Human Dorsal Hand Vein Segmentation Method Based on GR-UNet Model

Zhike Zhao[1], Wen Zeng[2], Kunkun Wu[3], Xiaocan Cui[4]
School of Electrical Engineering, Henan University of Technology, Zhengzhou, China[1, 2]
Henan Baichangyuan Medical Technology Co., LTD, Zhengzhou, China[3]
Xinxiang First People's Hospital, Xinxiang, China[4]

***Abstract*—To solve the issue of inaccurate segmentation accuracy of human dorsal hand veins (HDHV), we propose a segmentation method based on the global residual U-Net (GR-Unet) model. Initially, a visual acquisition device for dorsal hand vein imaging was designed utilizing near-infrared technology, resulting in the creation of a dataset comprising 864 images of HDHV. Subsequently, a Bottleneck from the deep residual network-50 (ResNet50) was integrated into the U-Net model to enhance its depth and alleviate the problem of vanishing gradients. Furthermore, a global attention mechanism (GAM) was introduced at the junction to improve the acquisition of global feature information. Additionally, a weighted loss function that combines cross-entropy loss and Dice loss was employed to address the imbalance between positive and negative samples. The experimental results indicate that the GR-Unet model achieved accuracies of 78.82%, 88.03%, 93.92%, and 97.5% in terms of intersection over union, mean intersection over union, mean pixel accuracy, and overall accuracy, respectively.**

***Keywords*—Human dorsal hand veins; GR-UNet; near infrared technology; deep residual network-50; global attention mechanism; loss function**

I. INTRODUCTION

Venipuncture is a critical procedure for blood collection, transfusion, and infusion in clinical medicine, with the hand serving as the primary site for this practice. The conventional technique for venipuncture on the dorsal aspect of the hand involves direct venipuncture, during which the patient clenches their fist and utilizes a pressure band to engorge the vein. Subsequently, medical personnel assess the optimal site for venipuncture based on clinical experience before performing the procedure. However, due to variations in the distribution of human veins, achieving accurate venous puncture through manual means alone can be challenging, particularly in children, the elderly, and obese individuals. These populations often present with thinner venous vessels, poorer vascular elasticity, and increased perivenous fat, which complicates venous positioning in the context of venipuncture [1]. Therefore, enhancing the imaging of hand veins using computer-assisted methods may be crucial for improving venous positioning.

Venous positioning is a prerequisite for successful venipuncture. Numerous researchers have conducted extensive studies to accurately locate venous vessels. Ma et al. [2] processed images of the dorsal hand vein using contrast-limited adaptive histogram equalization and multi-scale detail fusion algorithms, subsequently weighting and superimposing the two to enhance the vein images. Although this method addressed the issue of detail loss associated with the histogram equalization algorithm, it remained ineffective in managing vein images that contained high-frequency noise. Kuang et al. [3] employed multi-scale Gaussian blurring to denoise the L component extracted from the LAB color space representation of the image, and utilized guided filtering for illuminance estimation, combined with adaptive thresholding for dynamic adjustment to enhance the vein image. However, the algorithm's adaptability to various application environments required further improvement. Besra et al. [4] proposed a vein segmentation algorithm based on the repeated line tracking method, which achieved vein segmentation by superimposing the trajectory lines in the vein images tracked from different starting points. Nonetheless, this method was highly dependent on image clarity and involved a substantial number of arithmetic operations for image processing, making it challenging to meet the demands of real-time detection and recognition. Yakno et al. [5] employed contrast-limited adaptive histogram equalization and fuzzy adaptive gamma transform to enhance vein image processing. They subsequently weighted the two methods and applied a matched filter with a first-order derivative of the Gaussian function to achieve vein image segmentation. However, this approach required multiple transformations and integrations of the image, leading to high computational complexity and limited effectiveness in managing high-brightness areas. In contrast, Yang et al. [6] utilized a six-dimensional Gabor filter for vein image enhancement and a Markov random field for vein image segmentation. Nonetheless, this method demonstrated poor performance in handling low-contrast vein images or those containing high-frequency noise.

With the rapid development of deep learning theory, achieving vein image segmentation using convolutional neural networks has become increasingly feasible [7]. In 2015, Long et al. [8] proposed Fully Convolutional Networks, which processed pixel-level data by employing full convolution instead of full connectivity, thereby allowing for the processing of image inputs of any size. However, this approach did not adequately address the relationships between pixels when processing them independently, leading to incomplete feature information extraction. In the same year, Ronneberger et al. [9] introduced the U-Net network algorithm for image segmentation, which effectively addressed challenges in low-resolution image processing and has since gained widespread application in the field of medical image segmentation. The direct application of the U-Net network model to venous vessel

segmentation still presents several limitations: (1) A disproportionately small sample set leads to overfitting during network training; (2) The single encode-decode structure may cause the network training to reach a performance ceiling, making it challenging to extract additional effective features; (3) Although the U-Net decoding stage integrates local feature information through skip connections, simple concatenation does not fully merge features of varying scales. To address these issues, researchers have proposed enhanced algorithms based on the U-Net model by modifying the backbone and incorporating an attention mechanism. He et al. [10] introduced an algorithm that leverages the U-Net model alongside the attention mechanism, which improved the identification of multi-scale features and enhanced segmentation accuracy, although it was less effective in segmenting narrow venous vessels. Lefkovits et al. [11] proposed a hybrid approach combining unsupervised and supervised techniques for dorsal hand vein segmentation, experimentally demonstrating that the segmentation results of eight U-Net variants surpassed those of traditional image methods, with the ResNet-U-Net model exhibiting particularly significant performance. Gao et al. [12] proposed a semantic segmentation model called AT-U-Net, which integrates the Non-Local attention mechanism into the U-Net architecture to enhance feature extraction capabilities. This approach addresses the challenge of long-distance dorsal hand vein puncture; however, the limitation in capturing venous vessel images from afar may result in inadequate extraction of global feature information. Additionally, high-frequency noise and blurred vessel edges can adversely affect segmentation accuracy. Chen et al. [13] incorporated a Gabor convolution kernel into the U-Net framework and utilized inverted residual blocks to achieve model lightweighting, thereby mitigating the semantic information loss attributed to channel issues. Nonetheless, this algorithm depends heavily on the extraction of shallow features, and the parameters of the Gabor filter require manual adjustment, which adds to the algorithm's complexity.

In contrast to the limitations of traditional image segmentation algorithms, neural networks are more adept at managing complex image features and capturing nonlinear relationships within images, thereby achieving enhanced segmentation accuracy and robustness. Consequently, to improve the extraction of HDHV, this paper proposes a segmentation algorithm based on GR-Unet. The Bottleneck portion of the ResNet50 serves as the backbone network, while GAM is integrated into the skip connections, along with pre-trained weights to expedite feature extraction. The network's capacity to learn from defective pixels is further augmented through a weighted loss function that combines cross-entropy and Dice loss. Experiments conducted on a self-constructed dorsal hand vein dataset validate the effectiveness of the proposed method.

II. Design Scheme

The design scheme of this paper encompasses several key components: image acquisition, image preprocessing, data enhancement, dataset construction, GR-UNet model development, network training, and experimental testing sessions, as illustrated in Fig. 1. Specifically, image preprocessing involved grayscale normalization, global threshold binarization, the centroid method for extracting regions of interest (ROI), averaging filtering, and contrast-limited adaptive histogram equalization (CLAHE). The GR-UNet model utilized ResNet50 as its backbone network, incorporating the GAM module and a modified loss function at the skip connections.

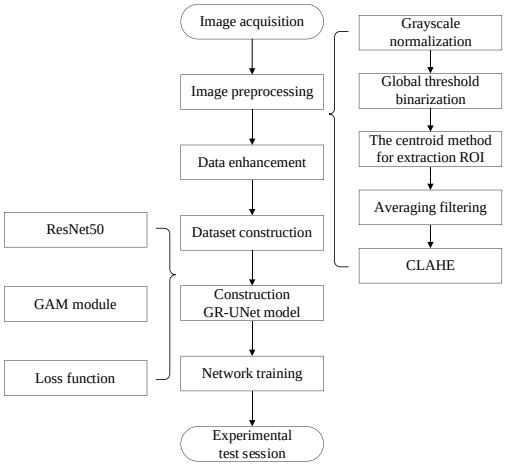


Fig. 1. The technical roadmap of this study.

III. Dorsal Hand Vein Dataset

A. Dorsal Hand Vein Collection System Design

Due to the differential absorption of near-infrared (NIR) light by human tissues, hemoglobin in human veins can absorb NIR light in the range of 700 nm to 1100 nm [14]. Consequently, images within this spectral band (700 nm to 1100 nm) can be captured using a camera equipped with a suitable filter. However, during the experimental process, results indicated that while NIR light could penetrate 5 to 10 mm into human tissues [14], it was less effective in capturing images from individuals with higher fat content in their hands. Existing studies have demonstrated that the use of 850 nm NIR light can effectively highlight the contour structure of veins [15-16], whereas adipose tissue exhibits optimal absorption characteristics at the 940 nm NIR wavelength band [17]. Therefore, this paper proposes a dual-band measurement superposition method, which involves utilizing a 940 nm NIR light source for top irradiation combined with an 850 nm NIR light source for bottom irradiation to facilitate the acquisition of human hand vein images. The schematic representation of the designed human hand vein image acquisition system is illustrated in Fig. 2.

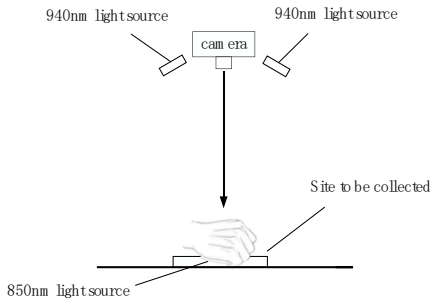


Fig. 2. Schematic diagram of human hand back image acquisition equipment.

In this study, a total of 72 male and female subjects were experimentally recruited, with ages ranging from 3 to 60 years. Among these participants, 30 were minors and 42 were adults. Each subject had images captured of their left and right hand veins, resulting in a total of 144 images, each with a resolution of 640x480 pixels. To enhance the diversity of the image samples, an additional 864 images of human hand veins were generated through rotation and mirroring of the collected images. The images of human dorsal hand vein obtained after NIR filtering is shown in Fig. 3.

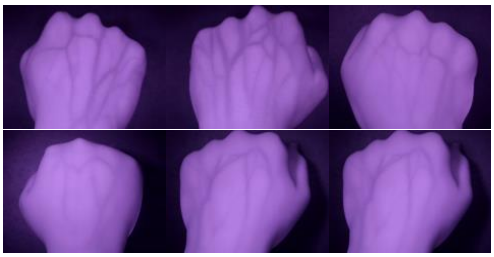


Fig. 3. The images of the human dorsal hand vein obtained after NIR filtering.

B. Image Preprocessing

Due to the interference of realistic conditions, including variations in illumination and the differing thickness of the dorsal hand vein among individuals, the contrast of the collected dorsal hand vein images was suboptimal and contained various artifacts. This compromised the clarity of the boundary between the veins and the background, leading to potential misjudgment in identifying the veins. Consequently, to accurately segment the venous vessels on the dorsal hand vein, it is essential to preprocess the vein images to effectively correct for uneven illumination and eliminate shadows and artifacts. The specific steps of image preprocessing are as follows:

Step 1: Grayscale normalization was employed to standardize the grayscale values of the dorsal hand vein image to a range of 0 to 255, with the calculation formula presented in Eq. (1).

$$P(x,y)=\frac{\left[R(x,y)-R_{\min}(x,y)\right]\times 255}{R_{\max}(x,y)-R_{\min}(x,y)} \tag{1}$$

Wherein, $R(x,y)$ denotes the gray value of the dorsal hand vein image, $R_{\min}(x,y)$ and $R_{\max}(x,y)$ denote the minimum and maximum values of the gray value of the dorsal hand vein image, respectively, and $P(x,y)$ denotes the normalized gray value of the dorsal hand vein image.

Step 2: The dorsal hand vein images were binarized using the global thresholding method, resulting in the dorsal part of the hand appearing in white against a black background.

Step 3: The centroid method $G(x_i, y_j)$ was applied to the binary image to calculate the centroid, and a ROI measuring 288×288 pixels was extracted using this centroid as the reference point. The formulas for the centroid method calculations are provided in Eq. (2) and Eq. (3).

$$x_i=\frac{\sum_{i=0}^{h}\sum_{j=0}^{w}\left[i\times f(i,j)\right]}{\sum_{i=0}^{h}\sum_{j=0}^{w}f(i,j)} \tag{1}$$

$$y_i=\frac{\sum_{i=0}^{h}\sum_{j=0}^{w}\left[j\times f(i,j)\right]}{\sum_{i=0}^{h}\sum_{j=0}^{w}f(i,j)} \tag{3}$$

Where, h and w denote the height and width of the image, respectively, and $f(x,y)$ denotes the pixel point in the white area.

Step 4: An averaging filter was applied to the ROI to reduce the Gaussian noise present in the image.

Step 5: The CLAHE algorithm was utilized to enhance the contrast of the ROI. Finally, the processed image of the dorsal hand vein is presented in Fig. 4.

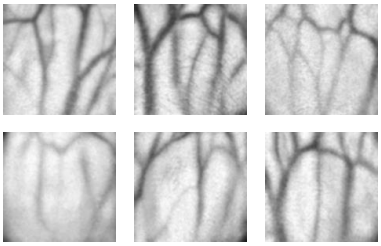


Fig. 4. Preprocessed image results of dorsal hand vein.

C. Dataset Construction

The preprocessed images of HDHV were labeled on a pixel-by-pixel basis, identifying the dorsal hand vein vessels. This process established the labeled images and ultimately constructed the dataset. Eighty percent of the images were randomly assigned to the training set, while ten percent were designated for the validation set and the remaining ten percent for the test set. Randomly selected original and labeled images from the dataset are presented in Fig. 5.

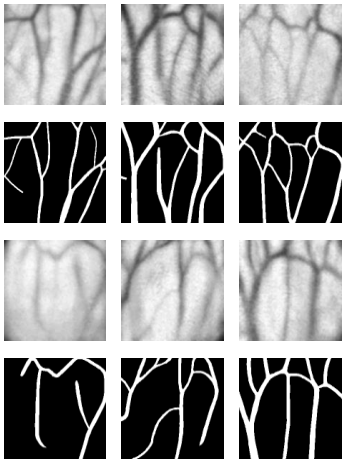


Fig. 5. Original and labeled images.

IV. METHODS

A. Basic Network Model

The U-Net model, developed by Ronneberger et al. [9], was initially applied as a convolutional neural network for cell segmentation in medical images. It features a U-shaped symmetric structure divided into encoding and decoding networks, as illustrated in Fig. 6. The left side comprises the encoding network, which consists of four subsampled modules. Each module contains two consecutive 3×3 convolutional layers, the ReLU activation function, and a 2×2 max-pooling layer. The primary purpose of the encoding network is to extract information from shallow features, as the size of the feature map decreases while the number of feature channels increases. Conversely, the right side contains the decoding network, which is upsampled by stacking 2×2 transposed convolutional layers. After each upsampling, it is spliced with the corresponding feature layer from the encoding network using skip connections, thereby fusing feature information from both deeper and shallower layers. This combined information is then processed by convolutional blocks to gradually recover the size and spatial dimensions of the input image. Ultimately, the segmentation result is obtained through a 1×1 convolutional layer.

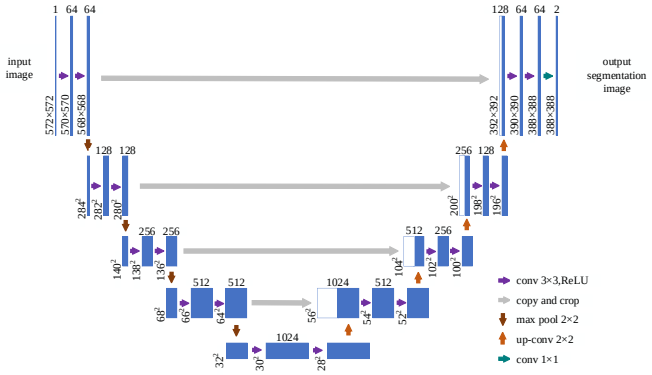


Fig. 6. U-Net model structure.

This paper proposes a GR-UNet model structure divided into two parts: encoding and decoding, as illustrated in Fig. 7. The encoding component utilizes the bottleneck part of the ResNet50 network, effectively mitigating issues related to vanishing gradients and network degradation. In the decoding phase, five up-sampling operations are conducted using bilinear interpolation, followed by two 3×3 convolutions with ReLU activation functions after each up-sampling to eliminate the aliasing effects present in the feature maps. Additionally, the GAM attention mechanism is incorporated at the skip connections to enhance the extraction of global feature information. This allows the network to focus more on the features of HDHV across both spatial and channel dimensions, thereby assigning greater weight coefficients within the network.

B. The ResNet50 Network

To obtain deeper features, this paper replaces the encoding component of the U-Net network with the Bottleneck module from the ResNet50 architecture. The global average pooling layer and the fully connected layer are removed, resulting in the construction of the ResNet50-UNet network. This modified architecture is employed to perform the vein segmentation task. The ResNet50-UNet network effectively combines the strengths of both the U-Net and ResNet50 architectures, thereby mitigating issues related to vanishing gradients and network degradation. Additionally, it fully leverages shallow features to enhance the efficiency of feature extraction when analyzing HDHV.

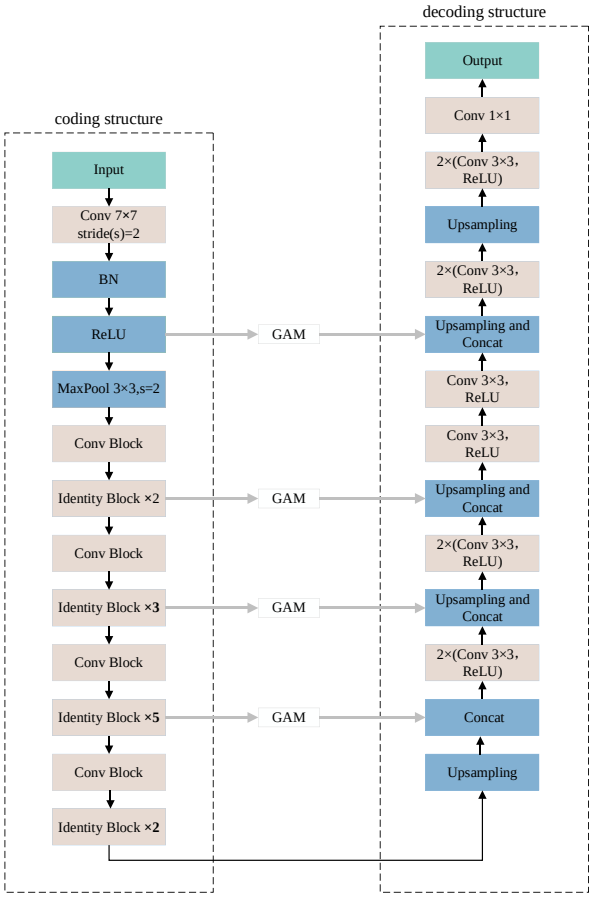


Fig. 7. GR-UNet model structure.

The ResNet50 network was proposed by He et al. [18] and demonstrates excellent performance across various visual application domains, including classification, object detection, and semantic segmentation. The architecture of the ResNet50 network primarily consists of an input layer, a residual block, a global average pooling layer, and a fully connected layer, as illustrated in Fig. 8. The residual mapping component on the left side of the Conv Block undergoes three convolution processes: a 1×1 convolution (with a stride of 2), a 3×3 convolution (with a stride of 1), and another 1×1 convolution (with a stride of 1). Following each convolution, a Batch Normalization (BN) layer is employed to normalize the features. The BN layer associated with the first two convolutions is succeeded by a ReLU activation function, which helps mitigate the vanishing gradient problem. In the right direct mapping component, a 1×1 convolution (with a stride of 2) and a BN layer normalization are performed, after

which the feature map output from the residual mapping is added pixel-by-pixel to the feature map output from the direct mapping. The combined output is then processed through a ReLU activation function, resulting in a feature map that is half the size of the input feature map. The Identity Block is analogous to the Conv Block; however, the left side maintains the same structure as the Conv Block, with all convolutional layers having a stride of 1. The right side features a direct mapping that does not include the 1×1 convolution operation or BN layer normalization, allowing the inputs to be added directly to the outputs, which subsequently pass through the ReLU activation function. The Conv Block is designed to reduce the size of the feature map, while the Identity Block serves to deepen the network.

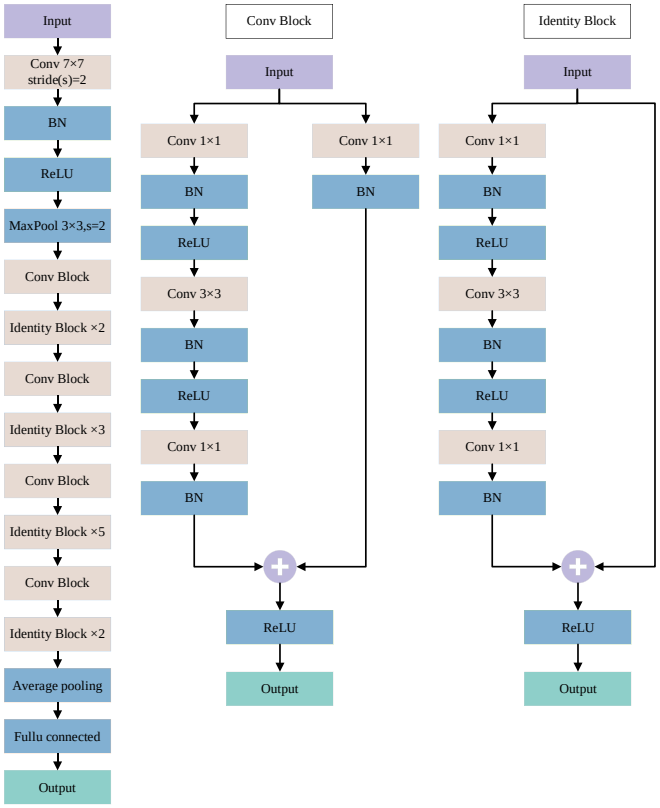


Fig. 8. ResNet50 network structure.

C. The GAM Attention Module

The decoder construction of U-Net closely resembles that of the encoder. Popular upsampling operations, such as transpose convolution, are local operations [19]. However, they do not adequately mine the in-depth feature information of the context, leading to incomplete information being conveyed during skip connections and subsampling. Consequently, this results in a failure to obtain rich feature information. To address this issue and achieve a more comprehensive understanding of HDHV, the GAM attention module has been introduced at the skip junction. This enhancement facilitates the acquisition of global feature information, allowing for further optimization and processing of deep feature information to better capture the global contextual information. The GAM attention mechanism is designed to minimize information loss while amplifying global dimensional interaction features [20]. It is a redesign based on the CBAM [21] attention mechanism, which also employs spatial and channel attention mechanisms. Its structure is illustrated in Fig. 9.

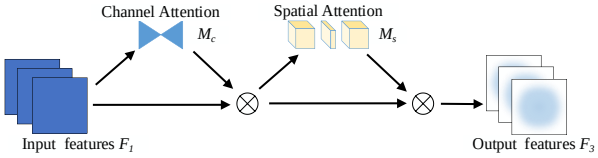


Fig. 9. GAM attention module.

In the whole process, the input features are $F_1 \in R^{C\times H\times W}$, F_2 is the intermediate state, and F_3 is the output features, and the relationship between them is shown in Eq. (4) and Eq. (5):

$$F_2 = M_C(F_1) \otimes F_1 \tag{2}$$

$$F_3 = M_s(F_2) \otimes F_2 \tag{3}$$

Where, M_C and M_s are channel feature maps and spatial feature maps respectively, and $\otimes$ is the multiplication operation.

In the channel attention submodule, a three-dimensional arrangement is initially applied to the input data to preserve its information across three distinct dimensions. Subsequently, the cross-dimensional channel and spatial dependencies are enhanced by a two-layer Multilayer Perceptron (MLP), which performs a dimensionality transformation, reverting the dimensions to their original state. Finally, the output is processed through the Sigmoid activation function, allowing for the fusion of information across different dimensions by implementing the dimensional transformation. The structure of the channel attention submodule is illustrated in Fig. 10.

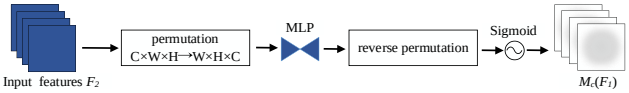


Fig. 10. Channel attention submodule.

In the spatial attention submodule, the convolution operation is primarily performed on the input feature graph F_2. Initially, information from the spatial layers is fused using two 7×7 convolutional layers to enhance the learning of spatial features. The output is then processed through the Sigmoid activation function to enhance the integration of data across different dimensions. The structure of the spatial attention submodule is illustrated in Fig. 11.

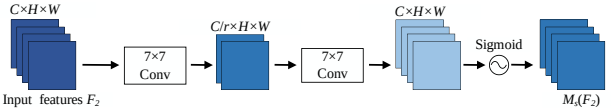


Fig. 11. Spatial attention submodule.

D. Improved Loss Function

To address the imbalance between positive and negative samples, this paper improves the original loss function by employing a weighted combination of Cross-Entropy Loss (CE

Loss) and Dice Loss. The calculation of the loss function is detailed in Eq. (6) to Eq. (8).

$$Loss = CE\_Loss + Dice\_Loss \tag{6}$$

$$CE\_Loss = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C} y_{i,c}\log\left(p_{i,c}\right) \tag{7}$$

$$Dice\_Loss = 1 - \frac{2\times\left(\sum_{i=1}^{N} P_i y_i\right)}{\sum_{i=1}^{N} P_i + \sum_{i=1}^{N} y_i} \tag{8}$$

Where CE_Loss is the cross-entropy loss function, $Dice\_Loss$ is the Dice loss function, N is the number of samples, C is the number of classes, $y_{i,c}$ is the actual label of sample i belonging to class c, $p_{i,c}$ is the probability of sample i belonging to class c, P_i and y_i are the predicted value of the model and the real label respectively.

V. Experiments

A. The Experimental Environment

The hardware configuration used for network training and testing in this article is as follows: 12th Gen Intel(R) Core(TM) i7-12700H (2.3 GHz), 16 GB memory, NVIDIA GeForce RTX 3060, CUDA12.1, Windows 11, Python3.8 and Pytorch2.3. The hyperparameters of the network are: learning rate 0.0001, number of epochs 200, batch size 4, and optimizer Adam.

B. Evaluation of Indicators

Dorsal hand vein classification can be viewed as classifying each pixel in the vein image, with each pixel classification is described by a confusion matrix. The four elements of TP, FN, TN, and FP confusion matrix, where TP is for correctly predicting hand dorsal vein samples as hand dorsal veins, FN for incorrectly predicting hand dorsal vein samples as non-hand dorsal veins, TN for correctly predicting non-hand dorsal vein samples as non-hand dorsal veins, and FP for incorrectly predicting non-hand dorsal vein samples as hand dorsal veins, and the confusion matrix is shown in Table I.

In this paper, the evaluation metrics used for the experimental results are Intersection Over Union (IOU), Mean Intersection Over Union (MIOU), Mean Pixel Accuracy (MPA) and Accuracy for categories. The formulas for IOU, MIOU, MPA and Accuracy are shown in Eq. (9) to Eq. (12), where n denotes the number of categories.

$$IOU = \frac{TP}{TP + FP + FN} \tag{4}$$

$$MIOU = \frac{1}{n}\sum_{i=1}^{n}\frac{TP}{TP + FP + FN} \tag{5}$$

$$MPA = \frac{1}{n}\sum_{i=1}^{n}\frac{TP}{TP + FP} \tag{6}$$

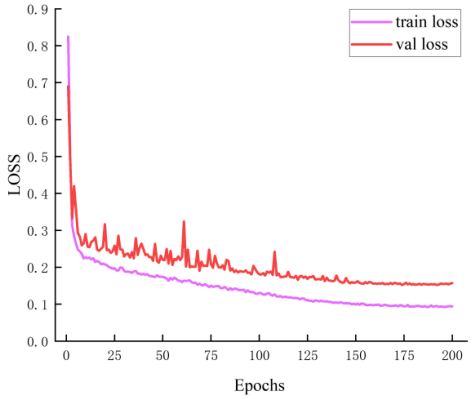
$$Accuacy = \frac{TP + TN}{TP + TN + FP + FN} \tag{12}$$

TABLE I. Confusion Matrix of Dorsal Hand Vein Segmentation

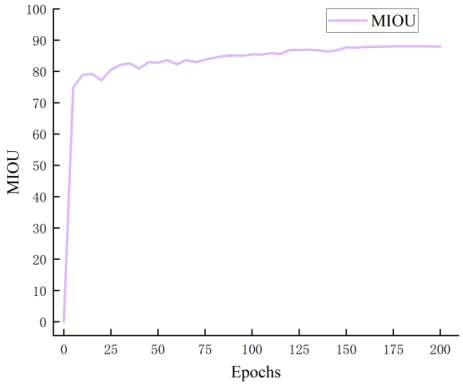
Projected results	Real results	
	hand dorsal vein	Non-dorsal hand veins
hand dorsal vein	TP	FP
Non-dorsal hand veins	FN	TN

C. Training Results of GR-UNet Model

In this paper, the attention mechanism and residual network were introduced to improve U-Net, while a weighted loss function was used to further prevent the imbalance of positive and negative samples. Finally, the improved model was used to train the detection on a self-constructed dataset of hand dorsal veins. The training results of the model are shown in Fig. 12, which indicates that as the number of iterations increased, the accuracy of the model increased while the loss value decreased, and when the loss curve tended to stabilize, the network converged and the training ended. Model accuracy reached a high steady state after 160 iterations.



(a) Loss value.



(b) MIOU.

Fig. 12. Training results of GR-UNet model.

D. Comparative Experiments on Network Models with Different Backbones

In order to verify the impact of backbone feature extraction network on the performance and accuracy of U-Net model, Mobile-UNet, VGG16-UNet and ResNet50-UNet were respectively used for comparative experiments. The U-Net models with different backbone were evaluated by IOU, MIOU, MPA and Accuracy indexes. Results are shown in Table II, where the model performs best when ResNet50 was used as the backbone network.

E. Results of Ablation Experiment

Ablation experiment is a widely used analytical method in machine learning and deep learning to assess the impact of a component of an algorithm, model, or system on overall performance by systematically removing or modifying that component.

The GR-UNet proposed model was experimentally analyzed using a self-constructed dataset of human dorsal hand vein images. An ablation study was conducted utilizing the GR-UNet model to assess the effectiveness of the improved ResNet50 backbone, the incorporation of the GAM attention mechanism at the skip connections, and the enhanced loss function. As demonstrated in Table III, compared to the U-Net model, the IOU of the GR-UNet model increased by 5.28%, the MIoU increased by 2.96%, the MPA increased by 4.14%, and accuracy improved by 0.61%, ultimately reaching 97.5%.

According to the test efficiency statistics of different models presented in Table IV, it is evident that the U-Net model exhibits the shortest processing time; however, its segmentation recognition performance is suboptimal. The implementation of the ResNet50 backbone network, coupled with the integration of the GAM attention mechanism, resulted in increased test time due to the augmented complexity of the network. Nonetheless, this modification significantly enhanced segmentation recognition accuracy. Consequently, the consideration of high-performance computers for model training will be pursued to further improve model efficiency.

F. Comparative Experiments with Different Models

To further validate the segmentation performance of the GR-UNet network model, the U-Net model, ResNet50-UNet model, PSPNet model, and Deeplabv3+ model were employed to assess the recognition capabilities on HDHV dataset. Following image preprocessing, a HDHV image with a consistent size of 288×288 pixels was obtained. The labeled image for training was generated through pixel-by-pixel labeling, with the labeling results presented in a binarized format, where the vein region is indicated in white.

The segmentation recognition results obtained from various network models are illustrated in Fig. 13. The U-Net model was ineffective in extracting weak edges of narrow vessels, leading to instances of local vein fracture. The edge positioning accuracy of the PSPNet model was suboptimal, and its smoothness was inadequate. Although the Deeplabv3+ model outperformed the PSPNet model, it still encountered issues with fracture separation at the connectivity. The ResNet50-UNet addressed the inaccurate localization of edge features seen in the traditional U-Net; however, it continued to exhibit local vein feature breakage. The GR-UNet model, which integrated the ResNet50 network and the GAM attention mechanism, effectively combined the residual structure's performance for feature extraction with the GAM attention mechanism's capability to capture global feature information. Additionally, it improved the loss function to mitigate the imbalance between positive and negative samples, further enhancing the model's segmentation efficacy. Consequently, when compared to other models, the GR-UNet demonstrated superior segmentation performance, excelled in extracting vein vessel boundaries and low-contrast regions, exhibited stronger feature recognition capabilities, and showed good resistance to interference in the segmentation of different HDHV. This further validated the accuracy of the improved network model presented in this paper for dorsal hand vein segmentation.

TABLE II. COMPARISON OF NETWORK MODELS WITH DIFFERENT BACKBONES

Mobile	VGG16	ResNet50	Decoder module	IOU	MIOU	MPA	Accuracy
√	—	—	√	63.31%	79.32%	83.31%	95.69%
—	√	—	√	75.45%	86.09%	92.92%	97.03%
—	—	√	√	77.66%	87.38%	93.35%	97.36%

TABLE III. THE ABLATION EXPERIMENT

Network	ResNet50	GAM	Loss	IOU	MIOU	MPA	Accuracy
Network a	—	—	—	73.54%	85.07%	89.78%	96.89%
Network b	√	—	—	77.66%	87.38%	93.35%	97.36%
Network c	√	√	—	78.81%	88.04%	93.58%	97.52%
Proposed method	√	√	√	78.82%	88.03%	93.92%	97.5%

TABLE IV. TEST EFFICIENCY OF DIFFERENT MODELS

Method	Test time/s	FPS
U-Net	0.0165	60.61
ResNet50-UNet	0.0336	29.76
GR-UNet	0.0579	17.27

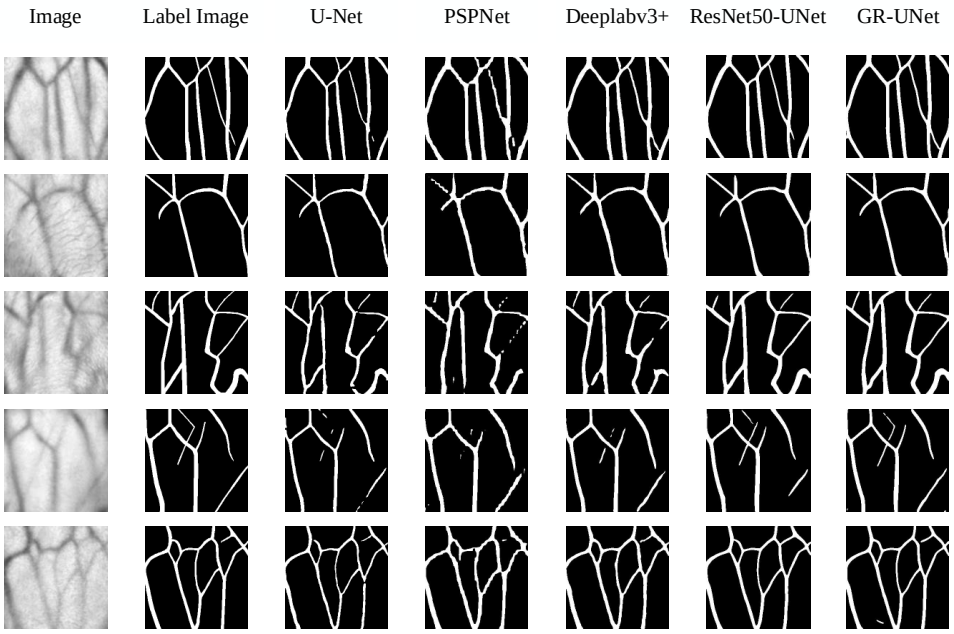


Fig. 13. The results of human back vein segmentation were compared by different models.

In order to verify the performance of the models proposed in this paper, the comparison results of the detection accuracy of each model are shown in Table V. The experimental results showed that the GR-UNet model proposed in this paper outperformed other models in terms of IOU, MIOU, MPA and Accuracy values.

TABLE V. COMPARISON OF DIFFERENT MODELS

Method	IOU	MIOU	MPA	Accuracy
PSPNet	70.29%	83.08%	91.51%	96.24%
Deeplabv3+	73.43%	84.95%	91.82%	96.78%
UNet	73.54%	85.07%	89.78%	96.89%
ResNet50-UNet	77.59%	87.3%	93.52%	97.3%
GR-UNet	78.82%	88.03%	93.92%	97.5%

VI. CONCLUSION

In this paper, a human dorsal hand veins dataset was established by designing a hand dorsal vein visualisation acquisition device based on near-infrared imaging technology. Based on the U-Net model architecture, a GR-UNet model combining attention mechanism and residual network was designed. The model used the residual structure of ResNet50 to enhance the feature extraction capability and effectively solved the vanishing gradient problem. The introduction of the GAM attention mechanism at skip connections optimized the processing of high-level feature information and effectively captured the global contextual features of the network. In addition, a weighted optimization loss function using cross-entropy loss and Dice loss to prevent the positive and negative sample imbalance problem further optimized the performance and stability of the network model.

Compared with the semantic segmentation models of PSPNet, Deeplabv3+, UNet and ResNet50-UNet, the proposed model achieved 78.82% in IOU, 88.03% in MIOU, 93.92% in MPA and 97.5% in Accuracy, respectively. It was superior to other semantic segmentation models, and significantly improved the segmentation accuracy of dorsal hand vein. However, the FPS of the model after completing the training was too low compared to the original model and did not reach the desired value, and a lightweight design of the model will be considered to improve the performance of the model in the subsequent research and there were still some limitations when dealing with more complex backgrounds and low-contrast vein regions, the subsequent work will expand the dataset to improve the generalisation ability and segmentation effect of the model in different age groups, genders and special populations.

ACKNOWLEDGMENT

This work was funded by the Natural Science Project of Zhengzhou Bureau of Science and technology (22ZZRDZX07) and the Scientific Research Special Project of National Clinical Research Base of Traditional Chinese Medicine of Henan Health Commission (2021JDZX2092).

REFERENCES

[1] Saeed A, Chaudhry M R, Khan M U A, Saeed M U, A.Ghfar A, Yasir M N, et al. Simplifying vein detection for intravenous procedures: A comparative assessment through near-infrared imaging system. International Journal of Imaging Systems and Technology, vol. 34, no. 3, p. e23068, 2024.

[2] Ma J, Ye B, Wang S. Vein Image Enhancement Based on CLAHE and Multi-scale Detail Fusion.SEMICONDUCTOR OPTOELECTRONICS, vol. 41, no. 05, pp. 738-742, 2020.

[3] Kuang H, Guan F, Ma X, Liu X. Adaptive threshold vein enhancement method based on illuminance estimation. IEEE 5th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), 2021, pp. 2591-2595.

[4] Besra B, Mohapatra R K. Extraction of segmented vein patterns using repeated line tracking algorithm. 2017 Third International Conference on Sensing, Signal Processing and Security (ICSSS), 2017, pp. 89-92.

[5] Yakno M, Mohamad-Saleh J, Ibrahim M Z. Dorsal hand vein image enhancement using fusion of CLAHE and fuzzy adaptive gamma. Sensors, vol. 21, no. 19, p. 6445, 2021.

[6] Yang G, Xu X. A FEATURE EXTRACTION METHOD FOR NIR VEIN IMAGE. Computer Applications and Software, vol. 40, no. 04, pp. 199-203+216, 2023.

[7] Zhang H, He L, Wang D. Deep reinforcement learning for real-world quadrupedal locomotion: A comprehensive review. Intelligence & Robotics, vol. 2, no. 3, p. 27597, 2022.

[8] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 3431-3440.

[9] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18. Springer International Publishing, 2015, pp. 234-241.

[10] He T, Guo C, Jiang L, Liu H. Automatic venous segmentation in venipuncture robot using deep learning. 2021 IEEE International Conference on Real-time Computing and Robotics (RCAR), 2021, pp. 614-619.

[11] Lefkovits S, Emerich S, Lefkovits L. Boosting Unsupervised Dorsal Hand Vein Segmentation with U-Net Variants. Mathematics, vol. 10, no. 15, p. 2620, 2022.

[12] Gao X, Zhang G, Zhou F, Yu D. Location Decision of Needle Entry Point Based on Improved Pruning Algorithm. Laser & Optoelectronics Progress, vol. 59, no. 24, pp. 154-166, 2022.

[13] Chen L, Lv M, Cai J, Guo Z, Li Z. U-Net-Embedded Gabor Kernel and Coaxial Correction Methods to Dorsal Hand Vein Image Projection System. Applied Sciences, vol. 13, no. 20, p. 11222, 2023.

[14] Zhao D, Tian Y, Chen H, Zhao Z, Chen Y, Yuan Y. Detection of Dorsal Hand Vein Based on Improved YOLO Nano and Embedded System. Chinese Journal of Biomedical Engineering, vol. 41, no. 06, pp. 691-698, 2022.

[15] Shu Z, Xie Z, Zhang C. Dorsal hand vein recognition based on transmission-type near infrared imaging and deep residual network with attention mechanism. Optical Review, vol. 29, no. 4, pp. 335-342, 2022.

[16] Abd Rahman A B, Juhim F, Chee F P, Bade A, Kadir F. Near infrared illumination optimization for vein detection: hardware and software approaches. Applied Sciences, vol. 12, no. 21, p. 11173, 2022.

[17] Ruan L, Yin Z, Zhou S, Zheng W, Lu W, Zhang T, et al. Vein visualization enhancement by dual-wavelength phase-locked denoising technology. Journal of Innovative Optical Health Sciences, vol. 17, no. 3, p. 2350033, 2024.

[18] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770-778.

[19] Gao H, Yuan H, Wang Z, Ji S. Pixel transposed convolutional networks. IEEE transactions on pattern analysis and machine intelligence, vol. 42, no. 5, pp. 1218-1227, 2019.

[20] Wu T, Ku T, Zhang H. Research for image caption based on global attention mechanism. Second target recognition and artificial intelligence summit forum, 2020, p. 11427.

[21] Woo S, Park J, Lee J Y, Kweon I S. Cbam: Convolutional block attention module. Proceedings of the European conference on computer vision (ECCV), 2018, pp. 3-19.