Date of publication December 15, 2022.

CGAN-DA: A Cross-Modality Domain Adaptation Model for Hand-Vein Biometric-based Authentication

SHUQIANG YANG[1], YIQUAN WU[2], MOUNIM A. EL-YACOUBI[3], XIN JING [2],AND HUAFENG QIN.[2]

[1] The college of physical and electronic information, Luoyang Normal University, HeNan 471934, China
[2]The Chongqing Key Laboratory of Intelligent Perception and BlockChain Technology, Chongqing Technology and Business University, Chongqing 400067, China.

[3] SAMOVAR, Telecom SudParis, Institut Polytechnique de Paris, 91120 Palaiseau, France

Corresponding author: Huafeng Qin (e-mail: qinhuafengfeng@163.com).

This work was supported in part by National Natural Science Foundation of China under Grant 61976030, the fellowship of China Postdoctoral Science Foundation (Grant No. 59676651E), the Scientific Innovation 2030 Major Project for New Generation of AI under Grant NO. 2020AAA0107300, Ministry of Science and Technology of the People's Republic of China, the key project of Henan Province science and technology attack (Grant No. 222102210301), the funds for creative research groups of Chongqing Municipal Education Commission CXQT21034 under Grant CXQT21034, National Natural Science Foundation of Chongqing under Grant cstc2018jcyAX0057, Chongqing Talent Program under Grant CQYC201903246, and in part by the Scientific and Technological Research Program of Chongqing Municipal Education Commission under Grants KJQN201900848, KJQN201800814 and KJQN202000841.

ABSTRACT Palm-vein recognition has been the focus of large research efforts over the last years. However, despite the effectiveness of deep learning models, in particular Convolutional Neural Networks (CNNs), in automatically learning robust feature representations, thereby obtaining good accuracy, such good performance is usually obtained at the expense of annotating a large training dataset. Labeling vein images, however, is an expensive and tedious process. Although handcrafted schemes for data augmentation usually increase slightly performance, they are unable to cover complex variations inherently characterizing such images. To overcome this issue, we propose a new unsupervised domain adaptation model, called CycleGAN-based domain adaptation (CGAN-DA), that extracts discriminant representation from the palm-vein images, without requiring any image labeling. Our CGAN-DA models allows a conjoint adaptation, at the image and feature levels. Specifically, in order to enhance the extracted features' domain-invariance, image appearance is transformed across two domains, palm-vein domain and retinal domain. We employ several adversarial losses namely a segmentation loss and a cycle consistence loss to train our model without any annotation from the target domain (palm-vein images). Our experiments on the public CASIA palm-vein dataset demonstrates that our models significantly outperforms the start of the art in terms of verification accuracy.

INDEX TERMS Palm-vein Authentication, Domain Adaptation, Generative adversarial network, Convolutional Neural Network.

I. INTRODUCTION

With the rapid application of internet technology and the increasing trend of online fraud cases, the privacy and security have been received more and more attention. Traditional identity methods, such as passwords, ID card, and keys, suffer from the drawbacks: the ID card may be lost, the passwords may be forgotten or stolen. To solve this problem, biometric recognition technology has been widely investigated and become a hot research topic in past years. Biometric recognition technology refers to use human physiological or behavioral characteristics for personal identity authentication. Currently, commonly used biometric features can be divided into two categories: Biometric verification harnessing modalities such as fingerprint [1], iris [2] and face [3] is a mature technology reflected in the deployment of various solutions. These modalities, however, are not only easy to collect without user consent, but their fake versions have been seamlessly applied to spoof such biometric systems. Because intrinsic modalities, like palm-vein [4], [5], finger-vein [6], and dorsal hand-vein [7], are concealed beneath the skin and

are thus difficult to spoof, they have become the focus of wide research effort over the recent years. Intrinsic traits offer higher security and privacy in critical applications as the vein patterns are generally not visible in visible light. They can, nonetheless, be collected through infrared illumination with a wavelength of 850 nm [4], [6], [7]. An additional advantage is that several studies [8], [9] have demonstrated that the blood vessel network is unique for each individual [8], and can distinguish even identical twins [8], [10], which shows the high person specificity, of vein biometrics. These reasons explain why person authentication based on vein biometrics has been so widely investigated in the last decade [4], [6], [7], [11]–[23].

A. RELATED WORKS

Despite its appealing characteristics, vein recognition accuracy may be compromised by several factors affecting the captured image quality including lighting, temperature, acquisition device, as well as user habits [4], [6], [11], [24]–[26]. The factors above may produce noisy regions with irregular shadow that ultimately will make it hard to distinguish actual vein patterns from the background, decreasing thereby matching verification accuracy. To tackle these issues, several endeavors have been made to effectively segment the palm-vein patterns; We can split them into two main categories:

(1) Handcrafted texture extraction approaches: Observing that the cross-profile of vein patterns shows a valley-like shape, some researchers have proposed some mathematical models to detect these valleys. Line tracking methods [11], [12], and curvature-based measures [13]–[16] are some of these techniques. Other works, making the assumption that the vein vessels appear as line-like textures in a predefined neighborhood region, propose Gabor filters [6], [17], matched filters [18], wide line detector [19] and neural networks [20], to extract vein textures.

(2) Deep learning-based texture extraction methods: By contrast to handcrafted methods, segmentation methods relying on Deep Neural Networks (DNNs), in particular CNNs [21]–[23], [27], [28], make no *a priori* assumption on the vein pattern distribution, as they can extract vein patterns in an end-to-end manner. Qin et al. [21] were among the first to propose a CNN to predict the probability of each pixel to belong to a vein pattern. Subsequently, to correct mislabeled data, an iterative deep neural network [22] was proposed to extract the hand vein pattern. A generative adversarial network (GAN) [23] has also been proposed, for the same task, to extract the finger-vein texture.

The handcrafted methods [4], [6], [7], [11]–[20] mentioned above assume that the veins structures usually show valleys and line segment-like shapes. These $prior$ assumptions may not hold in real application, however, as the vein pixel values may generate more complicate shape distributions than ones above. The performance of such vein texture extraction methods, therefore, may be compromised. As deep learning-based methods [21]–[23] are able to automatically learn robust feature representation for the vein texture segmentation without requiring any **prior** assumption, they usually outperform handcrafted approaches in terms of verification accuracy. Some researchers have brought them into medical image segmentation tasks [29]–[32], such as brain segmentation, retina image segmentation, and neuronal membranes segmentation. DNNs like CNNs [31], [33]], DBNs (Deep Belief Networks) [34], and Auto-Encoders (AE) [33], when fed with a large annotated training dataset, can extract more robust features than handcrafted approaches. Unlike image segmentation in the medical field, however, pixel-level ground-truth labels in biometric are not available in general, in particular for vein imaging, which prevents from training pixel-based supervised training models to solve the segmentation problem. The lack fine-grained pixel annotations is explained by the huge cost for collecting and annotating pixel-wise large biometric datasets. To tackle this issue, some studies [21], [23] have proposed the use of several segmentation baseline models or their fusion to generate the vein and background labels at the pixel level, that serve to train DNNs which, in turn, seek to improve the initial vein segmentation performance. Although increasing the number of baselines to combine may lead to more accurate initial labels, these pseudo annotations may still include several mislabelled pixels. Additionally, a large number of initial baselines may be hard to obtain, especially if we bear in mind that the accuracy of each should be good enough to benefit the combination scheme. These reasons explain why existing DNNs used for vein segmentation [13], [23], may not be highly robust in real-life verification tasks. Owing to these issues, learning a more robust feature representation for vein segmentation is till a challenging problem.

Domain adaptation is a smart data generation technique that has been investigated in several medical image segmentation tasks [32], [35]–[37]. Domain adaptation is a peculiar scheme of transfer learning that leverage labeled data in one (or more) relevant source domain to carry out new tasks in a target domain. This allows DNNs to reach competitive accuracy on unlabeled target data, by exploiting annotations from the source domain only. Previous studies have categorized domain shift mainly into two classes: the first, image adaptation, seeks to align image appearance between domains with pixel-to-pixel transformation. This allows to alleviate domain shift at the input level for deep learning models. The second, unsupervised domain feature adaptation, seeks, by contrast, to extract domain invariant DNN features, irrespective of the difference between the input domains in terms of appearance.

As pointed out above, DNN-based techniques have obtained good performance for medical image segmentation, in tasks like MR prostate segmentation or retinal segmentation, based on annotated training data. As retinal images (source) and hand-vein images (target) both correspond to vessel images, and their domains are hence related, knowledge inferred from the retinal image segmentation task can be transferred to the hand-vein segmentation task. In real-life tasks, nonetheless, a domain shift or pixel distribution change

always exists between the two domains, which may hamper segmentation accuracy. Domain adaptation is able to tackle this problem by drawing the source and target domains closer than they were initially.

B. OUR WORK

In this paper, we propose a novel unsupervised domain adaptation model, called CycleGAN-based domain adaptation (CGAN-DA), which we apply to adapt domain shift for cross-modality vein image segmentation. Instead of a single adaptation procedure, we perform conjointly two adaptation procedures, namely image adaptation and feature adaptation (Fig.1). Our contributions can be summarized as follows: 1) Our proposal is the first t o a ccommodate domain adaptation to hand-vein texture segmentation for vein biometrics. Concretely, we propose, for the vein segmentation task, a CycleGAN-based domain adaptation scheme, named CGAN-DA, that conjointly combines image domain adaptation and feature domain adaptation. Unlike DCNN segmentation models, like CNN+FCN [21] and FV-GAN [23], that seek learning feature representations based on pixel labels output by verification baselines, our scheme reduces the gap between the source (retinal image) and target (hand-vein image) domains, so as the hand-vein patterns are extracted by leveraging learning knowledge inferred from the densely annotated retinal images. Consequently, our model is not related to the performance of the baselines. Our experiments demonstrate that our model significantly o utperforms the state of the art in terms of verification accuracy.

2) To extract robust vein patterns, We propose CGAN-DA, a CycleGAN-based domain adaptation (DA) approach leveraging a a cycle-consistency GAN and a GAN segmentation model. CGAN-DA performs a synergistic fusion of adaptations, both image based (retinal image to hand-vein image) and feature based (hand-vein image to hand-vein feature). To degrade domain shift, we first transform the annotated source images (real retinal images) into the appearance of images drawn from the target domain (hand-vein image), based on a generative adversarial network with cycle-consistency constraint. We then train a segmentation model on the generated target-like images, and perform feature adaptation to decrease further the remaining domain shift. In the proposed CGAN-DA scheme, the feature encoder is shared, to enable it conjointly transforming image appearance and extracting domain-invariant representations for the segmentation task.

3) We have performed rigorous experiments on a large public dataset to assess the performance of our vein segmentation model. The results we obtain show that the proposed model does not only effectively extract the vein patterns from raw hand vein images without any hand-vein pixel-based annotation, but it significantly outperforms the state of the art in terms of authentication accuracy.

II. THE PROPOSED MULTI-SCALE RECYCLE GAN METHOD

Domain adaptation operates under the assumption of known relationship between the target and source domains, and that this knowledge can be transferred in one step from the latter to the former. Owing to factors like image quality, illumination, physiology, pose and object shape, nonetheless, a domain shift, i.e. change in input distribution, always occurs between the two domains, implying thereby solely a low overlap between the two domains (retinal image and hand-vein image). Carrying out domain adaptation in one step for hand-vein segmentation is, therefore, doomed to failure. To tackle this issue, we propose a latent domain that relates the two domains by smoothly transferring knowledge for vein segmentation, whereby synthesized retinal images are stylized as hand-vein images. Fig. 1 shows a graphical representation of our unsupervised domain adaptation approach for hand-vein image segmentation. We perform our image and feature adaptations under the framework of a novel learning scheme that reduces the performance gap emanating from domain shift in an effective way.

A. IMAGE ADAPTATION

The retinal image dataset [38] consists of retinal images and their associated manual vasculature annotations, that produce reliable segmentation knowledge to be transferred to other domains like hand vein segmentation. Because of the large difference between source domain (original retinal images) and target domain (original hand-vein images), which may impair vein segmentation accuracy, we propose a cycle GAN that seeks minimizing the shift between these two domains. We denote $\{x_i^s, y_i^s\}_{i=1}^N$ as a set of annotated samples from source domain $\{X^s\}$ (retinal image) and $\{x_i^t\}_{j=1}^M$ as a set of unlabeled samples from target domain $\{X^t\}$ (hand-vein image), respectively. Our goal is to transform the source images x^s into the appearance of target ones x^t, knowing that the latter exhibits different visual appearance from the former due to domain shift. We reduce this shift between the source and target domains by image appearance alignment.

To this end, we consider Cycle GAN [39], a highly successful model in the pixel-to-pixel unpaired image transformation task, by designing two generators G and F and two discriminators D_s and D_t. The mapping function $G : x_s \rightarrow x_t$ aims at transforming the source images into target-like ones $G(x_s) = x_{s\rightarrow t}$ while D_t seeks to distinguish between the real target images x_t and translated images $x_{s\rightarrow t}$. In the target domain, a minimax two-player game is set based on G and D_t, which are trained according to the following object function optimization.

$$L_{adv}(G, D_t) = \mathbb{E}_{x_t\sim\mathbb{P}_g}[log[D_t(x_t)]] + \mathbb{E}_{x_s\sim\mathbb{P}_r}[log[1 - (D_t(G(x_s)))]] \quad (1)$$

It is worth noting that it is hard to train the classic GAN as the latter seeks minimizing the divergences that may be not continuous w.r.t the generator's parameters. To stabilize

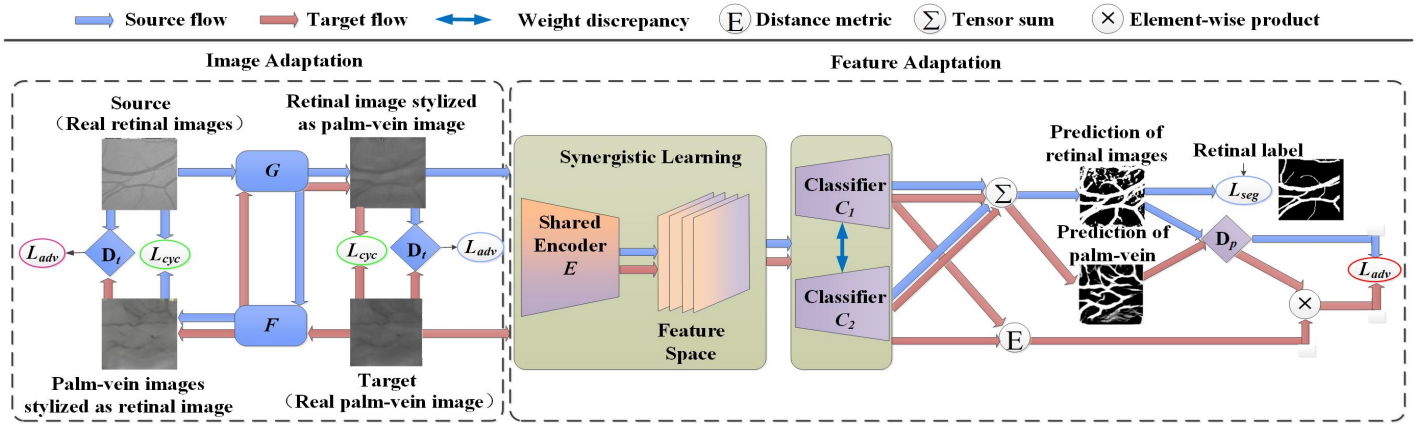

FIGURE 1. Framework of our unsupervised domain adaptation framework. Our image adaption schemes aims to transform the source images towards the appearance of target ones, while the feature adaption trains a network to learn an invariant-feature representation between the two domains for the segmentation task. The network for image adaption is a cycle-Consistent adversarial network, including two discriminators and two generators G and F. The generator G serves for the source-to-target image transformation and the generator F serves for the target-to-source image transformation. The network for feature adaption consists of an encoder E, two classifiers C_1 and C_2 and a discriminator D_p. Two feature maps from encoder E are input into the two classifiers C_1 and C_2 for image segmentation. The blue and red arrows indicate the source and target data flows, respectively. To further reduce the domain shift, the discrepancy of the two prediction maps is employed to produce a local alignment score map. This map evaluates the category-level alignment degree of each feature and is used to adaptively weight the raw adversarial loss map.

training, we consider the WGAN-GP loss [40] with gradient norm penalty for random samples $\hat{x} \sim \mathbb{P}_{\hat{x}}$ on both mapping functions for training. This loss is defined as:

$$L_{adv\_w}(G, D_t) = \mathbb{E}_{x_s \sim \mathbb{P}_g}[D_t(G(x_s))] - \mathbb{E}_{x_t \sim \mathbb{P}_r}[(D_t(x_t))] + \lambda \mathbb{E}_{\hat{x} \sim \mathbb{P}_{\hat{x}}}[(\|\Delta_{\hat{x}} D_t(\hat{x})\|_2 - 1)^2] \quad (2)$$

where $\mathbb{P}_r$ is the data distribution and $\mathbb{P}_g$ indicates the model distribution implicitly defined by $\widetilde{x} = G(x)$ and $x \sim p(x)$ (the input x of the generator is sampled from noise distribution). $\mathbb{P}_{\hat{x}}$ is a distribution sampling uniformly along straight lines between pairs of points sampled from $\mathbb{P}_r$ and $\mathbb{P}_g$, and λ is a penalty factor of gradient norm. The constraint is added as a penalty on the gradient norm in adversarial loss, which effectively alleviates the gradients vanishing nor exploding. Therefore, the WGAN-GP has better stability for training.

We also use a Cycle Consistency Loss as the regularization term to further reduce the space of possible mapping functions so that x_s is mapped into the source domain, i.e. $x_s \to G(x_s) \to F(G(x_s)) \approx x_s$. Similarly, for each image x_t from the target domain, G and F should also satisfy backward cycle consistency, i.e. $x_t \to F(x_t) \to G(F(x_t)) \approx x_t$. Then the pixel-wise cycle-consistency loss L_{cyc} is defined as follows

$$L_{cyc}(G, F) = \mathbb{E}_{x_t \sim \mathbb{P}_g}[||G(F(x_t) - x_t)||_1] + \mathbb{E}_{x_s \sim \mathbb{P}_r}[[||F(G(x_s) - x_s)||_1]] \quad (3)$$

Our image adaptation scheme allows transforming the source images x_s into target-like images $G(x_s) = x_{s \to t}$ with semantic contents preserved, based on the two losses defined above, i.e., adversarial loss and cycle-consistency loss. This pixel-to-pixel transformation has the potential to transform $G(x_s) = x_{s \to t}$ into the data distribution of the target domain. The resulting synthesized images can then be harnessed to train a neural network classifier for the segmentation task in the target domain.

B. FEATURE ADAPTATION

To attenuate the domain shift between the source and target domains, existing adaptation techniques transform the images from the source domain into realistic target-like images. Then we train a network based on these resulting target-like images to produce an effective pixel-to-pixel segmentation on the target data. Owing to the wide gap between the retinal and hand-vein image domains, however, such a domain adaptation is usually ineffective. To tackle this feature adaptation problem, we design an additional GAN that seeks mitigating the remaining shift between the synthesized target images $x_{s \to t}$ and the actual target ones x_t. To further bridge domain gap, we introduce Category-level Adversaries [36] to promote the alignment of the two domains' feature distributions, with the objective of enforcing local feature consistency during the global alignment process.

For the prediction of segmentation tasks from U, we propose, as shown in Fig. 1, a two-classes discriminator D_p to classify the outputs of U corresponding to $x_{s \to t}$ or x_t. Unspired by the standard co-training algorithm [41], the generator U is split into a feature extractor E and two classifiers C_1 and C_2, where E extracts features from input images and C_1 and C_2 predict the probability of the features generated from E belonging to predefined classes (background and vessel). As suggested in the co-training scheme [42], a cosine distance loss is considered to favor diversity of the weights of C_1 and C_2, which promotes distinct views/classifiers to make different predictions for vein features. Finally, the two diverse prediction tensors p_1 for C_1 and p_2 for C_2 are combined as feature map p for vein vessel prediction. Specifically, after

the features from the adapted image are extracted $x_{s\to t}$, the feature maps E($x_{s\to t}$) are input to the two classifiers C_1 and C_2 to output the pixel-level ensemble prediction $p = C_1(E(x_{s\to t})) + C_2(E(x_{s\to t}))$ for segmentation masks. For the sample pairs with the ground-truth label of $(x_{s\to t}, y_s)$, the segmentation loss is defined as:

$$L_{seg}(E, C_1, C_2) = H(y_s, p) + \beta Dice(y_s, p) \tag{4}$$

where the H represents cross-entropy loss, the second term is the Dice loss, and β is the trade-off hyper-parameter balancing them. The hybrid loss function is designed to solve the class imbalance in image segmentation.

As shown in the standard co-training algorithm [42], to provide two different feature representations, the two classifiers C_1 and C_2 should have possibly diverse parameters. Therefore, cosine similarity is employed to enforce the divergence of the weights of the convolutional layers of classifiers C_1 and C_2, which results in the following weight discrepancy loss:

$$L_W(C_1, C_2) = \frac{\overrightarrow{\Omega}_1 \cdot \overrightarrow{\Omega}_2}{\| \overrightarrow{\Omega}_1 \| \quad \| \overrightarrow{\Omega}_2 \|} \tag{5}$$

where $\overrightarrow{\Omega}_1$ and $\overrightarrow{\Omega}_2$ are produced by flattening and concatenating the weights of the convolution filters of C_1 and C_2 .

As shown in Fig 1, a binary discriminator D_p is employed to classify the outputs. The discrepancy between predictions p_1 and p_2 is employed to weight the adversarial loss. Therefore, the traditional adversarial loss is changed to obtain

$$\begin{aligned} L_{adv}(U, D_p) = -\mathbb{E}_{x_{s\to t}\sim\mathbb{P}_o}[log[D_p(U(x_{s\to t}))]] - \\ \mathbb{E}_{x_t\sim\mathbb{P}_r}[(\gamma\mathbb{C}(p_1, p_2) + \epsilon) log[1 - (D_p(U(x_t)))]] \end{aligned} \tag{6}$$

where p_1 and p_2 are predictions obtained from C_1 and C_2, respectively, $\mathbb{C}(x, y)$ denotes the cosine distance between x and y, and the parameter ϵ is the weight for the adversarial loss. We experimentally fix γ and ϵ to 10 and 0.4. First, we select initial values and then train our model for domain adaption. The results for image adaption and feature adaption are analyzed. If continuous and smooth vein patterns are extracted, the corresponding values are determined as optimal values. Otherwise, we increase or reduce the values of γ and ϵ with an interval of 2 and 0.1 until the vein patterns are effectively segmented from the original images.

C. SYNERGISTIC LEARNING

The image adaptation and feature adaption are fused to obtain a synergistic learning diagram for vein segmentation. Given the above loss terms, the overall objective of our framework can be written as

$$\begin{aligned} L = L_{adv}(G, D_s, x_t, x_s) + L_{adv}(G, D_t, x_s, x_t) + \\ \lambda_{cyc} L_{cyc}(G, F) + \lambda_{seg} L_{seg}(E, C_1, C_2) + \\ \lambda_W L_W(C_1, C_2) + \lambda_{adv} L_{adv}(U, D_p) \end{aligned} \tag{7}$$

where λ_{cyc},λ_{seg},λ_W,λ_{adv} are trade-off parameters controlling the importance of each term. In the experiments, the values of these parameters are 10, 1, 0.01, and 0.001, respectively.

FIGURE 2. The retinal image samples from the DRIVE dataset: (a) original image, (b) manual segmentation vasculature and (c) mask.

D. NETWORK ARCHITECTURE OF THE MODULES

This section details the network architecture of each module in the proposed framework, as shown in Fig.1. As the CycleGAN [39] has shown promising results for neural style transfer and object transfiguration, we adopt it for image adaptation to minimize the gap between the source and the target domains. The generative network consists of 3 convolutional layers with 64, 128 and 256 feature maps in each layer, 9 residual blocks [43] with 256 feature maps in each layer, and three fractionally convolutions with 256, 128 and 64 feature maps in each layer. For the first three convolutional layers, the kernel sizes are 7×7, 3×3, and 3×3 and the last three layers and the residual blocks employ 3×3 convolutional kernels for feature extraction. The ReLU is used as activation function. For the discriminator networks D_t and D_s, we design 70×70 PatchGANs [39], [44], that predict whether the 70×70 overlapping image patches are real or fake. The networks includes 5 convolutional layers with 64, 128, 256, 512 and 1 feature maps for each layer, respectively. We employ the convolutional kernels with a size of $4\times$ 4 and a stride of 2 in the first three layers and $2\times$ 2 convolutional kernels with a stride of 1 in last two layers, respectively. For the first four layers, each convolutional layer is followed by a layer normalization and a active function of leaky ReLU with parameter 0.2. As mentioned in [39], [44], there are fewer parameters in the patch-level discriminator architecture, compared to a full-image discriminator. This discriminator can be applied to any image with arbitrary size in a fully convolutional manner [44].

For feature adaption, the network architecture comprises a generator G and a discriminator D. G consists of one feature encoder E and two classifiers C_1 and C_1. E extracts features from input images and C_1 and C_1 classify the features from E into one of the predefined semantic classes. E comprises 32 residual bottlenecks, each consisting of three convolutional layers, followed by three batch normalization layers. The two classifiers C_1 and C_1 have the same network architecture with four convolutional layers. We forward an image into the generator G which outputs two feature maps. The discriminator has a typical CNN architecture that takes an input image of size 128×128 and outputs one decision: is this image sample of the the source domain or is it from the target domain? In this network, there are five convolutional layers with a kernel size of 4×4. To reduce spatial dimensionality and computation cost, 2 stride convolutions instead of pooling layers are employed to each convolution layer. The Leaky ReLU activation function is employed for all convolutional layers except the output layer.

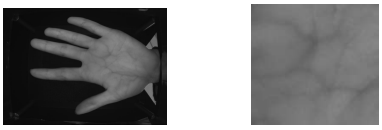

FIGURE 3. The palm-vein samples: (a) original image, and (b) ROI.

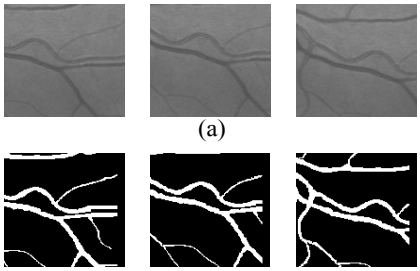

FIGURE 4. Generated retinal patches: (a) Three cropped retinal patches and (b) corresponding ground-truth vasculatures

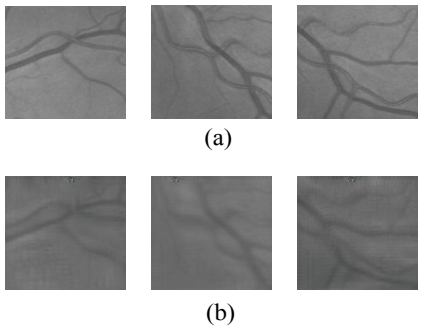

FIGURE 5. The result for image adaption. (a) The retinal patches and (b) The synthesised retinal patches stylized as palm-vein images

III. EXPERIMENT AND RESULTS

To test our approach, we have carried out rigorous experiments on a known public hand-vein dataset. All retinal images with ground truth from dataset [38] and half of hand-vein images from the hand-vein dataset are employed to train the proposed model, as shown in Fig.1. For testing, we take a hand-vein image as input of our trained model that outputs a probability map with the same size. The resulting map is further encoded using a threshold of 0.5. We match the vein patterns stored in the resulting binary image for verification by matching approach [21]. We compare our approach with the Maximum principle curvature [16], Repeated line tracking [11], Gabor filters [6], Hessian phase [4], and CNN [21] techniques. For CNN training, as described in [21], we label the palm-vein pixels based on several handcrafted segmentation methods [6], [11], [16] and construct a training set accordingly. The performances of all methods are shown in the following experiments.

A. DRIVE DATASET

The retinal image set named DRIVE [38] contains 40 retinal images, which have been divided into training and test sets. For each training image, there is a single manual segmentation vasculature. For the test images, two ground truth vasculatures are obtained by manual labeling. In addition, a mask image is available for each retinal image to extract the region of interest. The original images are in RGB with a resolution of 768 by 584 pixels (Fig. 2).

B. CASIA DATASET

The CASIA Multi-Spectral Palm-print Image Dataset [45] consists of 7200 palm images from 100 different people. The images, collected by a self-designed multiple spectral imaging device, are captured through six different wavelength bands, in two separate sessions. For each session, each subject provided left and right hands and 3 image samples are collected from each hand. So, there are 12 (3 images $\times$ 2 hands $\times$ 2 sessions) images from one subject. As our work focuses on hand-vein verification, the images collected under the 850 nm wavelength are employed in our experiments. Totally, there are 1200 images (100 subjects $\times$ 2 hands $\times$ 2 sessions $\times$ 3 images) from 100 subjects. To facility matching, the region of interest (ROI) image is extracted by a preprocessing method [22] and the resulting images are further subject to normalization. We obtain, in this way, images with a resolution of 128 $\times$ 128, as shown in Fig.3.

C. EXPERIMENTAL SETUP

There is a large vein pattern difference between the retinal image obtained in the two datasets (as shown in Fig.2 (a) and Fig.3(a)). To reduce the gap, all retinal images are transformed into gray-scale images and we randomly crop the resulting images to 100×100 and scale them to 128×128 so that the width of the veins in the two images become similar (as shown in Fig.3(b) and Fig.4(a)). In this way, we have generated 9000 retinal patches (Fig.4(a)) and their corresponding ground truth images (Fig.4(b)) from 40 original retinal images with manually labeled vasculatures images, respectively. For each retinal image, we have generated about 225 cropped images. For CASIA, half of the palm-vein images, i.e. 600 images (50 subjects $\times$ 2 hands $\times$ 2 sessions $\times$ 3 images) are employed for training and the remaining for test. As a result, the training set includes 9000 DRIVE retinal patches along with their 9000 ground truth patches and 600 CASIA palm-vein images for our model training. The performance of all methods is evaluated on the test set with 600 palm-vein images.

D. VISUAL ASSESSMENT

In this section, we visually analyze and assess the performance of various approaches, so that more insights into our proposed approach can be derived. Our approach consists of image adaption and feature adaption. The image adaption model aims at aligning the image appearance between different domains with pixel-to-pixel transformation, in such a way that the distribution of retinal images becomes similar

to the palm-vein images' distribution. For feature adaption, our model aims to extract the vein network from palm-vein images after the appearances between the input domains becomes similar. Fig.5 shows the experimental results for image adaption. From Fig.5 and Fig.3(b), we observe that the retinal images have been transferred into the style of palm-vein images, which reduces the domain shift between the two domains. In particular, the generated images are more blurred than the original retinal images. Also, the distribution style of the vascular veins in the synthesised images is more similar to the ones in the palm-vein images. Fig.6 shows the extracted vein networks from a palm-vein image by the benchmarking approaches and ours. Compared to existing approaches, our approach, without using any annotation from the target domain (palm-vein), achieves better segmentation results, as the vein patterns are more connected, smooth and less noisy. By contrast, there are more noise and isolated regions in the segmented images obtained from Repeated line tracking, Hessian phase and Maximum principle curvature. The reason may be that these three approaches compute the curvature of the cross-sectional profile to detect the vein patterns. However, the curvature is usually sensitive to noise, which may affect the quality of the segmentation results. Gabor filters can extract more connected and smooth vein patterns but may generate over-segmentation regions where some non-vein regions are mislabeled as vein patterns. The CNN is capable of learning robust a feature representation. However, similar to Repeated line tracking, Hessian phase and Maximum principle curvature, the vein patterns show noise poor connectivity for the CNN-based approach. Such results may be attributed to the following facts. The ground truths are generated based on four baselines according to a voting scheme (a threshold of 0.5 for each pixel). Some incorrect vein pixels in the resulting ground truths may occur accordingly. Therefore, the CNN model [21] trained on such ground-truth may achieve weak segmentation performance on similar patches. Compared to the CNN model [21], our approach learns rich segmentation knowledge based on the retinal dataset where accurate ground truths are provided for segmentation. As a result, by minimizing the gap between the two domains (palm-vein and retinal images), our approach improves domain-invariance of the extracted features towards the segmentation task.

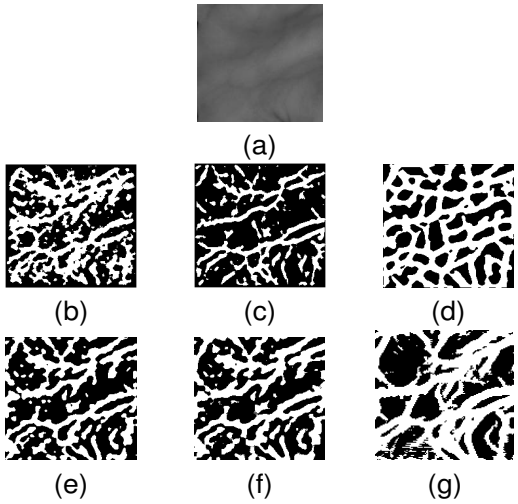

FIGURE 6. Segmentation results of the different approaches. (a) Original palm-vein image; (b) vein patterns extracted from (a) using Repeated line tracking; (c) vein patterns extracted from (a) using Maximum principle curvature; (d) vein patterns extracted from (a) using Gabor filter; (e) vein patterns extracted from (a) using Hessian phase; (f) vein patterns extracted from (a) using CNN; and (g) vein patterns extracted from (a) using the proposed approach.

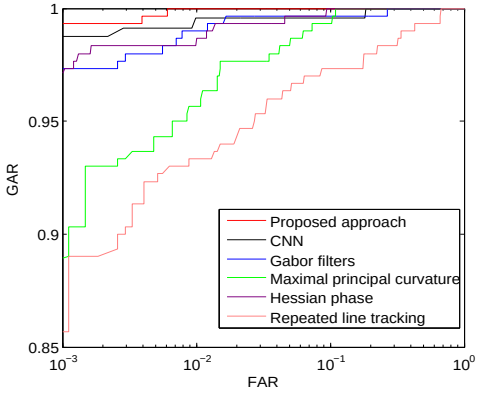

FIGURE 7. Receiver operating characteristics of various methods.

E. VERIFICATION RESULTS

We have performed extensive experiments to verify the verification performance of the proposed approach on the CASIA dataset; collected from both sessions. First, the deep learning based approach, i.e CNN [21], and hand-crafted approaches [4], [6], [11], [16] are employed to segment the vein texture. Then, the resulting binary vein image is matched by the method in [21]. The test set includes 600 vein images (50 subjects × 2 hands × 2 sessions × 3 images) associated with 100 hands. All images are captured in two separate sessions. In our experiments, the first 3 h and-vein images from the first session are selected as training data and the remaining 3 images from the second session are used as testing data, respectively. The genuine scores are produced by matching the images from the same hands. Similarly, we match different hands to compute the impostor scores. In this way, 300 (100 × 3) genuine scores and 178200 (6 × 6 × 100 × 99 / 2) imposter scores are obtained, respectively. Computing such impostor matching scores, however, is time consuming. To reduce computation cost, similar to work [22], we randomly split all hands into 10 groups and then compute the impostor matching scores of each group, respectively. For each group, we calculate the matching scores of the first three images against the remaining three images to generate 270 impostor scores. In this way, we obtain 2700 (270 × 10 groups) imposter scores for the 10 groups. The False Rejection Rate (FRR) and the False Acceptance Rate (FAR) are computed according to the genuine scores and the impostor scores. The Equal Error Rate (EER) is the error

TABLE 1. Result (%) of different methods

Method	EER
Repeated line tracking [11]	4.00
Maximum principle curvature [16]	2.33
Gabor filters [6]	1.00
Hassian phase [4]	1.33
CNN [21]	0.74
The proposed approach	0.52

rate when FAR is equal to FRR. Table 1 has shows the verification accuracy of the various approaches described in the previous section and the corresponding receiver operating characteristics (ROC) curve (the FAR against the the Genuine acceptance rate (GAR=1-FRR)) is illustrated in Fig.7.

The experimental results (Table 1 and Fig.7) demonstrate that our approach achieves the lowest EER and outperforms all the benchmarking approaches considered in our work. The handcrafted approaches, e.g. Repeated line tracking, Maximum principle curvature, Gabor filters, Hessian phase approaches and CNN achieve high verification errors, i.e. 4.00% EER, 2.33% EER, 1.00% EER and 1.33 % EER, respectively, while our model obtains the lowest verification error, i.e. 0.52% on the CASIA dataset. The corresponding ROC curve (as shown in Fig.7) also shows that our model obtains the highest GAR at different FAR regions, compared to existing works. Overall, the deep learning based models (i.e. CNN and our GAN-based approach) obtain much lower verification error w.r.t the segmentation approaches based on handcrafted feature extraction. Such a performance may be explained by the fact that latter approaches leverage image processing to explicitly extract low level features (e.g. edge) based on descriptors defined according to human expertise. This may overlook, however, some key information which is related to vein segmentation. The deep learning models, by contrast, objectively learn high-level features directly related to discriminating vein patterns in end-to-end way without any human's feature selection, based on the loss function and back propagation. Also, we observe furthermore that our CGAN-DA model outperforms CNN in terms as it significantly reduces the verification error. This may be attributed to the training of the CNN based on ground truth produced by four automatic vein segmentation baselines [4], [6], [11], [16], that might provide, therefore, some incorrect labels as ground truth to the CNN, and this may occur even when the four baselines are combined, as shown in [21]. Therefore, training the CNN on such incorrect labels degrades inevitably the overall performance. Our model avoids this issue by extracting the vein patterns with no palm vein ground truth, and significantly outperforms the CNN-based approach (33% decrease of the EER). The reason is twofold/ First, the retinal images rely on accurate annotations made by domain experts. Second, our model manages to reduce the domain shift, harnessing thereby the rich segmentation knowledge from retinal images and transferring it to extract the hand-veins in an effective way. Overall, the proposed approach is capable of reducing the shift between the source and target domain and improve the accuracy of vein recognition. The image prepossessing of patch division is also important because the thickness of vascular in eyes and hands are different, which can not been scaled by the domain adaption model.

IV. CONCLUSIONS

This paper proposes a novel domain adaption approach to extract the palm-vein texture based on retinal vascular ground truth instead of any palm-vein pixel annotation. Our model is able to conjointly reduce the appearance shift across the two domains by image adaption and minimize domain-invariant feature learning by feature adaptation. The two adaptive schemes are conducted by adversarial learning to exploit their mutual benefits for reducing domain shift. We test our method on unpaired retinal images to palm-vein ones by minimizing domain shift from medical images to biometric ones. The experimental results on a large public dataset shows that the proposed approach outperforms various benchmarking palm-vein segmentation methods, and achieves state-of-the-art verification results.

Although the verification performance achieved by our approach for hand-vein verification outperforms the state of the art, several improvements can be envisaged in the future. First, we can employ our model as baseline to generate ground-truth vein images in order to train existing deep learning models for vein segmentation. Second, we will carry out more experiments on other large vein datasets to further verify the efficiency of our model. Third, we currently encode the output of deep learning models with a threshold of 0.5 to extract the vein patterns from the background. Such a threshold may not be suitable for vein encoding as it is not objectively related to the reduction of verification error but determined by our *prior* knowledge. A search approach will be investigated to find the optimal threshold for verification. Fourth, we will explore more domain adaption approaches to vein pattern segmentation.

REFERENCES

[1] A. Jain, L. Hong, and R. Bolle, "On-line fingerprint verification," IEEE trans. pattern analysis and machine intelligence, vol. 19, no. 4, pp. 302–314, 1997.
[2] J. Daugman, "How iris recognition works," IEEE Trans. on Circuits and Systems for Video Technology, vol. 14, no. 1, pp. 21–30, 2004.
[3] M. A. Turk and A. P. Pentland, "Face recognition using eigenfaces," IEEE Computer Society, pp. 586–587, 1991.
[4] Y. Zhou and A. Kumar, "Human identification using palm-vein images," IEEE trans. on information forensics and security, vol. 6, no. 4, pp. 1259–1274, 2011.
[5] H. Qin, M. El Yacoubi, Y. Li, and C. Liu, "Multi-scale and multi-direction gan for cnn-based single palm-vein identification," IEEE Trans. on Information Forensics and Security, vol. PP, pp. 1–1, 02 2021.
[6] A. Kumar and Y. Zhou, "Human identification using finger images," IEEE Trans. on image processing, vol. 21, no. 4, pp. 2228–2244, 2011.
[7] A. Kumar and K. V. Prathyusha, "Personal authentication using hand vein triangulation and knuckle shape," IEEE Transactions on Image processing, vol. 18, no. 9, pp. 2127–2136, 2009.
[8] T. Tanaka and N. Kubo, "Biometric authentication by hand vein patterns," vol. 1, pp. 249–253, 2004.

[9] A. K. Jain, A. Ross, and S. Prabhakar, "An introduction to biometric recognition," IEEE Transactions on circuits and systems for video technology, vol. 14, no. 1, pp. 4–20, 2004.
[10] M. Shahin, A. Badawi, and M. Kamel, "Biometric authentication using fast correlation of near infrared hand vein patterns," Int. Journal of Biological and Medical Sciences, vol. 2, no. 3, pp. 141–148, 2007.
[11] N. Miura, A. Nagasaka, and T. Miyatake, "Feature extraction of finger-vein patterns based on repeated line tracking and its application to personal identification," Machine vision and applications, vol. 15, no. 4, pp. 194–203, 2004.
[12] H. Qin, L. Qin, and C. Yu, "Region growth-based feature extraction method for finger-vein recognition," Optical Engineering, vol. 50, no. 5, p. 057208, 2011.
[13] H. Qin, X. He, X. Yao, and H. Li, "Finger-vein verification based on the curvature in radon space," Expert Systems with Applications, vol. 82, pp. 151–161, 2017.
[14] L. Yang, G. Yang, Y. Yin, and X. Xi, "Finger vein recognition with anatomy structure analysis," IEEE Transactions on Circuits and Systems for Video Technology, vol. 28, no. 8, pp. 1892–1905, 2017.
[15] H. Qin, L. Qin, L. Xue, X. He, C. Yu, and X. Liang, "Finger-vein verification based on multi-features fusion," Sensors, vol. 13, no. 11, pp. 15 048–15 067, 2013.
[16] Miura, Naoto and Nagasaka, Akio and Miyatake, Takafumi, "Extraction of finger-vein patterns using maximum curvature points in image profiles," IEICE TRANSACTIONS on Information and Systems, vol. 90, no. 8, pp. 1185–1194, 2007.
[17] C.-B. Yu, H.-F. Qin, Y.-Z. Cui, and X.-Q. Hu, "Finger-vein image recognition combining modified hausdorff distance with minutiae feature matching," Interdisciplinary Sciences: Computational Life Sciences, vol. 1, no. 4, pp. 280–289, 2009.
[18] S. Chaudhuri, S. Chatterjee, N. Katz, M. Nelson, and M. Goldbaum, "Detection of blood vessels in retinal images using two-dimensional matched filters," IEEE Transactions on medical imaging, vol. 8, no. 3, pp. 263–269, 1989.
[19] B. Huang, Y. Dai, R. Li, D. Tang, and W. Li, "Finger-vein authentication based on wide line detector and pattern normalization," pp. 1269–1272, 2010.
[20] Z. Zhang, S. Ma, and X. Han, "Multiscale feature extraction of finger-vein patterns based on curvelets and local interconnection structure neural network," vol. 4, pp. 145–148, 2006.
[21] H. Qin and M. A. El-Yacoubi, "Deep representation-based feature extraction and recovering for finger-vein verification," IEEE Trans. on Information Forensics and Security, vol. 12, no. 8, pp. 1816–1829, 2017.
[22] 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. 34 823–34 837, 2019.
[23] W. Yang, C. Hui, Z. Chen, J.-H. Xue, and Q. Liao, "Fv-gan: Finger vein representation using generative adversarial networks," IEEE Trans. Information Forensics and Security, vol. 14, no. 9, pp. 2512–2524, 2019.
[24] H. Qin and M. A. El-Yacoubi, "Finger-vein quality assessment based on deep features from grayscale and binary images," Int. Journal of Pattern Recognition and Artificial Intelligence, vol. 33, no. 11, p. 1940022, 2019.
[25] H. Qin and M. A. El Yacoubi, "Finger-vein quality assessment by representation learning from binary images," in Neural Information Processing. Springer International Publishing, 2015, pp. 421–431.
[26] H. Qin, C. Gong, Y. Li, X. Gao, and M. A. El-Yacoubi, "Label enhancement-based multiscale transformer for palm-vein recognition," IEEE Transactions on Instrumentation and Measurement, vol. 72, pp. 1–17, 2023.
[27] H. Qin and M. A. El-Yacoubi, "End-to-end generative adversarial network for hand-vein recognition," pp. 47–60, 2022.
[28] H. Qin and P. Wang, "Finger-vein verification based on lstm recurrent neural networks," Applied Sciences, vol. 9, no. 8, p. 1687, 2019.
[29] Y. Zhang, L. Yang, J. Chen, M. Fredericksen, D. P. Hughes, and D. Z. Chen, "Deep adversarial networks for biomedical image segmentation utilizing unannotated images," International conference on medical image computing and computer-assisted intervention, pp. 408–416, 2017.
[30] Q. Dou, C. Ouyang, C. Chen, H. Chen, and P.-A. Heng, "Unsupervised cross-modality domain adaptation of convnets for biomedical image segmentations with adversarial loss," preprint arXiv:1804.10916, 2018.
[31] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," Int. Conference on Medical image computing and computer-assisted intervention, pp. 234–241, 2015.
[32] C. Chen, Q. Dou, H. Chen, J. Qin, and P. A. Heng, "Synergistic image and feature adaptation: Towards cross-modality domain adaptation for medical image segmentation," Proceedings of the AAAI Conference on Artificial Intelligence, vol. 33, pp. 865–872, 2019.
[33] D. Ciresan, A. Giusti, L. M. Gambardella, and S. Juergen, "Deep neural networks segment neuronal membranes in electron microscopy images," Advances neural information processing systems, pp. 2843–2851, 2012.
[34] H. Jang, S. M. Plis, V. D. Calhoun, and J. H. Lee, "Task-specific feature extraction and classification of fmri volumes using a deep neural network initialized with a deep belief network: Evaluation using sensorimotor tasks," Neuroimage, vol. 145, no. Pt B, 2016.
[35] D. Zhang, G. Huang, Q. Zhang, J. Han, and Y. Yu, "Cross-modality deep feature learning for brain tumor segmentation," Pattern Recognition, vol. 110, no. 11, p. 107562, 2020.
[36] Y. Luo, L. Zheng, G. Tao, J.-q. Yu, and Y. Yang, "Taking a closer look at domain shift: Category-level adversaries for semantics consistent domain adaptation," IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2502–2511, 06 2019.
[37] H. Tong, C. Shen, T. Zhi, G. Dong, and Y. Yan, "Knowledge adaptation for efficient semantic segmentation," IEEE CVPR, 2019.
[38] J. Staal, M. Abramoff, M. Niemeijer, M. A. Viergever, and B. V. Ginneken, "Ridge-based vessel segmentation in color images of the retina." IEEE Trans. Medical Imaging, vol. 23, no. 4, pp. 501–509, 2004.
[39] J. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in 2017 IEEE Int. Conference on Computer Vision (ICCV), 2017, pp. 2242–2251.
[40] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. C. Courville, "Improved training of wasserstein gans," 2017, pp. 5767–5777.
[41] Z.-H. Zhou and M. Li, "Tri-training: Exploiting unlabeled data using three classifiers," IEEE Transactions on knowledge and Data Engineering, vol. 17, no. 11, pp. 1529–1541, 2005.
[42] Z. H. Zhou and M. Li, "Tri-training: exploiting unlabeled data using three classifiers," IEEE Transactions on Knowledge and Data Engineering, vol. 17, no. 11, pp. 1529–1541, 2005.
[43] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770–77.
[44] P. Isola, J. Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in IEEE Conference on Computer Vision Pattern Recognition, 2016.
[45] (2005) Casia ms palmprint v1 database. [Online]. Available: [Available: http://www.cbsr.ia.ac.cn/MSPalmprint

SHUQIANG YANG received Bsc from Luoyang Normal Uinversity in 2000, and received Msc from Chongqing University of Technology in 2009. Now he is an associate professor in School of Physics and Electronic Information, Luoyang Normal University. His main research fields and specialties are signal acquisition processing, IOT wireless network, automotive electronics, automatic control and biometric recognition.

YIQUAN WU received his Bachelor's degree in Computer Science and Technology from Chongqing Technology and Business University. He is pursuing his Master's degree in Chongqing Key Laboratory of Intelligence Perception and BlockChain Technology at Chongqing Technology and Business University, China. His research interests include machine learning and palm-vein identification.

MOUNIM_AEL-YACOUBI PhD, University of Rennes, France, 1996) was with the Service de Recherche Technique de la Poste (SRTP) at Nantes, France, from 1992 to 1996, where he developed software for Handwritten Address Recognition that is still running in Automatic French mail sorting machines. He was a visiting scientist for 18 months at the Centre for Pattern Recognition and Machine Intelligence (CENPARMI) in Montreal, Canada, and then an associated professor (1998-2000) at the Catholic University of Parana (PUC-PR) in Curitiba, Brazil. From 2001 to 2008, he was a Senior Software Engineer at Parascript, Boulder (Colorado, USA), a world leader company in automatic processing of handwritten and printed documents (mail, checks, forms), for which he developed real-life software for address and check recognition. Since June 2008, he is a Professor at Telecom SudParis, University of Paris Saclay. His main interests include Machine Learning, Human Gesture and Activity recognition, Human Robot Interaction, Video Surveillance and Biometrics, Information Retrieval, and Handwriting Analysis and Recognition.PhD, University of Rennes, France, 1996) was with the Service de Recherche Technique de la Poste (SRTP) at Nantes, France, from 1992 to 1996, where he developed software for Handwritten Address Recognition that is still running in Automatic French mail sorting machines. He was a visiting scientist for 18 months at the Centre for Pattern Recognition and Machine Intelligence (CENPARMI) in Montreal, Canada, and then an associated professor (1998-2000) at the Catholic University of Parana (PUC-PR) in Curitiba, Brazil. From 2001 to 2008, he was a Senior Software Engineer at Parascript, Boulder (Colorado, USA), a world leader company in automatic processing of handwritten and printed documents (mail, checks, forms), for which he developed real-life software for address and check recognition. Since June 2008, he is a Professor at Telecom SudParis, University of Paris Saclay. His main interests include Machine Learning, Human Gesture and Activity recognition, Human Robot Interaction, Video Surveillance and Biometrics, Information Retrieval, and Handwriting Analysis and Recognition.(

XIN JING eceived his Bachelor's degree in Internet of Things Engineering from Pass college of Chongqing Technology and Business University. He is pursuing his Matser's degree in Chongqing Key Laboratory of Intelligent Perception and BlockChain Technology at Chongqing Technology and Business University,China. His research interests include finger-vein identification,Deep Learning and Data Augmentation.eceived his Bachelor's degree in Internet of Things Engineering from Pass college of Chongqing Technology and Business University. He is pursuing his Matser's degree in Chongqing Key Laboratory of Intelligent Perception and BlockChain Technology at Chongqing Technology and Business University,China. His research interests include finger-vein identification,Deep Learning and Data Augmentation.r

HUAFENG QIN received B.Sc in School of Mathematics and Physics and M.Eng in College of Electronic and Automation from Chongqing University of Technology and a Ph.D. degree in College of Opto-Electronic Engineering from Chongqing University. He was a visiting student for 12 months at Nanyang Technological University and then a postdoctoral researcher for two years at Universit´e Paris-saclay. Currently, he is a professor with the Chongqing Key Laboratory of Intelligent Perception and BlockChain Technology at Chongqing Technology and Business University, China. His research interests include Biometrics (e.g vein, face and gait) and machine learning.