

Received April 5, 2021, accepted May 22, 2021, date of publication May 26, 2021, date of current version June 4, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3084037

# DGLFV: Deep Generalized Label Algorithm for Finger-Vein Recognition

ZHIYONG TAO<sup>1</sup>, HAOTONG WANG<sup>2</sup>, YALEI HU<sup>3</sup>, YUEMING HAN<sup>1</sup>,  
SEN LIN<sup>4</sup>, AND YING LIU<sup>1</sup>

<sup>1</sup>School of Electronic and Information Engineering, Liaoning Technical University, Huludao 125105, China

<sup>2</sup>Graduate School of Information, Production and Systems, Waseda University, Tokyo 169-8050, Japan

<sup>3</sup>Monitoring Branch, Tiandi (Changzhou) Automation Company Ltd., Changzhou 213001, China

<sup>4</sup>School of Automation and Electrical Engineering, Shenyang Ligong University, Shenyang 110159, China

Corresponding author: Zhiyong Tao (xyzmail@126.com)

This work was supported in part by the National Key Research and Development Program of China under Grant 2018YFB1403303, and in part by the Science and Technology Tackling Key Project of the Educational Department of Liaoning Province under Grant LJ2020ZD005.

**ABSTRACT** With the growing demand for information security, finger-vein recognition has become widespread. However, the robustness of the recognition process becomes a major problem. When identifying unseen categories with traditional finger-vein recognition systems, a few issues remain, such as recognition interference and low efficiency. This paper proposes a Deep Generalized Label Finger-Vein (DGLFV) model to extract feature maps and achieve high-accuracy recognition. The largest rectangular finger-vein region is extracted through image semantic segmentation and the advanced bidirectional traversing and center diffusion method for the known categories. Then we generalize all the unseen categories actively as Class  $C + 1$  to reduce interference from unregistered users. Furthermore, an adaptive threshold acquisition algorithm is proposed for Label Receiver Operating Characteristic (LROC), so that the procedures of classification, recognition, and verification are unified. Apart from Shandong University Homologous Multi-modal Traits (SDUMLA-HMT), we have conducted additional experiments on our self-built database, Finger Veins of Signal and Information Processing Laboratory (FV-SIPL). The recognition accuracy of the approach proposed in this paper has reached 99.25% and 99.08% testing on FV-SIPL and SDUMLA-HMT, with a low error rate at 1.481% and 2.228% and little time consumption of 0.157s for a single image, which is better than most state-of-the-art finger-vein recognition methods.

**INDEX TERMS** DGLFV, finger-vein recognition, generalized category, FV-SIPL, SDUMLA-HMT.

## I. INTRODUCTION

Finger-vein recognition belongs to the field of biometrics, similar to some other types, such as faces, fingerprints, irises, and palmprints. Individuals are usually identified through personal characteristics. With the rapid development of technologies, people pay much more attention to security certification. However, deceitful and fabricated systems have emerged one after another, leading to countermeasure systems. They may face a challenging situation when biometric recognition is applied in security prevention and other scenarios. The recognition of finger veins has been a research hotspot in biometrics because of its uniqueness and immutability [1].

The associate editor coordinating the review of this manuscript and approving it for publication was Wei Wang<sup>1</sup>.

## A. RELATED WORK

The majority of finger-vein recognition systems mainly contain two modes: verification (1: 1) and recognition (1: N). Generally, these models are divided into two types: learning and non-learning. In non-learning models, Gabor filters [2] are considered as an effective extractor of vein lines. Kumar and Zhou [3] employed a judicious combination of morphological operations and even Gabor filters to achieve shape features. Some other researchers employ Local Binary Patterns (LBP) or Line Local Binary Patterns (LLBP) for the extraction of binary vein texture features [4]; the method like Scale Invariant Feature Transformation (SIFT) [2] is adopted to extract the features of specific points. In terms of matching strategies, most people choose the Hamming distance [4] and the Euclidean distance [5] to convert the features into vectors and enhance the recognition robustness. Yang *et al.* [6]

analyzed the anatomy structure and vein extraction (ASAVE) combined with integration matching strategy for performance improvement.

However, the above algorithms are susceptible to the severe impact of some problems, such as low picture quality, non-finger region inter vein-texture block loss, and low robustness. Therefore, people have gradually introduced some learning-based algorithms [7], which enjoy strong adaptability and are not prone to the environment. Reference [8] proposed using Gabor filters to extract features from finger veins and Support Vector Machines (SVM) for classification. They introduced Principal Component Analysis (PCA) to obtain the essential information and neural network for classification matching [9]. In learning models, Region of Interest (RoI) extraction is an essential step for non-finger region elimination, noisy data preprocessing, and further training. [10] improved the RoI extraction of finger vein in knuckle region. The feature images were transformed into particle features after feature enhancement, and the layered hypersphere method was used for recognition.

The block execution of the system increased research complexity. Recently, Convolutional Neural Network (CNN) has become more widely accepted in the field of biometrics. Deep feature extraction has improved the performance and robustness of finger-vein recognition. Ahmad Radzi *et al.* [11] proposed a 4-layer hybrid down-sampling CNN to classify finger veins. Reference [7] proposed to utilize Visual Geometry Group (VGG), in which the input was a set of pictures combining the testing samples with the training ones. Also, it aimed to acquire the correlation between the two ones and finally output the binary classification results. As shown in [12], deep convolutional networks play a crucial role in proper feature extraction. The recognition accuracy directly reflects the security level. Xie and Kumar [14] succeeded in finger-vein authentication using the CNN and supervised discrete hashing. Except for extraction method improvement, [14] extracted and recovered the finger-vein feature through the prior knowledge to enhance the feature quality.

Although such methods performed well in general finger-vein recognition tasks, it is possible that the unseen finger vein from unregistered people can be recognized as some label. There is a need for subjective finger-vein feature extraction and generalization.

## B. CONTRIBUTIONS

Focusing on the recognition accuracy and generalization capability, we propose a deep neural network called Deep Generalized Label Finger-Vein (DGLFV). Our main contributions are as follows:

### 1) MASK-RCNN AND PROPOSED ADAPTIVE THRESHOLD ACQUISITION ALGORITHM FOR MASK SEGMENTATION

Firstly, the image semantic segmentation model based on deep learning is applied to the RoI extraction of finger veins. We use the advanced bidirectional traversing and center diffusion method to extract the maximum RoI rectangle for the

irregular mask of finger veins. Based on Mask Region Convolutional Neural Network (Mask-RCNN) Mask-RCNN [15], a deeper convolution model was proposed for RoI extraction. Then we select the Regional Production Network (RPN) [16] to scout out the interest area. After that, the appropriate areas of finger veins were captured through the maximum rectangle to facilitate subsequent image enhancement and recognition. The successful extraction rate can reach 100% on Shandong University Homologous Multi-modal Traits (SDUMLA-HMT) [17]. In addition, an adaptive threshold acquisition algorithm is proposed to complete the intelligent matching strategy, which further improves the robustness.

Different decision-making algorithms are utilized according to different systems, which can improve the overall performance. In this paper, the LROC verification threshold self-adaption acquisition algorithm for the DGLFV model is proposed. We utilize the *Youden* index to find the best reception threshold for each class, and the maximum refers to the LROC separation threshold.

### 2) CLASS GENERALIZATION AND DGLFV CONSTRUCTION

In practice, new samples of unseen categories are the central factor that affects the recognition accuracy. The traditional passive comparison in finger-vein systems will cause serious recognition interference, low recognition efficiency, and bad recognition effects when identifying unseen samples. Reference [7] can deal with the verification of new finger veins, but lack of practice and much time consumption. It is necessary to select the primary comparison samples and the time consumption of N comparisons in the recognition mode. Our model focuses on not only the accuracy and time consumption but also generalization capability.

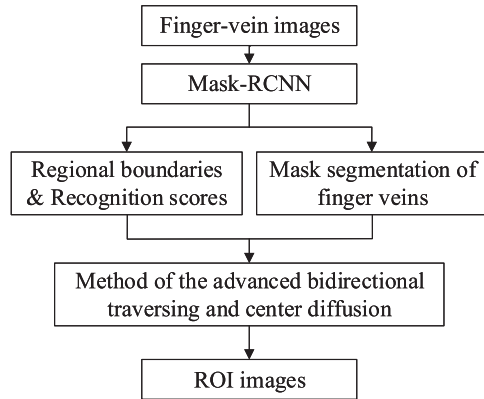
Based on well-extracted RoI and generalization, DGLFV has gained a low error rate, accounted for 1.481% and 2.228%. We have conducted additional experiments on our database called FV-SIPL to verify the effectiveness further. The experimental results are better than most of the latest finger-vein recognition models, with significant application value.

## II. PROPOSED ALGORITHM

### A. FINGER-VEIN RoI EXTRACTION ALGORITHM

Because of finger postures and environmental factors, some inaccurate segmentation problems exist in traditional finger-vein RoI extraction methods, such as direct method [18] and finger-knuckle positioning extraction [19]. This paper introduces the concept of image semantic segmentation to improve adaptability and robustness. Compared with traditional image segmentation algorithms, it can complete more complicated segmentation tasks. The RoI extraction process is shown in Fig.1.

1) FINGER-VEIN MASK EXTRACTION BASED ON MASK-RCNN  
The structure of Mask-RCNN is shown in Fig.2. We extract feature maps of original images by ResNet-101 [20], with



**FIGURE 1.** In this paper, Mask-RCNN is implemented to segment finger veins semantically and obtain the masks, region boundaries, and recognition category probabilities. According to the advanced bidirectional traversing and center diffusion method, the model has access to good regions of interest (RoI) by the maximum rectangular extraction.

shared weights. Afterward, the RPN network offers optimal bounding boxes and recognition probabilities. We normalized multi-target particular region mapping features through the RoI correction algorithm. In this way, part of the features has been connected to the fully connected layer, giving errors between the probabilities and bounding results; others serve as the input of the fully convolutional neural network, from which the mask outputs by threshold suppression.

- 1) We extract feature maps through ResNet-101 and pass them to the RPN model. Region Proposal Network (RPN) adopts the feature pyramid network and anchor working mechanism. The multi-scale feature maps are extracted by the convolutional layer of the neural network; then, we inversely reverse the smallest feature maps by up-sampling. Finally, it fuses with the original features to construct multi-scale semantic features, which are considered as the unpredicted target.
- 2) After giving all possible bounding boxes, we should determine whether the results are the target or background. Therefore, we assign a binary-class label to each anchor and introduce  $IoU$ , the intersection, and the union ratio of the anchor selection boxes and the ground. In (1),  $S_o$  is the intersection area of two boxes, and  $S_U$  is the union area.

$$IoU = S_o/S_U \tag{1}$$

Then we label all the boxes with more than 0.7  $IoU$  as positive, less than 0.3 as negative, while others not including targets will be corrected. If there are no positive samples (all the  $IoU$  are less than 0.7), we will label the anchor with the max  $IoU$ . In this paper, we define the loss function as (2). The former is classification loss, which calculates the error between bounding boxes and the truth; the latter is regression

loss of the yielded coordinate and the ground.

$$L(\{p_i\}, \{t_i\}) = \frac{1}{N_{cls}} \sum_i L_{cls}(p_i, p_i^*) + \lambda \frac{1}{N_{reg}} \sum_i p_i^* L_{reg}(t_i, t_i^*) \tag{2}$$

In this expression,  $i$  is the serial number of the anchors in each mini-batch.  $p_i$  denotes the probability of the target, and  $p_i^*$  is the label of 0 or 1.  $t_i$  denotes the four parameters of the prediction boxes, and  $t_i^*$  is the parameter of the real boxes. We define  $L_{cls}$  and  $L_{reg}$  as the classification and regression loss. Here,  $p_i^* L_{reg}$  represents the regression only for the positive-labeled samples (negative-labeled samples  $p_i^* = 0$ ). At last,  $p_i$  and  $t_i$  are the outcomes of classification and regression, respectively.

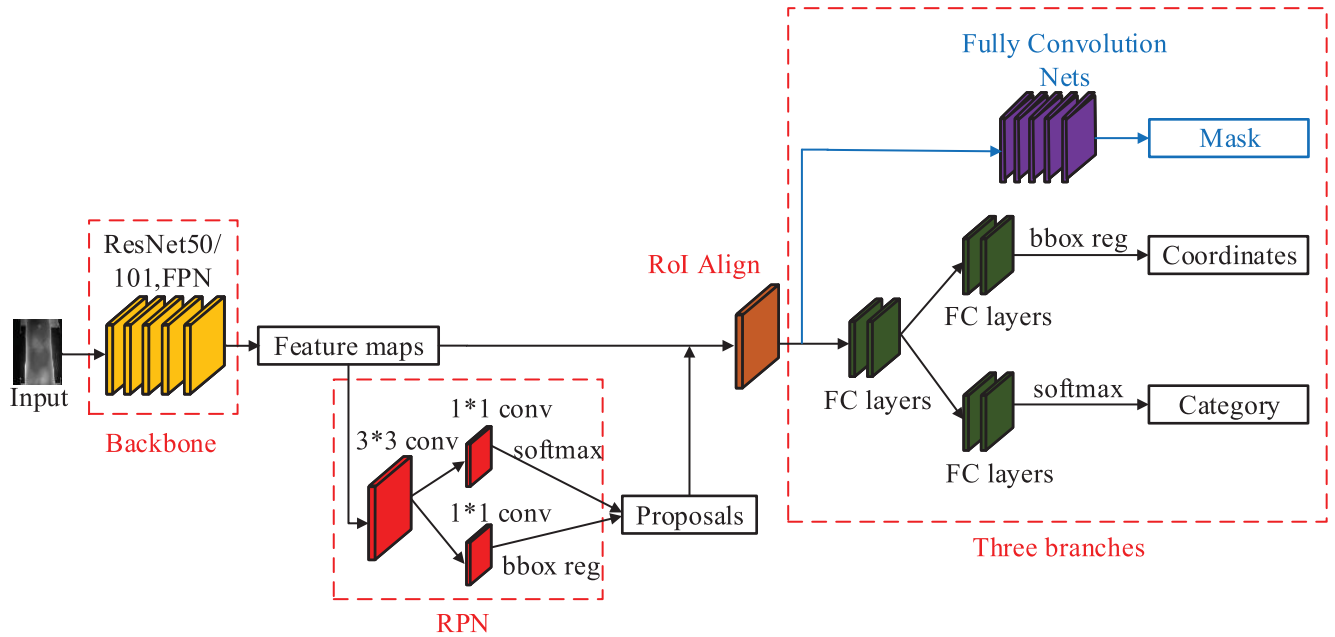
- 3) Concerning the region proposal suggestion and RPN prediction results, we have access to the ideal images with a shared convolution layer, called the correction output. Compared with Faster-RCNN [16], the RoI correction in Mask-RCNN cancels the quantization and proposes bilinear interpolation to obtain floating coordinates. After obtaining the correction map through Mask-RCNN, a pooling layer is added to normalize the feature maps.
- 4) A solid-size area feature map is generated through the RoI correction pooling layer. When adjusting the output of Conv5 in ResNet-101, we obtain a shared feature map for mask withdrawing and recognition. They also share feature maps to the fully connected layer, and the final result is obtained by Softmax and Sigmoid for regression. Mask-RCNN uses a multitasking loss function, defined as (3):

$$L = L_{cls} + L_{reg} + L_{mask} \tag{3}$$

Here, the classification loss of bounding boxes is  $L_{cls}$ , the regression loss of the box position is  $L_{reg}$ , and the loss of the mask area is  $L_{mask}$ .

## 2) FINGER-VEIN RoI EXTRACTION BASED ON THE ADVANCED BIDIRECTIONAL TRAVERSING AND CENTER DIFFUSION METHOD

With the regular identity of the finger-vein mask provided by Mask-RCNN, it needs to be further cropped into a rectangular RoI to transform into the recognition model. To retain as many features as possible, we need to extract the largest rectangle [21], [22] in the irregular region. This paper employs the advanced bidirectional traversing and center diffusion method to extract the rectangular RoI. Suppose the four vertical coordinates of the rectangle are  $P_1(x_1, y_1)$ ,  $P_2(x_2, y_1)$ ,  $P_3(x_1, y_2)$  and  $P_4(x_2, y_2)$ , then they form a rectangular area:  $R(x_1, x_2, y_1, y_2)$ , whose acreage is  $S_R = (x_2 - x_1)(y_2 - y_1)$ .  $Max(S_R)$  refers to the objective function.



**FIGURE 2.** Here is the advanced Mask-RCNN framework. We extract feature maps through the backbone structure of ResNet-101 and pass them to the RPN model. Region Proposal Network (RPN) adopts the feature pyramid network and anchor working mechanism to receive multi-scale proposals. A fully connected layer is used with corrected RoI region results to give out possibilities and bounding error. The final results will be obtained by Softmax and sigmoid regression.

**Algorithm 1** Horizontal Maximum Rectangle Area

**Input:** Set the initial rectangle coordinates as  $R(x_c, x_c + 1, y_c - 1)$   
**Output:** Horizontal maximum rectangle's  $R(x_c - 1, x_c + 1 + j, y_c - 1, y_c + 1)$   
 Set the initial rectangle coordinates as  $R(x_c, x_c + 1, y_c - 1)$   
**while**  $R(x_c, x_c + 1, y_c - 1)$  **do**  
     Increase the per-unit length of  $i$   
**end while**  
**while**  $R(x_c - 1, x_c + 1 + j, y_c - 1, y_c + 1 \in Q(q_n))$  **do**  
     Increase the per-unit length of  $j$   
**end while**

Irregular masks are made up of dots  $q_n(1, 2 \dots m)$ . We assume the enclosed area as  $Q(q_n)$ , so the constraints are as follows:

- The four vertices of the rectangle must be in the irregular area, which means  $P_k \in Q(q_n)$ ;
- Irregular mask points are limited to be outside the bounding rectangle ( $q_n \notin R$ ).

Mask-RCNN can give the optimal recognition frame when extracting the mask, regarded as region D ( $Q \in D$ ). Amongst the structure, the central point  $C(x_c, y_c)$  of the optimal recognition is deemed to be the cortex of the irregular mask. Therefore, this paper diffuses the rectangle to the horizontal and vertical directions from the center of  $C(x_c, y_c)$ , and optimizes the objective function w.r.t. the conditions.

The algorithm we proposed can be divided into two parts: the first is to take a single-direction maximal area rectangle in

**Algorithm 2** Lateral Degradation Process

**Input:** Horizontal maximum rectangle's  $R(x_c - 1, x_c + 1, y_c + 1)$   
**Output:** Horizontal degeneration coordinate and area dictionary  $Dict(h)$   
**for**  $k = 1, 2, 3, \dots$  **do**  
     **while**  $R(x_c, x_c + 1, y_c - 1)$  **do**  
         **if**  $R(x_c - 1 - i + K, x_c + 1 + j, y_c - 1, y_c + 1 \in Q(q_n))$   
         **then**  
              $Dict(h)^k$  Fixed horizontal axis, use Algorithm 1 to achieve a longitudinal maximum area rectangle  
             **else if**  $Dict(h)^k < Dict(h)^{k-\{1, \dots, 9\}}$  **then**  
                 break  
             **end if**  
         **end while**  
     **end for**

two directions. For example, when taking a maximum lateral area rectangle, we fix the vertical axis a length of 1 and traverse it to the maximum lateral boundary. At that time, to choose the maximum bounding, we are mainly confined to be outside or above the target's largest rectangle. Step 1 (the rectangle algorithm for finding the maximum area in the horizontal direction) is shown in Algorithm 1, with the same theory of the vertical one.

The second one is the degradation algorithm in Algorithm 2. With the exposure of the maximal rectangles, there needs to degenerate from these two directions. For instance, increasing the vertical axis value is an excellent

way when the lateral direction is degraded. Meanwhile, we put the coordinate and the area of degradation into the dictionary  $Dict(h)$ . Similarly, the dictionary for the vertical axis is  $Dict(v)$ . The degradation procedure will not continue until the current area covers ten times less than that of the previous. Finally, the dictionaries are compared to find out overlapping data, which should be sorted by the area's size. Then we achieve the maximum rectangle of the irregular masks. Algorithm 2 illustrates the extraction processing of the dictionary  $Dict(h)$  during lateral degradation.

## B. GENERALIZED CATEGORY FINGER-VEIN RECOGNITION MODEL BASED ON DEEP LEARNING

Finger-vein recognition is divided into classification and verification. The training-based algorithms are proved to be of high classification robustness. However, constructing these algorithms requires prior knowledge. In practice, the untrained samples will bring out severe interference with the classification results. Even though verification is secure, we should consider matching templates and the time loss of repeated identifications. We aim to solve the problems that unseen categories cannot be classified in traditional finger-vein algorithms, and passive comparison would be rejected and inefficient. Hereafter, we propose a recognition method based on a convolutional neural network to identify unseen samples and generalize them into a class. It is named the Deep Generalized Label Finger-vein (DGLFV) model, which utilizes the deep Inception-ResNet V2 [23] and constructs a neural network to identify unseen finger-vein categories. Based on this, we can explore the Softmax classification results. Consequently, the Label Receiver Operating Characteristic (LROC) is proposed to convert the classification results into verification ones. While ensuring the high accuracy of the known category, we achieve the aim of generalizing the unseen categories into one class ( $C+1$ ).

### 1) BASIC MODEL OF DEEP LEARNING

After introducing the convolutional calculation, the artificial neural network utilizes local interconnection and shared weights to make itself more consistent with biological neurons' sparse characteristics. CNN directly inputs the original pictures without much attention to image preprocessing. Further, a deeply hidden network can extract multi-level features [24] and obtain recognition results through subsequent calculation. At this time, only a complete CNN needs to be trained. Such an end-to-end idea enables CNN to train a unified objective function. The model with data contains more adjustment opportunities and increases the overall fitness.

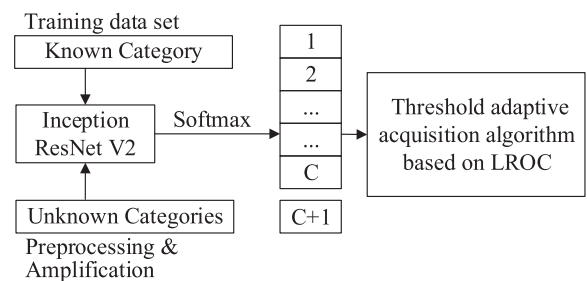
Inception-ResNet V2 is a model combining the ideas of Inception and ResNet. The Inception network was developed by GoogleNet [25], which tends to reduce parameters and invest parallel and asymmetric convolution, causing a widening network and effective feature extraction [26]. The ResNet network introduced a residual calculation methodology to strengthen the network's depth and improve performance [20]. Increasing the network's width or depth

will improve the performance and achieve a better result if keeping them in parallel. In this paper, we choose the Inception-ResNet V2 model, which achieves optimal performance by balancing the number of layers in the network and filters in each layer. Table 1 shows the detailed structure of each module in Inception-ResNet V2, from input to output. The input is a three-channel image with a size of  $224 \times 224$ , and the output is the classification probabilities. Inception-ResNet V2 proposed three types of modules: Stem, Inception-ResNet, and Reduction, in which both parallel and asymmetric convolution kernel structures are covered. It decreased the computation complexity with small enough information loss.

Large-scale data is one of the foundations of CNN training. Due to extraction costs, the number of stored images cannot afford CNN's training within huge parameters. Fine-tuning of weights can be performed using pre-trained models in ImageNet. Also, the amount of training data can be expanded by translating and rotating, which can prevent overfitting and enhance robustness.

### 2) DGLFV MODEL

This paper proposes a novel finger-vein recognition algorithm, DGLFV, based on a deep convolutional neural network. This model realizes an active identification of unseen categories then generalizes them into a class. The recognition accuracy has been improved through deepening more layers and formulating related matching strategies. The establishment and training process is illustrated in Fig.3. First, we preprocess and amplify the known and other categories of finger-vein images; There is a modification of the parameters in Inception-ResNet V2 and training two datasets, where  $C + 1$  is regarded as an unseen class; Based on the overall Softmax results, a matching strategy is developed to obtain each class's verification threshold.



**FIGURE 3.** We amplify the training dataset and modify the parameters in Inception-ResNet V2, where  $C + 1$  denotes an unseen class. Finally, based on the overall Softmax results, a matching strategy is developed to obtain the verification threshold for each class.

The unseen classes come from unregistered users rather than the finger-vein databases with labels. For this reason, the testing results for specific categories are approximately 0, ideally. By contrast, when using Softmax on CNN, the probability sum should be 1, and the probabilities of every unseen class are  $1/C$ .  $C$  is the number of registered user classes. The Softmax classifier adopts an approximate optimal strategy,

TABLE 1. Inception-ResNet V2 module framework.

Layer Names	Input	Conv1	Conv2	Conv3	Conv4	Output
Inception-ResNet-A	Relu	$1 \times 1 \times 32$	$3 \times 3 \times 32$	$3 \times 3 \times 64$	$1 \times 4 \times 384$ (Linear)	Sum & Relu
		$1 \times 1 \times 32$	$3 \times 3 \times 32$			
		$1 \times 1 \times 32$				
Direct Connection						
Inception-ResNet-B	Relu	$1 \times 1 \times 128$	$1 \times 7 \times 160$	$7 \times 1 \times 192$	$1 \times 1 \times 1154$ (Linear)	Sum & Relu
		$1 \times 1 \times 192$				
		$1 \times 1 \times 192$				
Direct Connection						
Inception-ResNet-C	Relu	$1 \times 1 \times 192$	$1 \times 3 \times 224$	$3 \times 1 \times 256$	$1 \times 1 \times 2048$ (Linear)	Sum & Relu
		$1 \times 1 \times 192$				
		$1 \times 1 \times 192$				
Direct Connection						
Reduction-A	Layers	$1 \times 1 \times 256$	$3 \times 3 \times 256$	$3 \times 3 \times 384$ (Stride 2)		Filter concat
		$3 \times 3 \times 384$ (Stride 2)				
		$3 \times 3$ MaxPool (Stride 2)				
Reduction-B	Layers	$1 \times 1 \times 256$	$3 \times 3 \times 288$	$3 \times 3 \times 320$ (Stride 2)		Filter concat
		$1 \times 1 \times 256$		$3 \times 3 \times 288$ (Stride 2)		
		$1 \times 1 \times 256$		$3 \times 3 \times 384$ (Stride 2)		
		$3 \times 3$ MaxPool (Stride 2)				

which can lead to a high-probability approximate event. It is triggered by the characteristics of Softmax. There is still low security and low accuracy when applying a threshold separation strategy. Therefore, we propose a method for generalizing unseen categories. Adding category  $C + 1$  (unseen categories) when the known category is identified, the approximate optimal result is attributed to Class  $C + 1$ . It is a condition that Class  $C + 1$  must have strong generalization identities.

In this paper, we have trained the new class through large-scale finger-vein images after the known ones. Consequently, the generalization ability of the unseen categories has been enhanced. Then we train and conclude all sorts of finger veins into Class  $C + 1$  for generalization. Account for the dataset expansion, the effects have been improved, which will be verified in the later experiment. To prevent over-fitting, researchers usually adopt the regularization item method, which can level up the generalization and the robustness of the model. When adding Class  $C + 1$ , The objective function  $J(\theta)$  will be shown as (4), in which  $R(\theta)$  is a regularization term. For Class  $C + 1$  models, the primary role of adding regularization is to minimize probability peaks and selecting Class  $C$  for unseen categories.

$$\begin{aligned} \min_{\theta} J(\theta) &= \frac{1}{C+1} \sum_{k=1}^C L(y^{(k)}, h^{(k)}) + \lambda \cdot R(\theta) \\ &\approx \frac{1}{C} L(y^{(k)}, h^{(k)}) + L(y^{(C+1)}, h^{(C+1)}) + \lambda \cdot R(\theta) \end{aligned} \quad (4)$$

Recognition decision is a vital part of the vein-recognition algorithms. Different decision-making algorithms are proposed according to different systems, which can improve the overall performance. In this paper, the LROC verification

threshold self-adaption acquisition algorithm for the DGLFV model is proposed. Following the training outcomes of Softmax, LROC analysis is performed for every single category. We utilize the *Youden* index to find the best reception threshold for each class, and the maximum refers to the LROC separation threshold.

The decision-making algorithm determines that the recognition system is divided into two parts: the first is the rough identification, which means that the category of the maximum probability directly obtained by Softmax is served as the results; the other is fine verification, using the best acceptance threshold to confirm whether the verification is accepted. Verification is to judge whether the sample belongs to a category. In this paper, the output of Softmax,  $P(y_k|x)$ , is set to the verification probability. Whether to accept or reject depends on the verification threshold of  $T_k$ . Algorithm 3 shows the LROC verification threshold self-adaption acquisition algorithm. Combined with ROC, we evaluate every class at the beginning. When a specific category is prone to be positive, all the others are counterexamples. Under different stimulus (acceptance thresholds), different test results were gained by the measures of True Positive Rate ( $\delta_{TPR}^k$ ) and False Positive Rate ( $\delta_{FPR}^k$ ). These results were applied to evaluate the overall selection effect of such acceptance and rejection in different situations.

Assuming that there are  $P$  positive examples and  $N$  negative examples in the classification system, the calculation formulas of False Positive Rate ( $\delta_{FPR}^k$ ) and True Positive Rate ( $\delta_{TPR}^k$ ) are illustrated as (5).

$$\delta_{FPR}^k = \frac{FP}{N}, \delta_{TPR}^k = \frac{TP}{P} \quad (5)$$

$\delta_{FPR}^k$  is the false positive rate, indicating the probability that the negative case is divided into a positive one.

### Algorithm 3 Verification Threshold Self-Adaption Acquisition Algorithm Based on LROC

**Input:** Softmax testing output matrix  $G_{N \times (C+1)}$  and ground Category  $H_N$

**Output:** Verification thresholds  $T_{C+1}$  for each category

**for**  $k \in \{1, 2, \dots, C + 1\}$  **do**

    Extract column  $g_{N \times 1}^k$  from  $G_{N \times (C+1)}$  and arrange it in reverse order

**while** Set the initial threshold  $t = 0$  **do**

        Calculate  $\delta_{FPR}^k$  and  $\delta_{TPR}^k$  based on where  $t$  is ranked in  $g_{N \times 1}^k$  and  $H_N$

        Increase  $t$ , which is the value closest to the ground truth in  $g_{N \times 1}^k$

**end while**

    Find out the best *Youden*<sup>k</sup> index and according to  $\delta_{FPR}^k$  and  $\delta_{TPR}^k$ , then according to this,  $t$  is the best threshold  $T_k$

    Output the overall threshold matrix  $T_{C+1}$

**end for**

Besides,  $\delta_{TPR}^k$  represents the probability that a positive case can be divided successfully. In this case, the *Youden*<sup>k</sup> index is used to find the optimal threshold. Because TPR is as crucial as FPR, the *Youden*<sup>k</sup> index is usually used to evaluate the authenticity of the screening results, whose most reasonable 'incentives' can also be found. When a set of  $\delta_{FPR}^k$  and  $\delta_{TPR}^k$  occurs, the *Youden*<sup>k</sup> index is located according to (6). The larger the index, the better the effect of reception or rejection. At this time, the optimal threshold  $T_k$  is the value  $t$  of  $\delta_{FPR}^k$  and  $\delta_{TPR}^k$  corresponding to the best *Youden*<sup>k</sup> index.

$$Youden^k = \max(\delta_{TPR}^k - \delta_{FPR}^k) \quad (6)$$

### III. EXPERIMENTS

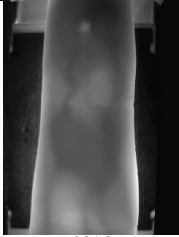

We have carried out the following experiments to verify the advantages of the finger-vein recognition algorithm we proposed in this paper. The experiments are mainly split into two parts: firstly, the RoI extraction algorithm based on Mask-RCNN is used to extract RoI and decide the training and testing parts; then finger-vein recognition experiments are performed on DGLFV.

#### A. TWO DATABASES FOR EXPERIMENTS

The finger-vein databases used in this paper are SDUMLA-HMT [17] and FV-SIPL, shown in Table 2. The former was collected by the Group of Machine Learning and Applications of Shandong University. FV-SIPL is the finger-vein database we built in Signal and Information Processing Laboratory, Mine Safety Data Research Center.

1) SDUMLA-HMT: The dataset was produced by the Group of Machine Learning and Applications of Shandong University. This database contains a variety of finger-vein images of 106 people. Each person offered the index finger, middle finger, and ring finger of two hands. Then six images were extracted for each finger. There is a total of 3,816 sheets, with a size of  $240 \times 320$ .

TABLE 2. Introduction of two finger-vein databases we used.

	SDUMLA-HMT	FV-SIPL
Sample images		
Number of images	3816	1296
Number of people	106	27
Number of palms	2	1
Number of fingers	6	4
Number of categories (images in each category)	636(6)	108(12)

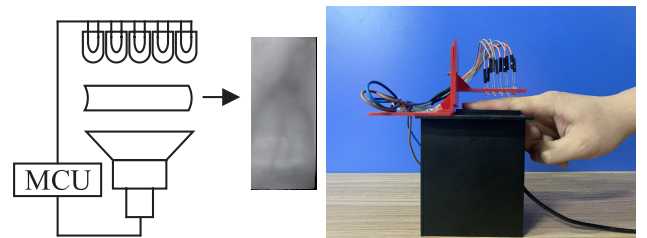


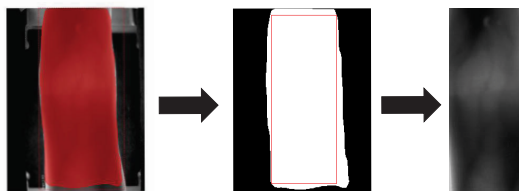
FIGURE 4. The picture on the left is the forming principle of finger veins. The infrared light is above the finger, and the infrared camera is located below. The finger-vein images can be formed between them. The camera transmits the received data to the micro control unit (MCU), responsible for storing, calculating, and matching. MCU and the camera are smaller than general, so the device is more portable. The right shows the homemade finger-vein collector of FV-SIPL.

The images in this database contain many non-finger areas. Also, different images of the same finger were covered, which are flipped over.

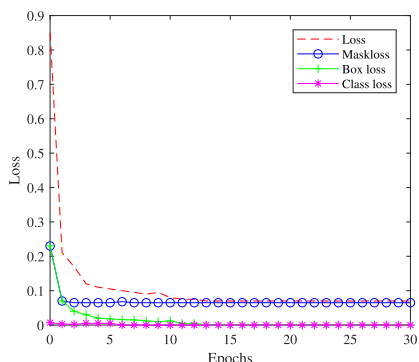
2) FV-SIPL: We collected the database by a self-made finger-vein collection device. The equipment has an 850nm near-infrared LED, IR camera, and PC software structure, using a single-sided LED projection. We constructed the database from 27 people and collected the index finger, middle finger, ring finger, and little finger of everyone's right hand. Each finger extracted 12 images, and there is a total of 1296 copies ( $176 \times 415$ ). The database enjoys high quality, little shadow, and highlight issues, little non-finger areas. More specific the different images of the same finger are not flipped over a lot. Fig. 4 shows a single-sided light imaging process of collection and our finger-vein collector.

#### B. EFFECT ANALYSIS OF THE PROPOSED RoI EXTRACTION ALGORITHM

We verified the finger-vein mask from Mask-RCNN with the RoI extraction using advanced bidirectional traversing and center diffusion method. Then compared them with the direct extractor and finger-knuckle positioning method. Finally, we tested the effectiveness and compared the ratio of the



**FIGURE 5.** The finger-vein ROI image extraction process. The left picture is the outputs given by Mask-RCNN; the middle is the binary image extracted according to the mask and the largest rectangle obtained by the advanced bidirectional traversing and center diffusion method in this paper; moreover, the right is the actual extracted ROI image.



**FIGURE 6.** The Mask-RCNN loss value trend. Because of the pre-trained weight, the initial loss value is minimal. After 15 steps of training, the model loss value does not change and is already in the best fitting. The SDUMLA-HMT dataset was chosen for this experiment.

largest rectangular area with existing algorithms. Fig.5 shows the extracting process of the finger-vein ROI regions.

It contains many non-finger-vein regions, which can quickly reflect ROI extraction performance. Before training Mask-RCNN, this paper randomly selected 100 images as the training set and annotated the pictures through tagging software. Because of the numerous parameters in Mask-RCNN, retraining the network requires tremendous data support. Thus, we utilize transfer training [28]–[30] to reduce input. The weights trained by the COCO dataset [30] are employed as the initials. There are no finger veins in this database, so we need to make an adjustment based on finger-vein datasets and make it the optimal extractor. In the end, we decreased the value of the loss function to complete the extracting process with fewer inputs. During training, the learning rate is set to 0.001 for the whole network. Fig.6 shows the loss tendency during training, covering the total loss, mask regression loss, classification loss, and bounding box regression loss.

To prove the effectiveness of the ROI extraction algorithm based on Mask-RCNN proposed in this paper, we picked out 500 testing samples. The successful extraction refers to the number of images from which we can extract ROI regions successfully, which means more than 95% of the total ROI can be extracted as ROI, as shown in Table 3. When extracting ROI, the algorithm proposed in this paper realized a significant number of successful extractions. Also, the effective rate in later selections reaches 100%. It is significantly higher than

**TABLE 3.** ROI extraction performance comparison.

Methods	Testing	Successful Extraction	Effective Extraction	Effective Ratio/%
The direct method	500	494	471	94.2
The finger-knuckle positioning method	500	494	490	98
The method in this paper	500	500	500	100

**TABLE 4.** Comparison of the maximum rectangular area ratios.

Ratio(s)/%	Pic1	Pic2	Pic3	Pic4	Effective Ratio
Traversing center diffusion	60.78	82.44	75.73	79.54	74.62
Largest Inscribed Rectangles in Convex Polygons	80.78	81.78	81.44	80.79	81.20
Bidirectional traversing and center diffusion method (in our work)	82.97	83.47	77.61	82.66	81.68

the direct method of 94.2% and the finger-knuckle positioning methodology of 95%, with extremely strong robustness.

Our experiments have been compared with the traversing and center diffusion method [31] and Largest Inscribed Rectangles in Convex Polygons [22] to prove the reliability of the advanced bidirectional traversing and center diffusion method we proposed. This method starts from the top, bottom, left, and right to the calculated center, compared to the previous area. Moreover, Largest Inscribed Rectangles in Convex Polygons mainly covers the approximation method. When adding it to the experiment, mask patterns have been approximated as a polygon near the original irregular one. As is shown in (7), the measurement parameter is the ratio of the largest rectangular area  $R(s)$ , the pixel number of extracted rectangular is  $S_{rect\_pixel}$ , and that of masks is  $S_{mask\_pixel}$ .

$$R(s) = \frac{S_{rect\_pixel}}{S_{mask\_pixel}} \tag{7}$$

We randomly selected four pictures from the testing set and tested the maximum rectangular area’s proportion using three different methods, as in Table 4.

From the experimental results, our work shows a stable and excellent extraction effect. In contrast, the traversing and center diffusion method easily fall into a local optimum and lead to inconsistent results. The Largest Inscribed Rectangles in Convex Polygons misses some regions when approximating the polygon, which will affect the result.

With the well-labeled data and transfer learning, Mask-RCNN converges and fits the finger-vein identification area. Also, it outputs better loss trends. We compared the maximum rectangular area ratio with traditional algorithms and verified the effectiveness of the advanced bidirectional traversing and center diffusion method. Based on the mask extracted by Mask-RCNN, the method in this paper can better avoid local optimum, extract as many finger-vein regions as possible, and remove the background. In the process



of testing, the average ratio is optimal. By comparing the successful extraction ratio and effectiveness rate of SDUMLA-HMT, it proves that the method in this paper enjoys high accuracy and robustness. The effective extraction rate can reach 100%.

**C. EXPERIMENT ON THE EFFECT OF FINGER-VEIN RECOGNITION**

**1) TRAINING MODEL**

The images collected from FV-SIPL and SDUMLA-HMT are divided into three types: the known category training set, unseen category training set, and testing set. For instance, we picked up eight images from each of the 80 categories in FV-SIPL as a known category training set. Then 20 out of the remaining 28 categories are placed in the unseen training set, and the rest eight categories are chosen to be tested. In addition to the above 20 types of images, the unseen training set also contains images of the finger-vein database published by Malaysian Polytechnic University, Tianjin Key Laboratory of Intelligent Signal and Image Processing, generally refer to as  $U_n\_Data$ . In conclusion, the dataset is divided into two parts: the unseen class and the data with labels. Besides, the division results of SDUMLA-HMT are shown in Table 5.

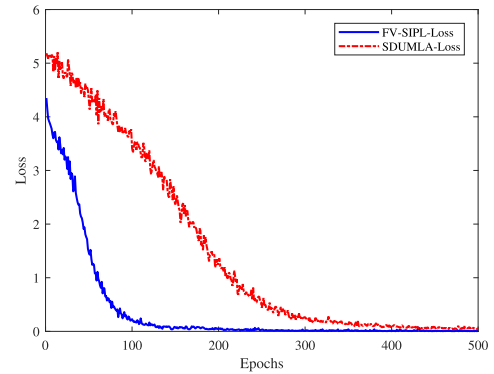
**TABLE 5. Database Partition. Except for splitting the normal training and testing dataset, we considered unseen categories to verify generalization capability of our proposed method.**

Categories	SDUMLA-HMT	FV-SIPL
Known categories for training	$400 \times 4 = 1600$	$80 \times 8 = 640$
Unseen categories for training	$136 \times 6 + U_n\_Data = 816 + U_n\_Data$	$20 \times 12 + U_n\_Data = 240 + U_n\_Data$
Testing set	$100 \times 6 + 400 \times 2 = 1400$	$8 \times 12 + 80 \times 4 = 416$

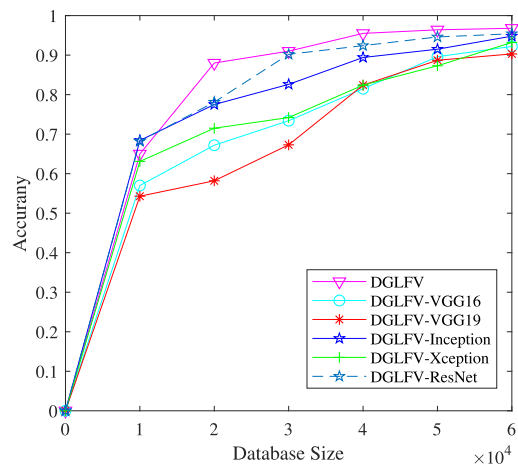
The dataset should be expanded to prevent the network from over-fitting, increase the robustness of the network, and adapt to the needs of network parameters. Firstly, we extracted the RoI region and resized them to  $224 \times 224$ . The images are then normalized, randomly rotated, translated, and imported into the CNN network for training and testing. In specific applications, such as using FV-SIPL to train the network, one batch is 64, and a total of 500 steps of training were carried out. With 1296 pictures sequentially extracted in a loop, we achieved 64 pictures and conducted preprocessing. At this time, the number of images in the whole training network will reach 32,000. Fig.7 shows the changing trend of the loss function of DGLFV when training SDUMLA-HMT and FV-SIPL.

**2) EXPERIMENT ON THE RECOGNITION EFFECT OF UNSEEN CATEGORY**

The unseen training category of different scales will cause more or less influence on the results of the vein recognition, We must determine the data volume of the unseen category training dataset, which ensures the high accuracy



**FIGURE 7. Here are the trends of training loss for SDUMLA-HMT and FV-SIPL. The image quality of SDUMLA-HMT is lower than FV-SIPL, with more shadowing data and a more extensive posture adjustment range. So the experiment on SDUMLA-HMT enjoys significant initial loss and slower training converges. The final training loss proves that the DGLFV model we proposed has converged and good recognition ability.**



**FIGURE 8. The chart displays the impact of unseen label database size on accuracy. Experiments are performed using a test set of unseen categories divided from the data set. With more training data, the recognition results are better. For the DGLFV model proposed in this paper, the amount of training data is increased to about 30,000. The recognition of unseen categories tends to a certain stability.**

of subsequent experiments. We are clear that the unseen category testing dataset is different from the training so that it can be closer to the actual scenario. Based on the testing set mentioned above, we used some conventional and classic models for comparison, such as VGGNet [12], Inception [23], ResNet [20], which obtained excellent results in games held by ImageNet. The experimental results are shown in Fig.8.

According to those two databases, our novel algorithm based on the Inception-ResNet V2 approach reach an excellent accuracy level of over 99%.

Table 6 shows the identification effects when testing the new categories on DGFLV and other usual models. To test the performance and robustness to unseen finger-vein categories, we conducted different experiments based on various feature extraction methods. It shows that our DGFLV model based on Inception-ResNet V2 enjoys the highest recognition accuracy, and VGG could not give more improvement than ResNet.

**TABLE 6. Recognition effects of multiple networks on unseen label finger veins.**

Methods(Accuracy/%)	DGFLV	DGFLV-VGG19	DGFLV-VGG16	DGFLV-Inception	DGFLV-Xception	DGFLV-ResNet
FV-SIPL	96.82	90.37	92.23	94.82	93.27	95.48
SDUMLA-HMT	96.71	90.93	92.61	93.94	93.51	96.63

**TABLE 7. Overall finger-vein recognition effects of multiple networks.**

Methods(Accuracy/%)	DGFLV	DGFLV-VGG19	DGFLV-VGG16	DGFLV-Inception	DGFLV-Xception	DGFLV-ResNet
FV-SIPL	99.25	95.83	96.29	96.53	98.02	97.03
SDUMLA-HMT	99.08	95.76	96.83	98.36	98.06	98.79

3) EXPERIMENT ON THE OVERALL ACCURACY OF FINGER-VEIN RECOGNITION

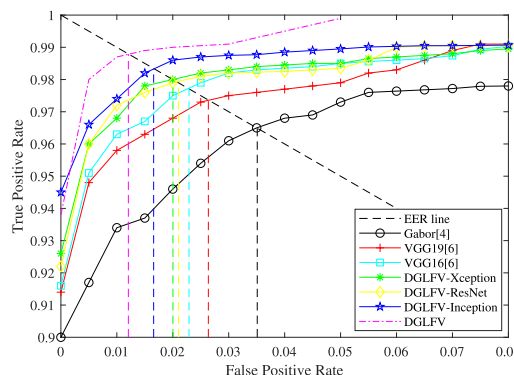
There are two key evaluation indicators for measuring the effect of finger-vein recognition: the accuracy rate and the ROC curve. In the beginning, we compared the overall accuracy of the model constructed with VGGNet [12], Inception [23], and ResNet [20]. We measured the gross accuracy of two kinds of databases (the unseen and known testing set). The comparison results are shown in Table 7.

Although the database has unseen categories, the model proposed in this paper still achieved an excellent recognition effect and generalization capability. To further reflect this algorithm’s recognition performance, we compare its accuracy with other recent finger-vein recognition technology using the same database (SDUMLA-HMT). The results are presented in Table 8.

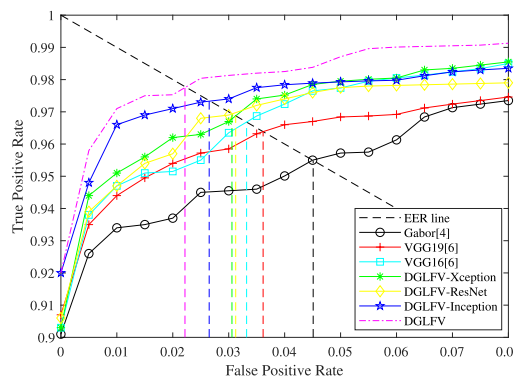
4) EER EXPERIMENT OF THE WHOLE FINGER VEIN

Compared with the recognition accuracy, the ROC curve is a more meaningful indicator for evaluating the effect. In the ROC curve, the true positive rate (TPR) and the false positive rate (FPR) are the vertical and horizontal axes. The equal error rate (EER) is a point at which FPR is equal to the false-negative rate (FNR), and the sum of FNR and TPR is 1. Because FNR can be converted by TPR, we generally regard the horizontal axis of the point where the ROC curve intersects with the EER line as the EER value. It is customary to use EER to measure the recognition effect. The smaller the EER value, the better the overall recognition effect [40]. The Softmax output can be obtained after inputting the testing set to the training model. Take the maximum probability as the verification one, and compare it with the verification threshold during training to decide whether to accept or reject. After multiple batches, we can receive TPR and FPR and establish the ROC curve.

The comparison experiments cover [2], [7], and the recognition of other classic models using the same recognition method in this paper. Except for the Gabor-based recognition algorithm [2] and that based on VGG19 and VGG16 [7], we replaced the basic model proposed by Inception [23], ResNet [20] and Xception [41] for comparison. They are named as DGLFV-Inception, DGLFV-ResNet, and DGLFV-Xception. Fig.9 and Fig.10 show the results on FV-SIPL



**FIGURE 9. With the dataset of FV-SIPL, the algorithm in this paper remains the lowest equal error rate (EER) even after introducing an unseen category of finger veins. Compared with the reference [2] and [7], the equal error rate (EER) decreased by 2.049% and 1.095%, respectively, and the related classic models based on DGLFV were improved. Besides, the Inception-ResNet V2 model used in this paper has better recognition accuracy than other network models, with an EER of 1.481%.**



**FIGURE 10. Experimental results on SDUMLA-HMT show that the model based on DGFLV has a better recognition effect than most existing algorithms. EER has decreased by 2.192% and 1.225%, compared with [2] and [7].**

and SDUMLA-HMT, Table 9 demonstrates the statistical test results.

5) EXPERIMENT ON THE EFFECTIVENESS OF THE LROC ALGORITHM

DGLFV can utilize the LROC verification threshold self-adaption acquisition algorithm to transform the classification recognition probability into verification. Nevertheless,

**TABLE 8.** Our proposed algorithm is compared with state-of-the-art for finger-vein recognition technology research. In the case that the finger-vein recognition system becomes more complicated after the introduction of unseen finger-vein categories, the finger-vein recognition algorithm proposed in this paper still has reached the state-of-art level of over 99 %, compared with the research results of finger-vein recognition in recent years. Besides, there are different data partitioning in these models. We select and compare the best ratios with the results their papers claimed.

Paper	Published Year	Feature Extraction Method	Database	Journal or Publisher	Accuracy/%
Kumar <i>et al.</i> [3]	2011	Gabor filter with morphological processing	HKPU (105 categories)	IEEE Transactions on image processing	90.08
Yu L., <i>et al.</i> [32]	2013	Polydirectional Local Line Binary Pattern	SDUMLA-HMT	2013 International Conference on ICT Convergence (ICTC) IEEE	99.21
Xie S J., <i>et al.</i> [33]	2014	Block-based average absolute deviation (AAD) features	SDUMLA-HMT	Cognitive Computation	97.76
Hoang T V., <i>et al.</i> [34]	2015	MFRAT & GridPCA	SDUMLA-HMT	2015 Seventh International Conference on Knowledge and Systems Engineering(KSE)	95.67
Li Z Z., <i>et al.</i> [35]	2016	SPCF	SDUMLA-HMT	IEEE International Joint Conference on Biometrics	92.71
Qiu S., <i>et al.</i> [36]	2016	Dula-sliding window + location + Pseudo-elliptical transformer + 2D-PCA	SDUMLA-HMT	Expert Systems with Applications	97.61
Wen J L., <i>et al.</i> [37]	2017	CNN (based on AlexNet)	SDUMLA-HMT	IEEE Conference on Industrial Electronics and Applications	99.53
Banerjee A., <i>et al.</i> [38]	2018	CLAHE+directional dilation	SDUMLA-HMT	Multimedia Tools and Applications	90.72
Rig D., <i>et al.</i> [39]	2019	CNN (Proposed CNN)	SDUMLA-HMT	IEEE Transactions on Information Forensics and Security	98.90
In this paper	-	CNN (based on Inception-ResNet V2)	SDUMLA-HMT OWN(FV-SIPL)	-	99.25 99.08

**TABLE 9.** Comparison of DGLFV and other model identification capabilities. The finger-vein recognition model based on DGLFV still outperforms the existing algorithms. When using the same training method, it is found that Inception-ResNet V2 network can make progress.

Methods	EER/%	
	FV-SIPL	SDUMLA-HMT
Gabor [2]	3.530	4.420
VGG19 [7]	2.576	3.453
VGG16 [7]	2.290	3.340
DGLFV-Xception	1.908	3.025
DGLFV-ResNet	2.003	3.280
DGLFV-Inception	1.658	2.551
DGLFV	1.481	2.228

in [7], based on CNN, there is a need to repeat the verification according to the number of categories. In practice, DGLFV will significantly improve the recognition efficiency, shown in Table 10.

There are 400 known categories in SDUMLA-HMT. The time spent on DGLFV recognition is approximately 0.0125 times as much as that of [7].

**D. RESULT ANALYSIS**

For finger-vein recognition, we determine the size of the untold category training set and the method first. After that,

**TABLE 10.** Comparison of recognition efficiency. We present the time consumption of our algorithm recognizing a single picture based on different feature extraction models in SDUMLA-HMT.

Paper	Published Year	Feature Extraction Method	Time Consumption(s)
Hyung G H., <i>et al.</i> [7]	2017	CNN (based on VGG19)	8.134
Hyung G H., <i>et al.</i> [7]	2017	CNN (based on VGG16)	7.798
In this paper	-	CNN (Inception-ResNet V2)	0.157

we test the effects of other categories and the overall dataset on DGLFV. The results show that the DGLFV model proposed in this paper has an excellent recognition effect on unseen finger-vein categories and enjoys more advantages than other novel CNN recognition models. Compared with other networks, DGLFV has achieved the state-of-art recognition accuracy level of over 99% for the entire finger-veins, including unseen categories. It has reached the lowest EER value, indicating that the model enjoys higher security and accuracy. When identifying unseen categories, different from the passive comparison rejection used in the traditional finger-vein recognition method, the DGLFV model uses an

active way to generalize the test samples into separate unseen categories directly. Although it costs some time to train the model, the recognition speed can be significantly improved. The results of the LROC algorithm also show that the DGLFV model has a high finger-vein recognition efficiency.

#### IV. CONCLUSION

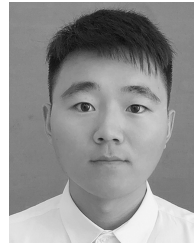
To address the recognition problem of interference and low efficiency when identifying unseen categories with traditional methods, we propose a novel generalized finger-vein recognition model called DGLFV, with good accuracy and practical value. We apply an advanced Mask-RCNN-based mask extraction algorithm to obtain accurate masks, both for the known and unseen categories. Then an advanced bidirectional traversing and center diffusion method is proposed to obtain the finger-vein RoI. In terms of the unseen categories, this model utilizes a deep neural network to train them with a large-scale image database and realizes active recognition. Finally, a recognition decision algorithm based on Softmax is constructed to obtain the verification thresholds of each category actively. The FV-SIPL dataset we constructed can be accessible by emailing us, and later it will be open-source online. Through experiments on the FV-SIPL and SDUMLA-HMT databases, the recognition accuracy of DGLFV in this paper has reached 99.25% and 99.08%. EER values reach 1.481% and 2.228%, respectively. From the tables we present, our model reaches the state-of-art recognition level. The total time consumption of our structure has been improved at 0.157s, as shown in Table 10.

We have not explored more experiments in a real scene or different products. For future research, we will concentrate on the extraction efficiency of RoI regions in complicated backgrounds and recognition model generalization. Also, using transfer learning and implementing more high-precision algorithms to finger-vein recognition is another research trend. This kind of algorithms will influence personal information security and protection as a main branch of biometrics. We will further work on accuracy enhancement without increasing time consumption. The possible application research to micro-computing or bit-level systems will provide widespread finger-vein recognition.

#### REFERENCES

- [1] K. N. Srinivas and R. Arumugam, "Finite element analysis combined circuit simulation of dynamic performances of switched reluctance motors," *Electr. Power Compon. Syst.*, vol. 30, no. 10, pp. 1033–1045, Oct. 2002.
- [2] J. Peng, N. Wang, A. A. A. El-Latif, Q. Li, and X. Niu, "Finger-vein verification using Gabor filter and SIFT feature matching," in *Proc. 8th Int. Conf. Intell. Inf. Hiding Multimedia Signal Process.*, Jul. 2012, pp. 45–48.
- [3] A. Kumar and Y. Zhou, "Human identification using finger images," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 2228–2244, Apr. 2012.
- [4] B. A. Rosdi, C. W. Shing, and S. A. Suandi, "Finger vein recognition using local line binary pattern," *Sensors*, vol. 11, no. 12, pp. 11357–11371, Nov. 2011.
- [5] 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. 15048–15067, Nov. 2013.
- [6] 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.
- [7] 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.
- [8] S. Khellat-kihel, R. Abrishambaf, N. Cardoso, J. Monteiro, and M. Benyettou, "Finger vein recognition using Gabor filter and support vector machine," in *Proc. Int. Image Process., Appl. Syst. Conf.*, Nov. 2014, pp. 1–6.
- [9] C. He, Z. Li, L. Chen, and J. Peng, "Identification of finger vein using neural network recognition research based on PCA," in *Proc. IEEE 16th Int. Conf. Cognit. Informat. Cognit. Comput. (ICCI\*CC)*, Jul. 2017, pp. 456–460.
- [10] J. Yang, J. Wei, and Y. Shi, "Accurate ROI localization and hierarchical hyper-sphere model for finger-vein recognition," *Neurocomputing*, vol. 328, pp. 171–181, Feb. 2019.
- [11] S. A. Radzi, M. Khalil-Hani, and R. Bakhteri, "Finger-vein biometric identification using convolutional neural network," *Turkish J. Electr. Eng. Comput. Sci.*, vol. 24, pp. 1863–1878, Mar. 2016.
- [12] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*. [Online]. Available: <https://arxiv.org/abs/1409.1556>
- [13] C. Xie and A. Kumar, "Finger vein identification using convolutional neural network and supervised discrete hashing," *Pattern Recognit. Lett.*, vol. 119, pp. 148–156, Mar. 2019.
- [14] 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.
- [15] K. He, G. Georgia, D. Piotr, and G. Ross, "Mask R-CNN," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 2, pp. 386–397, Feb. 2020.
- [16] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017.
- [17] Y. Yin, L. Liu, and X. Sun, "SDUMLA-HMT: A multimodal biometric database," in *Proc. Chin. Conf. Biometric Recognit.*, 2011, pp. 260–268.
- [18] W. Jiawan, "Research of the finger vein recognition algorithm based on subspace analysis and LBP," Ph.D. dissertation, Dept. Inf. Eng. School, Hangzhou Dianzi Univ., Hangzhou, China, 2016.
- [19] M. A. Hui and K. Wang, "A region of interest extraction method using rotation rectified finger vein images," *CAAI Trans. Intell. Syst.*, vol. 7, no. 3, pp. 230–234, 2012.
- [20] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 770–778.
- [21] X. Xin-Hua, L. Dong, Z. Xiang-Qian, Z. Dong-Yan, and Z. Jian-Jun, "Detecting maximum inscribed rectangle area of target object based on image processing," *Sci. Technol. Eng.*, vol. 15, no. 17, pp. 193–197, 2015.
- [22] C. Knauer, L. Schlipf, J. M. Schmidt, and H. R. Tiwary, "Largest inscribed rectangles in convex polygons," *J. Discrete Algorithms*, vol. 13, pp. 78–85, May 2012, doi: [10.1016/j.jda.2012.01.002](https://doi.org/10.1016/j.jda.2012.01.002).
- [23] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2818–2826.
- [24] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 1798–1828, Aug. 2013.
- [25] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1–9.
- [26] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, "Inception-v4, Inception-ResNet and the impact of residual connections on learning," in *Proc. 31st AAAI Conf. Artif. Intell.*, 2017, pp. 1–12.
- [27] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010.
- [28] S. Veluchamy and L. R. Karlmarx, "System for multimodal biometric recognition based on finger knuckle and finger vein using feature-level fusion and K-support vector machine classifier," *IET Biometrics*, vol. 6, no. 3, pp. 232–242, May 2017.
- [29] A. Argyriou, M. Pontil, Y. Ying, and C. A. Micchelli, "A spectral regularization framework for multi-task structure learning," in *Proc. Adv. Neural Inf. Process. Syst.*, J. C. Platt, D. Koller, Y. Singer, and S. T. Roweis, Eds. Red Hook, NY, USA: Curran Associates, 2008, pp. 25–32. [Online]. Available: <http://papers.nips.cc/paper/3187-a-spectral-regularization-framework-for-multi-task-structure-learning.pdf>

- [30] T.-Y. Lin et al., "Microsoft COCO: Common objects in context," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2014, pp. 740–755.
- [31] X. Xinhua et al., "Detecting maximum inscribed rectangle area of target object based on image processing," *Sci. Technol. Eng.*, vol. 15, no. 17, pp. 193–197 and 217, Jun. 2015.
- [32] Y. Lu, S. J. Xie, S. Yoon, and D. S. Park, "Finger vein identification using polydirectional local line binary pattern," in *Proc. Int. Conf. ICT Converg. (ICTC)*, Oct. 2013, pp. 61–65.
- [33] S. J. Xie, S. Yoon, J. Yang, Y. Lu, D. S. Park, and B. Zhou, "Feature component-based extreme learning machines for finger vein recognition," *Cognit. Comput.*, vol. 6, no. 3, pp. 446–461, Sep. 2014.
- [34] H. T. Van, T. T. Thai, and T. H. Le, "Robust finger vein identification base on discriminant orientation feature," in *Proc. 7th Int. Conf. Knowl. Syst. Eng. (KSE)*, Oct. 2015, pp. 348–353.
- [35] L. Zhou, G. Yang, Y. Yin, L. Yang, and K. Wang, "Finger vein recognition based on stable and discriminative superpixels," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 30, no. 6, Jul. 2016, Art. no. 1650015.
- [36] 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.
- [37] W. Liu, W. Li, L. Sun, L. Zhang, and P. Chen, "Finger vein recognition based on deep learning," in *Proc. 12th IEEE Conf. Ind. Electron. Appl. (ICIEA)*, Jun. 2017, pp. 205–210.
- [38] A. Banerjee, S. Basu, S. Basu, and M. Nasipuri, "ARTEM: A new system for human authentication using finger vein images," *Multimedia Tools Appl.*, vol. 77, no. 5, pp. 5857–5884, Mar. 2018.
- [39] R. Das, E. Piciucco, E. Maiorana, and P. Campisi, "Convolutional neural network for finger-vein-based biometric identification," *IEEE Trans. Inf. Forensics Security*, vol. 14, no. 2, pp. 360–373, Feb. 2019.
- [40] X. Tianyang, H. Xiaowei, and L. Sen, "A near infrared finger vein recognition approach based on wavelet grayscale surface matching," *Laser Optoelectron. Prog.*, vol. 53, no. 4, 2016, Art. no. 041005.
- [41] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1251–1258.



**YALEI HU** received the master's degree in electronics and information technology from Liaoning Technical University, in 2020.

He is currently working as a Hardware Engineer with Monitoring Branch, Tiandi (Changzhou) Automation Company Ltd., China. He helped with dataset collecting, experiments, and analysis. His research interests include image processing and image recognition.



**YUEMING HAN** is currently pursuing the master's degree in electronics and communication systems with Liaoning Technical University.

She helped with the experiments and analysis. Her research interests include image processing and image recognition.



**ZHIYONG TAO** received the Ph.D. degree from Liaoning Technical University.

He worked as a Visiting Scholar with the Clausthal University of Technology, Germany, in 2006. Since 2008, he has been working as an Associate Professor and a Master Tutor with Liaoning Technical University. He designed the deep generalized label algorithm for finger-vein recognition (DGLFV) model, analyzed the result of experiments, and wrote the original article. His

main research interests include the Internet of Things, machine learning, and biometric recognition, with a wealth of project and engineering experience.



**SEN LIN** received the bachelor's and master's degrees from Liaoning Technical University, in 2003 and 2006, respectively, and the Ph.D. degree from the Shenyang University of Technology, in 2013. He is currently an Associate Professor and a Master Supervisor with Shenyang Ligong University. He is a Postdoctoral Research Fellow with the Chinese Academy of Sciences, Shenyang Institute of Automation. He helped with the revision of the original article. His

research interests include image processing, machine vision, and biometric recognition.



**HAOTONG WANG** received the M.S. degree in communication and information system from Liaoning Technical University, China, in 2019. He is currently pursuing the Ph.D. degree with the Graduate School of Information, Production and Systems, Waseda University, Japan.

He helped to construct the finger-vein recognition systems. He helped with dataset collecting, experiment designing, and analysis. Portions of this research were done while he was a master's student with Liaoning Technical University. His current research interests include pattern recognition, natural language processing, and machine learning.



**YING LIU** received the B.Eng. degree from the Heilongjiang Institute of Science and Technology, in 2006, and the M.Eng. and Ph.D. degrees from Jilin University, in 2008 and 2011, respectively.

She is currently an Associate Professor with Liaoning Technical University. She helped with the experiments and analysis. Her research interests include wireless localization and non-contact identification.

...