when the white Gaussian noise $\sigma = 10$, the images are most likely the outside box image. We use the noisy image as training data. The structure of the WD model is described in part B of METHOD III.

We randomly split Database I into a training set of 1,200 (200×6) images and a validation set of 300 (50×6) images. To avoid overfitting, we perform data augmentation in two ways: (1) randomly rotate the image by $\pm 5°$, and (2) flip the image horizontally or vertically with a probability of 0.5.

The initial parameters of the WD model are set as follows: the learning rate equals 0.001, beta_1 is set to 0.9, beta_2 is set to 0.999, and epsilon is 1×10^{-8}. The number of batches is selected as 64. The number of epochs is 200 for the trained model.

TABLE 3. Quantitative results of different state-of-the-art methods evaluated on Database I.

	PSNR	SSIM
BM3D	37.3918	0.9243
LSM_NLR	36.6352	0.8825
NL_means	35.7773	0.8349
ECND	37.3781	0.9196
DNCNN	37.4073	0.9201
WD model	**37.5328**	**0.9248**

2) COMPARISON WITH OTHER DENOISING METHODS ON DATABASE I

The WD model compared denoising evaluation indicator and visually quantitative results with some state-of-the-art denoising methods. Two evaluation indicators are adopted: peak signal-to-noise ratio (PSNR) and structural similarity (SSIM).

Quantitative results are reported in Table 3. It shows the average PSNR and SSIM metric values of state-of-the-art methods and our method.

The PSNR of the WD model is higher than the other compared methods from 0.1255 to 1.7555. The SSIM metric value is 0.9248, higher than the other compared methods from 0.0005 to 0.0899. These results demonstrate the efficiency of the WD model.

To visually demonstrate the results obtained by these methods on Database I, we randomly choose two sample images for testing. A comparison of the quantitative results is presented in Fig.15.

3) TESTING THE WD MODEL ON DATABASE II

We also illustrate the results of the WD model and several popular methods on Database II, a real noisy image database. Fig. 16 shows the comparison of visually qualitative results.

C. SER MODEL

We selected the parameters of the SER model from the experiments on Database II. For some state-of-the-art methods applied in different image capture environments, we compare the proposed method with some classical methods that have been applied in palm vein recognition in Dataset II and Dataset III.

We use accuracy (Acc) and equal error rate (EER) as the evaluation metrics, which are widely used in biometrics [23], [24]. Accuracy is the ratio of the number of correct predictions to the number of test samples. EER is the equilibrium point where FRR equals FAR. FRR is the ratio that a biometric system incorrectly rejects to the number of genuine samples. FAR is the ratio that a biometric system falsely accepts to the number of forged samples [55]. A lower EER indicates better performance.

1) EXPERIMENTS ON DATABASE II

We randomly split Database II into a training set (530 × 7 images) and a validation set (530 × 3 images). To avoid overfitting, we perform data augmentation in two ways: (1) randomly rotate the image by ±5° and (2) change brightness, contrast and saturation by 0.5 separately. The parameters in the network are set as Table 1. The other parameters are set as follows. The input image size is 128 *pixels* × 128 *pixels*. The learning rate is 0.01. Cross-entropy is adopted as the loss. The number of batches is 32. The number of epochs is 100.

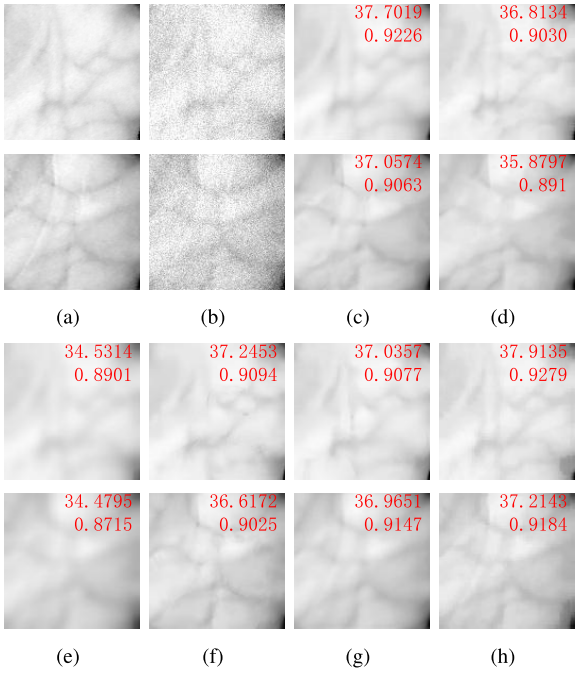


FIGURE 15. Comparison of quantitative results on database I. (a) GT. (b) Noise image. (c) BM3D. (d) LSM_NLR. (e) NL_means. (f) ECND. (g) DNCNN. (h) WD model.

The wavelet denoising ResNet is compared with some other classical deep learning palm vein recognition algorithms [25], [27], [29], [40], and [30]. We compare the training and validation process of our method with these methods on Database II, as shown in Fig. 17.

In Fig. 17, we note that the overall classification accuracy increases as the number of iterations increases. All of the accuracies are nearly stable during epoch 100. Our method is the most stable during the training process and validation processes. By contrast, Fig. 18 denotes the loss change trend on Database II.

Fig. 18 demonstrates that the overall validation accuracy increases until it tends to stabilize as the number of iterations increases. We load the model of the methods and compare their validation accuracies. The validated accuracy of each network is shown in Table 4.

Table 4 shows that the proposed method outperforms all the other five methods by a margin ranging from 0.91% to 4.35% on accuracy. Such a performance gain demonstrates that the wavelet denoising ResNet is a promising direction for outside box and contactless palm vein recognition.

We also compare the ROC curve and EER with some deep learning algorithms that have been used in palm vein

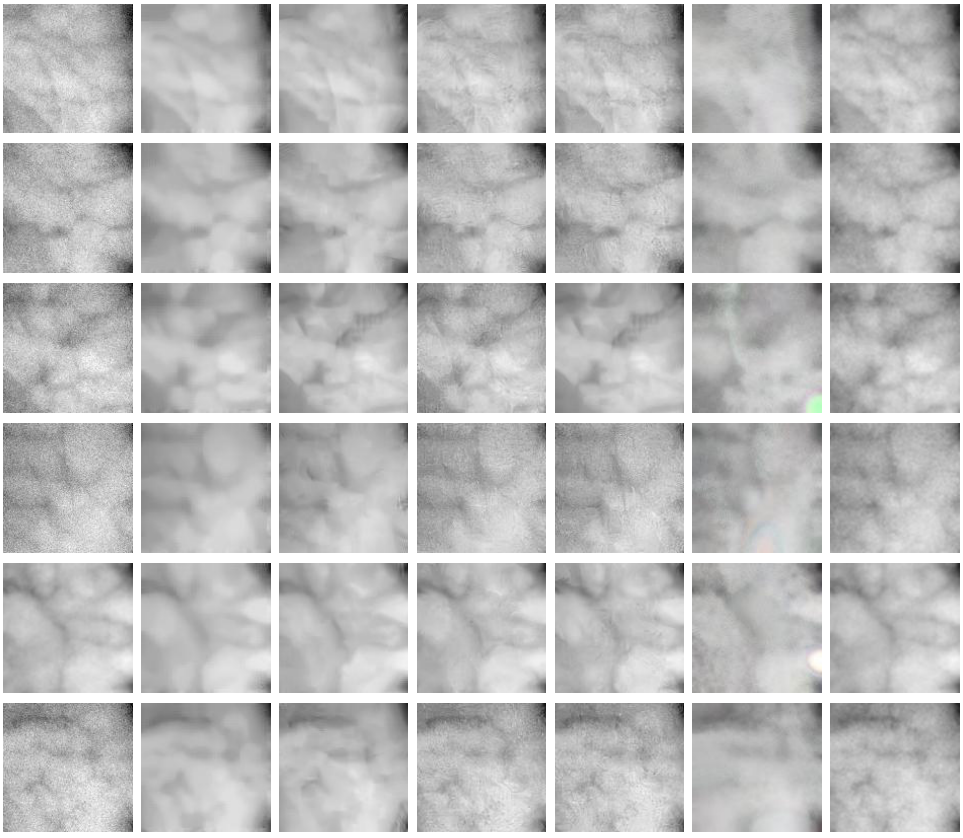


FIGURE 16. Comparison of visually qualitative results on dataset II. (a) Original images. (b) BM3D. (c) LSM_NLR. (d) DNCNN. (e) ECND. (f) DIP. (g) WD model.

TABLE 4. Comparison of accuracy with classical methods on database II.

Reference	Method	Accuracy
[29]	VGG16	0.9413
[27]	Inception V3	0.9757
[25]	modified-Inception Resnet	0.9712
[40]	Resnet18	0.9684
[30]	VGG19	0.961
Ours	WD Resnet	**0.9848**

TABLE 5. Comparison of EER with classical methods on database II.

Reference	Method	EER(%)
[29]	VGG16	3.66
[27]	Inception V3	2.1
[25]	modified-Inception Resnet	2.2
[40]	Resnet18	2.64
[30]	VGG19	2.51
Ours	WD Resnet	**0.88**

recognition. We perform one-to-many matching among test images ($100 \times 6 = 600$ images). First, we take test images into the well-trained proposed network. Then, through the output of block 5, we obtain a feature vector with 2,048 dimensions $V^{(i)} = \left[p_1^{(i)}, p_2^{(i)}, \ldots, p_{2048}^{(i)},\right]^T$ corresponding to each palm vein ROI image $I^{(i)}$. We match the distance between different images with the Euclidean distance of the feature vector $V^{(i)}$. The intra-class matching number is 1,500 ($100 \times C_6^1 \times C_5^1/2$), and the inter-class matching number is 178,200 ($100 \times 99 \times C_6^1 \times C_6^1/2$). We calculate FAR and FRR as [23], [24], [54], [55], and we draw the ROC curve. We draw the ROC curves for the other comparison methods, as shown in Fig. 19. By adjusting the recognition threshold to 0.5975, FAR equals FRR. We record the EER of the proposed method. In the same way, we record the EER of the comparison methods. The results are shown in Table 5.

Table 5 demonstrates that the wavelet denoising ResNet method outperforms all the other five compared methods by a large margin ranging from 1.22% to 2.78% on EER. The reason is that the proposed method fuses effective handcrafted features into the deep learning network by residual learning technology and enhances the classification features by eigenvalue recalibration.

2) RECOGNITION PERFORMANCE ON DATABASE III

This section tests the performances of the proposed network applied to Database III. The trained parameters on II are directly applied to this database. Table 6 demonstrates

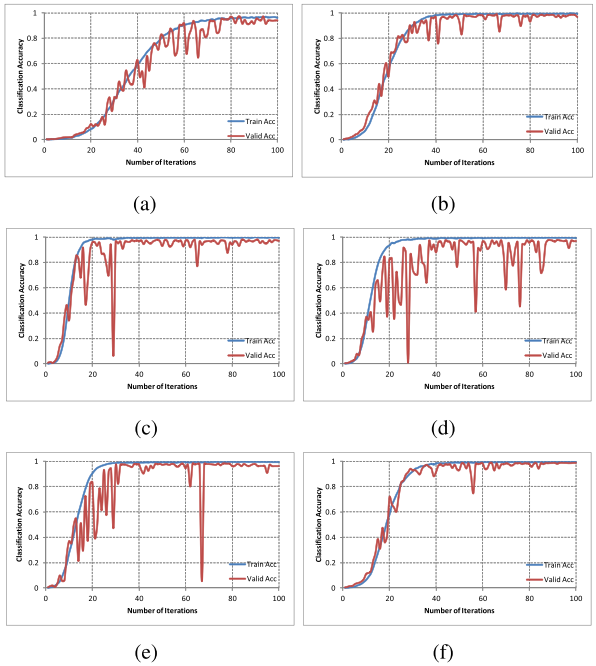


FIGURE 17. Training and validation sets accuracy change trend of comparison methods on database II. (a) Ref [29]. (b) Ref [27]. (c) Ref [25]. (d) Ref [40]. (e) Ref [30]. (f) proposed.

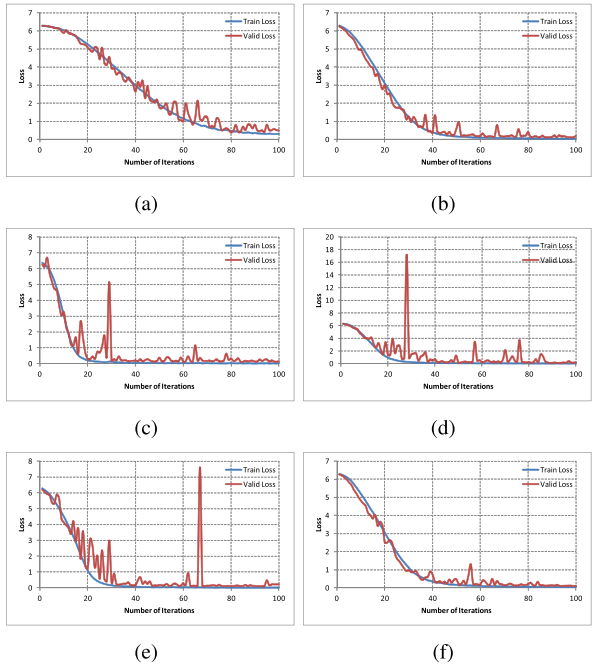


FIGURE 18. Training and validation sets loss change trend of comparison methods on database II. (a) Ref [29]. (b) Ref [27]. (c) Ref [25]. (d) Ref [40]. (e) Ref [30]. (f) proposed.

the comparison of accuracy with classical methods on Database III.

Table 6 demonstrates that the proposed method outperforms the other five methods by a margin ranging from 0.68% to 1.59% accuracy on Database III. The reason is that the image quality in Database III is better than Database II.

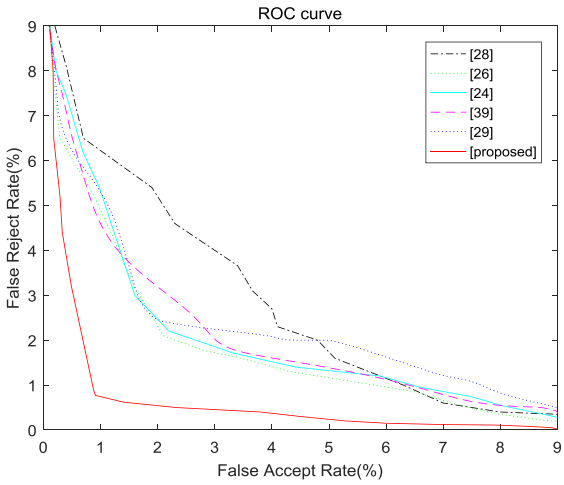


FIGURE 19. Comparison of ROC curves with classical methods on database II.

TABLE 6. Comparison of accuracy with classical methods on database III.

Reference	Method	Accuracy
[29]	VGG16	0.9711
[27]	Inception V3	0.9875
[25]	modified-Inception Resnet	0.9900
[40]	Resnet18	0.9902
[30]	VGG19	0.9741
Ours	WD Resnet	**0.9970**

TABLE 7. Comparison of EER with classical methods on database III.

Reference	Method	EER(%)
[29]	VGG16	2.3
[27]	Inception V3	1.88
[25]	modified-Inception Resnet	1.02
[40]	Resnet18	0.87
[30]	VGG19	2.0
Ours	WD Resnet	**0.41**

With the same one-to-many matching methods, we compare the ROC curve and EER with some deep learning algorithms on Database III. We draw the ROC curves of these methods in Fig. 20. The EER of these networks is shown in Table 7. Table 7 demonstrates that the wavelet denoising ResNet method outperforms the other five compared methods by a margin ranging from 0.46% to 1.89% on EER.

According to Table 7, the proposed network achieves the best performance. However, the performance is slightly better than [25] and [40]. This is because the quality of the image in this database is better than that in Database II. Database III captures part of the palm, and Database II contains the whole hand. The height of the hand during capture was higher in Database II than in Database III. The images in Database II were affected by visible light more than the images in Database III. Visible light brings more scattering noise and blurring noise. The potential of the proposed method cannot fully exploit the advantages.

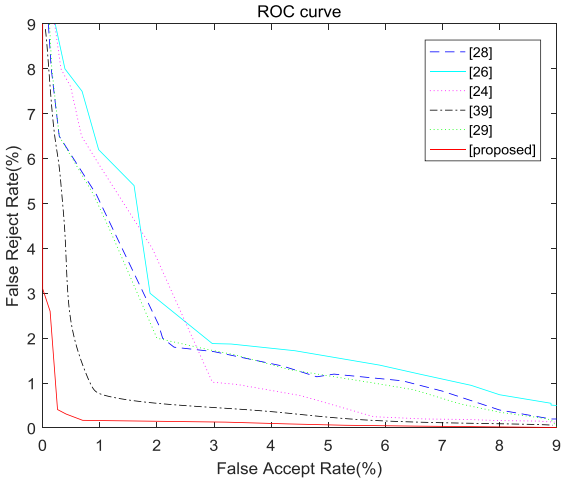


FIGURE 20. Comparison of ROC curves with classical methods on database III.

TABLE 8. Ablation experiment result on database II.

Methods	EER(%)
Without the WD model	2.06
Resnet 18 without modify	1.61
The whole net	0.88

TABLE 9. Average time consumption of our system on database II.

Process	Average time consumption(s)
the WD model	0.0838
Se-resnet18 part	0.1801
Total	0.2638

D. ABLATION EXPERIMENT AND TIME CONSUMPTION

To evaluate the effectiveness of individual components, the following ablation studies on the proposed network are conducted on Database II.

We conduct two comparative experiments: recognition without the WD model, and the SER model without the squeeze-and-excitation module. The experiment results are shown in Table 8.

The WD model increases the recognition result by 1.18%. The modification ResNet18 also slightly increases the recognition performance.

Time consumption is another indicator to measure the recognition system. The time consumption components of our proposed network include the WD model and the SER model. We run 100 images to calculate the average time consumption on Database II in our experimental environment. The result is shown in Table 9. Based on Table 9, the time consumption is acceptable for a real system.

V. CONCLUSION

In the past, the majority of the research captured palm vein images inside boxes or in dark environments. This kind of image environment constrains the practicality of palm vein recognition. Our research uses the power of deep learning to denoise the influence of visible light and enhances the performance of contactless recognition. This deep learning network includes two models: the WD model and SER model. The WD model is adopted to remove the noise of skin scattering and optical blurring from palm vein images. This strategy increases the weight of an effective handcrafted feature in the deep learning network. The WD model compared denoising evaluation indicator and visually quantitative results with some state-of-the-art denoising methods. The PSNR of WD model is 37.5328, which is higher than the other compared methods from 0.1255 to 1.7555. The SSIM metric value is 0.9248, higher than the other compared methods from 0.0005 to 0.0899. Second, the SER model overcomes rotation, position translation and scale transformation by selectively emphasizing classification features and weakening less useful features. At the same time this model addresses the problem of vanishing or exploding gradients, and degradation. A self-built inside box palm vein image database (Database I) and an outside box palm vein image database (Database II) are set up to train and validate the network. A public outside box and contactless palm vein database (Database III) is also employed to demonstrate the superiority and robustness of our network. The proposed deep learning network decreases the EER to 0.88 in Database II and 0.41 in Database III. This work explores the practice of palm vein recognition.

REFERENCES

[1] A. K. Jain, A. Ross, and S. Prabhakar, "An introduction to biometric recognition," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 1, pp. 4–20, Jan. 2004.

[2] P. MacGregor and R. Welford, "Veincheck: Imaging for security and personnel identification," *Adv. Imag.*, vol. 6, no. 7, pp. 52–56, 1991.

[3] W. Wu, S. J. Elliott, S. Lin, S. Sun, and Y. Tang, "Review of palm vein recognition," *IET Biometrics*, vol. 9, no. 1, pp. 1–10, 2019.

[4] R. R. Anderson and J. A. Parrish, "The optics of human skin," *J. Investigative Dermatology*, vol. 77, no. 1, pp. 13–19, Jul. 1981.

[5] A. F. Akbar, T. A. B. Wirayudha, and M. D. Sulistiyo, "Palm vein biometric identification system using local derivative pattern," in *Proc. 4th Int. Conf. Inf. Commun. Technol. (ICoICT)*, May 2016, pp. 1–6.

[6] K.-S. Wu, J.-C. Lee, T.-M. Lo, K.-C. Chang, and C.-P. Chang, "A secure palm vein recognition system," *J. Syst. Softw.*, vol. 86, no. 11, pp. 2870–2876, Nov. 2013.

[7] Y. Aberni, L. Boubchir, and B. Daachi, "Palm vein recognition based on competitive coding scheme using multi-scale local binary pattern with ant colony optimization," *Pattern Recognit. Lett.*, vol. 136, pp. 101–110, Aug. 2020.

[8] 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.

[9] L. Mirmohamadsadeghi and A. Drygajlo, "Palm vein recognition with local texture patterns," *IET Biometrics*, vol. 3, no. 4, pp. 198–206, Dec. 2014.

[10] P. Tome and S. Marcel, "Palm vein database and experimental framework for reproducible research," in *Proc. Int. Conf. Biometrics Special Interest Group (BIOSIG)*, Sep. 2015, pp. 1–7.

[11] E. Piciucco, E. Maiorana, and P. Campisi, "Palm vein recognition using a high dynamic range approach," *IET Biometrics*, vol. 7, no. 5, pp. 439–446, Sep. 2018.

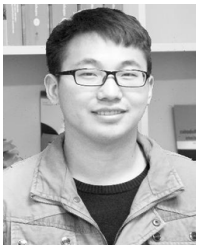
[12] W. Kang and Q. Wu, "Contactless palm vein recognition using a mutual foreground-based local binary pattern," *IEEE Trans. Inf. Forensics Security*, vol. 9, no. 11, pp. 1974–1985, Nov. 2014.

[13] 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.
[14] 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, Sep. 2017.
[15] A. M. Al-juboori, W. Bu, X. Wu, and Q. Zhao, "Palm vein verification using Gabor filter," *Int. J. Comput. Sci. Issues*, vol. 10, no. 1, p. 678, 2013.
[16] 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, May 2014, Art. no. e97548.
[17] 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.
[18] M. Micheletto, G. Orru, I. Rida, L. Ghiani, and G. L. Marcialis, "A multiple classifiers-based approach to palmvein identification," in *Proc. 8th Int. Conf. Image Process. Theory, Tools Appl. (IPTA)*, Nov. 2018, pp. 1–6.
[19] F. Rizki, T. A. B. Wirayuda, and K. N. Ramadhani, "Identity recognition based on palm vein feature using two-dimensional linear discriminant analysis," in *Proc. 1st Int. Conf. Inf. Technol., Inf. Syst. Electr. Eng. (ICITISEE)*, Aug. 2016, pp. 21–25.
[20] S. Elnasir and S. M. Shamsuddin, "Proposed scheme for palm vein recognition based on linear discrimination analysis and nearest neighbour classifier," in *Proc. Int. Symp. Biometrics Secur. Technol. (ISBAST)*, Aug. 2014, pp. 67–72.
[21] Z. Pan, J. Wang, Z. Shen, X. Chen, and M. Li, "Multi-layer convolutional features concatenation with semantic feature selector for vein recognition," *IEEE Access*, vol. 7, pp. 90608–90619, 2019.
[22] 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.
[23] J. Wang, Z. Pan, G. Wang, M. Li, and Y. Li, "Spatial pyramid pooling of selective convolutional features for vein recognition," *IEEE Access*, vol. 6, pp. 28563–28572, 2018.
[24] H. Qin, M. A. El Yacoubi, J. Lin, and B. Liu, "An iterative deep neural network for hand-vein verification," *IEEE Access*, vol. 7, pp. 34823–34837, 2019.
[25] L. Zhang, Z. Cheng, Y. Shen, and D. Wang, "Palmprint and palmvein recognition based on DCNN and a new large-scale contactless palmvein dataset," *Symmetry*, vol. 10, no. 4, p. 78, Mar. 2018.
[26] S. Lefkovits, L. Lefkovits, and L. Szilágyi, "Applications of different CNN architectures for palm vein identification," in *Proc. Int. Conf. Modeling Decis. Artif. Intell.* Cham, Switzerland: Springer, 2019, pp. 295–306.
[27] S. Chantaf, A. Hilal, and R. Elsaleh, "Palm vein biometric authentication using convolutional neural networks," in *Proc. Int. Conf. Sci. Electron., Technol. Inf. Telecommun.* Cham, Switzerland: Springer, 2018, pp. 352–363.
[28] S. Trabelsi, D. Samai, A. Meraoumia, K. Bensid, and A. Taleb-Ahmed, "An improved multispectral palmprint system using deep CNN-based palm-features," in *Proc. Int. Conf. Adv. Electr. Eng. (ICAEE)*, Nov. 2019, pp. 1–6.
[29] M. Wulandari, Basari, and D. Gunawan, "On the performance of pre-trained CNN aimed at palm vein recognition application," in *Proc. 11th Int. Conf. Inf. Technol. Electr. Eng. (ICITEE)*, Oct. 2019, pp. 1–6.
[30] 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.
[31] 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.
[32] 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)*, Jan. 2019, pp. 1–8.
[33] R. S. Kuzu, E. Maiorana, and P. Campisi, "Vein-based biometric verification using densely-connected convolutional autoencoder," *IEEE Signal Process. Lett.*, vol. 27, pp. 1869–1873, 2020.
[34] M. A. Ferrer, A. Morales, and A. Díaz, "An approach to SWIR hyperspectral hand biometrics," *Inf. Sci.*, vol. 268, pp. 3–19, Jun. 2014.
[35] B. Li and Y. He, "An improved ResNet based on the adjustable shortcut connections," *IEEE Access*, vol. 6, pp. 18967–18974, 2018.
[36] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 448–456.
[37] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," 2015, *arXiv:1511.07122*. [Online]. Available: http://arxiv.org/abs/1511.07122
[38] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*. [Online]. Available: http://arxiv.org/abs/1409.1556
[39] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1026–1034.
[40] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
[41] X. Ou, P. Yan, Y. Zhang, B. Tu, G. Zhang, J. Wu, and W. Li, "Moving object detection method via ResNet-18 with Encoder–Decoder structure in complex scenes," *IEEE Access*, vol. 7, pp. 108152–108160, 2019.
[42] C. Xiang, L. Zhang, Y. Tang, W. Zou, and C. Xu, "MS-CapsNet: A novel multi-scale capsule network," *IEEE Signal Process. Lett.*, vol. 25, no. 12, pp. 1850–1854, Dec. 2018.
[43] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7132–7141.
[44] 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.
[45] Y. Hao, Z. Sun, T. Tan, and C. Ren, "Multispectral palm image fusion for accurate contact-free palmprint recognition," in *Proc. 15th IEEE Int. Conf. Image Process.*, Oct. 2008, pp. 281–284.
[46] D. Zhang, Z. Guo, G. Lu, L. Zhang, and W. Zuo, "An online system of multispectral palmprint verification," *IEEE Trans. Instrum. Meas.*, vol. 59, no. 2, pp. 480–490, Feb. 2010.
[47] R. Kabacinski and K. Kowalski, "Vein pattern database and benchmark results," *Electron. Lett.*, vol. 47, no. 20, pp. 1127–1128, 2011.
[48] L. Zhang, L. Li, A. Yang, Y. Shen, and M. Yang, "Towards contactless palmprint recognition: A novel device, a new benchmark, and a collaborative representation based identification approach," *Pattern Recognit.*, vol. 69, pp. 199–212, Sep. 2017.
[49] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," *IEEE Trans. Image Process.*, vol. 16, no. 8, pp. 2080–2095, Aug. 2007.
[50] T. Huang, W. Dong, X. Xie, G. Shi, and X. Bai, "Mixed noise removal via Laplacian scale mixture modeling and nonlocal low-rank approximation," *IEEE Trans. Image Process.*, vol. 26, no. 7, pp. 3171–3186, Jul. 2017.
[51] C. Sutour, C.-A. Deledalle, and J.-F. Aujol, "Adaptive regularization of the NL-means: Application to image and video denoising," *IEEE Trans. Image Process.*, vol. 23, no. 8, pp. 3506–3521, Aug. 2014.
[52] V. Lempitsky, A. Vedaldi, and D. Ulyanov, "Deep image prior," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 9446–9454.
[53] C. Tian, Y. Xu, L. Fei, J. Wang, J. Wen, and N. Luo, "Enhanced CNN for image denoising," *CAAI Trans. Intell. Technol.*, vol. 4, no. 1, pp. 17–23, Mar. 2019.
[54] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising," *IEEE Trans. Image Process.*, vol. 26, no. 7, pp. 3142–3155, Jul. 2017.
[55] W. Kang, Y. Lu, D. Li, and W. Jia, "From noise to feature: Exploiting intensity distribution as a novel soft biometric trait for finger vein recognition," *IEEE Trans. Inf. Forensics Security*, vol. 14, no. 4, pp. 858–869, Apr. 2019.

WEI WU received the Ph.D. degree in measuring technology and instruments from the Shenyang University of Technology, Liaoning, China, in 2013. From 2016 to 2017, she was a Visiting Scholar with Purdue University, USA. She is currently an Associate Professor in Information and Engineering Department, Shenyang University. She currently holds a postdoctoral position with the State Key Laboratory of Robotics, Shenyang Institute of Automation (SIA), Chinese Academy of Sciences (CAS). Her main research interests include biometrics and deep learning.

QIANG WANG received the M.S. degree from Tianjin Normal University, in 2008, and the Ph.D. degree in pattern recognition and intelligent systems from the University of Chinese Academy of Sciences, China, in 2020. He is currently a Research Associate with the Key Laboratory of Manufacturing Industrial Integrated, Shenyang University. His research interests include deep learning, feature selection, and image restoration.

SIQUAN YU received the B.Sc. and M.Sc. degrees in automatic control from Liaoning Shihua University, China, in 2011 and 2015, respectively. He is currently pursuing the joint Ph.D. degree with Northeastern University, China, and the State Key Laboratory of Robotics, Shenyang Institute of Automation, Chinese Academy of Sciences. His research interests include image/video representation and deep learning.



QIONG LUO is currently pursuing the Ph.D. degree with the Shenyang Institute of Automation, Chinese Academy of Sciences. His research interests include image processing and low rank matrix recovery.

SEN LIN received the B.Sc. and M.Sc. degrees from Liaoning Technical University, in 2003 and 2006, respectively, and the Ph.D. degree from the Shenyang University of Technology, in 2013. He is currently an Associate Professor and a Master Supervisor with Shenyang Ligong University. He is a Postdoctoral Researcher with the Shenyang Institute of Automation, Chinese Academy of Sciences. His main research interests include biometric, machine vision, and deep learning.

ZHI HAN (Member, IEEE) received the B.Sc., M.Sc., and Ph.D. degrees in applied mathematics from Xi'an Jiao tong University (XJTU), China, in 2005, 2007, and 2012, respectively. From 2009 to 2011, he was a joint Ph.D. Student of statistics at the University of California, Los Angeles (UCLA), USA. He is currently a Professor with the State Key Laboratory of Robotics, Shenyang Institute of Automation (SIA), Chinese Academy of Sciences (CAS). His research interests include image/video modeling, low-rank matrix recovery, and deep neural networks.

YANDONG TANG (Member, IEEE) received the B.Sc. and M.Sc. degrees in mathematics from Shandong University, Jinan, China, in 1984 and 1987, respectively, and the Ph.D. degree in applied mathematics from the University of Bremen, Bremen, Germany, in 2002. He is currently a Professor with the Shenyang Institute of Automation, Chinese Academy of Sciences, Beijing, China. His current research interests include numerical computation, image processing, and computer vision.