Finger Vein Verification Algorithm Based on Fully Convolutional Neural Network and Conditional Random Field

JUNYING ZENG[ID][1], **FAN WANG**[ID][1], **JIANXIANG DENG**[ID][1], **CHUANBO QIN**[ID][1],
YIKUI ZHAI[ID][1,2], **(Member, IEEE), JUNYING GAN**[ID][1],**(Member, IEEE),**
AND VINCENZO PIURI[ID][2], **(Fellow, IEEE)**
[1]Department of Intelligent Manufacturing, Wuyi University, Jiangmen 529020, China
[2]Departimento di Information, Universita Degli Studi di Milano, 20133 Milano, Italy

Corresponding author: Chuanbo Qin (tenround@163.com)

ABSTRACT Owing to the complexity of finger vein patterns in shape and spatial dependence, the existing methods suffer from an inability to obtain accurate and stable finger vein features. This paper, so as to compensate this defect, proposes an end-to-end model to extract vein textures through integrating the fully convolutional neural network (FCN) with conditional random field (CRF). Firstly, to reduce missing pixels during ROI extraction, the method of sliding window summation is employed to filter and adjusted with self-built tools. In addition, the traditional baselines are endowed with different weights to automatically assign labels. Secondly, the deformable convolution network, through replacing the plain counterparts in the standard U-Net mode, can capture the complex venous structural features by adaptively adjusting the receptive fields according to veins' scales and shapes. Moreover, the above features can be further mined and accumulated by combining the recurrent neural network (RNN) and the residual network (ResNet). With the steps mentioned above, the fully convolutional neural network is constructed. Finally, the CRF with Gaussian pairwise potential conducts mean-field approximate inference as the RNN, and then is embedded as a part of the FCN, so that the model can fully integrate CRF with FCNs, which provides the possibility to involve the usual back-propagation algorithm in training the whole deep network end-to-end. The proposed models in this paper were tested on three public finger vein datasets SDUMLA, MMCBNU and HKPU with experimental results to certify their superior performance on finger-vein verification tasks compared with other equivalent models including U-Net.

INDEX TERMS CRF, FCN, finger-vein verification, U-Net.

I. INTRODUCTION

In recent years, biometric technology has been gaining more and more attention from the public due to the ascendingly demanding requirement in safety and accuracy of biometric system. At present, various biological characteristics such as fingerprints [1], palm-print [2], finger-vein [3], [4], hand-vein [5], palm-vein [6], face [7], iris [8], voice [9], gait [10], signa-

ture [11] have been included in recognition and verification. Among these choices, finger vein proves to be the current research hotspot due to its enticing strengths, such as non-contact collection, liveness detection, anticounterfeit and low cost. Nevertheless, finger vein verification is simultaneously confronted with rigorous challenges. In practice, variables such as environmental illumination [12], [13], ambient temperature [3], [13], [14], light scattering [15] and user behavior [12] and so on can abate the contrast of finger vein images, and thus creating ambiguous areas between vein and non-vein

The associate editor coordinating the review of this manuscript and approving it for publication was Chenguang Yang[ID].

regions. Generally, veins and backgrounds in these regions show worse separation. Therefore, it is extremely hard to extract accurate vein features with mathematical model, even with manual label on the actual veins.

Recent studies [16] have shown that the depth features learned from the deep convolutional neural network (DCNN) possess better generalization and expression ability, which prompts us to utilize the DCNN to extract the features of finger vein patterns. Therefore, this paper proposes a new fully convolutional neural network, which makes full use of the advantages of recurrent neural network [17], residual network [18], deformable convolution network [19], U-Net [20] and conditional random field [21], also with comprehensive consideration for the label consistency, spatial dependence of pixel prediction and complex venous shapes, to achieve accurate and effective segmentation in vein textures.

The main contributions of this paper are as follows:

(i). We proposed an improved ROI extraction method and venous pixel labeling approach. For the image with missing pixels, the method of sliding window summation is employed to filter and adjusted manually with self-built tools. Besides, the traditional baselines are endowed with different weights to automatically assign labels.

(ii). We presented a new fully convolutional neural network. Deformable convolution and residual recurrent convolution are introduced based on the fundamental structure of U-Net model, through which, not only the receptive fields can be adjusted adaptively according to the size and shape of the veins, but also the model can provide better depth features through effective feature accumulation.

(iii). We integrated the fully convolutional neural network with conditional random field, embedded as a part of fully convolutional neural network to realize end-to-end training, which aims to refine the rough pixel level label prediction in the fully convolutional neural network.

The remainder of this paper is organized as follows: Section 2 grants an introduction to the related work and motivation while Section 3 elaborates the proposed methods. Section 4 explains the experiments conducted, followed by the experimental results and analysis presented in section 5. Ultimately, a conclusion of this paper and future work in this field are summarized in section 6.

II. RELATED WORK AND MOTIVATION

In 2000, Japanese scholar Kono *et al.* pioneered with an effective feature extraction method of finger vein recognition, which could gather finger vein images with near-infrared light [22]. However, because of the uneven distribution of the brightness, the subtle contrast between the vein and the background, the high complexity of the venous structures, the noise pollution in the obtained finger vein images, coupled with the inevitable influence of individual differences [23], the finger vein verification remains a challenging task. Therefore, researchers are faced with the task to develop effective methods to extract appropriate features.

In response to this problem, segmentation-based methods from myriad of researchers consecutively sprung up to extract vein texture features for finger-vein recognition. It can be roughly divided into the following two categories.

(1) Handcraft-based segmentation approaches. The similarity among these approaches lies in the employment of the existing mathematical models, serving to detect vein features based on various assumptions. For example, assuming that vein patterns can be approximated to linear textures in predefined neighborhood region, with the descriptors such as Gabor filter to extract the vein pattern. The representative methods in this category include the wide line detector [12], the Gabor filters [3], [24], [25] and the matched filters [26]. In addition, part of scholars discovered that the cross-sectional section of the vein pattern featured with valley shape, which stimulated the establishment of a large number of models to specialize in valley [27]–[30] detection during vein pattern extraction. However, in view of the sensitivity of curvature to valley, experts also devoted efforts into enhancing the vein patterns by calculating the mean curvature [13], difference curvature [31], maximum curvature points [14], and enhanced maximum curvature [32] of pixels in an image. In recent years, the advancement in anatomy has urged the inclusion of finger-vein structure characteristics such as continuity, directionality, smoothness and variability for finger-vein texture extraction in [33].

(2) Deep learning-based segmentation approaches. Different from handcraft approaches, deep learning-based approaches can draw vein patterns from a original image without assuming vein features distribution. More importantly, during medical image segmentation, these approaches verify splendid performance as in cell segmentation [20], pancreas segmentation [34], and retinal vessel segmentation [35]–[38]. In work [39], a feature extraction and recovery network, which outperformed manual feature extraction in terms of verification errors, was proposed. As an extension of this work, [40] was engaged with an novel method to extract the depth feature of vein based on both short-term and long-term memory recurrent neural network, while Yang *et al.* [41] introduced a finger vein segmentation model in light of the generative adversarial networks, which stands astonishing robustness to outliers and vessel breaks. Referring to [42]–[44], semantic segmentation convolutional neural network was optimized to directly abstract the actual finger-vein patterns from NIR finger images while Jalilian and Uhl [43] investigated the effects of fusion and combination training with different labels on finger vein recognition.

The above approaches, doubtlessly, have achieved comparatively sound performance in finger-vein recognition tasks, and yet the existed problems as follows remain to be countered. For instance, the assumptions in handcrafted approaches may fail to detect vein textures, for some pixel values are created by more complex distribution than that of valley or straight-line. Furthermore, the manual methods need to endure the lack of any prior knowledge from other images because of its independent segmentation. And

the final prediction results for handcrafted approaches are closely related to the selected features, which is empirical with artificial interference and possibly results in deviation. For the deep learning-based approaches [39], [42]–[44], these problems are alleviated to some extent since the vein features are directly acquired from raw images, thus minimizing the decision errors of the texture extraction without the attribute distribution assumptions. Meanwhile, via conducting training on a large number of patches based data from different images, abundant prior knowledge can be fully obtained to settle the class imbalance problem. However, the edges of the methods discussed above can only be testified in building the large and unknown models, which can mainly be attributed to the fact that the fixed structure of convolution neural network deprives itself the internal mechanism to deal with complex geometric transformation. Also, without considering the spatial dependencies of finger vein pixels, each pixel will only be segmented independently based on a predefined neighborhood region. Consequently, the existing approaches of finger vein segmentation are still engaged with an enormous space to improve themselves.

Recurrent neural network [17], which has been successfully applied to many domains such as speech recognition [45] and handwriting recognition [46], is able to obtain better feature representation through effective feature accumulation. Residual network [18] enables itself to train a deeper network model after solving the problem of gradient disappearance. Research has certificated the stronger suitability of deeper network for recognition and segmentation tasks. The improved residual networks in [48] can be used for mathematical proof of residual recurrent convolutional operations. The segmentation method based on fully convolutional neural network provides advanced results for image semantic segmentation [49]. The most popular network frameworks, U-Net [20], can achieve excellent segmentation results with only a few training samples thanks to its simultaneous use of global location and context information. In addition, the deformable convolution was proved to be effective for complex semantic segmentation or target detection tasks by adding an offset to the conventional grid sampling location of the standard convolution layer, helping free deform of the sampling grid in [19].

With the development of deep learning technology, probabilistic graphical models have become an interpretable modeling method for many real-world tasks with dependencies, proving to be an effective pixel-level labeling method. The conditional random field of semantic annotation, as one of the most successful graphical models in computer vision [21], [50]–[52], owns its kernel concept to transform the label assignment problem into a probabilistic inference problem, which includes the hypothesis of label consistency among similar pixels. The specific example can be referred to [21], in which an end-to-end training semantic segmentation model was obtained through embedding the conditional random field into the convolutional neural network.

Inspired by recurrent neural network [17], residual network [18], deformable convolution network [19], U-Net [20] and conditional random field [21], we propose a new fully convolutional neural network for finger vein segmentation. The network takes the classic U-Net as its basic structure, the randomly cropped venous patches as the input, and the complete segmentation map acquired by reasoning as the output. The U-shaped structure utilizes up-sampling to improve the output resolution for the purpose of extracting context information and combining the down-sampling features to achieve more precise localization. After normal convolution operation, deformable convolution and residual recurrent convolution are introduced to extract and retain deeper and more complex features. The fully convolutional neural network of this paper constructed through the above process is termed as FCN. Following the former step, we draw the output of FCN as the unary energy of the conditional random field to refine the rough output, and formulate mean-field approximate inference for the conditional random fields with Gaussian pairwise potentials as recurrent neural network. We name the network as CRF-RNN designed by this process. The whole structure optimizes the advantages of fully convolutional neural network and probabilistic graphical models, successfully realizing an end-to-end training by employing back-propagation algorithm and random gradient descent process.

III. PROPOSED METHOD

To study the complex feature representations of finger vein patterns and spatial dependencies information, this paper suggests a new finger vein segmentation model for finger-vein feature extraction. Above all, the ROI areas are extracted from original finger vein images, with the improved ROI extraction method to lessen the plausible pixel missing. Meanwhile, six baselines are endued with different weights to label vein pixels from training sets and test sets, followed by inputting the randomly cropped patches of ROI images into fully convolutional neural networks (FCN) so as to acquire rough segmentation outputs. For one thing, via engaging the FCN, the deformable convolution replaces the standard convolution in the traditional U-Net to adaptively adjust the receptive fields, enabling the capture of veins characteristics in different shapes, sizes and directions. For another, the introduction of residual recurrent convolution effectuates depth mining and feature accumulation. Next, fed with the ROI images and the outputs of the FCNs, the conditional random field (CRF-RNN), embedded as a part of the fully convolutional neural network rather than the traditional post-processing process, realizes the refinement of the previous rough outputs, in which the conventional back-propagation algorithm can be deployed for end-to-end training. Finally, after reasoning the image of the test set, we recombine all the outputs to produce a complete segmentation map. The framework of the proposed approach above can refer to Fig. 1.
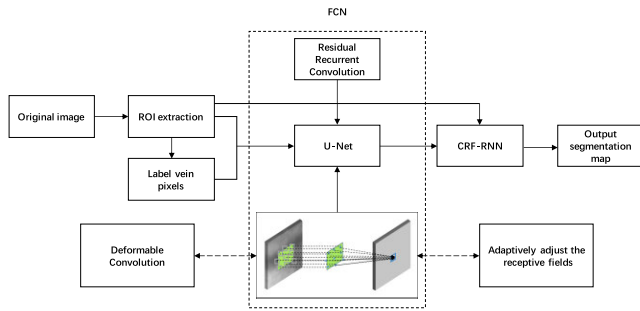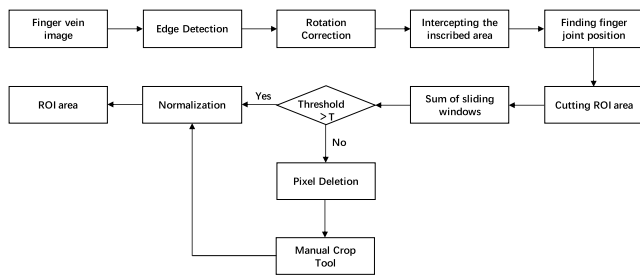
FIGURE 1. The framework of the proposed approach.

FIGURE 2. ROI extraction flow chart in this paper.

A. ROI EXTRATION

To ensure less time consumption and incremental accuracy in the process, we extract the ROI areas from original images according to [53], with the algorithm flow chart presented in Fig. 2. The original finger vein images are sequentially subject to four steps to cut the ROI areas: which are detecting edges, correcting rotation, intercepting the inscribed area of the finger, and locating the finger joint position. However, special intrinsic weakness in this method that fails to detect the edge of the images with missing pixels precisely is far from satisfaction. Thus, this paper presents a sliding window summation method to further screen the extracted inaccurate ROI regions by comparing with the set threshold. For the filtered images, this paper also provides a tool that can crop ROI at will to conduct manual adjustments. Eventually, we normalize the extracted ROI areas to a specific size.

B. LABEL VEIN PATTERNS

Following the pattern in [39], [43], we begin with image enhancement for each input finger vein image to detect vein patterns more precisely. Then six baselines, such as Repeated line tracking [27] (REP), Wide line detector [12] (WLD), Gabor filter [25] (Gabor), Maximum curvature points [14] (MC), Mean curvature [13] (MMC), Enhanced maximum curvature [32] (EMC) are employed to divide vein pattern into segments, resulting in six binary images. Each binary image's values (0 and 1 represent background and vein pixels accordingly) are viewed as labels of corresponding pixels in the input image. The difference between work [39] and the discussed method lies in that the six baselines are complementary to each other. For example, the enhanced maximum curvature can well segment the trunk of the vein, while the wide line detector can decollate the details of the vein. To mitigate primary and secondary influences and enhance the effects of the benchmark segmentation, we give different weights for each baseline by experience. The probability of each pixel (x, y) becoming vein feature $P(x, y)$ can be calculated by: (1), as shown at the bottom of this page,

where i is the number of baselines. Then the binary image is acquired by the adaptive thresholding operation. The label $L(x, y)$ of pixel (x, y) is assigned as follows:

$$L(x, y) = \begin{cases} 1 & if\ P(x, y) = 1 \\ 0 & if\ P(x, y) = 0 \end{cases} \quad (2)$$

C. NETWORK ARCHITECTURE

The network architecture of our model which adopts a u-shaped structure with encoders and decoders on both sides, is elaborated in Fig. 3. After a normal convolution operation, the deformable convolution and residual recurrent convolution are carried out. Deformable convolution adapts to the size and shape of veins by training the receiving field and sampling locations. The feature accumulation with different time steps in the residual recurrent convolution ensures a stronger feature representation, facilitating the extraction of deeper features. The copying and cropping unit in the classic U-Net is replaced by cascading operation, which allows this method to fully absorb the context information and achieve precise localization. The internal details of the model are described below.

We take the first repetitive structure as an example to illustrate the components of deformable convolution and residual recurrent convolution. Fig. 4(a) shows the components of deformable convolution each of which is composed of a convolution offset layer, a convolution layer, a batch normalization layer [52] and an activation layer. Convolution offset layer is the core concept of deformable convolution. In each code stage, the local, dense and adaptive receptive fields are learned by deformable convolutions to establish models for finger veins of various shapes and scales. Fig. 5 visualizes the detailed design. Offsets, generated by additional convolution layers are added to grid positions commonly used in standard convolution. We deploy a 3×3 kernel with grid size 1; therefore, the conventional convolution grid $\mathcal{R}$ can be expressed as:

$$\mathcal{R} = \{(-1, -1), (-1, 0), \ldots, (0, 1), (1, 1)\} \quad (3)$$

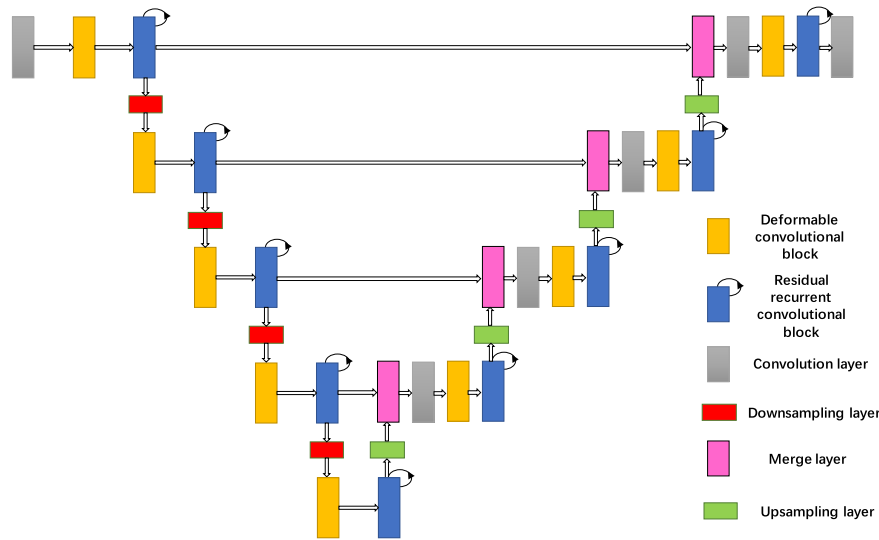$$P(x, y) = \frac{1.5 * EMC + 0.8 * MC + 0.6 * MMC + 0.5 * REP + 1 * Gabor + 0.5 * WLD}{i} \quad (1)$$

FIGURE 3. Network architecture of the proposed method.

For each position n_0 on the output feature map y, we have

$$y(n_0) = \sum_{n_i \in \mathcal{R}} w(n_i) * x(n_0 + n_i) \tag{4}$$

where x represents the input feature map, w means the weights of sampled value and n_i denotes the locations in $\mathcal{R}$. While in deformable convolution, the offset Δn_i is ultilized to augment the regular grid $\mathcal{R}$. So the Eq. (4) evolves into

$$y(n_0) = \sum_{n_i \in \mathcal{R}} w(n_i) * x(n_0 + n_i + \Delta n_i) \tag{5}$$

The positions of the sampling points are also changed into $n_i + \Delta n_i$, which is usually not an integer. Thus we utilize bilinear interpolation to determine the non-integer sampling values. Compared with conventional convolution, deformable convolution may result in some computational cost aiming to perform in a more local and adaptive manner.

The output of deformable convolution is transferred forward as the input of residual recurrent convolution. The detailed operation of residual recurrent convolution is displayed in the dotted box on the right of Fig. 4 (b), where the recurrent convolution operation includes a convolution layer and three subsequence convolutions layers. Suppose that the input sample in the lth layer of the residual recurrent convolution is x_l, and that the network output of the pixel (i, j) in the kth feature map of the recurrent convolution layer at the time step t is $O^l_{ijk}(t)$, the output can be formalized as:

$$O^l_{ijk}(t) = (w^s_k)^T * x^{s(i,j)}_l(t) + (w^r_k)^T * x^{r(i,j)}_l(t-1) + b_k \tag{6}$$

$x^{s(i,j)}_l(t)$, $x^{r(i,j)}_l(t-1)$ are the inputs of the standard and recurrent convolutions of the lth layer respectively. w^s_k, w^r_k proves the weights of the standard and recurrent convolutions of the kth feature map respectively, and b_k denotes the bias. A ReLU activation function f will follow each recurrent
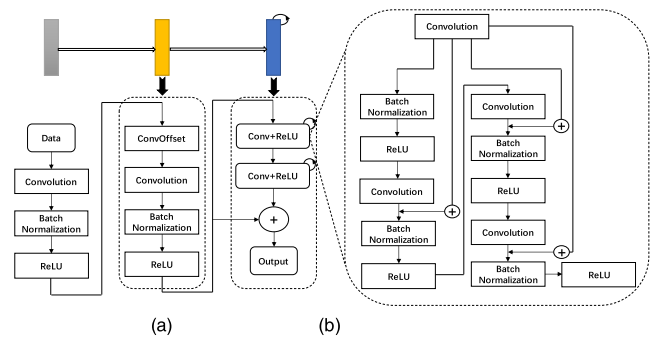

FIGURE 4. The components of the deformable convolution (a) and the residual recurrent convolution (b).
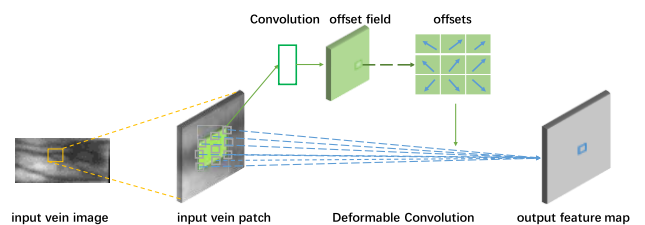

FIGURE 5. Detailed design of the deformable convolution.

convolution layer, so we have:

$$F(x_l, w_l) = f\left(O^l_{ijk}(t)\right) = max(0, O^l_{ijk}(t)) \tag{7}$$

$F(x_l, w_l)$ represents the outputs from the lth recurrent convolution layer. The final outputs x_{l+1} of the residual recurrent convolution are transferred through the residual units shown in Fig. 4(b) and can be calculated following the formula below:

$$x_{l+1} = x_l + F(x_l, w_l) \tag{8}$$

In (8), x_l stands for the input of the residual recurrent convolution, i.e., the output from the deformable convolution mentioned above. In the encoding and decoding convolution unit of the network, the x_{l+1} sample is employed as the input for the instant subsequent down-sampling or up-sampling layers.

After the last residual recurrent convolution operation of the up-sampling in the network, we take its output as the unary potential energy of the dense conditional random field, and finally the objective function of label assignment x can be calculated as:

$$E(x) = \sum_i \psi_\mu(x_i) + \sum_{i<j} \psi_p(x_i, x_j) \qquad (9)$$

$\psi_\mu(x_i)$ is a unary potential function for measuring the probability of the pixel point i pertaining to the category label x_i when the pixel point i equals to y_i, which is provided by the fully convolutional neural network shown in Fig. 3. $\psi_p(x_i, x_j)$ is a pairwise potential function, employed to compute the probability of simultaneous occurrence of x_i and x_j. The pairwise potential function furnishes a smoothing term associated with image features that encourages the assignment of similar labels to pixels with similar attributes. Consistent with the work of [48], the pairwise potential function is modeled as a weighted Gaussian function:

$$\psi_p(x_i, x_j) = \mu(x_i, x_j) \sum_{m=1}^M w^{(m)} k_G^{(m)}(f_i, f_j) \qquad (10)$$

$k_G^{(m)}$ proves a Gaussian weight term used to measure the similarity between the feature vectors f_i and f_j of pixel points i and j. The function $\mu(x_i, x_j)$ is helpful in capturing the compatibility between different pairs of labels. Minimizing the objective function E (x) produces the most probable label predictions.

Same as the work [21], the iterative process of the mean-field in dense condition random fields is decomposed into common CNN operations by conveying message, weighting filter outputs, transforming compatibility, adding unary potentials and normalization shown in Fig. 6. Given an image I, unary potential energy U obtained by the fully convolutional neural network and the edge probability Q_{in} from the previous iteration, the output Q_{out} can be expressed by:

$$Q_{out} = f_\theta(U, Q_{in}, I) \qquad (11)$$

where f_θ represents the above decomposition operation, $\theta = \{w^{(m)}, \mu(x_i, x_j)\}$ stands for the above Gaussian kernel weight and compatibility function.

Multiple mean-fields iterations are reasoned as a recurrent neural network shown in Fig. 7, which is implanted in as a part of the fully convolutional neural network. We define the RNN-like structure obtained by this process as CRF-RNN, and the network behavior is given by:

$$H_1(t) = \begin{cases} softmax(U), & t = 0 \\ H_2(t-1), & 0 < t \le T \end{cases} \qquad (12)$$

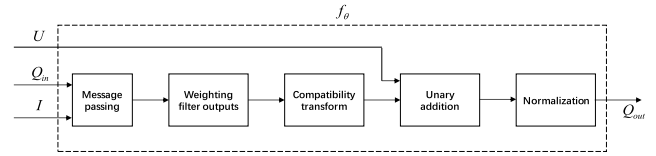$$H_2(t) = f_\theta(U, H_1(t), I), \, 0 \le t < T \qquad (13)$$

FIGURE 6. Visualization process of the deformable convolution.
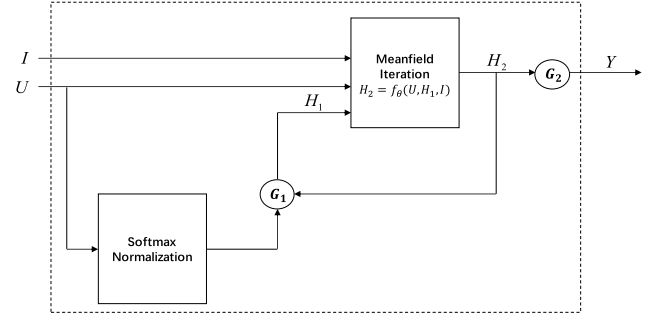

FIGURE 7. The CRF-RNN Network. Gating functions G_1 and G_2 are given by the above formula.

$$Y(t) = \begin{cases} 0, & 0 \le t < T \\ H_2(t), & t = T \end{cases} \qquad (14)$$

Here T denotes the number of mean-field iterations.

In the fully convolutional neural network phase, the pixel level labels are predicted without taking vein structures and spatial dependencies into account. Then, probabilistic graphical model is established during the CRF-RNN phase based on conditional random fields. Consequently, the whole model draws the advantages of fully convolutional neural network and conditional random field, effectuating an end-to-end training by utilizing back-propagation algorithm and stochastic gradient process. During the forward propagation of the network, as soon as the calculation ingresses the CRF-RNN through FCN, the data need to take T iterations before leaving the loop created by the RNN, so the refinement of vein pixels only occurs in this loop. The error differentials, during the backward pass, are passed forward once the output departs from the cycle. In addition, in the whole training process of the network, the weight of Gauss kernel, the compatibility function of label and other CRF parameters can be optimized automatically.

IV. EXPERIMENT
A. DATASETS

The experiment was conducted on three public finger vein datasets. The first one was the SDUMLA-HMT dataset [55] created by the School of Computer Science and Technology, Shandong University in China. The dataset contains 3,816 finger images from 106 individuals, including 61 males and 45 females, with their age ranging from 17 to 31. Data from the index finger, middle finger and ring finger of each person's left and right hands were collected separately, and each finger was repeatedly collected for 6 times. Each image has a resolution of 320×240 pixels stored in bmp format.

The second finger vein image dataset, the MMCBNU_6000 [3], was constructed by the Chonbuk National University,

which contains 6,000 finger images from 83 males and 17 females, 100 individuals in total. Subjects at the age of 16 to 72 come from more than 20 countries in Asia, Europe, Africa and Americas, most of whom are between the age of 21 to 32. Images from the index finger, middle finger and ring finger of each person's left and right hands were collected, and each finger was repeatedly collected for 10 times. The original finger vein image provided by the dataset stays with a pixel resolution of 640×480, while the ROI images provided with a resolution of 128×60.

The third dataset was the Hong Kong Polytechnic University finger vein image dataset, HKPU [56], which contains 3,132 finger images from 156 individuals, with images from each person's index finger and middle finger collected 6 times separately. Among them, the first 105 people provided 2,520 finger images (105 people $\times$ 2 fingers $\times$ 6 images $\times$ 2 phrases), which were collected in two separate phases, engaging intervals from one month to six months (an average of 66.8 days). The rest of the subjects only offered image data once. Each image possesses a resolution of 513×256 pixels stored in bmp format. Moreover, masks and enhanced images by local histogram equalization in 488×218 pixels can also be found in the dataset.

B. ROI EXTRACTION

SDUMLA dataset extracts directly from the original image while HKPU from the enhanced image. Nonetheless, results shows that some collected images from HKPU dataset suffer from pixel deficiency, which can be explained in Fig. 8(a). The upper half of Fig. 8(c) gives evidence that the ROI extraction results obtained by the method of [53] are in an undesirable situation. Thus, after imitating [53] to preliminarily extract the finger circumscribed regions (in order to keep more information for the images with missing pixels), sliding window summation method is engaged to filter the finger vein images so that the mentioned pixel deficiency above can be favorably eliminated. As shown in Fig. 8(b), we fix the size of sliding window to 20×20, the stride to 20, and the threshold to $20 \times 20 \times 40$ based on our previous experience. On the condition that the sum of pixels in the sliding window is smaller than that of the threshold, it is assertive that pixel missing areas exist in images. Moreover, a tool capable of freely cropping and normalizing is provided for these images with inaccurate ROI extraction, with its final extraction result as an example exhibited at the bottom half of Fig. 8(c). At length, the ROI resolutions of SDUMLA and HKPU are adjusted to 270×120 and 130×70 respectively, while MMCBNU_6000 dataset adopts the ROI images provided in its own dataset with a resolution of 128×60.

C. LABEL VEIN PATTERNS

The extracted ROI (Fig. 9(a)) is initially processed by Contrast Limited Adaptive Histogram Equalization (CLAHE) (Fig. 9(b)) to facilitate vein patterns labeling. Fig. 9(c) explicitly presents from the top to the bottom the vein binary images obtained from the six baseline segmentation methods respec-
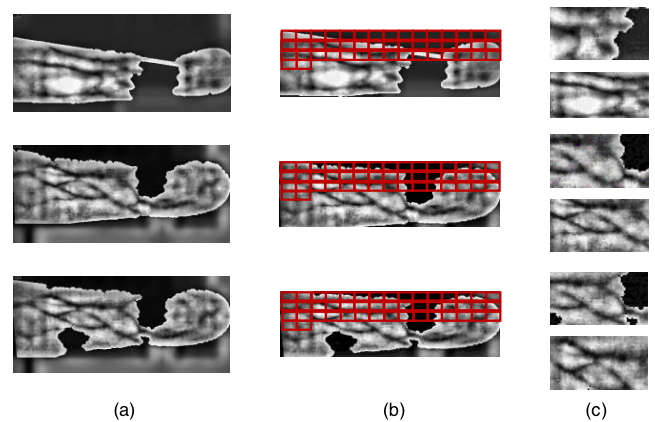

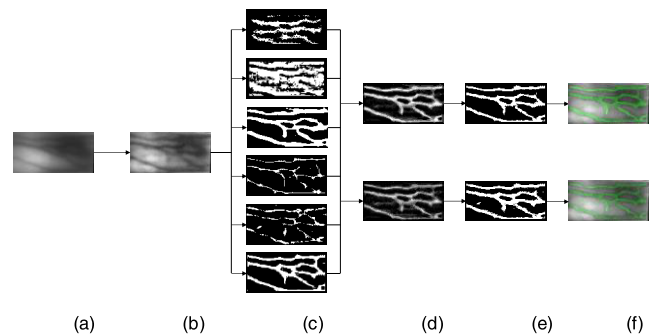**FIGURE 8.** ROI extraction example of pixel missing images.

FIGURE 9. Example of vein pattern labeling.

tively, which are repetitive line tracking [27], wide line detector [12], Gabor filter [25], maximum curvature point [14], mean curvature [13] and enhanced maximum curvature [32].

Subsequently, we take an operation to compare the differences between the literature [39] and our method, which are demonstrated in Fig. 9(d-f), with the upper images deriving from our proposed labeling method, and the lower part from the labeling method of work [46]. In the selection of weights, as described in Section 3.2, some baselines possess the strength in detecting the trunk of veins while the other excel in recognizing the details of veins. To fully exploit each baseline's specialty, different initial weights are assigned for different baselines to obtain rough labels, which will be utilized to do the finger vein segmentation experiment and to fine-tune the weight of the label in the light of the segmentation experiment results in the following step. The segmentation experiment will be repeated unless an acceptable segmentation effect has occured. According to this principle, we can acquire the image (Fig. 9 (d)) after superimposing and averaging the different weights assigned to each baseline and ultimately obtain the binary image (Fig. 9(e)) through the adaptive binarization operation. Fig. 9(f) refers to the vein image of the original image covered by the labeled pixel, according to which, a fact that setting different weights for different benchmarks can produce more accurate labeling results than simply stacking and averaging do is readily

observed. What is more, the adaptive binarization operation also takes on its effectiveness in filtering out redundant noise.

D. EXPERIMENTAL CONFIGURATION

We compared the proposed model with array of equivalent models including U-Net, trained these models from scratch, and initialized the weights randomly. In order to ensure a rapid convergence and prevent overfitting, for one thing, we engaged a dynamic method to automatically change the learning rate (with the default initial one at 0.001) according to the loss value in the training process. The learning rate would decrease by tenfold when the lost value remained stable after p_e epochs, and the training process would cease when the lost value barely changed after q_e epochs. Here p_e and q_e were empirically set to 4 and 20 separately. For another, we trained the model with the randomly cropped patches. We set different training strategies according to different resolutions of ROI image in three datasets. For SDUMLA, MMCBNU_6000 and HKPU datasets, to reduce computational complexity and guarantee local characteristics, the sizes of the patches were set to 48 × 48, 24 × 24 and 24 × 24 respectively. It has been mentioned in [42] that when manual labels are employed in network training, the network outperforms other ones when the number of labels is 20. Benefited from this discovery, we fix the number of images and labels of the training set to 20, and randomly select 10,000, 5,000 and 5,000 patches for each of the three datasets SDUMLA, MMCBNU_6000 and HKPU (a total of 200,000, 100,000, 100,000 patches) for network training. For the test images, the size of the patches was consistent with the one of the training images. In addition, multiple continuous overlapping patches with a stride of 5 pixels in both width and height were extracted, and the vein probability was obtained by averaging probabilities over all prediction patches covering the pixel. In case that the computer runned out of memory, we tested all the remaining images in batches and simply averaged all batches to get the final evaluation metrics.

For CRF-RNN, the parameter setting followed that of work [21], and Potts model was arranged to initialize its compatibility transformation parameters. In the training process, the number of mean-field iterations T was set to 5, while during the test it was increased to 10. The final number of classifications was changed to 2 because only the vein and background categories were included.

In terms of finger vein verification, we copied the experimental setup of [57], and obtained a differential image from two images which was input into the pre-trained VGGNet-16. In order to verify the effectiveness of segmentation results on verification performance, similar settings were imposed on the segmented images and ROI images in SDU, MMCBNU and HKPU datasets. There are 636 (106 × 6), 600 (100 × 6) and 522 (105 × 4 + 51 × 2) categories in the three datasets accordingly, half of which were randomly selected as training sets while the other half as test sets. In the process of training, in order to solve class imbalance, the number of

imposter matching images (A_{318}^2, A_{300}^2, A_{261}^2) in the training sets was selected to be consistent with that of the authentic matching images ($318A_6^2$, $300A_{10}^2$, $261A_6^2$), while all the authentic matching and imposter matching images of the test sets were tested during the test. Lastly, we calculated the average accuracy by interchanging these images for training and testing after two-fold cross validation. It should be noted that all experiments were conducted under the Tensorflow and Keras frameworks using an NVIDIA GTX Titan XP GPU.

E. PERFORMANCE EVALUATION METRICS

We deployed several metrics to evaluate our model in the experiment of finger vein segmentation: Accuracy (AC), Sensitivity (SE), Specificity (SP), F-measure (F1), Jaccard similarity (JS). In addition, since the area under curve (AUC) and receiver operating characteristic (ROC) pay frequent visit in medical image segmentation tasks, they had also been taken into account. We took EER as the evaluation metric in the finger vein verification experiment.

V. EXPERIMENTAL RESULTS AND ANALYSIS

In this section, we explore the effects of deformable convolution and residual recurrent convolution on finger vein segmentation, and conduct a systematic comparison between the performance of the proposed model and the current mainstream segmentation model. Besides, the influence of conditional random field on the performance of fully convolutional neural network has also been the main focus in this section. Apart from the results from various evaluation metrics, the experiment also presents the results of the segmentation, thus facilitating the observation of the subtle segmentation differences. Last but not least, we study the effect of finger vein segmentation results on verification performance.

A. FINGER VEIN VERIFICATION BASED ON SDUMLA DATASET

We commenced with exploring the performances of deformable convolution and residual recurrent convolution on finger vein segmentation with the segmentation performance metrics of adding deformable convolution and residual recurrent convolution illustrated in Table 1. We can conclude from the graphic that the deformable convolution improves the performance of traditional U-Net to some extent, while the additional involvement of residual recurrent convolution bares out a more satisfactory performance of replacing conventional convolution in traditional U-Net with deformable convolution. Fig. 10(a) describes the ROC curve of finger vein segmentation with the addition of deformable convolution and residual recurrent convolution, proving that the closer the ROC curve is to the upper left corner in ROC coordinates, the higher the accuracy of the model is. Another phenomenon can be discovered that the ROC curve of the model proposed in this paper centers mostly on the upper-left section and occupies the most areas under the curve
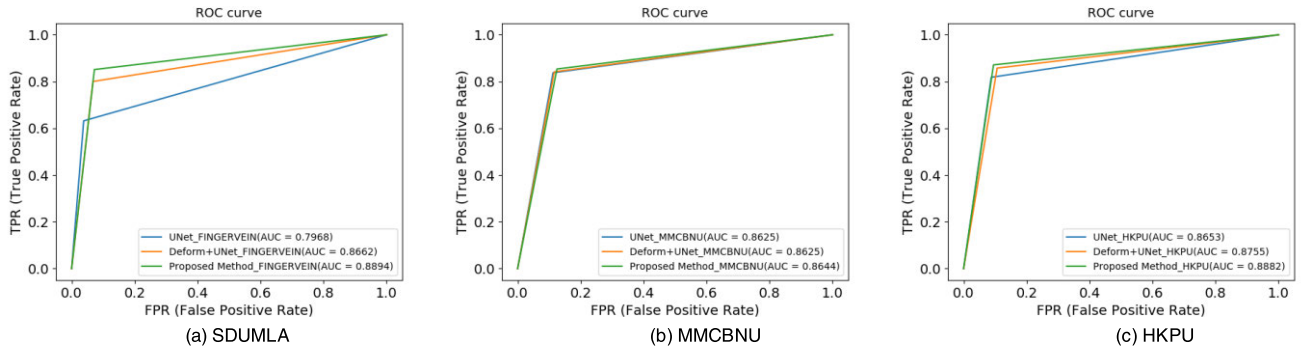
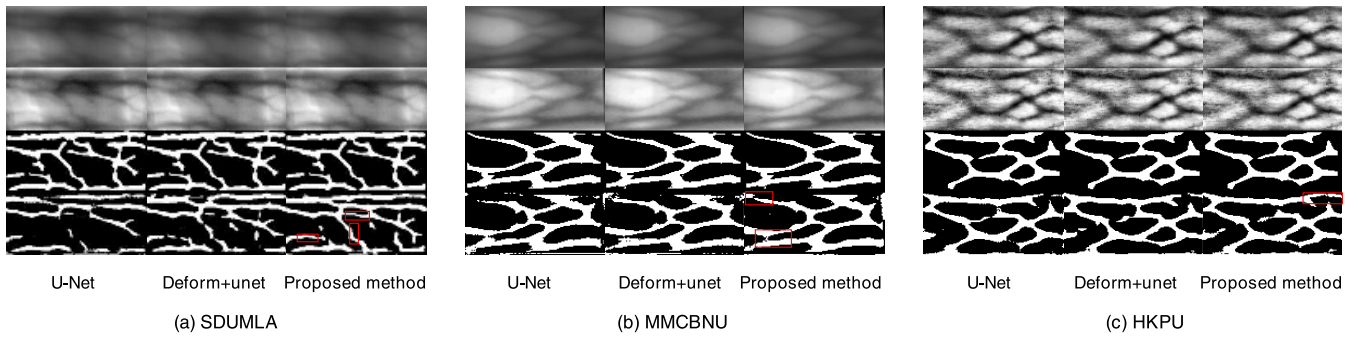**FIGURE 10.** ROC curves for adding deformable convolution and residual recurrent convolution on three datasets.

FIGURE 11. Example of segmentation results for adding deformable convolution and residual recurrent convolution on three datasets.
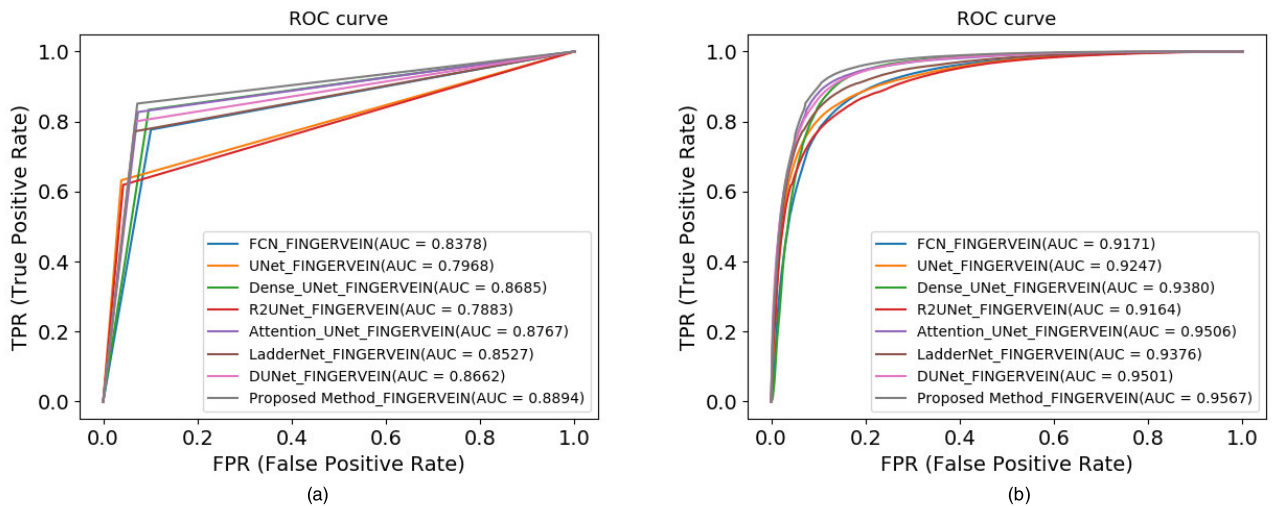

FIGURE 12. ROC curves of different models on SDUMLA dataset.

(AUC). An example of segmentation results is illustrated in Fig. 11(a) with the left column extracted from traditional U-Net segmentation, the middle column from U-Net model segmentation, in which normal convolution is replaced by deformable convolution, and the right column from network segmentation after adding residual recurrent convolution on the former basis. And the four rows respectively stands for the original ROI image, enhanced image, label and predicted

segmentation result. It can be seen that in the tiny vessel regions, U-Net shows its limitations in handling details, and although deformable convolution can capture finger veins of myriad of shapes and sizes, showing its powerful geometric modeling ability, it still owns insufficiency in deep mining and feature accumulation capacities, certified by some discontinuous lines in some places. Howbeit after the addition of residual recurrent convolution, the deeper
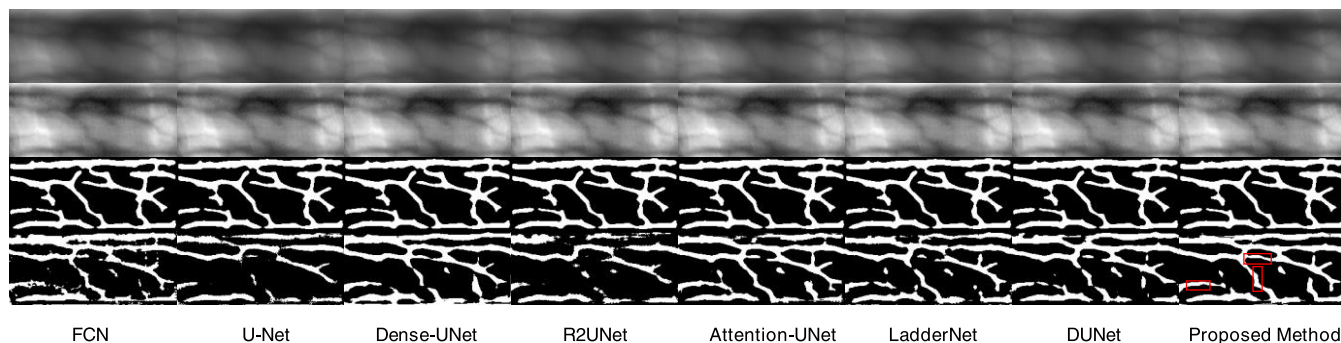
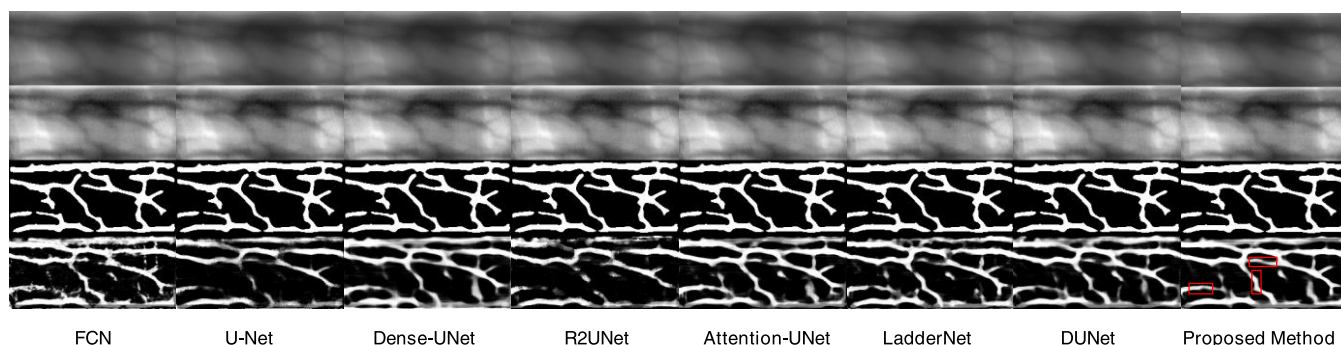**FIGURE 13.** Examples of segmentation results for various models on SDUMLA dataset (without CRF-RNN).

FIGURE 14. Examples of segmentation results for various models on SDUMLA dataset (with CRF-RNN).

TABLE 1. Segmentation performance for adding deformable convolution and residual recurrent convolution on sdumla dataset.

Methods	SE	SP	AC	F1	JS	AUC
U-Net[20]	0.6320	**0.9616**	0.8713	0.7291	0.8713	0.7968
Deform+unet	0.8001	0.9322	0.8961	0.8083	0.8961	0.8662
Proposed method	**0.8512**	0.9276	**0.9067**	**0.8333**	**0.9067**	**0.8894**

TABLE 2. Segmentation performance of various models on sdumla dataset (without CRF-RNN).

Methods	Year	SE	SP	AC	F1	JS	AUC
FCN[49]	2015	0.7768	0.8988	0.8654	0.7597	0.8654	0.8378
UNet[20]	2015	0.6320	**0.9616**	0.8713	0.7291	0.8713	0.7968
Dense-UNet[35]	2018	0.8334	0.9036	0.8843	0.7979	0.8843	0.8685
R2UNet[36]	2018	0.6189	0.9577	0.8648	0.7150	0.8648	0.7883
Attention-UNet[34]	2018	0.8271	0.9262	0.8991	0.8178	0.8991	0.8767
LadderNet[37]	2018	0.7716	0.9339	0.8894	0.7927	0.8894	0.8527
DUNet[38]	2018	0.8001	0.9322	0.8961	0.8083	0.8961	0.8662
Proposed Method	2020	**0.8512**	0.9276	**0.9067**	**0.8333**	**0.9067**	**0.8894**

effective features are excavated, accumulated and retained, which helps connect the veins with discontinuous lines and thus a better segmentation result can be obtained.

Then we shifted our objects to systematically compare the segmentation performance of the proposed model with that of the current mainstream segmentation models. The experimental metrics depicted in Table 2 show that the proposed model outperforms other equivalent models in most aspects. It is worth notice that the proposed model achieves the highest accuracy among the eight equivalent models,

reaching 0.9067. We further evaluate the model with ROC curve, as shown in Fig. 12 (a). We can see that the AUC of the model proposed in this paper reaches its highest value at 0.8894. Fig. 13 illustrates the predicted segmentation results, containing from left to right the examples of segmentation results of FCN [49], UNet [20], Dense-UNet [35], R2UNet [36], Attention-UNet [34], LadderNet [37], DU-Net [38] and the proposed method, which certifies that the proposed fully convolutional neural network obtains clearer vein segmentation results. It can detect the weak vein vessels that may be lost or tied, and also connect the discontinuous vessels, enabling the a more effective saving of details.

In succession, we evaluated the influences of conditional random field on finger vein segmentation. In case of losing the generality, we likewise intercalated CRF-RNN into other seven models to refine the output. Various performance metrics expounded in Table 3 mainifest that the proposed model still occupies the first place in most performance metrics. Another worth-noting matter is that the area under the ROC curve in Fig. 12(b) has been greatly enlarged, mainly for the reason that when CRF-RNN is trained with stochastic gradient descent, the FCN component and the CRF component can cooperate to deliver the optimal output of the whole network, thus improving the overall performance of the model. However, the other performance metrics of all eight models are not improved much, which can be mainly attributed to that the patch-based training solves the class imbalance problem of the test sets. Considering the complex spatial dependence, the smoothness and consistency of the

TABLE 3. Segmentation performance of various models on sdumla dataset (with CRF-RNN).

Methods	Year	SE	SP	AC	F1	JS	AUC
FCN[49]	2015	0.7771	0.8987	0.8654	0.7598	0.8654	0.9171
UNet[20]	2015	0.6323	**0.9616**	0.8714	0.7292	0.8714	0.9247
Dense-UNet[35]	2018	0.8335	0.9036	0.8844	0.7980	0.8844	0.9380
R2UNet[36]	2018	0.6191	0.9573	0.8646	0.7147	0.8646	0.9164
Attention-UNet[34]	2018	0.8283	0.9261	0.8993	0.8184	0.8993	0.9506
LadderNet[37]	2018	0.7716	0.9339	0.8894	0.7927	0.8894	0.9376
DUNet[38]	2018	0.8002	0.9322	0.8961	0.8084	0.8961	0.9501
Proposed Method	2020	**0.8516**	0.9276	**0.9068**	**0.8335**	**0.9068**	**0.9567**

TABLE 4. Verification performance of various models on sdumla dataset (with CRF-RNN).

Methods	Year	EER(%)
Original ROI	—	9.397
FCN[49]	2015	11.056
UNet[20]	2015	9.219
Dense-UNet[35]	2018	7.834
R2UNet[36]	2018	10.147
Attention-UNet[34]	2018	7.359
LadderNet[37]	2018	7.90
DUNet[38]	2018	6.705
Proposed Method	2020	**5.827**

TABLE 5. Segmentation performance for adding deformable convolution and residual recurrent convolution on mmcbnu dataset.

Methods	SE	SP	AC	F1	JS	AUC
U-Net[20]	**0.8601**	0.8693	0.8665	0.7964	0.8665	0.8625
Deform+unet	0.8431	0.8813	0.8697	0.7971	0.8697	0.8625
Proposed method	0.8380	**0.8871**	**0.8722**	**0.7992**	**0.8722**	**0.8644**

TABLE 6. Segmentation performance of various models on mmcbnu dataset (without CRF-RNN).

Methods	Year	SE	SP	AC	F1	JS	AUC
FCN[49]	2015	0.8145	0.8755	0.8570	0.7756	0.8570	0.8437
UNet[20]	2015	**0.8601**	0.8693	0.8665	0.7964	0.8665	0.8625
Dense-UNet[35]	2018	0.8075	0.8720	0.8524	0.7686	0.8524	0.8397
R2UNet[36]	2018	0.8272	0.8828	0.8659	0.7892	0.8659	0.8545
Attention-UNet[34]	2018	0.8217	**0.8879**	0.8678	0.7905	0.8678	0.8548
LadderNet[37]	2018	0.8406	0.8732	0.8633	0.7887	0.8633	0.8532
DUNet[38]	2018	0.8431	0.8813	0.8697	0.7971	0.8697	0.8625
Proposed Method	2019	0.8380	0.8871	**0.8722**	**0.7992**	**0.8722**	**0.8644**

FIGURE 15. EER curves under different models.

TABLE 7. Segmentation performance of various models on mmcbnu dataset (with CRF-RNN).

Methods	Year	SE	SP	AC	F1	JS	AUC
FCN[49]	2015	0.8112	0.8763	0.8565	0.7744	0.8565	0.9186
UNet[20]	2015	**0.8536**	0.8751	0.8686	0.7977	0.8686	0.9296
Dense-UNet[35]	2018	0.8075	0.8720	0.8524	0.7686	0.8524	0.9159
R2UNet[36]	2018	0.8249	0.8842	0.8662	0.7891	0.8662	0.9284
Attention-UNet[34]	2018	0.8207	**0.8888**	0.8681	0.7907	0.8681	0.9250
LadderNet[37]	2018	0.8300	0.8764	0.8623	0.7853	0.8623	0.9245
DUNet[38]	2018	0.8415	0.8836	0.8708	0.7981	0.8708	0.9304
Proposed Method	2019	0.8374	0.8876	**0.8724**	**0.7993**	**0.8724**	**0.9313**

TABLE 8. Verification performance of various models on mmcbnu dataset (with CRF-RNN).

Methods	Year	EER(%)
Original ROI	—	0.807
FCN[49]	2015	1.266
UNet[20]	2015	0.732
Dense-UNet[35]	2018	0.888
R2UNet[36]	2018	0.718
Attention-UNet[34]	2018	0.685
LadderNet[37]	2018	0.721
DUNet[38]	2018	0.406
Proposed Method	2020	**0.364**

Finally, we inquired into the effect of segmentation results on verification performance. The experimental results are elaborated in Table 4 and Figure 15. It can be seen that the segmentation performance closely relates to the verification performance while our proposed method obtains the minimum EER. After that, we also discover that the segmentation results based on FCN and R2UNet possess worse verification performance than the original ROI does, while the other segmentation results shows the opposite. It is proved that these two methods can reduce the inherent distinguishability of the image itself.

B. FINGER VEIN VERIFICATION BASED ON MMCBNU DATASET
Consistent with the SDUMLA dataset evaluation method, three steps were taken. We initiated with investigating effects of deformable convolution and residual recurrent convolution

label assignment, the pixel can further adjust the predicted value according to the global information, thus resulting in the increased overall performance, and the finely adjusted corresponding pixels. The predicted segmentation results are presented in Fig. 14, which implies that the conditional random field does take the spatial dependence of pixels into account, so that the probability of similar pixels with same label assignment increases, and the segmentation results can be more easily judged by humans. In addition to detecting the weak vein vessels, the proposed method supplements the pixels at the vascular rupture as well. Experiments verify that the end-to-end training of FCN and CRF raises the accuracy of the system.
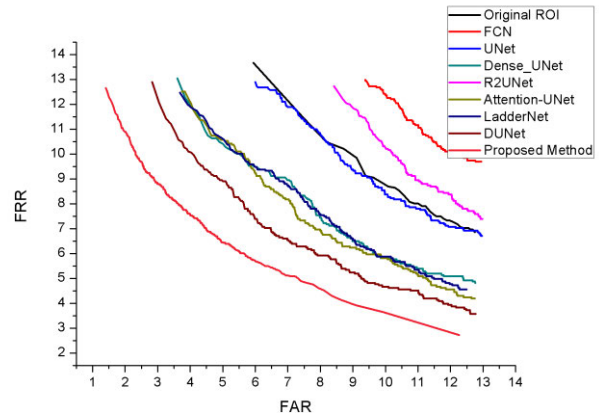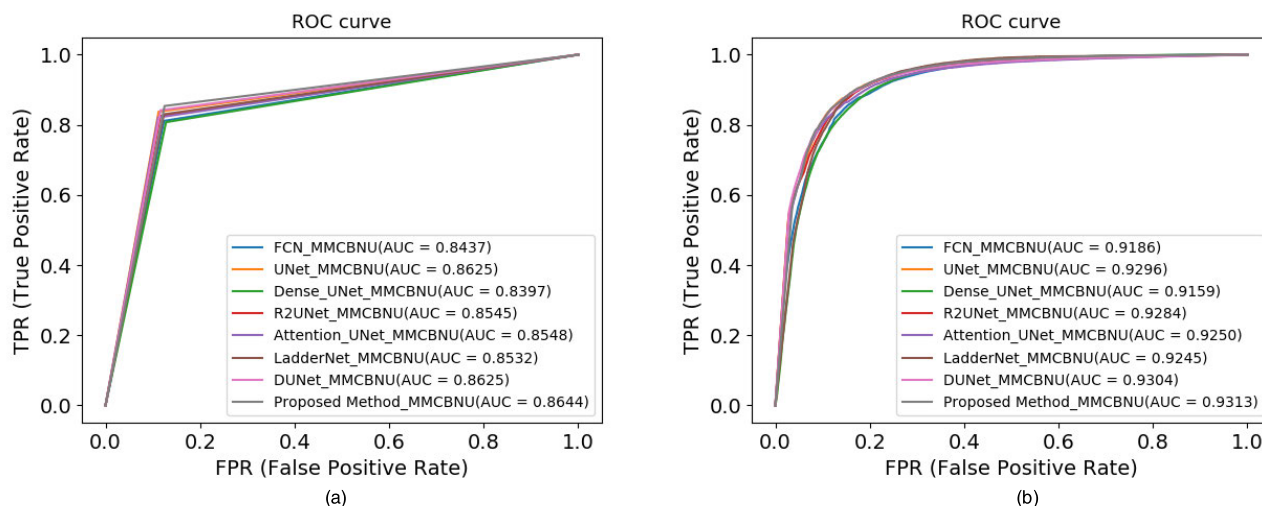
FIGURE 16. ROC curves of different models on mmcbnu dataset.
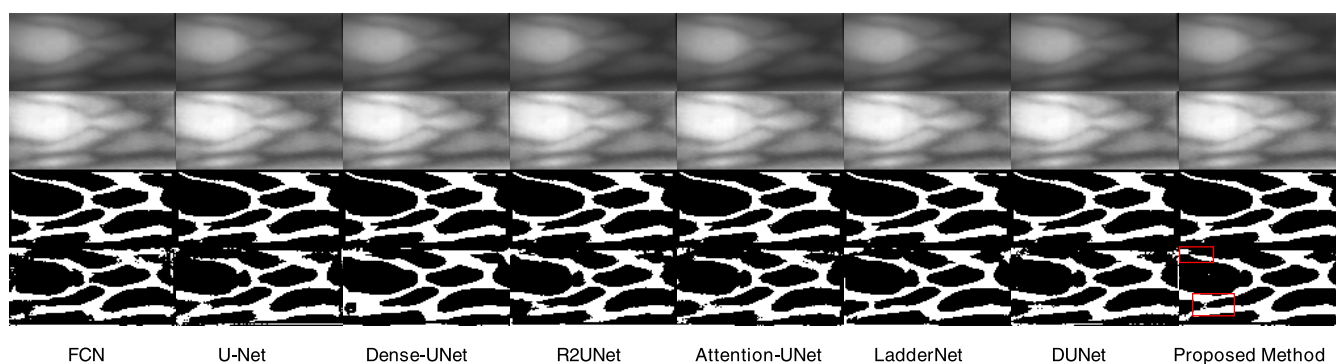

FIGURE 17. Examples of segmentation results for various models on mmcbnu dataset (without CRF-RNN).
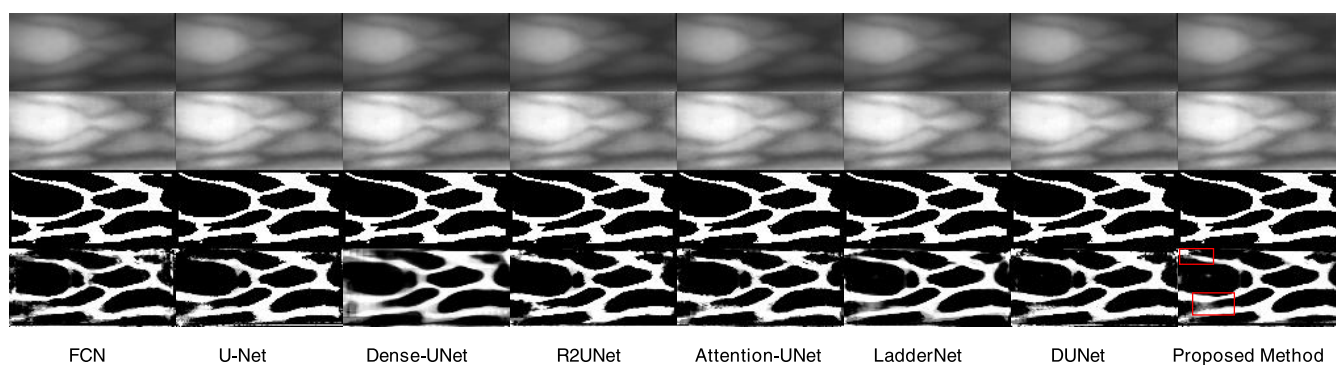

FIGURE 18. Examples of segmentation results for various models on mmcbnu dataset (with CRF-RNN).

on finger vein segmentation, then conducted a systematic comparison between the performance of the proposed model and those of the current mainstream segmentation models, and finally embed the conditional random field into recurrent neural network structure to further evaluate the performance of the model. In Table 5, metrics after adding deformable convolution and residual recurrent convolution does not show

much progress due to the quality differences among ROI images, some of which are clear enough to have their features learned while some of which are too obscure to have their vein patterns detected, thus reinforcing the difficulties for deep mining and feature accumulation of residual recurrent convolution. Fig. 10(b) depicts ROC curves with the addition of deformable convolution and residual recurrent convolution,

TABLE 9. Segmentation performance for adding deformable convolution and residual recurrent convolution on HKPU dataset.

Methods	SE	SP	AC	F1	JS	AUC
U-Net[20]	0.8182	0.9049	0.8876	0.7908	0.8876	0.8653
Deform+unet	0.8575	0.8930	0.8838	0.7931	0.8838	0.8755
Proposed method	**0.8713**	**0.9120**	**0.8962**	**0.8133**	**0.8962**	**0.8882**
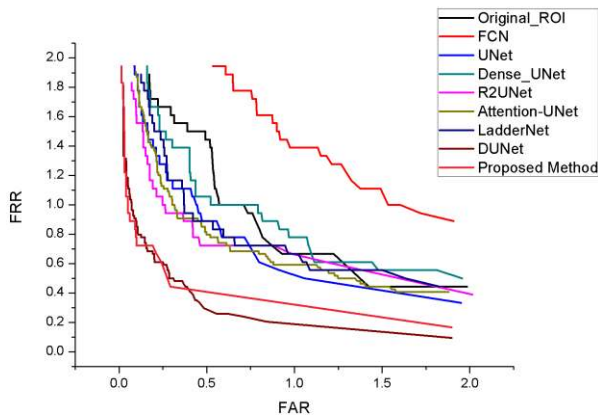
FIGURE 19. EER curves under different models.

TABLE 10. Segmentation performance of various models on HKPU dataset (without CRF-RNN).

Methods	Year	SE	SP	AC	F1	JS	AUC
FCN[49]	2015	0.8533	0.8968	0.8855	0.7946	0.8855	0.8753
UNet[20]	2015	0.8182	0.9049	0.8876	0.7908	0.8876	0.8653
Dense-UNet[35]	2018	0.8758	0.8829	0.8810	0.7926	0.8810	0.8793
R2UNet[36]	2018	0.8057	0.9081	0.8815	0.7793	0.8815	0.8570
Attention-UNet[34]	2018	**0.8786**	0.8881	0.8857	0.7996	0.8857	0.8842
LadderNet[37]	2018	0.8764	0.8839	0.8819	0.7940	0.8819	0.8803
DUNet[38]	2018	0.8575	0.8930	0.8838	0.7931	0.8838	0.8755
Proposed Method	2019	0.8713	**0.9120**	**0.8962**	**0.8133**	**0.8962**	**0.8882**

TABLE 11. Segmentation performance of various models on HKPU dataset (with CRF-RNN).

Methods	Year	SE	SP	AC	F1	JS	AUC
FCN[49]	2015	0.8529	0.8978	0.8861	0.7955	0.8861	0.9341
UNet[20]	2015	0.8181	0.9051	0.8879	0.7912	0.8879	0.9288
Dense-UNet[35]	2018	0.8758	0.8829	0.8810	0.7926	0.8810	0.9353
R2UNet[36]	2018	0.8057	0.9083	0.8817	0.7795	0.8817	0.9221
Attention-UNet[34]	2018	**0.8783**	0.8901	0.8870	0.8015	0.8870	0.9328
LadderNet[37]	2018	0.8764	0.8842	0.8822	0.7943	0.8822	0.9322
DUNet[38]	2018	0.8574	0.8936	0.8842	0.7936	0.8842	0.9297
Proposed Method	2019	0.8713	**0.9124**	**0.8963**	**0.8135**	**0.8963**	**0.9448**

where their curves unfold greater approximation, and also better network performance after adding residual recurrent convolution. Fig. 11(b) illustrates an example of their segmentation results, which share astonishing similarity. Thanks to the convolution offset layer in deformable convolution, which outstrips other methods in learning characteristics of finger veins with different shapes, sizes and directions, the proposed model can slightly reduce the adjacent pixel values on the tied veins.

Table 6 expounds the performance metrics of the proposed model and the current mainstream segmentation model, where the proposed model outperforms the remaining subjects in most metrics with the accuracy at 0.8722. We notice that compared with all the performance metrics in SDUMLA dataset, corresponding data here without exception are superior. Reason for this can be retrospected to the inferior quality of some ROI images provided here, resulting in low image contrast and difficulty in detecting vein patterns, which will directly affect the accuracy of the label. Fig. 16(a) also portrays the ROC curve and the area under the curve of each model. Through observation, the proposed method owns the largest AUC at 0.8644, which proves its superior performance of recurrent neural network, residual network and deformable convolution network for the segmentation of vein and blood vessel. The predicted segmentation results are shown in Fig. 17, which are consistent with those in the SDUMLA dataset; each column stands for the examples of segmentation results of FCN [49], UNet [20], Dense-UNet [35], R2UNet [36], Attention-UNet [34], LadderNet [37], DUNet [38] and the proposed method respectively, while each row refers to original ROI image, enhanced image, label

and predicted segmentation result one by one. Since adding offset to the deformable convolution sampling position can make the vein adapt to different size, shape and direction changes, the proposed method can automatically correct the predicted segmentation results on the premise of inaccurate label, and supplement the missing vein pixels. With this corollary, it is not difficult to comprehend that the evaluation metric only represents the learning ability of the network, and the model with better learning ability does not necessarily satisfy the visual standards of people, despite of which, our model proves better than other equivalent models in both visual and quantitative analysis.

When we embed conditional random field and transfer the error in the reverse direction, the overall performance of the model can be improved due to the corresponding relationship between pixels. The most distinguished performance lies in that the ROC curve and the area under the curve in Fig. 16(b) has been expanded by about 7%, reaching 0.9313. Following the similar track of the SDUMLA dataset, the improvement of other metrics shown in Table 7, remains little. The reason for that may also ascribe to relative assumptions such as label consistency between similar pixels involved in conditional random fields, which regulates the adjustment between pixels without much change. The results of vein segmentation in Fig. 18 show the importance of compatibility transformation between pixels in conditional random fields as well.

In both Table 8 and Figure 19 with the illustration of EER under different models, our proposed method still achieves the most prominent verification performance. It is worth mentioning that the EER of all models is much smaller than that of SDUMLA dataset, which can be ascribed to the samll intra class difference and large inter class difference in MMCBNU dataset. Compared with SDUMLA dataset, in spite of the slightly lower segmentation metrics MMCBNU dataset contains better verification performance, which bears out that the dataset has more impact on the verification performance than on the segmentation results. The changing trend is similar to
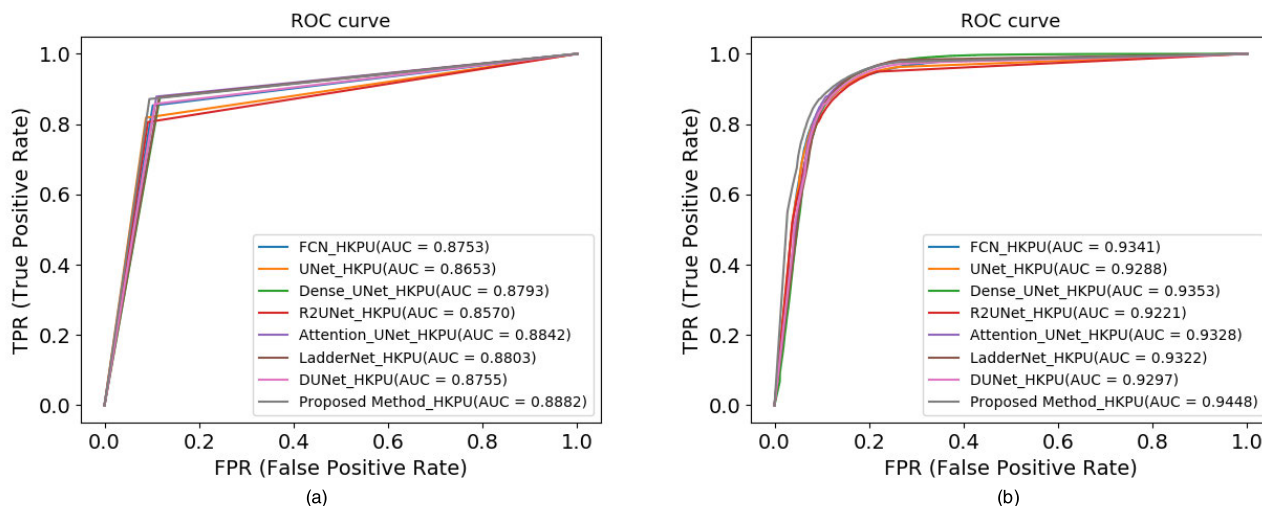
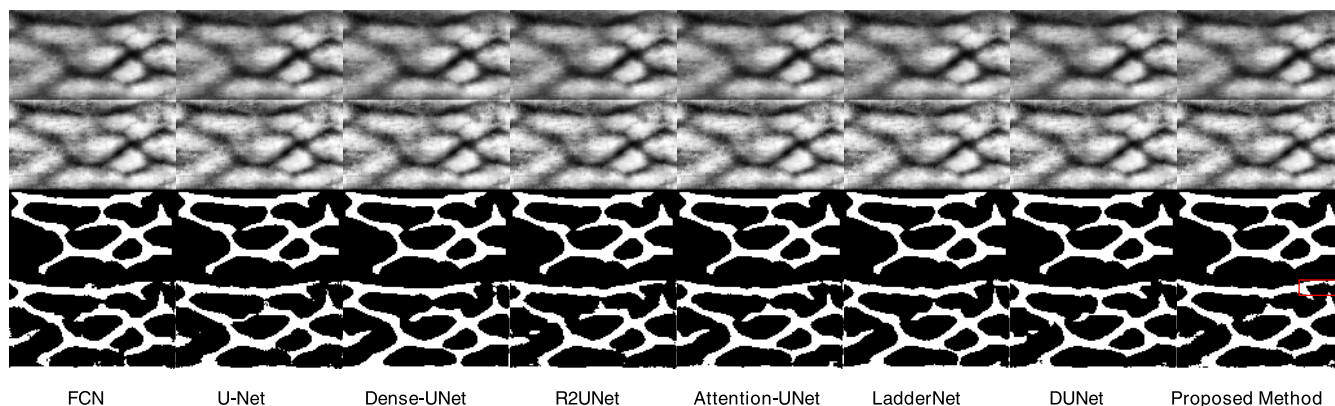**FIGURE 20.** ROC curves of different models on HKPU dataset.

FIGURE 21. Examples of segmentation results for various models on HKPU dataset (without CRF-RNN).
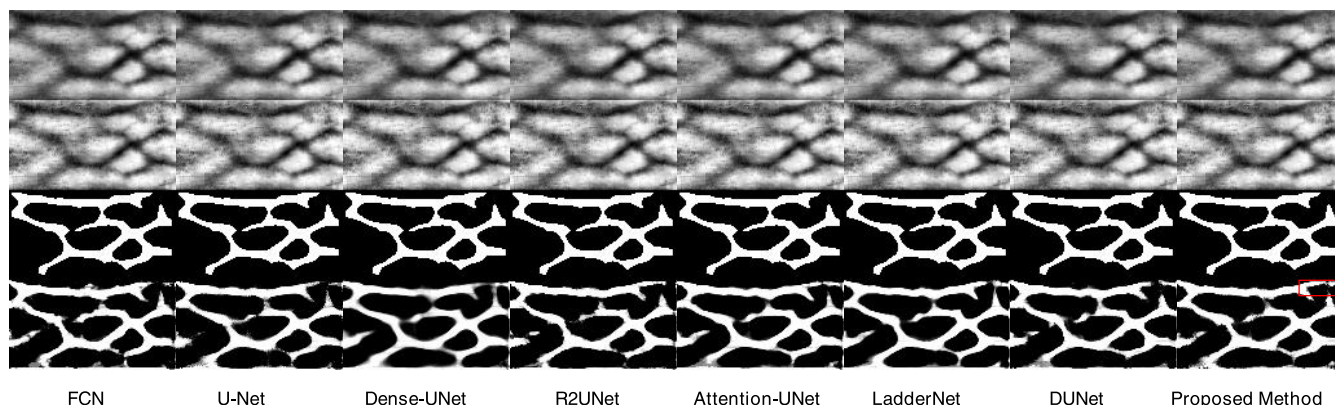

FIGURE 22. Examples of segmentation results for various models on HKPU dataset (with CRF-RNN).

the segmentation result, while FCN still gets the worst result, and gets worse verification performance than the original ROI together with Dense-UNet does. The performance of other models is better than that the original ROI.

C. FINGER VEIN VERIFICATION BASED ON HKPU DATASET

Table 9 concludes the test performance metrics after introducing deformable convolution and residual recurrent

convolution. Compared with the other two datasets, all metrics of the dataset are improved after adding residual recurrent convolution. The reason is that after receiving the features from deformable convolution, residual recurrent convolution not only accumulates and retains the effective features, but also further excavates deeper features. This is reflected in Fig. 11(c), where the residual recurrent convolution allows the veins to be connected at the venous rupture.

TABLE 12. Verification performance of various models on HKPU dataset (with CRF-RNN).

Methods	Year	EER(%)
Original ROI	—	6.415
FCN[49]	2015	7.088
UNet[20]	2015	5.681
Dense-UNet[35]	2018	4.570
R2UNet[36]	2018	5.96
Attention-UNet[34]	2018	5.045
LadderNet[37]	2018	5.556
DUNet[38]	2018	4.087
Proposed Method	2020	**2.372**
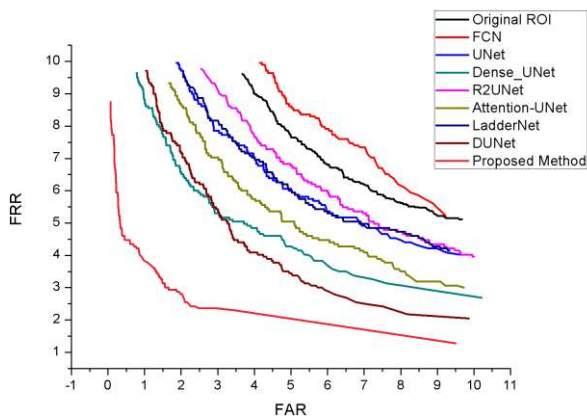
FIGURE 23. EER curves under different models.

Fig. 10(c) also shows their ROC curves, according to which there is no doubt that the model we proposed defeats the other ones.

When evaluating the performance of the proposed model and the current mainstream segmentation models, we find that the evaluation results based on the HKPU dataset share great similarities in changes with the other two datasets of SDUMLA and MMCBNU, and the proposed model still occupies the most excellent performance level. Table 10 shows the metrics of the non-embedded conditional random field. The fully convolutional neural network proposed ranks first with the accuracy of 0.8962, and most metrics are the highest among their corresponding columns. Fig. 20(a) also shows the highest AUC at 0.8882. The predicted segmentation results are summarized in Fig. 21. Even though the highly similar metrics of each model can diminish the judging accuracy of bare eyes, the vein pixels at the boundary can still be retained through the feature accumulation of limited steps. We notice that all the data of performance metrics of this dataset are between those of SDUMLA and MMCBNU. In fact, the image quality of MMCBNU and HKPU shows not much difference, because HKPU inputs the enhanced image to the original network and erased the low contrast of the low-quality image, which leads to these results.

The ROC curve in Fig. 20(b) has been greatly improved due to the effect of conditional random fields, with our proposed method reaching AUC of 0.9448. The performance metrics in Table 11 again certify the possibility of enhancing the performance of vein segmentation task on the basis of the fully convolutional neural network and conditional random fields. After analysing the segmentation results in Fig. 22, we can summarize that the conditional random field further refines the output of the fully convolutional neural network, and comprehensively considers the complex dependence of the venous space to produce more excellent segmentation performance.

The EER of each model is shown in Table 12 and Figure 23. The experimental results appear with the smallest EER of our proposed method at 2.732% yet the largest improvement compared with other models. Interesting enough, the performance of FCN performs worst no matter in segmentation results or in verification performance among all three datasets, while our proposed model gains the most outstanding results. Although several methods do not perform as well as ROI does, most of them get better results than the original ROI does, which proves the effectiveness of pattern extraction.

VI. CONCLUSION

In this paper, we propose a new fully convolutional neural network to fulfil the finger vein segmentation task in a pixel-wise manner. As an extension of U-Net, it can well integrate the advantages of recurrent neural network, residual network and deformable convolution network. Moreover, by adding an offset to the sampling grid of normal convolution, the convolution operation can adaptively adjust the receptive fields according to the size and shape of the vein blood vessels, so as to facilitate the capture of them. These more complex and deeper features are mined and accumulated by the recurre-nt neural network and residual network. In addition, conditional random field is introduced to refine the output of fully convolutional neural network by controlling smooth constraints such as image edge, appearance consistency and space consistency when assigning labels, which heightens the accuracy of the segmentation results. The conditional random field can be inferred as the embedment of the whole network model, while the conventional back propagation algorithm is used for end-to-end training. Every equivalent model is trained from scratch, with their performances compared with each other. The experimental results on three public finger vein datasets manifest that the fully convolutional neural network proposed in this paper earns superior segmentation results to other equivalent models, and the embedded conditional random field also further improves the performance of the system. Furthermore, we delve into the effect of segmentation results on verification performance. It is discovered that the quality of segmentation results can indirectly affect the performance of verification with keeping the changing trend consistent, while the quality of dataset can directly affect the performance of verification. In segmentation and verification,

the proposed methods receive the most satisfactory results.

In the future, we will design a classification network specifically for finger vein verification to further improve accuracy. We also expect to extend the proposed fully convolutional neural network to three dimensions, so as to obtain more accurate results in various medical image segmentation tasks.

REFERENCES

[1] A. Jain, L. Hong, and R. Bolle, "On-line fingerprint verification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 4, pp. 302–314, Apr. 1997, doi: 10.1109/34.587996.

[2] D. Zhang, W.-K. Kong, J. You, and M. Wong, "Online palmprint identification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 9, pp. 1041–1050, Sep. 2003, doi: 10.1109/TPAMI.2003.1227981.

[3] A. Kumar and Y. Zhou, "Human identification using finger images," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 2228–2244, Apr. 2012, doi: 10.1109/TIP.2011.2171697.

[4] L. Yang, G. Yang, X. Xi, K. Su, Q. Chen, and Y. Yin, "Finger vein code: From indexing to matching," *IEEE Trans. Inf. Forensics Security*, vol. 14, no. 5, pp. 1210–1223, May 2019, doi: 10.1109/TIFS.2018.2871778.

[5] A. Kumar and K. V. Prathyusha, "Personal authentication using hand vein triangulation and knuckle shape," *IEEE Trans. Image Process.*, vol. 18, no. 9, pp. 2127–2136, Sep. 2009, doi: 10.1109/TIP.2009.2023153.

[6] 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, doi: 10.1109/TIFS.2011.2158423.

[7] V. P. Kshirsagar, M. R. Baviskar, and M. E. Gaikwad, "Face recognition using eigenfaces," in *Proc. 3rd Int. Conf. Comput. Res. Develop.*, Maui, HI, USA, vol. 2, Mar. 2011, pp. 302–306, doi: 10.1109/ICCRD.2011.5764137.

[8] J. Daugman, "How iris recognition works," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 1, pp. 21–30, Jan. 2004, doi: 10.1109/TCSVT.2003.818350.

[9] J. Ramirez, J. C. Segura, J. M. Gorriz, and L. Garcia, "Improved voice activity detection using contextual multiple hypothesis testing for robust speech recognition," *IEEE Trans. Audio, Speech Lang. Process.*, vol. 15, no. 8, pp. 2177–2189, Nov. 2007, doi: 10.1109/TASL.2007.903937.

[10] Niyogi and Adelson, "Analyzing and recognizing walking figures in XYT," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Seattle, WA, USA, Jun. 1994, pp. 469–474, doi: 10.1109/CVPR.1994.323868.

[11] M. A. El-Yacoubi, M. Gilloux, and J.-M. Bertille, "A statistical approach for phrase location and recognition within a text line: An application to street name recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 2, pp. 172–188, Feb. 2002, doi: 10.1109/34.982898.

[12] B. Huang, Y. Dai, R. Li, D. Tang, and W. Li, "Finger-vein authentication based on wide line detector and pattern normalization," in *Proc. 20th Int. Conf. Pattern Recognit.*, Istanbul, Turkey, Aug. 2010, pp. 1269–1272, doi: 10.1109/ICPR.2010.316.

[13] W. Song, T. Kim, H. C. Kim, J. H. Choi, H.-J. Kong, and S.-R. Lee, "A finger-vein verification system using mean curvature," *Pattern Recognit. Lett.*, vol. 32, no. 11, pp. 1541–1547, Aug. 2011, doi: 10.1016/j.patrec.2011.04.021.

[14] N. Miura, A. Nagasaka, and T. Miyatake, "Extraction of finger-vein patterns using maximum curvature points in image profiles," *IEICE Trans. Inf. Syst.*, vols. E90–D, no. 8, pp. 1185–1194, Aug. 2007, doi: 10.1093/ietisy/e90-d.8.1185.

[15] J. Yang and Y. Shi, "Towards finger-vein image restoration and enhancement for finger-vein recognition," *Inf. Sci.*, vol. 268, pp. 33–52, Jun. 2014, doi: 10.1016/j.ins.2013.10.009.

[16] S. Zagoruyko and N. Komodakis, "Learning to compare image patches via convolutional neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Boston, MA, USA, Jun. 2015, pp. 4353–4361, doi: 10.1109/CVPR.2015.7299064.

[17] M. Liang and X. Hu, "Recurrent convolutional neural network for object recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Boston, MA, USA, Jun. 2015, pp. 3367–3375, doi: 10.1109/CVPR.2015.7298958.

[18] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 770–778, doi: 10.1109/CVPR.2016.90.

[19] J. Dai, H. Qi, Y. Xiong, Y. Li, G. Zhang, H. Hu, and Y. Wei, "Deformable convolutional networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Venice, Italy, Oct. 2017, pp. 764–773, doi: 10.1109/ICCV.2017.89.

[20] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2015, doi: 10.1007/978-3-662-54345-0_3.

[21] S. Zheng, S. Jayasumana, B. Romera-Paredes, V. Vineet, Z. Su, D. Du, C. Huang, and P. H. S. Torr, "Conditional random fields as recurrent neural networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Santiago, Chile, Dec. 2015, pp. 1529–1537, doi: 10.1109/ICCV.2015.179.

[22] M. Kono, H. Ueki, and S.-I. Umemura, "A new method for the identification of individuals by using of vein pattern matching of a finger," in *Proc. 5th Symp. Pattern Meas.*, 2000, pp. 9–12.

[23] S. Qiu, Y. Liu, Y. Zhou, J. Huang, and Y. Nie, "Finger-vein recognition based on dual-sliding window localization and pseudo-elliptical transformer," *Expert Syst. Appl.*, vol. 64, pp. 618–632, Dec. 2016, doi: 10.1016/j.eswa.2016.08.031.

[24] W. Yang, X. Huang, F. Zhou, and Q. Liao, "Comparative competitive coding for personal identification by using finger vein and finger dorsal texture fusion," *Inf. Sci.*, vol. 268, pp. 20–32, Jun. 2014, doi: 10.1016/j.ins.2013.10.010.

[25] J. Yang, Y. Shi, and J. Yang, "Finger-vein recognition based on a bank of Gabor filters," in *Proc. Asian Conf. Comput. Vis.*, Xi'an, China, Sep. 2009, pp. 374–383, doi: 10.1007/978-3-642-12307-8_35.

[26] S. Chaudhuri, S. Chatterjee, N. Katz, M. Nelson, and M. Goldbaum, "Detection of blood vessels in retinal images using two-dimensional matched filters," *IEEE Trans. Med. Imag.*, vol. 8, no. 3, pp. 263–269, Sep. 1989, doi: 10.1109/42.34715.

[27] N. Miura, A. Nagasaka, and T. Miyatake, "Feature extraction of finger-vein patterns based on repeated line tracking and its application to personal identification," *Mach. Vis. Appl.*, vol. 15, no. 4, pp. 194–203, Oct. 2004, doi: 10.1007/s00138-004-0149-2.

[28] T. Liu, J. B. Xie, W. Yan, P. Q. Li, and H. Z. Lu, "An algorithm for finger-vein segmentation based on modified repeated line tracking," *Imag. Sci. J.*, vol. 61, no. 6, pp. 491–502, Jul. 2013, doi: 10.1179/1743131X12Y.0000000013.

[29] P. Gupta and P. Gupta, "An accurate finger vein based verification system," *Digit. Signal Process.*, vol. 38, pp. 43–52, Mar. 2015, doi: 10.1016/j.dsp.2014.12.003.

[30] H. Qin, "Region growth–based feature extraction method for finger-vein recognition," *Opt. Eng.*, vol. 50, no. 5, May 2011, Art. no. 057208, doi: 10.1117/1.3572129.

[31] H. Qin, X. He, X. Yao, and H. Li, "Finger-vein verification based on the curvature in radon space," *Expert Syst. Appl.*, vol. 82, pp. 151–161, Oct. 2017, doi: 10.1016/j.eswa.2017.03.068.

[32] M. A. Syarif, T. S. Ong, A. B. J. Teoh, and C. Tee, "Enhanced maximum curvature descriptors for finger vein verification," *Multimedia Tools Appl.*, vol. 76, no. 5, pp. 6859–6887, Mar. 2017, doi: 10.1007/s11042-016-3315-4.

[33] L. Yang, G. Yang, Y. Yin, and X. Xi, "Finger vein recognition with anatomy structure analysis," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 28, no. 8, pp. 1892–1905, Aug. 2018, doi: 10.1109/TCSVT.2017.2684833.

[34] O. Oktay, J. Schlemper, L. Le Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz, B. Glocker, and D. Rueckert, "Attention U-net: Learning where to look for the pancreas," 2018, *arXiv:1804.03999*. [Online]. Available: http://arxiv.org/abs/1804.03999

[35] X. Li, H. Chen, X. Qi, Q. Dou, C.-W. Fu, and P.-A. Heng, "H-DenseUNet: Hybrid densely connected UNet for liver and tumor segmentation from CT volumes," *IEEE Trans. Med. Imag.*, vol. 37, no. 12, pp. 2663–2674, Dec. 2018, doi: 10.1109/TMI.2018.2845918.

[36] M. Zahangir Alom, M. Hasan, C. Yakopcic, T. M. Taha, and V. K. Asari, "Recurrent residual convolutional neural network based on U-Net (R2U-Net) for medical image segmentation," 2018, *arXiv:1802.06955*. [Online]. Available: http://arxiv.org/abs/1802.06955

[37] J. Zhuang, "LadderNet: Multi-path networks based on U-Net for medical image segmentation," 2018, *arXiv:1810.07810*. [Online]. Available: http://arxiv.org/abs/1810.07810

[38] Q. Jin, Z. Meng, T. D. Pham, Q. Chen, L. Wei, and R. Su, "DUNet: A deformable network for retinal vessel segmentation," *Knowl.-Based Syst.*, vol. 178, pp. 149–162, Aug. 2019, doi: 10.1016/j.knosys.2019.04.025.

[39] H. Qin and M. A. El-Yacoubi, "Deep representation-based feature extraction and recovering for finger-vein verification," *IEEE Trans. Inf. Forensics Security*, vol. 12, no. 8, pp. 1816–1829, Aug. 2017, doi: 10.1109/TIFS.2017.2689724.

[40] H. Qin and P. Wang, "Finger-vein verification based on LSTM recurrent neural networks," *Appl. Sci.*, vol. 9, no. 8, p. 1687, Apr. 2019, doi: 10.3390/app9081687.

[41] W. Yang, C. Hui, Z. Chen, J.-H. Xue, and Q. Liao, "FV-GAN: Finger vein representation using generative adversarial networks," *IEEE Trans. Inf. Forensics Security*, vol. 14, no. 9, pp. 2512–2524, Sep. 2019, doi: 10.1109/TIFS.2019.2902819.

[42] E. Jalilian and A. Uhl, "Finger-vein recognition using deep fully convolutional neural semantic networks: The impact of training data," in *Proc. IEEE Int. Workshop Inf. Forensics Secur. (WIFS)*, Hong Kong, Dec. 2018, pp. 1–8, doi: 10.1109/WIFS.2018.8630794.

[43] E. Jalilian and A. Uhl, "Improved CNN-segmentation-based finger vein recognition using automatically generated and fused training labels," *Handbook Vascular Biometrics*. Cham, Switzerland: Springer, 2019, pp. 200–223, doi: 10.1007/978-3-030-27731-4_8.

[44] E. Jalilian and A. Uhl, "Enhanced segmentation-CNN based finger-vein recognition by joint training with automatically generated and manual labels," in *Proc. IEEE 5th Int. Conf. Identity, Secur., Behav. Anal. (ISBA)*, Jan. 2019, pp. 1–8, doi: 10.1109/ISBA.2019.8778522.

[45] A. Graves, A.-R. Mohamed, and G. Hinton, "Speech recognition with deep recurrent neural networks," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Vancouver, BC, Canada, May 2013, pp. 6645–6649, doi: 10.1109/ICASSP.2013.6638947.

[46] A. Graves and J. Schmidhuber, "Offline handwriting recognition with multidimensional recurrent neural networks," in *Proc. 21st Int. Conf. Neural Inf. Process. Syst.*, Vancouver, BC, Canada, 2009, pp. 545–552, doi: 10.1007/978-3-540-74690-4_56.

[47] 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

[48] M. Z. Alom, M. Hasan, C. Yakopcic, T. M. Taha, and V. K. Asari, "Improved inception-residual convolutional neural network for object recognition," *Neural Comput. Appl.*, vol. 32, no. 1, pp. 279–293, Jan. 2020, doi: 10.1007/s00521-018-3627-6.

[49] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Boston, MA, USA, Jun. 2015, pp. 3431–3440, doi: 10.1109/CVPR.2015.7298965.

[50] P. Krähenbühl and V. Koltun, "Efficient inference in fully connected CRFs with Gaussian edge potentials," 2012, *arXiv:1210.5644*. [Online]. Available: http://arxiv.org/abs/1210.5644

[51] L. Ladicky, C. Russell, P. Kohli, and P. H. S. Torr, "Associative hierarchical CRFs for object class image segmentation," in *Proc. IEEE 12th Int. Conf. Comput. Vis.*, Kyoto, Japan, Sep. 2009, pp. 739–746, doi: 10.1109/ICCV.2009.5459248.

[52] H. Fu, Y. Xu, S. Lin, D. W. K. Wong, and J. Liu, "DeepVessel: Retinal vessel segmentation via deep learning and conditional random field," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2016, pp. 132–139, doi: 10.1007/978-3-319-46723-8_16.

[53] T. Su, "Research of finger vein recognition based on deep learning," M.S. thesis, South China Univ. Technol., Guangzhou, China, 2018, pp. 20–23.

[54] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," 2015, *arXiv:1502.03167*. [Online]. Available: http://arxiv.org/abs/1502.03167

[55] Y. Yin, L. Liu, and X. Sun, "SDUMLA-HMT: A multimodal biometric database," in *Proc. Chin. Conf. Biometric Recognit.* Berlin, Germany: Springer, 2011, pp. 260–268, doi: 10.1007/978-3-642-25449-9_33.

[56] Y. Lu, S. J. Xie, S. Yoon, Z. Wang, and D. S. Park, "An available database for the research of finger vein recognition," in *Proc. 6th Int. Congr. Image Signal Process. (CISP)*, Hangzhou, China, Dec. 2013, pp. 410–415, doi: 10.1109/CISP.2013.6744030.

[57] H. Hong, M. Lee, and K. Park, "Convolutional neural network-based finger-vein recognition using NIR image sensors," *Sensors*, vol. 17, no. 6, p. 1297, Jun. 2017, doi: 10.3390/s17061297.

JUNYING ZENG received the master's degree in physical electronic from Yunnan University, Yunnan, China, in 2005, and the Ph.D. degree in physical electronics from the Beijing University of Posts and Telecommunications, Beijing, China, in 2008. Since June 2008, he has been working with the Department of Intelligence Manufacturing, Wuyi University, Guangdong, China. He is currently an Associate Professor with Wuyi University. His research interests include image understanding, deep learning, and signal processing.

FAN WANG received the B.S. degree from the Qilu University of Technology (Shandong Academy of Sciences), in 2016. He is currently pursuing the master's degree with the Department of Intelligence Manufacturing, Wuyi University. His research interests include biometric identification and deep learning.
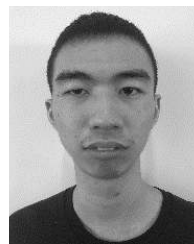
JIANXIANG DENG is currently pursuing the degree with the Department of Intelligence Manufacturing, Wuyi University. His research interests include image segmentation of medical imaging and deep learning.

CHUANBO QIN received the B.S. and M.S. degrees from WYU University, China, in 2004 and 2008, respectively, and the Ph.D. degree from the Department of Automation Science and Engineering, South China University of Technology, in 2015. He is currently a Lecturer with WYU University. His main research interests are in medical image segmentation and biometric recognition.

YIKUI ZHAI (Member, IEEE) received the bachelor's degree in optical electronics information and communication engineering and the master's degree in signal and information processing from Shantou University, Guangdong, China, in 2004 and 2007, respectively, and the Ph.D. degree in signal and information processing from Beihang University, in June 2013. Since October 2007, he has been working with the Department of Intelligence Manufacturing, Wuyi University, Guangdong. Since 2016, he has been a Visiting Scholar with the Department of Computer Science, University of Milan. He is currently an Associate Professor with Wuyi University. His research interests include image processing, deep learning, and pattern recognition.

JUNYING GAN (Member, IEEE) received the B.S., M.S., and Ph.D. degrees in electrical information engineering from Beihang University, in 1987, 1992, and 2003, respectively. She joined Wuyi University, Guangdong, China, in 1992, where she is currently a Full Professor. She is also the Executive Director of the Guangdong Image Graphics Association. She has published more than 50 journal articles. Her research interests include biometric extraction and pattern recognition. She has received several provincial technology awards.

VINCENZO PIURI (Fellow, IEEE) received the M.S. and Ph.D. degrees in computer engineering from the Politecnico di Milano, Italy. From 1992 to 2000, he was an Associate Professor with the Politecnico di Milano. From 1996 to 1999, he was a Visiting Professor with The University of Texas at Austin, USA. Since 2000, he has been a Full Professor with the University of Milan, Italy. From 2007 to 2012, he was the Department Chair with the University of Milan. From 2012 to 2016, he was a Visiting Researcher with George Mason University, USA. He founded a start-up company, Sensure srl, in the area of intelligent systems for industrial applications (leading it from 2007 to 2010). He was active in industrial research projects with several companies. His main research and industrial application interests are in intelligent systems, computational intelligence, pattern analysis and recognition, machine learning, signal and image processing, biometrics, intelligent measurement systems, industrial applications, distributed processing systems, the Internet-of-Things, cloud computing, fault tolerance, application-specific digital processing architectures, and arithmetic architectures. He is a Fellow of ACM.

• • •