Bipedal Thermographic Segmentation Based on Diabetic Foot Layout Guide for Journal of Physics: Conference Series using Microsoft Word

Zhi Zeng, Zhenjie Cao, Junxia Zhu, Dan Liu
Zzh406@hotmail.com, 741266385@qq.com, 1733019360@qq.com, 983095898@qq.com

College of Computer and Information Science, Chongqing Normal University, Chongqing 401331, China

Abstract: In this paper we propose an algorithm for segmentation of bipedal thermograms. To obtain the complete segmented foot region, we used a neural network trained using data obtained from active thermography. The foot region was segmented using improved U^2-Net network. The results of experimental studies and simulations are given in this paper. The results are as follows. The proposed scheme can effectively segment the foot regions in different situations. The accuracy rate is 0.987, the miss detection rate is 0.006 and the detection speed is increased by 19%. The practical needs of diabetic foot detection can be better met by improving the U^2-Net network segmentation algorithm.

Keywords: bipedal thermograms, active thermography, U^2-Net network, diabetic foot detection

1 Introduction

Diabetes is a chronic endocrine disorder caused by insufficient insulin production. This deficiency raises the level of glucose concentration in the blood, potentially damaging blood vessels and nerves. In 2021, approximately 537 million adults worldwide will have diabetes, an increase of nearly 50 million people worldwide since 2019.

In recent years, research scholars have used infrared thermography[2] to study changes in plantar foot temperature in diabetic patients. In asymmetric temperature analysis, the plantar foot temperature of the left foot was compared with that of the right foot. If the same complications were present in both feet, the diseased area could not be identified. When performing temperature distribution analysis, identifying the spatial distribution of temperature remains difficult due to the wide variation in temperature distribution, especially in diabetic patients, which makes the classification process more difficult. In addition, details of bipedal thermograms may be missing and the interpretation of plantar foot fever may be difficult.

The aim of this study is to automatically segment diabetic foot (with or without neuropathy) using plantar temperature. First, the radiation data needs to be converted to temperature, the temperature is grayscale converted to image according to a fixed temperature window[3] , tested by a modified U^2 -Net network model based on[4] to generate a mask map (i.e., binary map), and feature extraction is performed on the dataset by the mask map, and finally only the

ICICA 2022, December 02-04, Chongqing, People's Republic of China

DOI 10.4108/eai.2-12-2022.2328003

foot region is retained.

2 Image acquisition and preprocessing

The infrared thermal image device is used to obtain the thermal image sequence of the surface of the tested foot, and the thermal image sequence is stored in the general memory, and the data in the memory is imaged. The data in the memory is the radiation value. First, the radiation value is converted into the temperature, and the temperature is converted into the image gray according to the fixed temperature window. Then, the pixel gray value is normalized to compress the pixel value to 0-1.

3 Bipedal thermogram segmentation

3.1 Overall structure of U^2 -Net

Qin et al. proposed a U^2-Net network consisting of a two-layer nested U-block structure, mainly consisting of residual U-blocks (RSU) that extract multi-scale features within a stage and an outer U-block structure that connects the RSUs. This network design eliminates the need to use a backbone network for image classification, and can be trained from scratch to obtain excellent results and make the network deeper to obtain high-resolution feature maps without increasing the computational cost as much as possible. For bipedal heat map foot features, the network is built using Pytorch to test the effect of bipedal heat map segmentation, and the overall structure of RSU and U2 -Net is described below.

The overall structure of the network includes a 6-stage encoder, a 5-stage decoder, and a feature map fusion output module, each stage is populated by configured RSU as shown in Figure 1. As can be seen from Fig. 7, the left side of the network is a downsampling process, the first four stages are populated by RSU with layer parameters L of 7, 6, 5, and 4, respectively, and the feature map size is halved and recovered layer by layer within the stages,and the last two stages are populated by RSU configured with hole convolution, and the feature map size remains constant within the stages. In the encoding phase, the RSU are connected to each other by a 2×2 maximum pooling, and the feature map size becomes 1/32 of the original size. while the right side of the network is an upsampling process, the RSU configuration within each phase is the same as the left symmetric position, and the input is a cascaded union of the upper phase output and the left symmetric phase output. In the decoding stage, the RSUsare connected to each other by a bilinear interpolation operation to gradually reduce the feature map. Finally, the feature maps output from each RSU in the decoding stage are stitched together to obtain the final feature maps.

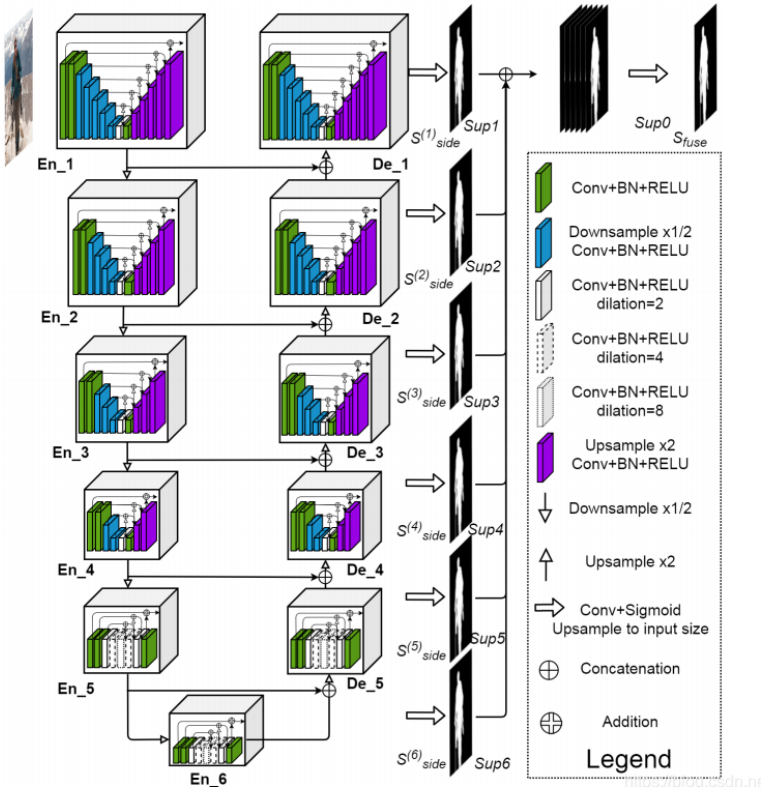


Figure 1 U^2 -Net overall structure

3.2 Loss function

In order to accelerate the convergence speed of neural networks and to solve the gradient disappearance problem that occurs during the training of deep networks, a deeply supervised [5] approach is used, i.e., an auxiliary classifier is added to the hidden layer of the network to supervise the backbone network. Therefore, the loss function of the semantic segmentation network is defined as

$$L = \sum_{m=1}^{M} w_{side}^{(m)} \ell_{side}^{(m)} + w_{fuse} \ell_{fuse} \tag{1}$$

Eq. (1) where: m is the mth lateral branch; $\ell_{side}^{(m)}$ is the loss of the lateral output feature map; ℓ_{fuse} is the loss of the final fused output feature map; and w_{fuse} and $w_{side}^{(m)}$ are the weights of each loss term. For each term, we use the standard binary cross-entropy to calculate the loss.

$$\ell = -\sum_{(r,c)}^{(H,W)} [P_{G(r,c)} \log P_{S(r,c)} + (1 - P_{G(r,c)}) \log(1 - P_{G(r,c)})] \tag{2}$$

where (r,c) (r,c) (r,c) (r,c) are the pixel coordinates and (H,W) are the image sizes: height and width. $P_{G(r,c)}$ and $P_{S(r,c)}$ denote the GT pixel values and the predicted significant probability maps, respectively. The training process tries to minimize the overall loss. During testing, we choose the final fusion result as the final saliency map of ℓ_{fuse} .

To better train the network, the sample image size was scaled down to 320×320 pixels and randomly cropped to 288×288 pixels before the training started. The semantic segmentation network uses Adam optimizer, in which the parameters take default values, and the Xavier method is used to initialize the convolution layer parameters, and the initial loss weights w_{fuse} and $w_{side}^{(m)}$ are set to 1. Since the foot region accounts for a relatively large portion of the whole image, some adjustments are made in the output feature map part of the network, and the side output feature maps are upsampled to the input image size and then stitched together in the channel dimension, and finally the final output feature maps are obtained by sigmoid activation function and 3×3 convolution to get the final output feature map. To start the training, the training dataset is fed into the neural network, and after 1,000 to 1,500 rounds of training (batch of 5), the training loss converges. At this point, the training optimal model file is saved and the model is loaded for prediction, and the obtained network evaluation metrics are shown in Table 1.

As can be seen from Table 1, the test set results show that although the model detects most of the foot region, there are still serious false detections and missed detections as well as the inability to segment the foot region completely. u^2 -Net neural network solves the problem that traditional neural networks cannot take into account local details and global contrast information, but the network is complex, has large depth, and tends to lose information of high importance but low frequency of occurrence, which is especially obvious when facing small data sets and small target detection problems are especially obvious. An analysis of the experimental results shows that the model does not pay enough attention to the foot region and may have overfitting. Therefore, direct use of the U^2 -Net network for foot region detection cannot achieve the expected results, and in other aspects, the U^2 -Net network model is large, has many parameters, and takes a long time for training and prediction, which does not meet the requirements of real-time detection well.

Table 1 U^2 -Net training results evaluation metrics

	HausdorFfdistance	DSC	IOU
training set	0.933	0.93	0.923
test set	0.924	0.896	0.91

4 Improvements

Through the analysis in Section 2.1, the U^2 -Net bipedal heat map detection scheme is improved by incorporating the residual circular convolution module[6] and the Attention

mechanism[7] in the decoding phase to improve the detection accuracy while reducing the network training and detection time.

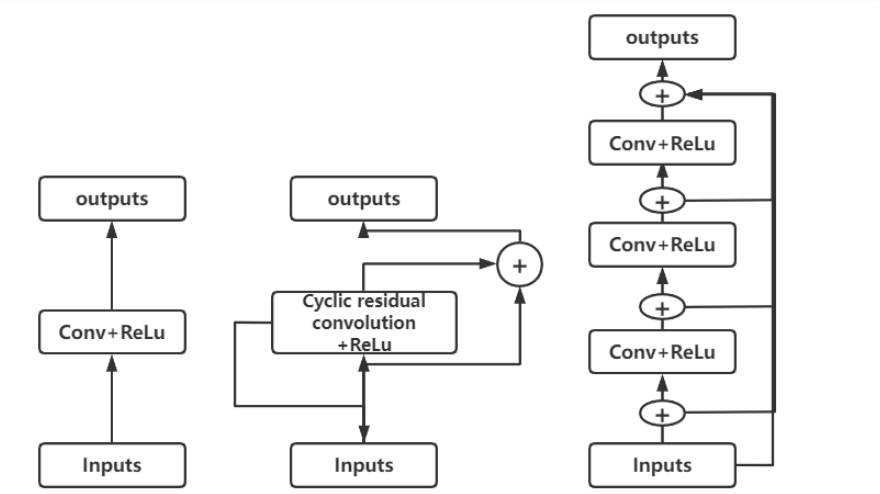


Figure 2 Cyclic residual convolution module

4.1 Cyclic residual convolution module

The classical U^2 -Net network architecture is widely used for segmentation of medical images, but due to the high complexity of the processed bipedal thermograms, the noise contained blends with the biped, and the different temperatures in different regions of the foot, sometimes it is not possible to segment the plantar region completely. In order to achieve higher accuracy with a model of low complexity and a balance between accuracy and complexity as a way to extract to more complex features. We take the classical U^2 -Net network architecture as the basis, and introduce the cyclic residual convolutional network, which contains several cyclic residual convolutional modules, and replace the ordinary convolutional modules in the original network with cyclic residual convolutional modules, so as to deepen the U^2 -Net network depth and extract more complex features.

In the classical UNET network architecture, its feature coding module is composed of 5 convolutions, each convolution block includes 2 convolution layers and 1 maximum pooling layer. The structure of the UNET convolution block is shown in Fig. 3. The method in this paper uses the cyclic convolution module structure to replace the convolution block in the UNET feature module. The feature coding module in this paper is composed of 5 convolution blocks, Each convolution block contains one convolution residual module (RRCB) and one MPL. Each cyclic residual module unit contains two cyclic convolution layers (RCLS). A single RCL contains three cyclic sub sequences, and its structure is shown in Fig. 3.

The structure of the ordinary convolution layer, the cyclic residual convolution module, and the cyclic subsequence is shown in Fig. 2, where T denotes the number of cycles, which is calculated as

$$\begin{cases} {}^{k}_{(l)}O_{ij}(t) = (w_f^k)^T x^l_{f(i,j)}(t) + (w_f^k)^T x^l_{r(i,j)}(t-1) + b_k \\ F(x_i) = f({}^{k}_{(l)}O_{ij}(t)) = \max(0, {}^{k}_{(l)}O_{ij}(t)) \\ x_{l+1} = x_l + F(x_l) \end{cases} \tag{3}$$

Where: *k is* the sequence of feature maps in the cyclic convolution layer,*l is the sequence* number of the cyclic convolution layer in the cyclic residual convolution module,*x is the* input feature map,w_r^k is the weight of the previous cyclic output in the *kth feature map*,w_f^k is the weight of *x* in the *kth* feature map,f is the activation function,*F* is the output feature of the cyclic convolution layer (RCL),*O* is the feature map of the cyclic subsequence output,*T* is the number of cycles,which determines the number of cycles per cyclic residual The convolution module contains several convolution modules,b_k is the bias compensation.

Compared to the convolution module in the classical U^2 -Net network architecture, the cyclic residual convolution module retains both the fewer parameters of the original network structure and allows the network to extract to more complex features.

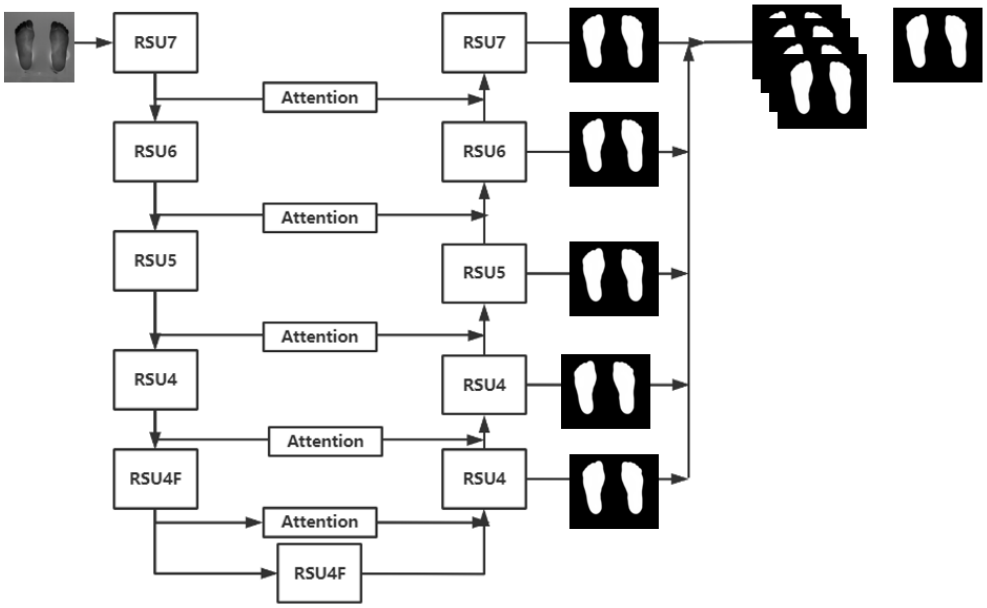


Figure 3 Structure diagram of the improved algorithm

4.2 Improving the network

The proportion of positive samples in the bipedal heat map image is too small, and to solve the problem of unbalanced positive and negative samples and improve the effect of small target detection, the Attention mechanism from the literature is introduced into the U^2 -Net network. In the decoding stage, the Attention mechanism is added to each RSU block for stitching upsampling to enhance the model's ability to process the features of the region of interest and

improve the segmentation effect. The number of image channels is reduced from 3 channels in the RSU7 input, through 32 channels in the intermediate layer to 64 channels in the output layer, and then through 128, 256, and 512 channels for upsampling in order to avoid the phenomenon of overfitting the data set. Adding the Attention mechanism yields the improved network, as shown in Figure 3.

5 Network training results

5.1 Experimental process

The experimental environment is Windows, the computer processor is Intel Core i7-9500, and the GPU is GTX1650. After repeated comparison and tuning experiments, we finally obtained the results of the improved network after training, comparing the training results of U^2 -Net and the results obtained by using U-Net network trained on the same dataset, the experimental results are shown in Table 2.

The experiments prove that the optimization strategy can indeed effectively improve the segmentation accuracy, outperforming the semantic segmentation network U^2 -Net and the traditional classical segmentation network U-Net, which proves the correctness of the improvement idea. To test the detection effect of the model on real scenes, five networks are used to detect the collected bipedal heat maps, and the results of network evaluation metrics comparison are shown in Table 2.As can be seen from Table 2, the improved network outperforms the other two networks in the original image detection of transparent parts, which again validates the correctness of Attention U^2 -Net improved idea in bipedal heat map detection.

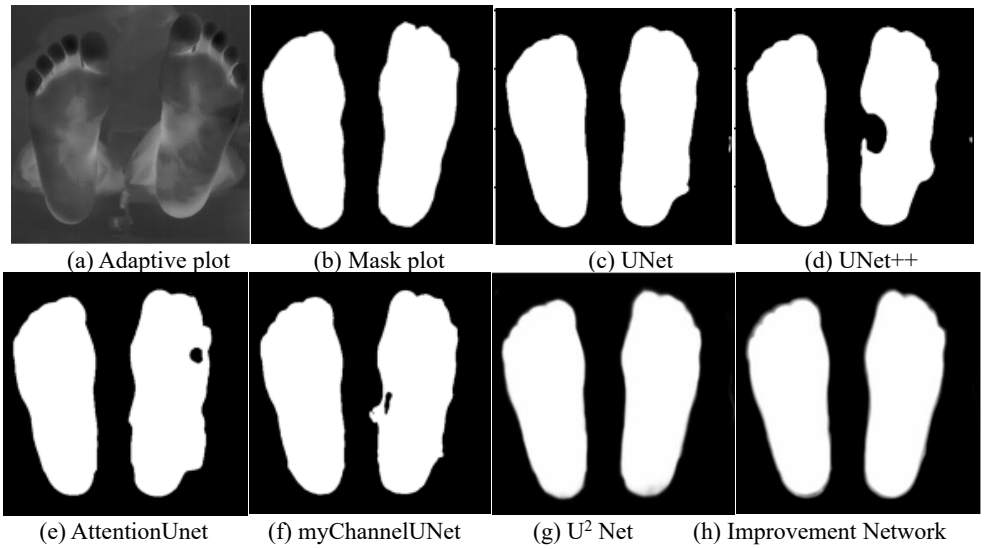
Table2 Comparison of the results of the improved network, U^2 -Net and U-Net series test metrics

	Hausdorffdistance	DSC	IOU
Improving the network	0.954	0.942	0.961
U^2 -Net	0.933	0.931	0.923
U-Net	0.914	0.853	0.896
MychannelU-Net	0.934	0.936	0.922
AttentionU-Net	0.914	0.923	0.916
U-Net++	0.924	0.921	0.914

To further evaluate the improved solution, the model size and detection time of the above-mentioned mult-iple networks were compared. the model size of U-Net network [9], MychannelU-Net [10], AttentionU-Net [11], U-Net++ [12] and U2 -Net networks were 51.2 MB and 181 MB, respectively, and the average time to detect an image was 0.45 s and the model size of the improved network is 172.7 MB, and the average time to detect an image is 0.76 s. Thus, it can be seen that the improved network has various performance improvements and can meet the practical needs of foot region detection.

5.2 Experimental results of bipedal thermogram detection

The bipedal thermograms were tested under different models and the results are shown in Figure 10.



(a) Adaptive plot (b) Mask plot (c) UNet (d) UNet++

(e) AttentionUnet (f) myChannelUNet (g) U^2 Net (h) Improvement Network

Figure 4 Graph of results

From the segmentation results of the bipedal heat map in Figure 10, it can be seen that there are a small number of false detections in the U^2 -Net, AttentionUnet, myChannelUNet, UNet++ and U-Net prediction maps, and some point defects and foot edge information are extracted, while the improved network performs better compared to both and the error detection has been eliminated.

6 Conclusion

A bipedal heat map segmentation method based on an improved U^2 -Net deep learning model is proposed for the automated bipedal heat map segmentation problem, and the following conclusions are obtained.

(1) Applying U^2 -Net neural network for bipedal thermogram segmentation, it has some segmentation ability for foot region in bipedal thermogram, but the foot segmentation results have defects, which are difficult to meet the requirements of practical segmentation.

(2) From the experimental results, it is shown that the proposed bipedal thermogram segmentation method can effectively improve the accuracy of foot edge segmentation in bipedal thermogram segmentation problem. This method can segment the images of various complex defect cases accurately, and the detection accuracy reaches 0.961, the detection effect is better than the direct application of U^2 -Net with U-Net series, and the detection speed is improved, and the final results verify the correctness and feasibility of the improved scheme.

References

[1] Gong Hongping, Raju Bista, Cha Panpan, Ran Yanhao, Ren Yan, Gao Eddie, Huang Hui, Ran Xingwu, Wang Chun. Clinical characteristics and prognosis analysis of hospitalized diabetic patients with high-risk foot[J]. West China Medicine,2022,37(03):408-413.

[2] Ramirez GarciaLuna Jose L., Bartlett Robert, Arriaga Caballero Jes us E., Fraser Robert D. J., Saiko Gennadi. Infrared Thermography in Wound Care, Surgery, and Sports Medicine: A Review [J]. Fro ntiers in Physiology,2022,13.

[3] Kastberger G, Stachl R. Infrared imaging technology and biological applications [J]. Behavior Research Methods, Instruments, & Computers, 2003, 35(3): 429-439.

[4] Xuebin Qin; Zichen Zhang; Chenyang Huang; Masood Dehghan; Osmar R. Zaiane; Martin Jagersand.U^2 -Net: Going deeper with nested U-structure for salient object detection[J]. Pattern Recognition,2020,106:107404.

[5] Tan X, Diao Yichao, Chen Xinjian, Shi Fei, Fan Ying, Xie Jiamin, Zhu Weifang. Deeply supervised feature aggregation network for high myopic streak damage[J]. Chinese Journal of Graphics, 2022,27(03):961-972.

[6] JIN Yan,XUE Zhi-Zhong, JIANG Zhi-Wei. A medical image segmentation algorithm based on recurrent residual convolutional neural network [J/O L]. Journal of Computer-Aided Design and Graphics:1-11 [2022-08-05].

[7] Gao Zhengxia. Attention mechanism based image application[J]. Science and Technology Innovation and Application,2021,11(21):167-169.

[8] Zhanyu Ma, Dongliang Chang, Xiaoxu Li. Channel Max Pooling Layer for Fi ne-Grained Vehicle Classification. [J]. CoRR, 2019:469-425

[9] Zhou ZhengSong, Chen XuMiao, Zhang HaoYu, Wan HongLi, Zhao JieYi, Zhang Tao, Wang XiaoYu. Application of Improved Unet Network in the Recognition and Segmentation of Hemorrhage Regions in Brain CT I mage [s]. [J]. Sichuandaxue. Yi xue ban Journal of Sichuan University. Medical science edition,2022,53(1).

[10] Zhang Yu, Xiao Zhiyong, Ding Yan, Zhu Xiaowei, Qin An, Chang Yanhua, Wu Pengxi, Zhou Fengsheng. The value of improved U-Net neural network model in automatic segmentation of thyroid nodules[J]. China Medical Equipment,2022,19(05):30-33.

[11] Shu, Xingzhe. Research on Attention U-Net virtual fitting method based on parallel convolutional kernel [J]. Software Engineering, 2022,25(06):1317.

[12] Wu S.-Y., Shen N., Meng Y.-H., Wang Z.-M. Human spine MRI image recognition based on U-Net++[J]. Journalof Guilin University of Electronic Science and Technology,2022,42(01): 36-42. DOI:10.16725/j.cnk i.cn45-1351/tn.202 2.01