

Research Article

Noncontact Palm Vein ROI Extraction Based on Improved Lightweight HRnet in Complex Backgrounds

Fen Dai,¹ Ziyang Wang ,¹ Xiangqun Zou,² Rongwen Zhang,¹ and Xiaoling Deng ^{1,3,4}

¹College of Electronic Engineering (College of Artificial Intelligence), South China Agricultural University, Guangzhou 510642, China

²Guangzhou Intelligence Oriented Technology Co. Ltd. No. 604, Tian'an Technology Development Building, Tian'an Hi-Tech Ecological Park No. 555, North Panyu Avenue, Panyu District, Guangzhou 511493, China

³Lingnan Modern Agriculture Guangdong Laboratory, Guangzhou 510642, China

⁴National International Joint Research Center of Precision Agriculture Aviation Application Technology, Guangzhou 510642, China

Correspondence should be addressed to Xiaoling Deng; dengxl@scau.edu.cn

Received 5 June 2023; Revised 2 December 2023; Accepted 22 December 2023; Published 17 January 2024

Academic Editor: Christoph Busch

Copyright © 2024 Fen Dai et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The extraction of ROI (region of interest) was a key step in noncontact palm vein recognition, which was crucial for the subsequent feature extraction and feature matching. A noncontact palm vein ROI extraction algorithm based on the improved HRnet for keypoints localization was proposed for dealing with hand gesture irregularities, translation, scaling, and rotation in complex backgrounds. To reduce the computation time and model size for ultimate deploying in low-cost embedded systems, this improved HRnet was designed to be lightweight by reconstructing the residual block structure and adopting depth-separable convolution, which greatly reduced the model size and improved the inference speed of network forward propagation. Next, the palm vein ROI localization and palm vein recognition are processed in self-built dataset and two public datasets (CASIA and TJU-PV). The proposed improved HRnet algorithm achieved 97.36% accuracy for keypoints detection on self-built palm vein dataset and 98.23% and 98.74% accuracy for keypoints detection on two public palm vein datasets (CASIA and TJU-PV), respectively. The model size was only 0.45 M, and on a CPU with a clock speed of 3 GHz, the average running time of ROI extraction for one image was 0.029 s. Based on the keypoints and corresponding ROI extraction, the equal error rate (EER) of palm vein recognition was 0.000362%, 0.014541%, and 0.005951% and the false nonmatch rate was 0.000001%, 11.034725%, and 4.613714% (false match rate: 0.01%) in the self-built dataset, TJU-PV, and CASIA, respectively. The experimental result showed that the proposed algorithm was feasible and effective and provided a reliable experimental basis for the research of palm vein recognition technology.

1. Introduction

Biometric technologies have shown great advantages and reliability in the field of security authentication and identification. Traditional biometric identification technologies, such as fingerprint recognition [1], face recognition [2], iris recognition [3], and palm print recognition [4–6], are widely used in real-life applications. Face recognition is the most widely used, but it may cause unsuccessful recognition due to make-up, beard, and wearing masks during the COVID-19 epidemic. Fingerprints and palm prints are easily forged and destroyed due to being exposed on the skin surface, leading to recognition security problems. Palm vein recognition, as a

new biometric identification technology, has many advantages and has received wide attention. Near-infrared light at 700–1,000 nm can be absorbed by hemoglobin in blood, which means that vein images are usually extracted under the irradiation of near-infrared light. Studies have shown that all individuals, including twins, have different palm veins, and that the vast majority of palm veins do not change radically with age. Because the veins are under the skin, they are less likely to be injured and less likely to be falsified. It was well known that palm veins contain more features than finger veins [7–9], so palm vein recognition is more secure and reliable than finger vein recognition. The first important step in the noncontact palm vein recognition process is the

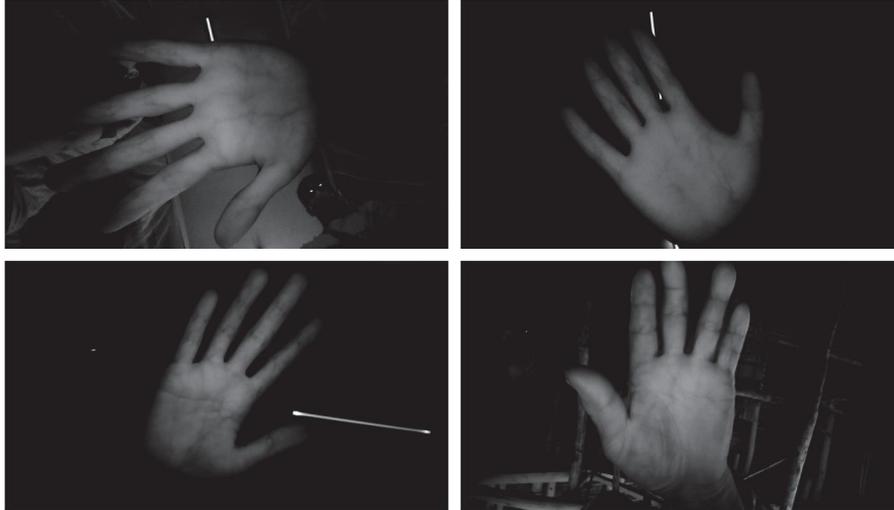


FIGURE 1: Some noncontact palm vein images in complex backgrounds.

localization of region of interest (ROI) in the palm vein image. Due to the noncontact acquisition method and the different acquisition environment, complex backgrounds, extra wrist parts, different angles, zooming, and panning are usually contained in the noncontact palm vein images, as shown in Figure 1. We need figure out a way to eliminate these interfering factors, so that we can extract the ROI in palm vein image accurately.

To solve the above problems, scholars had conducted a lot of research in recent years. Chai et al. [10] proposed a method to localize the ROI using the feature points of the hand, i.e., the valley points between two fingers. El Sayed et al. [11] proposed an ROI localization method based on threshold segmentation, morphological, and geometric operations. Kang and Wu [12] proposed an improved OTSU method to extract hand contours from grayscale palm vein images and then used the radial distance function between reference points and contour points to locate the peaks and valleys of the palm to extract palm vein ROI. Lin et al. [13] proposed a maximum inner tangent circle and center of mass based method to extract palm ROI. Yakno et al. [14] discussed the ROI extraction algorithm and proposed an improved algorithm for larger ROI extraction. Damak et al. [15] used hand boundary tracing by scanning contour lines to draw hand boundary distance contours, rotating the image so that the line connecting the first and third finger valleys became horizontal, and selecting four hand boundaries (vertical left limit, vertical right limit, horizontal lower limit, and horizontal upper limit) to create the ROI region. Cimen et al. [16] segmented the hand image and determined the boundaries of the hand surface area. Then, the whole image was scanned pixel by pixel from right to left and from top to bottom, and the first point that reached 255 pixels was found to be the tip of the bone, and a 256×256 pixel square region was selected as the ROI region by dropping 150 pixels at this point. Wu et al. [17] proposed to separate the palm of the hand using image binarization, and afterwards made vertical lines of the four fingers except the thumb, intersecting at eight points. The edge length of the ROI region was determined by the number of 255 pixels

contained between two adjacent points, and then the vein region was rotated so that it was parallel to the image boundary.

Ananthi et al. [18] proposed to apply OTSU (Otsu's method) to wrist-rejected palm vein images. Among the connected regions of the generated binary image, the maximum connected region represented the boundary of the palm region with fingers and the ROI was extracted from this palm region using an improved bounding rectangle strategy [18]. However, all the above methods usually required clean background in palm vein image, and when the image background was complex, it was difficult for the above methods to extract the vein ROI accurately.

With the rapid development of computer algorithms, deep learning networks became the mainstream algorithms for target detection. So far, many classical neural networks have been proposed, mainly represented by Fast R-CNN, Faster R-CNN, MaskR-CNN, and YOLOV3 [19–22]. Zhang et al. [23] proposed Tiny-YOLOV3 target detection algorithm with a target box and the keypoints would be determined by selecting the midpoint of the target box, and finally the ROI in palm vein image would be obtained after geometric calculation. Luo and Zhong [24] proposed an improved detection method based on Ruixin Zhang's method, in which the whole palm was detected in the first step by Tiny-YOLOV3, and then the keypoints coordinates were regressed by MobileNetV2. Although the method performed well in their self-built dataset, splitting the detection process into two parts was a bit tedious and increased the network computation. Sun et al. [25] proposed to extract hand features using HRNet network, which localized hand articulation points for gesture prediction and achieved good detection results. To localize the keypoints and extract the ROI in noncontact palm vein images as well as increase the computation speed, a lightweight keypoints detection network based on HRnet was proposed in this paper, which removed the redundant network branches, modified network structure, and adopted deep separable convolution for the residual block structure.

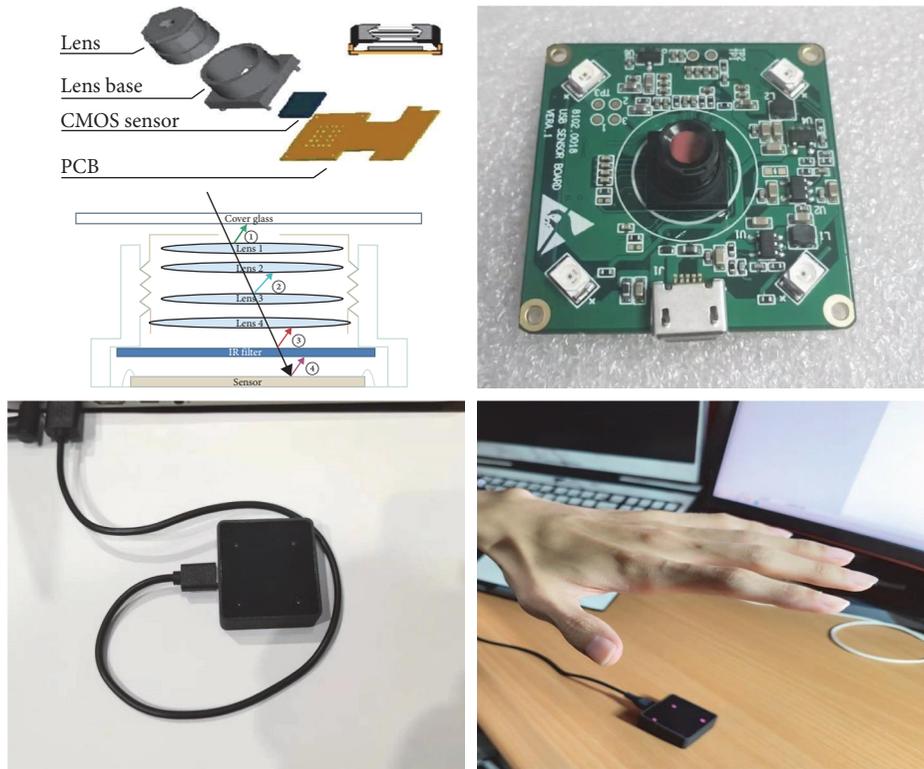


FIGURE 2: Shooting equipment and operating procedures.

2. Data Acquisition

In this paper, the self-built palm vein dataset was built by a self-built palm vein capture device. The collection device and the shooting process are shown in Figure 2 [26]. In May and June 2022, the collection was carried out on the campus of South China Agricultural University in Guangzhou, Guangdong Province, and the target population were students and staff of different ages. Five images were taken from five random angles with single and complex backgrounds. The shooting distance was between 15 and 20 cm, and a total of 3,000 palm vein images were collected. Each image size was $1,280 \times 720$.

To verify the effectiveness of the experimental algorithm, two public datasets of palm veins were used in the paper, namely, the CASIA (Chinese Academy of Sciences) and TJU-PV (Tongji University) datasets of palm veins. The two datasets consisted of 7,200 images with some background interference and 6,000 images with no background interference, respectively. The image size in the CASIA dataset was 768×576 , whereas the image size in the TJU-PV database was 800×600 , as shown in Figure 3. The acquired palm vein images were labeled with Labelme tool according to Pascal VOC dataset format. In the task of extracting ROI of palm vein images, four keypoints, i.e., four valleys between the five fingers, needed to be labeled, starting from the thumb and labeled sequentially from the numbers 1 to 4. The labeled image is shown in Figure 4.

To make the training dataset more ample as well as to increase the diversity of training samples, data enhancement

operations such as random angle inversion, brightness, and contrast adjustment were performed on the training dataset in the paper. The enhancement is shown in Figure 5.

3. Methods

A lightweight network based on improved HRnet was proposed to localize four keypoints in a palm vein image and corresponding ROI was extracted based on the keypoints by geometric methods for noncontact palm vein images. The flowchart of the proposed algorithm is shown in Figure 6. First, the four keypoints were located through the improved HRnet. Then, the left and right palms were distinguished according to cross product of vectors drawn from keypoints and then the obtained keypoint coordinates were used to locate the ROI of the palm vein image through the geometric operation. Finally, the eventual ROI was obtained using the affine transformation algorithm.

3.1. HRnet Algorithm. In pose estimation, the resolution of the image feature map was crucial. Usually pose estimation methods would adopt a serial method of reducing from high resolution to low resolution and then restoring high resolution to obtain a high-resolution feature map with strong semantic information. Compared with other networks, HRnet has two advantages: (1) HRnet connects multiple high resolutions with low-resolution branches in parallel, instead of in series, and improves interaction between branches with different resolutions. Therefore, the HRnet approach can maintain high resolution, instead of recovering high-resolution information

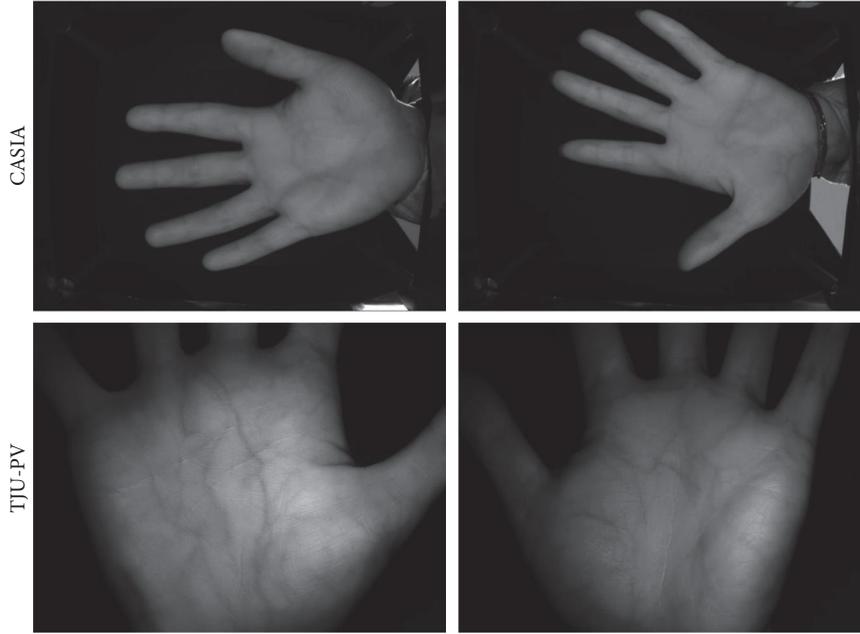


FIGURE 3: Two public palm vein dataset images, CASIA and TJU-PV.

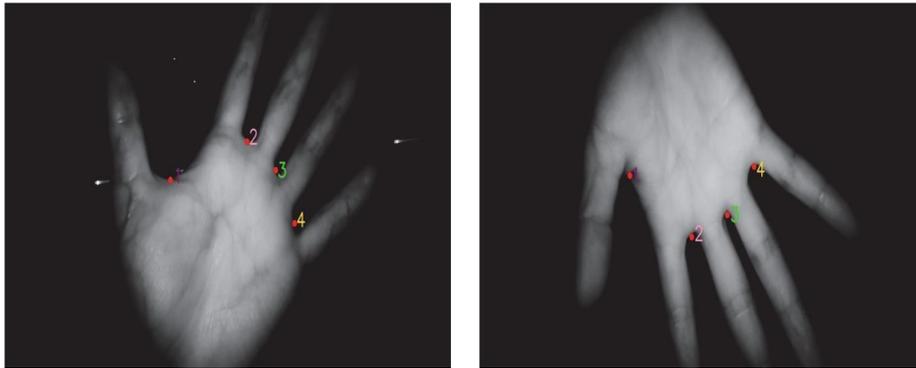


FIGURE 4: Image labeling effect.

through low-to-high process, and the output heatmap may be more spatially accurate. (2) Most other methods directly fuse low- and high-level feature maps. In contrast, HRnet performs repetitive multiscale fusion to enhance the high-resolution features and the high-resolution features are abundant enough to achieve pose estimation.

The network structure of HRnet was divided into two phases: the low-resolution phase and the high-resolution phase. The low-resolution phase generates feature maps at multiple resolutions, including original resolution, 1/2 resolution, 1/4 resolution, and 1/8 resolution. The high-resolution phase fuses the feature maps generated in the low-resolution phase to generate the high-resolution feature maps. A schematic diagram of the HRnet algorithm is shown in Figure 7. The bottleneck block was the bottleneck layer of ResNet, which was used to deepen the network, and the basic block was the general ResNet structure. Each basic convolutional block includes a batch normalization layer and an rectified linear unit (ReLU) layer, with

up and down representing upsampling and downsampling, respectively. The entire network generates reliable and location-sensitive high-resolution feature maps by iteratively fusing multi-resolution stream representations, and finally the number of feature map channels was determined according to the number of detected keypoint.

For training, the mean-square error was used as the loss function. The calculation formula was as follows:

$$\text{MSE} = \frac{\sum_{i=1}^n (y_i - y'_i)^2}{n}, \quad (1)$$

where y_i denotes the true value and y'_i denotes the predicted value. The mean-square-error loss function was a smooth function that was capable of minimizing the loss function using the gradient descent method. In the prediction part of the keypoints, a heat map was generated. The confidence level of whether there was a keypoint at that location was

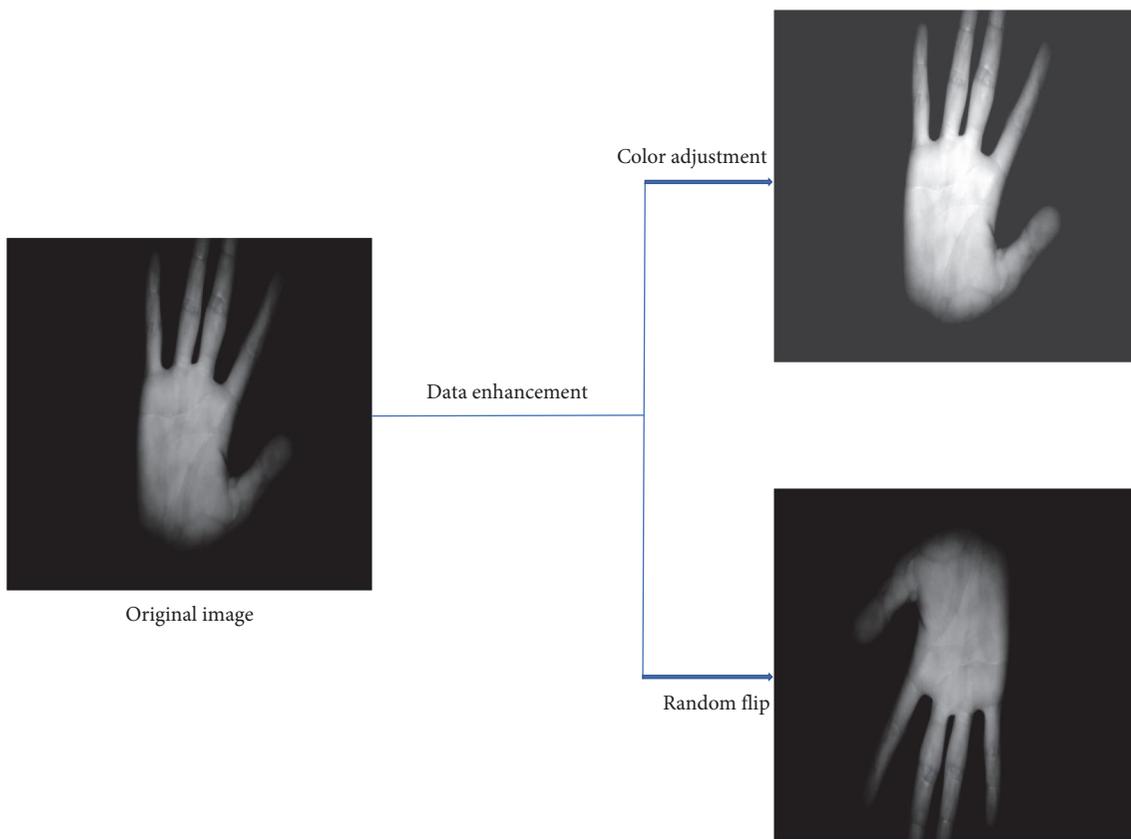


FIGURE 5: Image enhancement.

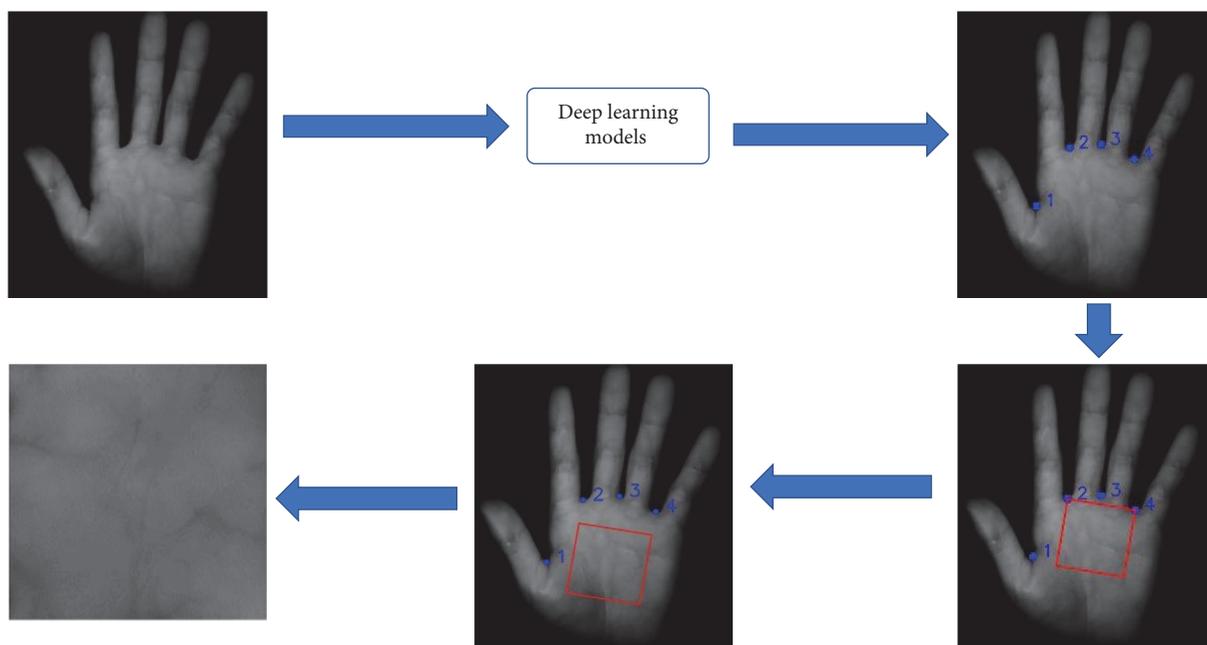


FIGURE 6: Algorithm flow chart.

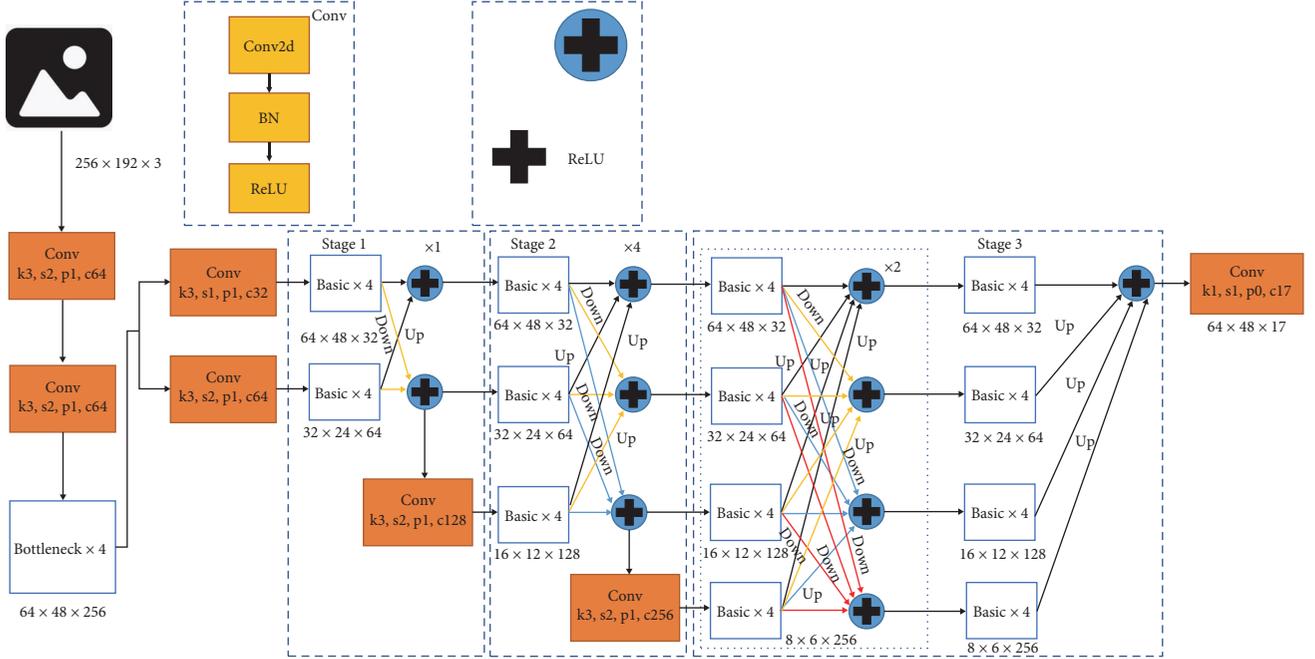


FIGURE 7: HRnet network structure.

output. The exact location of the keypoint was finally determined by setting a threshold value.

3.2. Improved HRnet Network. HRnet outperforms numerous target detection algorithms for keypoints detection tasks due to repetitive stacked multiresolution fusion, but such a network structure brings huge computational overhead and time overhead. It is unsuitable for real-time applications in embedded systems with limited computing power and storage capacity. To accelerate the palm vein ROI localization, the original HRnet was modified and used to detect the keypoints of palm vein image by the following methods:

- (1) The number of multiresolution fuse stacks was reduced and high-resolution features were maintained by fusing high-resolution features in the first two stages and low-resolution features in the last two stages.
- (2) Downsampling was achieved by controlling the convolution stride and pooling operation (MaxPool, Average-Pool), whereas upsampling was achieved by transposed convolution (Transposed Conv) operation.
- (3) The standard convolution of the original HRnet residual module was replaced by the depth-separable convolution (DSC).

The modified network architecture is shown in Figure 8.

3.3. Deeply Separable Convolution. Due to the standard convolution in which the convolution kernel acts on each channel of the input feature map, the computational amount is large. In this paper, the DSC was used to replace the standard convolution at the residual connection to reduce the number of network parameters. The DSC was a decomposable convolution structure, which can decompose the normalized

convolution into depth convolution and point-by-point convolution. This process can effectively reduce model parameters and computation. For a feature map with input size $H \times W$, the convolution kernel size was $K \times K$, C_{in} was the number of input feature map channels, and C_{out} was the number of output feature map channels. The standard convolution and DSC computation volume equations are

$$F_1 = H \times W \times K^2 \times C_{in} \times C_{out} \quad (2)$$

$$F_2 = H \times W \times K^2 \times C_{in} + H \times W \times C_{in} \times C_{out} \quad (3)$$

The ratio of the DSC to the standard convolution computation is

$$\frac{F_2}{F_1} = \left(\frac{1}{C_{out}} + \frac{1}{K^2} \right) \quad (4)$$

According to Equation (4), when an image input size was 12×12 , the number of input channels was three, the number of output channels was 128, and the convolution kernel size was 5×5 , as shown in Figure 9. The computation of the deep separable convolution was only 5% of the number parameters computed by the standard convolution.

3.4. Palm Vein ROI Extraction Based on Keypoints. The extraction of ROI was a critical step. To eliminate the effect of translation, rotation, and scaling, a normalization process for palm vein images ROI extraction based on keypoints was proposed in this paper. The proposed ROI extraction scheme (Algorithm 1) in the paper was as follows:

The selection process is shown in Figure 10.

1. Obtain three keypoints P1, P2, and P4;
2. Set the line through P2P4 as the X-axis;
3. Set the direction perpendicular to P2P4 and P1 as the positive direction of the Y-axis;
4. Set the midpoint Q of P2P4 as the origin (x_0, y_0) ;
5. Take $|P2P4|$ as the unit length of axis;
6. Construct the local coordinate system;
7. Obtain coordinate A of the ROI as $(x_0 - \frac{1}{2}|P2P4|, y_0 + \frac{1}{4}|P2P4|)$,
Obtain coordinate B of the ROI as $(x_0 + \frac{1}{2}|P2P4|, y_0 + \frac{1}{4}|P2P4|)$,
Obtain coordinate C of the ROI as $(x_0 + \frac{1}{2}|P2P4|, y_0 + \frac{5}{4}|P2P4|)$,
Obtain coordinate D of the ROI as $(x_0 - \frac{1}{2}|P2P4|, y_0 + \frac{5}{4}|P2P4|)$;
8. Connect ABCD in turn to get ROI.

ALGORITHM 1. ROI Extraction

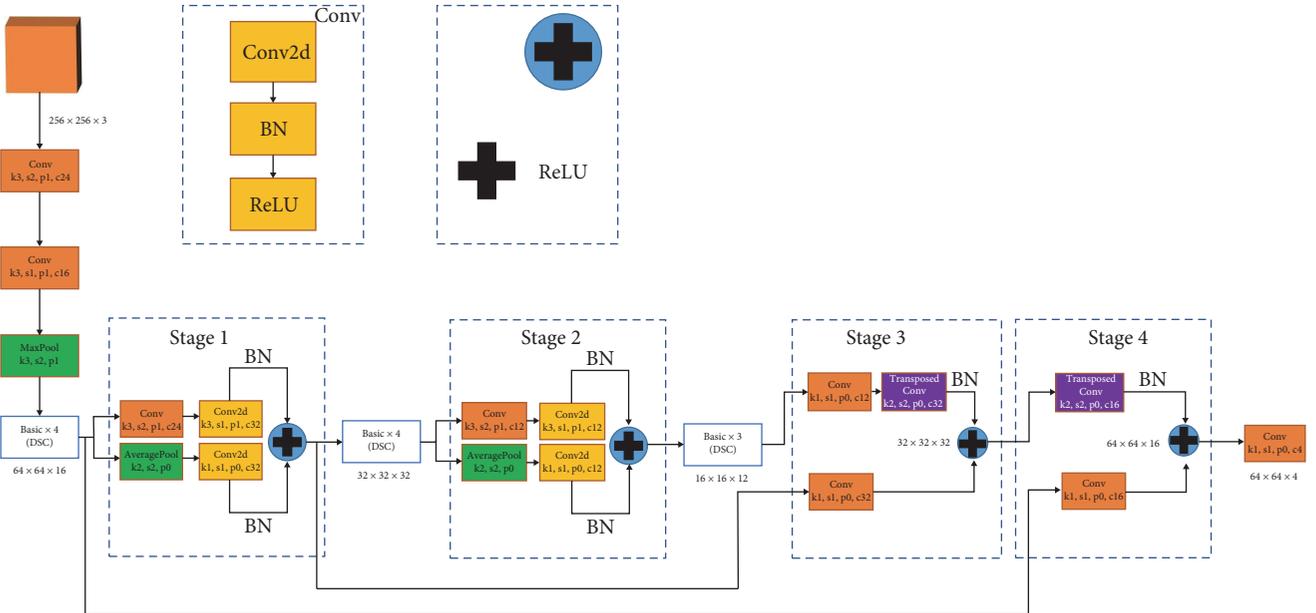


FIGURE 8: Improved HRnet network structure.

4. Experimental

4.1. Experimental Environment. The experiments were conducted in Ubuntu operating system, Python distribution, Anaconda (Python 3.7), Pytorch 1.8.2 deep learning framework, and cuda 10.2 accelerator. The CPU used in the experiments was Intel model i7-9,700 F, the GPU was NVIDIA Geforce RTX 3080 in the process of training the improved HRnet keypoints detection model, SGD was chosen as the optimizer of the network, the beta1 parameter was adjusted to 0.5, the beta2 was adjusted to 0.999, the batchsize was set to 32, the initial learning rate was set to 0.001, and the learning rate was adjusted by exponential decay. The details in self-built dataset and two public palm vein datasets are shown in Table 1. The self-built dataset was expanded to 4,500 images through the data enhancement.

4.2. Evaluation Indicators. The keypoints detection evaluation indicators differ from the target detection evaluation

indicators intersection over union. First, the Euclidean distance between the predicted point coordinate and the real label was calculated as follows:

$$d = \sqrt{(x_{dt} - x_{gt})^2 + (y_{dt} - y_{gt})^2} \quad (5)$$

In the above equation, (x_{dt}, y_{dt}) are the predicted coordinates and (x_{gt}, y_{gt}) are the true label coordinates

$$P(x) = \begin{cases} 1d < \text{threshold} \\ 0d > \text{threshold} \end{cases} \quad (6)$$

when d is greater than the given threshold, $P(x)$ is 1; otherwise, it was 0

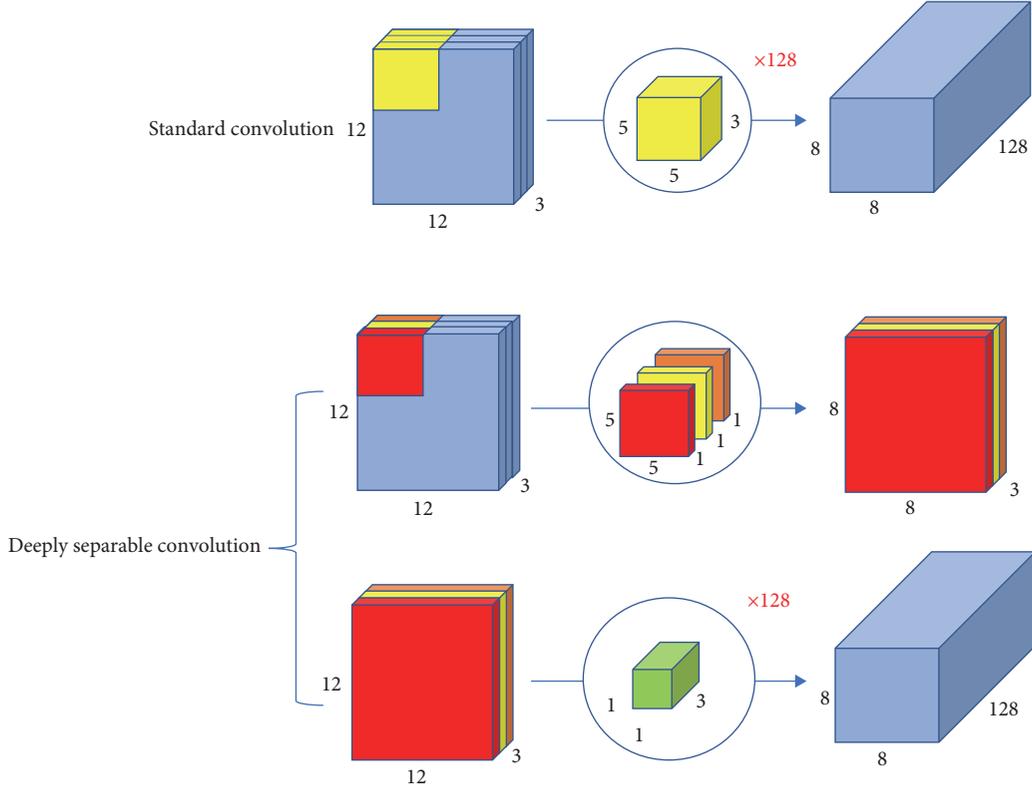


FIGURE 9: Deeply separable convolutional structures.

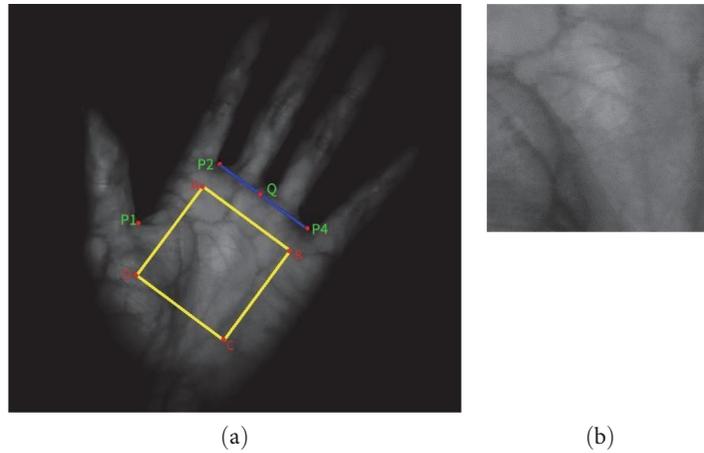


FIGURE 10: (a, b) Palm vein ROI extraction process.

TABLE 1: Experimental data setting.

Database	Train	Test	Total
CASIA	1,600	400	2,000
TJU-PV	4,800	1,200	6,000
Self-data	3,600	900	4,500

$$\text{Accuracy} = \frac{\sum_{i=1}^n P(x)}{n} \quad (7)$$

The ratio of the sum of the predicted correct keypoints to the number of all keypoints is recorded as accuracy.

4.3. Experimental Results of Keypoints Detection. The training set loss curve and the test set accuracy curve on the self-built dataset for 150 epochs are shown in Figure 11. The first step was to train the dataset from scratch using the original HRnet network. The second step was to initialize the improved HRnet model using the weights obtained from the initial training of the original HRnet network on the self-built dataset. The

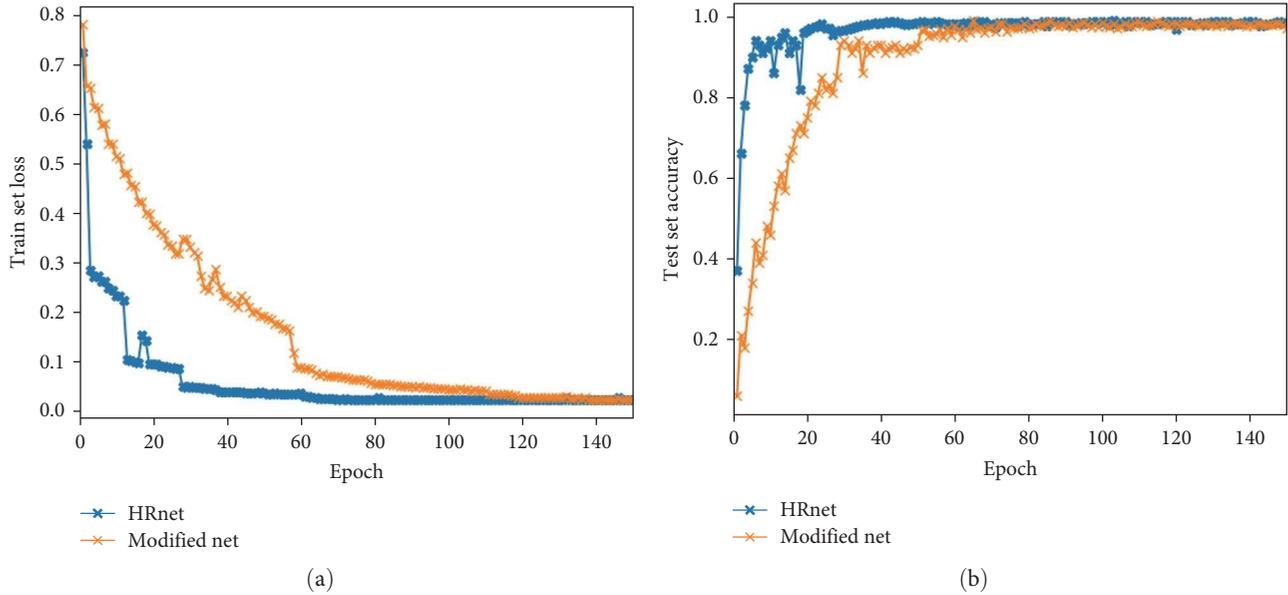


FIGURE 11: Model training loss and test accuracy curves: (a) training loss curve and (b) test accuracy curve.

TABLE 2: Accuracy of keypoint detection under different thresholds.

Threshold/class	CASIA	TJU-PV	Self-data
Threshold = 3	93.25%	96.27%	92.87%
Threshold = 4	95.43%	97.77%	94.39%
Threshold = 5	98.23%	98.74%	97.36%
Threshold = 6	99.26%	99.34%	98.72%

keypoints training loss of the proposed model and the original HRnet model is shown in Figure 11(a), and the keypoints detection accuracy of the test set (with a threshold value of five) is shown in Figure 11(b). It can be seen that the convergence of the HRnet was faster than the improved HRnet due to the deeper and more complex layers, but the improved network achieved the same prediction effect with much smaller parameter quantities.

Because the accuracy of palm vein keypoints detection was related to the value of the threshold, the keypoints localization accuracy according to different thresholds is provided in Table 2. The higher value the threshold was set, the more keypoints were considered valid and the higher accuracy would be achieved.

Regardless of the threshold setting, the keypoints localization model performed best in TJU-PV dataset due to its pure background. To determine the most appropriate threshold value, the effectiveness of the improved HRnet network for actual ROI extraction at different threshold values was verified. In Figure 12, it can be seen that the ROI region can be extracted accurately when threshold was 3, 4, and 5. When threshold was 6, the extraction effect was not so good because the no. 4 keypoint has a large deviation, which causes the whole ROI region to be shifted to the right. Finally, the threshold was set to 5 for calculating the keypoints localization of the palm vein images.

4.4. Comparison Experiments

4.4.1. Ablation Experiments for Keypoints Localization. The effects of different improvement methods were validated on keypoints detection, as shown in Table 3. From Table 3, it can be seen that the fusion of two layers of high-resolution and low-resolution features performed much better than the direct fusion of one layer of high-resolution and low-resolution features. Similarly, transpose convolution upsampling could also bring about an increase in accuracy. Through fusing two layer features and adopting transpose convolution upsampling, the keypoints detection accuracy reaches 98.24%. Furthermore, to further compress the model parameter size, depthwise separable convolution could be employed, resulting in 0.88% accuracy loss and 35% model size reduction

4.4.2. Comparison of Proposed Method and Traditional Machine Learning Methods for Keypoints Localization. Generally speaking, for traditional palm vein ROI extraction, the steps include image thresholding, manual segmentation, contour detection, keypoints localization, and ROI extraction. For the palm vein images with clean background, the ROI extraction comparison between the traditional methods and the proposed improved HRnet algorithm is shown in Figure 13. Both of them could perform well.

However, for palm vein images with complex backgrounds, the method of hand segmentation and hand boundary tracking

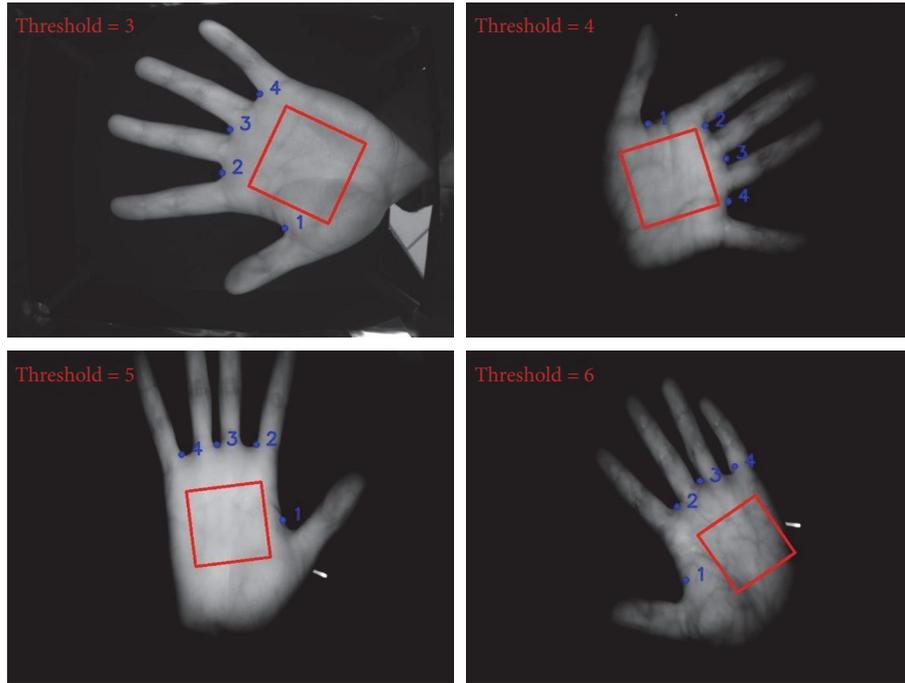


FIGURE 12: ROI extraction effect under different thresholds.

TABLE 3: Effect of different improvement methods on the accuracy of critical point detection.

Fusion of one-layer high and low resolutions	Fusions of two-layer high and low resolutions	Transpose convolution	Deeply separable convolution	Accuracy
√	—	√	—	94.35%
√	—	—	√	92.77%
√	—	√	√	93.17%
—	√	√	—	98.24%
—	√	—	√	96.83%
—	√	√	√	97.36%

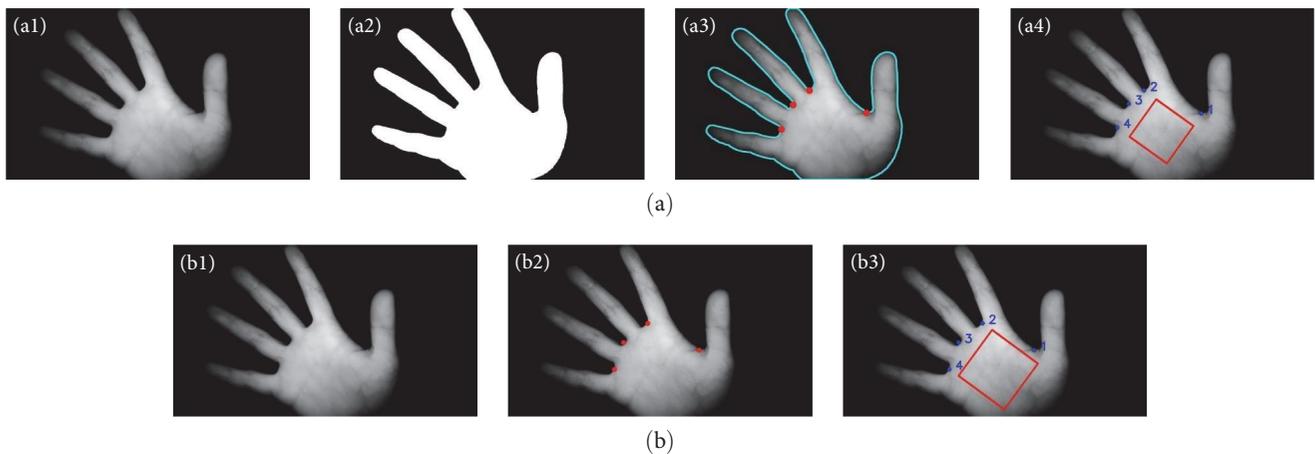


FIGURE 13: Comparison of ROI extraction effects in simple background. (a) Effect of traditional method of extraction: (a1) original image, (a2) binarization, (a3) valley point positioning, and (a4) ROI extraction. (b) Improved HRnet algorithm extraction effect: (b1) original image, (b2) keypoint positioning, and (b3) ROI extraction.

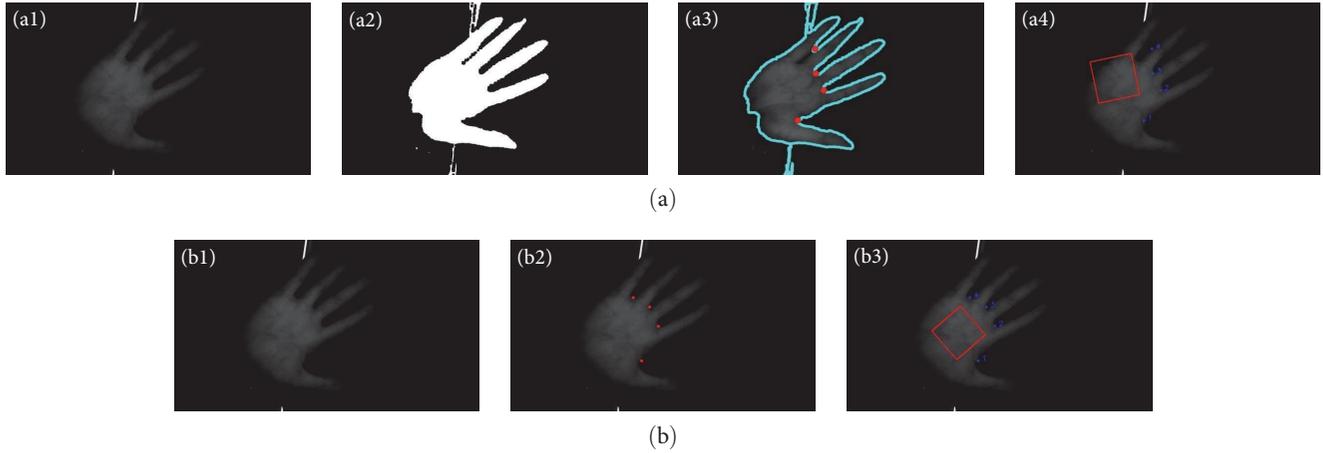


FIGURE 14: Comparison of ROI extraction effects in complex background. (a) Effect of traditional method of extraction: (a1) original image, (a2) binarization, (a3) valley point positioning, and (a4) ROI extraction. (b) Improved HRnet algorithm extraction effect: (b1) original image, (b2) keypoint positioning, and (b3) ROI extraction.

TABLE 4: Performance comparison of several target detection algorithms.

Algorithm	Image size	Params (M)	Accuracy	Speed (s)
VGG16	512×512	49.2	79.46%	0.5
Resnet-18	512×512	42.7	78.64%	0.23
Resnet-50	512×512	90.2	96.75%	0.7
HRnet	512×512	109	98.87%	1.3
Proposed	512×512	0.45	97.36%	0.0029

TABLE 5: Comparison with advanced vein ROI algorithms.

ROI algorithm	CASIA accuracy	TJU-PV accuracy	Self-data accuracy	Params (M)
Improved U-Net	97.78%	98.93%	97.25%	1.16
Proposed	98.23%	98.74%	97.36%	0.45

using OTSU thresholding method did not work well. This would lead to runaway of keypoints locations and make the subsequent palm vein ROI inaccurately localized, as shown in Figure 14.

4.4.3. Comparison of Proposed Method and Other Deep Learning Methods for Keypoints Localization. In Table 4, the proposed model was compared with the original HRNet as well as three other network models: VGG16, ResNet-18, and ResNet-50. It could be observed that when the input image size was fixed at 512×512 , the proposed network could achieve the accuracy of 97.36% with a model size of only 0.45 M and a runtime of only 0.029 s. In addition, the state-of-the-art vein ROI extraction algorithms, improved U-Net [26], and proposed method were compared in different datasets, as shown in Table 5.

4.4.4. Keypoints Detection for Irregular Hand Gestures. Irregular hand gesture, such as finger bending, finger closure, and wearing objects on hands could be seen in the palm vein datasets. In these three cases, the proposed HRnet algorithm

can also successfully extract the ROI of palm veins. The verification results are shown in Figure 15. In addition, the detection effect was verified under rotation, scaling, and fluorescent light backgrounds, as shown in Figure 16. These results provided important reference value for the practical application of palm vein recognition technology and help to improve the accuracy and stability of palm vein recognition.

4.5. Palm Vein Recognition Performance Experiment. After the extraction of ROI, the palm vein recognition was conducted in self-built dataset and two public datasets (CASIA and TJU-PV). The dataset was divided into training set and test set in 8:2 ratio. MobileFaceNet [27] was chosen as the network for feature extraction. Using cosine distance to calculate the similarity between feature vectors, followed by performing feature matching. The test results are shown in Figure 17.

Experiments were also conducted on the impact of different ROI sizes on palm vein recognition results in self-built dataset. By adjusting the size of the L parameter of the ROI extraction algorithm in Section 3.4, three different sizes of

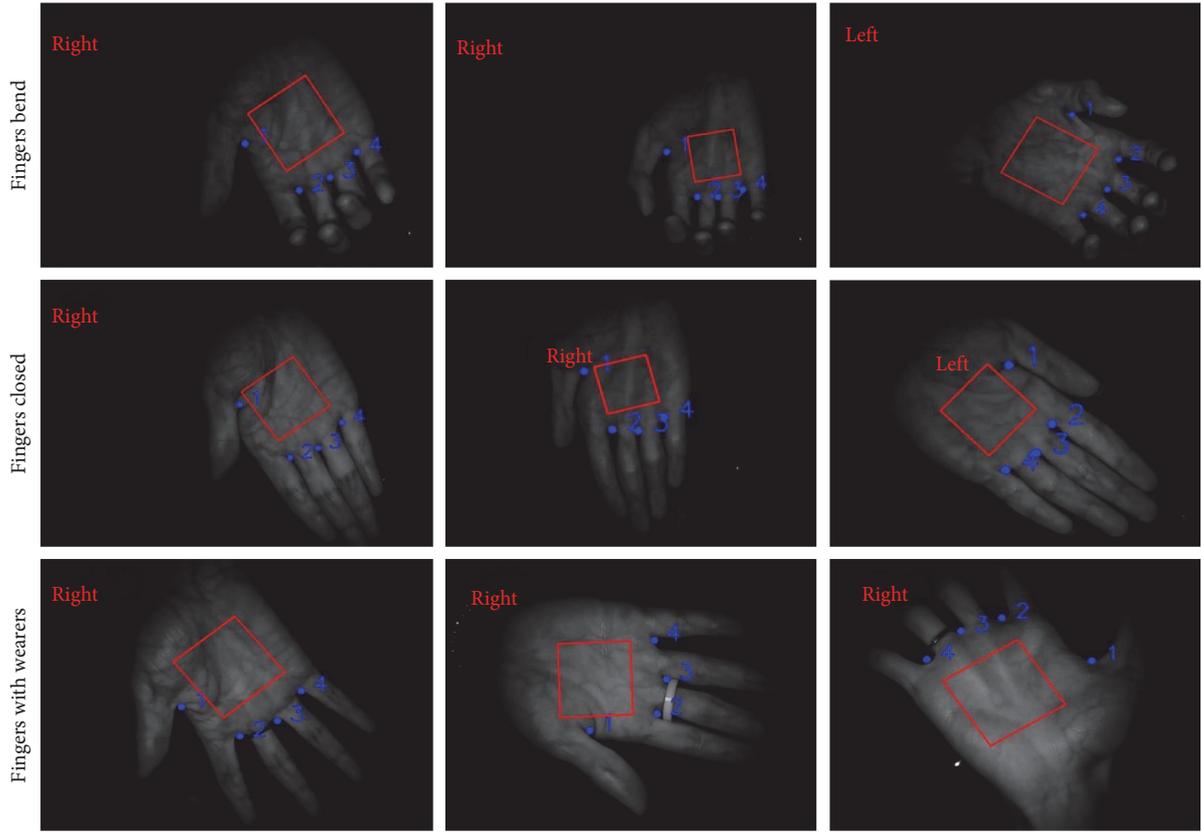


FIGURE 15: Extraction effects of three hand abnormalities.

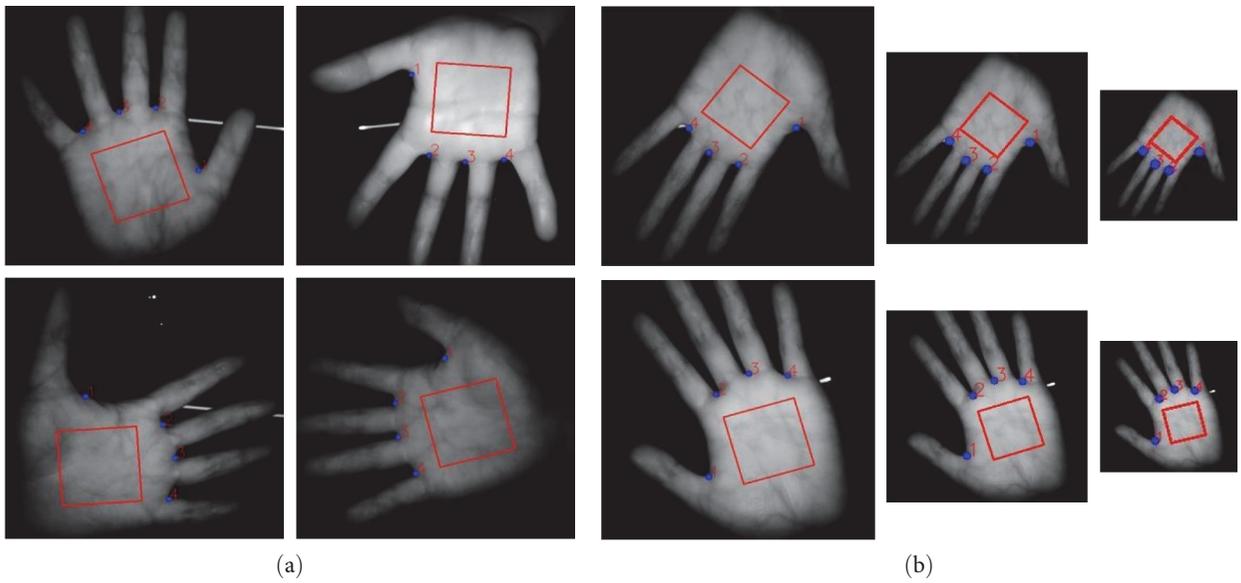


FIGURE 16: Continued.

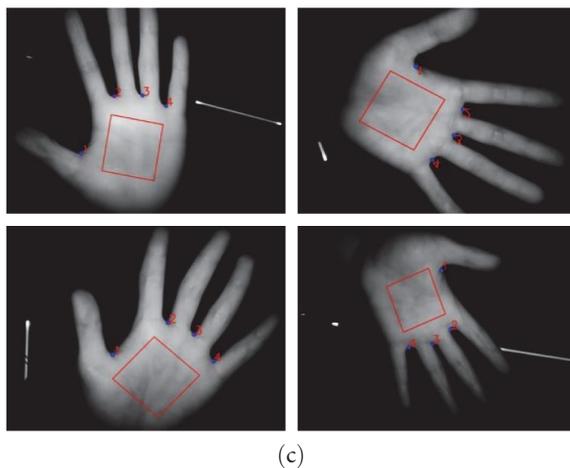


FIGURE 16: Detection of image deformation and background interference: (a) rotation, (b) scaling, and (c) fluorescent lamp.

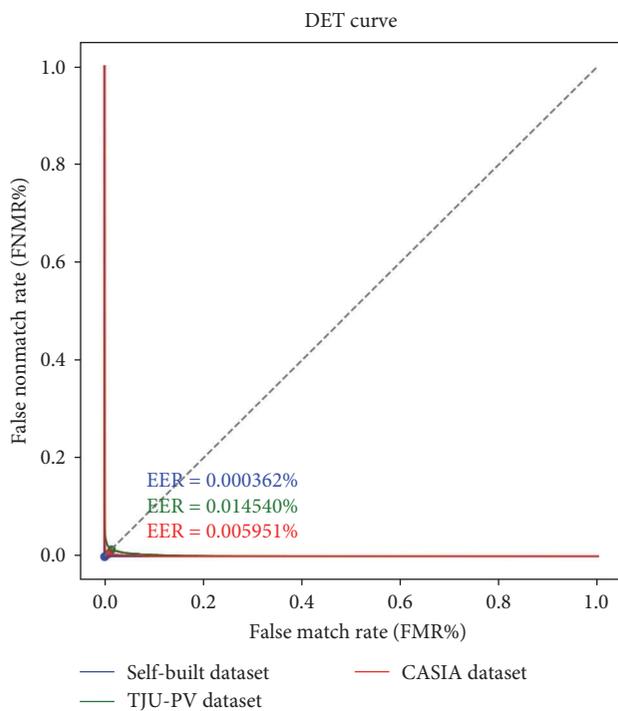


FIGURE 17: Detection-error tradeoff (DET) curve plotted between FMR against FNMR.

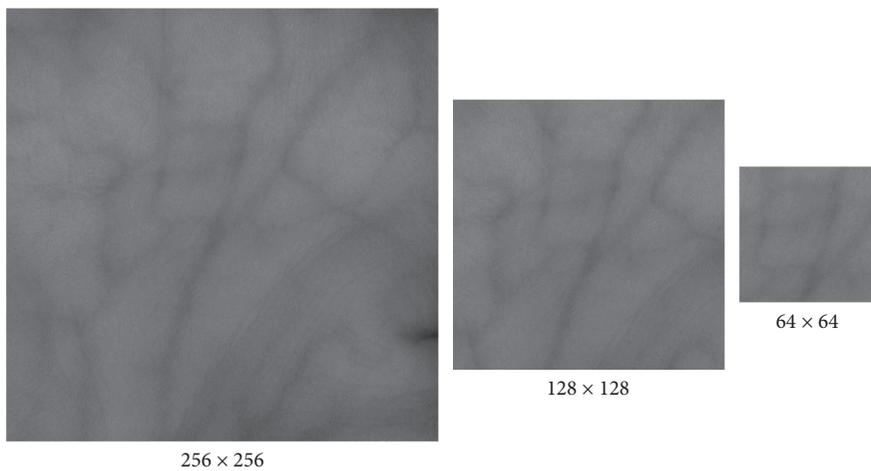


FIGURE 18: Palm vein ROI of three different sizes.

TABLE 6: The impact of different ROI sizes on palm vein recognition results.

Size of ROI	Performance (EER)	FNMR (FMR: 1%)	FNMR (FMR: 0.1%)	FNMR (FMR: 0.01%)
256 × 256	0.000362%	0.000000%	0.000000%	0.000001%
128 × 128	0.164162%	0.274622%	0.867673%	3.621233%
64 × 64	1.112907%	3.762512%	9.037083%	26.743207%

vein regions were selected, 256×256 , 128×128 , and 64×64 , respectively, as shown in Figure 18. In Table 6, the best performance was 256×256 , with an equal error rate (EER) of 0.00036%, when the vein region was adjusted downward to 64×64 due to the inclusion of smaller vein regions and large differences in vein characteristics between intraclass images, the EER was 1.11291%.

5. Conclusion

Fast and accurate extraction of the noncontact palm vein images ROI was the basis for subsequent palm vein recognition applications. In this paper, a lightweight ROI extraction algorithm was proposed based on improved HRnet for non-contact palm vein images, in which irregular gesture, complex backgrounds, and the problem of having items on the hands could be included. The experimental results showed that the method had good extraction accuracy, while the network model size was only 0.45 M and the running speed was only 0.029 s on a CPU with a clock speed of 3 GHz. The accuracy of keypoints detection reaches 97.36% on the self-built palm vein dataset and 98.23% and 98.74% on two public palm vein datasets, respectively. The EER of palm vein recognition was 0.000362%, 0.014541%, and 0.005951% and the false nonmatch rate (FNMR) was 0.000001%, 11.034725%, and 4.613714% (false match rate (FMR): 0.01%) in the self-built dataset, TJU-PV, and CASIA, respectively.

Experimental analysis showed that this method converged slowly during the training process, mainly due to the small differences in several target keypoints, the output error of the heat map, and the use of fewer convolutional channels. Future research will further consider the issue of balancing model parameter size. In addition, we will also study the resistance of spoofing attacks on palm vein recognition and the detection of finger defects and carry out subsequent palm vein recognition work.

Data Availability

The authors will supply the relevant data in response to reasonable requests.

Conflicts of Interest

The authors declare that there are no conflicts of interest between the authors, including financial, personal, or professional relationships, which could potentially influence the objectivity or integrity of this manuscript.

Authors' Contributions

Conceptualization: Fen Dai and Ziyang Wang; methodology: Ziyang Wang and Xiangqun Zou; software: Ziyang Wang and Rongwen Zhang; validation: Ziyang Wang, Xiangqun Zou, and Fen Dai; formal analysis: Ziyang Wang and Xiangqun Zou; investigation: Ziyang Wang and Xiangqun Zou; resources: Fen Dai, Xiangqun Zou, and Xiaoling Deng; data curation: Ziyang Wang and Xiangqun Zou; writing—original draft preparation: Ziyang Wang; writing—review and editing: Xiaoling Deng and Fen Dai; supervision: Fen Dai; project administration: Fen Dai; funding acquisition: Xiaoling Deng. All authors have read and agreed to the published version of the manuscript.

Acknowledgments

Special thanks to Chen Yifan and Lin Chenrui from the South China Agricultural University for their data collection, thanks to Wang Xi from Guangzhou No. 2 High School, and Jiang Yizhi from Guangzhou Zhixin High School for their data annotations. This research was funded by the IEC NSFC 191320-International Exchanges 2019 Cost Share (NSFC).

References

- [1] Z. Zhang, S. Liu, and M. Liu, "A multi-task fully deep convolutional neural network for contactless fingerprint minutiae extraction," *Pattern Recognition*, vol. 120, Article ID 108189, 2021.
- [2] M. Taskiran, N. Kahraman, and C. E. Erdem, "Face recognition: past, present and future (a review)," *Digital Signal Processing*, vol. 106, Article ID 102809, 2020.
- [3] Q. Hu, S. Yin, H. Ni, and Y. Huang, "An end to end deep neural network for iris recognition," *Procedia Computer Science*, vol. 174, pp. 505–517, 2020.
- [4] D. Zhong, X. Du, and K. Zhong, "Decade progress of palmprint recognition: a brief survey," *Neurocomputing*, vol. 328, pp. 16–28, 2019.
- [5] S. Zhao and B. Zhang, "Deep discriminative representation for generic palmprint recognition," *Pattern Recognition*, vol. 98, Article ID 107071, 2020.
- [6] T. Chai, S. Prasad, and S. Wang, "Boosting palmprint identification with gender information using deepnet," *Future Generation Computer Systems*, vol. 99, pp. 41–53, 2019.
- [7] 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, 2019.
- [8] X. Meng, J. Zheng, X. Xi, Q. Zhang, and Y. Yin, "Finger vein recognition based on zone-based minutia matching," *Neurocomputing*, vol. 423, pp. 110–123, 2021.

- [9] T. Chai, J. Li, S. Prasad, Q. Lu, and Z. Zhang, "Shape-driven lightweight CNN for finger-vein biometrics," *Journal of Information Security and Applications*, vol. 67, Article ID 103211, 2022.
- [10] T. Chai, S. Wang, and D. Sun, "A palmprint ROI extraction method for mobile devices in complex environment," in *2016 IEEE 13th International Conference on Signal Processing (ICSP)*, pp. 1342–1346, IEEE, 2016.
- [11] A. S. El Sayed, H. M. Ebeid, M. Roushdy, and Z. T. Fayed, "A method for contactless palm ROI extraction," in *2016 11th International Conference on Computer Engineering & Systems (ICCES)*, pp. 193–200, IEEE, 2016.
- [12] W. Kang and Q. Wu, "Contactless palm vein recognition using a mutual foreground-based local binary pattern," *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 11, pp. 1974–1985, 2014.
- [13] S. Lin, T. Xu, and X. Yin, "Region of interest extraction for palmprint and palm vein recognition," in *2016 9th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*, pp. 538–542, IEEE, 2016.
- [14] M. Yakno, J. Mohamad-Saleh, and B. A. Rosdi, "New technique for larger ROI extraction of hand vein images," in *2015 International Conference on BioSignal Analysis, Processing and Systems (ICBAPS)*, pp. 82–87, IEEE, 2015.
- [15] W. Damak, R. Boukhris Trabelsi, M. Alima Damak, and D. Sellami, "Dynamic ROI extraction method for hand vein images," *IET Computer Vision*, vol. 12, no. 5, pp. 586–595, 2018.
- [16] M. E. Cimen, O. F. Boyraz, M. Z. Yildiz, and A. F. Boz, "A new dorsal hand vein authentication system based on fractal dimension box counting method," *Optik*, vol. 226, Article ID 165438, 2021.
- [17] W. Wu, Q. Wang, S. Yu et al., "Outside box and contactless palm vein recognition based on a wavelet denoising resnet," *IEEE Access*, vol. 9, pp. 82471–82484, 2021.
- [18] G. Ananthi, J. R. Sekar, and S. Arivazhagan, "Human palm vein authentication using curvelet multiresolution features and score level fusion," *The Visual Computer*, pp. 1–14, 2022.
- [19] R. Girshick, "Fast r-cnn," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1440–1448, IEEE, 2015.
- [20] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: towards real-time object detection with region proposal networks," *Advances in Neural Information Processing Systems*, vol. 28, 2015.
- [21] K. He, G. Gkioxari, P. Dollar, and R. Girshick, "Mask r-cnn," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2961–2969, IEEE, 2017.
- [22] J. Redmon and A. Farhadi, "Yolov3: an incremental improvement," 2018.
- [23] Y. Zhang, L. Zhang, R. Zhang, S. Li, J. Li, and F. Huang, "Towards palmprint verification on smartphones," 2020.
- [24] K. Luo and D. Zhong, "Robust and adaptive region of interest extraction for unconstrained palmprint recognition," *Journal of Electronic Imaging*, vol. 30, no. 3, pp. 033005–033005, 2021.
- [25] K. Sun, B. Xiao, D. Liu, and J. Wang, "Deep high-resolution representation learning for human pose estimation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5693–5703, IEEE, 2019.
- [26] R. Zhang, X. Zou, X. Deng et al., "Fast and accurate ROI extraction for non-contact dorsal hand vein detection in complex backgrounds based on improved u-net," *Sensors*, vol. 23, no. 10, Article ID 4625, 2023.
- [27] F. Schroff, D. Kalenichenko, and J. Philbin, "Facenet: a unified embedding for face recognition and clustering," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 815–823, IEEE, 2015.