








Interactive Skin Wound Segmentation Based on Feature Augment Networks

Pengfei Zhang , Xinjian Chen , Ziting Yin , Xin Zhou , Qingxin Jiang, Weifang Zhu , Dehui Xiang , Yun Tang, and Fei Shi 

Abstract—Skin wound segmentation in photographs allows non-invasive analysis of wounds that supports dermatological diagnosis and treatment. In this paper, we propose a novel feature augment network (FANet) to achieve automatic segmentation of skin wounds, and design an interactive feature augment network (IFANet) to provide interactive adjustment on the automatic segmentation results. The FANet contains the edge feature augment (EFA) module and the spatial relationship feature augment (SFA) module, which can make full use of the notable edge information and the spatial relationship information between the wound and the skin. The IFANet, with FANet as the backbone, takes the user interactions and the initial result as inputs, and outputs the refined segmentation result. The proposed networks were tested on a dataset composed of miscellaneous skin wound images, and a public foot ulcer segmentation challenge dataset. The results indicate that the FANet gives good segmentation results while the IFANet can effectively improve them based on simple marking. Comprehensive comparative experiments show that our proposed networks outperform some other existing automatic or interactive segmentation methods, respectively.

Index Terms—Deep learning, feature augment network, interactive image segmentation, wound image.

I. INTRODUCTION

SKIN is the largest organ of the human body which contains a complex network of nerves and blood vessels, accounting

Manuscript received 4 July 2022; revised 21 March 2023; accepted 21 April 2023. Date of publication 26 April 2023; date of current version 3 July 2023. This work was supported in part by the National Natural Science Foundation of China under Grant 62271337 and in part by the National Key Research and Development Program of China under Grant 2018YFA0701700. (Pengfei Zhang and Xinjian Chen contributed equally to this work.) (Corresponding author: Fei Shi.)

Pengfei Zhang, Ziting Yin, Xin Zhou, Qingxin Jiang, Weifang Zhu, and Dehui Xiang are with the MIPAV Lab, the School of Electronic and Information Engineering, Soochow University, Suzhou 215006, China (e-mail: jasonzpf@foxmail.com; 1209347544@qq.com; 1658804632@qq.com; 498570343@qq.com; wfzhu@suda.edu.cn; xiangdehui@suda.edu.cn).

Fei Shi is with the MIPAV Lab, the School of Electronic and Information Engineering, Soochow University, Suzhou 215006, China, and also with the Fariver Innovation Technologies Company, Ltd., Suzhou 215151, China (e-mail: shifei@suda.edu.cn).

Xinjian Chen is with the MIPAV Lab, the School of Electronic and Information Engineering and the State Key Laboratory of Radiation Medicine and Protection, Soochow University, Suzhou 215006, China (e-mail: xjchen@suda.edu.cn).

Yun Tang is with the Suzhou Science and Technology Town Hospital, Suzhou 215153, China (e-mail: tangyun0212@163.com).

Digital Object Identifier 10.1109/JBHI.2023.3270711

for about 16% of the entire body weight [1]. As the first line of defense of the human body, it is vulnerable to physical and mechanical injury. According to the wound healing cycle, skin wounds can be divided into acute wounds and chronic wounds [2]. Acute wounds usually refer to traumas such as contusions, lacerations and cuts, including knife wounds, abrasions, gunshot wounds, and chemical injuries [3]. Chronic wounds mainly include venous ulcers of the lower extremities, pressure ulcers, arterial ischemic ulcers and diabetic ulcers [4]. Acute and chronic skin wounds have caused a heavy economic burden on the entire healthcare systems [8]. By 2024, the skin wound care market is expected to exceed \$22B [9]. Especially, chronic wounds have more influencing factors and disordered healing procedures, and therefore require long-term and meticulous care and treatment [10].

In the process of treatment, medical staff need to check and evaluate the wounds regularly so as to make proper treatment plans or adjustments. Currently, medical staff mainly evaluate wounds through spatial measurement and visual inspection. The status evaluation by visual inspection is based on texture, color, tissue composition, and severity stage [11]. Quantitative indicators such as the area, depth, maximum and minimum diameter of the wound [12] are also important basis for evaluation. However, manual measurement, such as by ruler or by thin film, is subjective, time-consuming, and will directly or indirectly contaminate the wound [13]. The way to solve these problems is a non-contact system using imaging and image processing techniques [14].

In this paper, we propose a deep learning method for automatic skin wound segmentation from images taken by common cameras or smart phones. The proposed feature augment network (FANet) is a U-shaped network embedded with two feature augment modules, for effective utilization of wound edge information and spatial relationship between the wound and the surrounding skin, respectively.

Although the proposed FANet can achieve decent segmentation results on average, there can still be inaccuracies near the edge or missing of small wound areas. In practice, there may be the need for further refinement of the segmentation under the guidance of medical staff. Simple interaction is much preferred in such cases considering their everyday working load. To this end, we further build upon the FANet and propose an interactive feature augment network (IFANet). The IFANet takes inputs including simple user interaction as scribbles, the initial segmentation result and the original image, and outputs the refined segmentation result.

Our main contributions are summarized as follows:

- 1) We propose a two-stage deep learning based skin wound segmentation framework which is convenient for clinical use. The first stage is automatic while the second one is for interactive refinement, both based on the same network structure.
- 2) We propose the edge feature augment (EFA) module to augment the edge features, and use a cascade of them to form an edge extraction branch in the network, making it pay more attention to the edges.
- 3) We propose the spatial relationship feature augment (SFA) module and embed it between the encoder and the decoder, to expand the field of view and improve spatial feature extraction.
- 4) We propose an interactive segmentation network, where the initial segmentation and the interactions are both transformed into distance maps, providing important guiding information in achieving improved segmentation results.

II. RELATED WORKS

A. Automatic Wound Segmentation

In recent years, computer vision technologies have been applied to automatic wound analysis. Fauzi et al. [18] proposed to generate a four-dimensional color probability map in a modified hue-saturation-value (HSV) space, on which thresholding or region growing were applied for tissue segmentation and label card detection. Dhane et al. [19] proposed a method based on fuzzy spectrum clustering to segment chronic wounds. Wang et al. [20] used active contour method to segment diabetic foot ulcer wounds. Wannous et al. [20] designed a support vector machine (SVM) classifier based on color and texture features following the J-value segmentation (J-SEG) algorithm. Hani et al. [22] proposed to apply independent component analysis (ICA) to preprocessed RGB images to generate hemoglobin-based images, which are used as inputs to K-means clustering to segment granulation tissue in the wound. The above are all traditional computer vision algorithms with limited segmentation accuracy and have only been verified on small datasets.

More recently, deep learning methods have been proposed to solve the wound segmentation problem. Lu et al. [23] proposed a model for wound area segmentation using an encoder-decoder based deep network following intensity and color correction. Jiao et al. [24] designed a deep learning segmentation framework based on Mask R-CNN adopting ResNet101 with atrous convolution as the backbone. Wang et al. [9] proposed a lightweight skin wound automatic segmentation network based on MobileNetV2, where an atrous spatial pyramid pooling block was added to the decoder to capture multi-scale feature information. In addition, Sarp et al. [25] proposed a hybrid wound boundary segmentation and tissue classification method using conditional generative adversarial network (cGAN). However, most of the above methods used existing network structures. Though some applied minor changes that focused on multi-scale features, the characteristics of wound images were not addressed. Some methods were only trained and tested on certain type of wounds,

and the generalization ability was unclear. In contrast, our model includes substantial modification to the existing network. It is designed by taking into consideration of the characteristics of wound images. Furthermore, it is cross-validated on a dataset with miscellaneous skin wounds, including images obtained from both public dataset and a local hospital, and also tested on a public foot ulcer segmentation challenge dataset.

B. Interactive Segmentation With Deep Learning

Although deep learning has achieved excellent performance in many image segmentation tasks, human interaction is still needed in some scenarios, to deal with difficult cases and to ensure higher accuracy. The type of interaction includes points, bounding boxes or scribbles. Xu et al.'s work [26] was among the earliest efforts in deep learning based interactive segmentation. They proposed to use a fully convolutional network (FCN) with interactive input as Euclidean distance maps transformed from foreground and background seeds for natural image segmentation. Liew et al. [27] improved the above method by adding a local regional refinement branch, generating local predictions based on pair of user clicks. Maninis et al. proposed deep extreme cut (DEXTR) [28], which used Gaussian heatmaps generated by four manually annotated extreme points to guide semantic segmentation with a Res-Net101-based network. A fifth point on the boundary could be added to further refine the segmentation. Xu et al. [29] also proposed Deep GrabCut which used Euclidean distance maps transformed from bounding boxes for interactive semantic segmentation. Most recent works emphasized reducing the number of interactions and improving the efficiency. Jang et al. [30] proposed BRS which performed backpropagation iteratively until all clicked pixels have correct labels. Sofiuk et al. further proposed f-BRS [31] which only refined the intermediate features to get more precise masks with faster speed compared with BRS. Chen et al. proposed CDNet [32] which used diffusion modules to propagate user clicks to conditioned destinations. Chen et al. also proposed FocalClick [33] which achieved high efficiency by predicting and updating the mask in localized areas according to user interaction. However, most of these works focused on interaction as guidance in the beginning of segmentation, and not all of them provided means in segmentation refinement. What's more, these methods were proposed for natural image segmentation and usually required large amount of training data.

For medical image segmentation, Wang et al. [34] proposed a two-stage interactive deep learning method with P-Net. A user-provided bounding box was first needed for preliminary segmentation, and then with scribble-based interaction, image-specific fine-tuning was performed to get a refined result. In [35], Wang et al. proposed another interactive method for medical image segmentation with P-Net combined with conditional random field, where scribble-based interactions were transformed into geodesic distance maps to guide the refinement of automatic segmentation results. These methods achieved good results for MRI segmentation. In this paper, following the idea of [35], we first propose a network designed for automatic wound segmentation and then reform it for interactive refinement with scribble inputs.

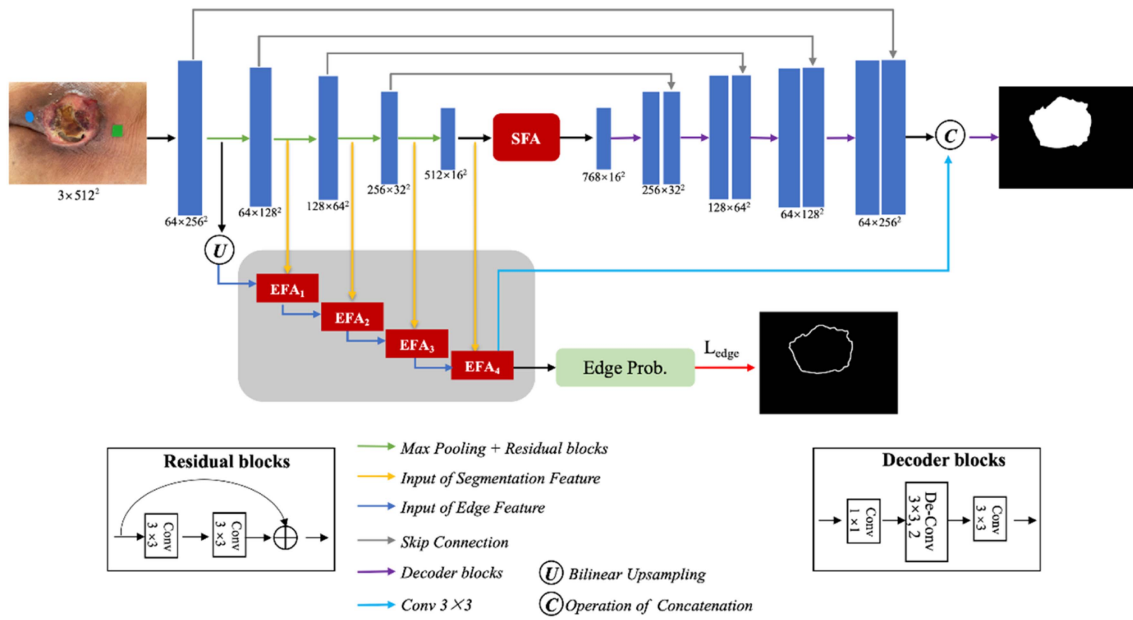


Fig. 1. Structure of the proposed FANet.

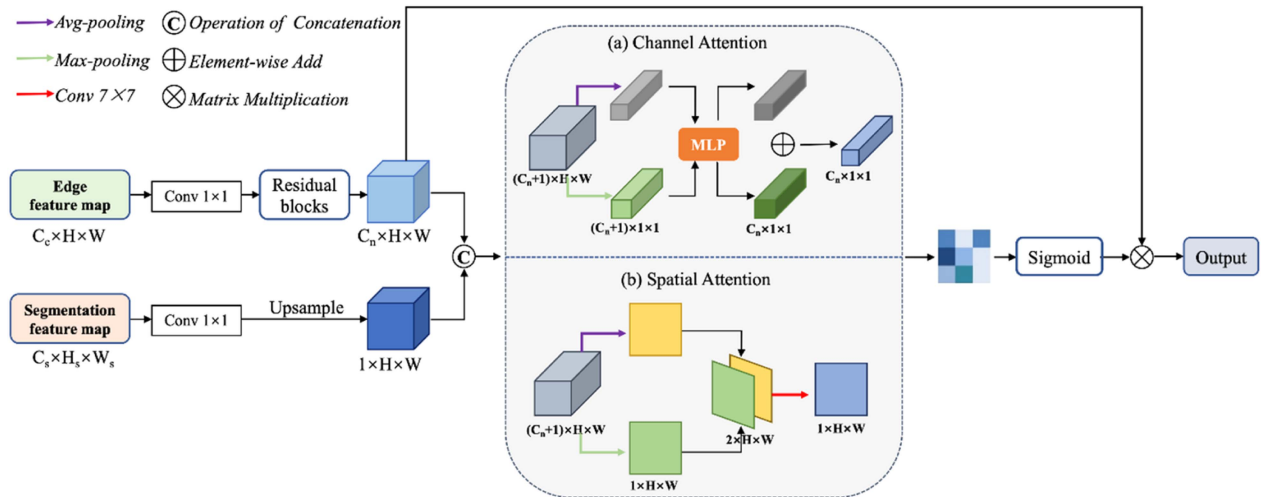


Fig. 2. Components of the edge feature augment (EFA) module.

III. METHODS

As shown in Fig. 1, the proposed FANet is a fully convolutional network based on the encoder-decoder architecture. The encoder is a ResNet34 [36] structure composed of residual blocks. The decoder is made of deconvolution blocks. The encoder and the decoder are connected by skip connections. In this section, the edge feature augment (EFA) module and the spatial-relationship feature augment (SFA) module embedded in FANet are specifically introduced. Then, the details of IFANet are given.

A. Edge Feature Augment Module

Associated with the wound healing process, in the skin wound images, the edge of wound often exhibits different features

from both the surrounding normal skin and the center of the wound. Therefore, learning the edge features is particularly important for wound segmentation. We design the EFA module to obtain significant edge features from the semantic features of the encoder. As shown in Fig. 1, the EFA modules are cascaded to form an edge extraction branch. Multiscale features from the encoder are sequentially infused into the branch to augment the edge features. The edge feature maps output from this branch are concatenated with features from the last network layer before final decoding, so that the edge can be enhanced in the final segmentation map.

As shown in Fig. 2, the EFA module uses spatial or channel attention mechanisms to obtain augmented edge feature maps. The inputs of EFA are the edge feature maps from the previous level and the segmentation feature maps from the encoder. The

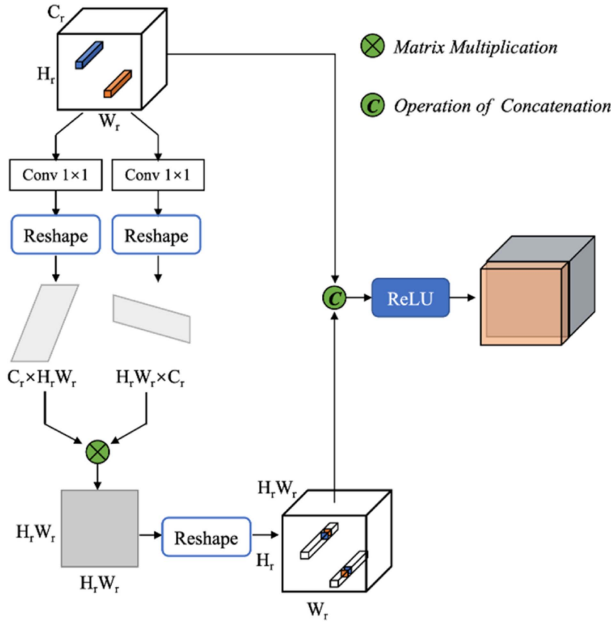


Fig. 3. Components of the spatial relationship feature augment (SFA) module.

two streams of feature maps are concatenated, and spatial or channel attention is applied to obtain the weights for recalibration. The processed edge feature maps are multiplied by the weights to obtain the augmented edge feature maps. In the EFA modules of the first two layers, the spatial attention mechanism is used, while for the last two layers, the channel attention mechanism is adopted. This distinction is made considering the fact that spatial information is more abundant in lower levels and semantic information is more important in higher levels, which may spread over different channels. Note that all EFA modules operate on the original resolution to retain as many edge details as possible.

To ensure that the edge information is emphasized in this branch, an edge probability map is also output in the end, and an edge loss is calculated for supervision.

B. Spatial Relationship Feature Augment Module

The wound to be segmented is always surrounded by the normal skin. In addition, when multiple wounds exist in one image, there is long-range spatial relationship between wounds. Therefore, the spatial relationship between objects and background are important features. We propose the SFA module to improve the model's learning ability in this aspect, and embed it between the encoder and the decoder.

In order to capture the global spatial relationship features, the SFA module adopts the idea of the non-local network [37], as shown in Fig. 3. The high-level feature maps of $C_r \times H_r \times W_r$ are fed into two 1×1 convolutions and reshaped to matrices of $H_r W_r \times C_r$ and $C_r \times H_r W_r$ respectively. The two are multiplied to obtain the $H_r W_r \times H_r W_r$ matrix, where each element represents the spatial relationship between two elements in the original input. It is reshaped and concatenated with the original

input, passed through the ReLU function to eliminate negative spatial relationships, to form the spatial relationship augmented features.

C. Interactive Feature Augment Network

As shown in Fig. 4, IFANet is constructed on the basis of FANet. The user interactions are scribbles in the false positive or false negative regions of the initial automatic segmentation result. The initial segmentation and the two types of scribbles are all transformed into Gaussian distance maps, acting as auxiliary information in the new segmentation process.

The Gaussian distance maps are calculated as:

$$G(x, y) = A \exp \left[-\frac{D^2(x, y)}{2\sigma^2} \right] \quad (1)$$

where

$$A = \frac{1}{\sum_{(x,y)} \exp \left[-\frac{D^2(x,y)}{2\sigma^2} \right]} \quad (2)$$

When calculating Gaussian maps G_f for false negative and G_b for false positive, $D(x, y)$ is the shortest Euclidean distance from (x, y) to all marked points, and the value of σ is empirically set as 10. When calculating the Gaussian map G_c for the initial segmentation, $D(x, y)$ is the shortest Euclidean distance from (x, y) to the center point of the initial segmented region, and the value of σ is empirically determined as 30.

In order to emphasize the interactive information, we concatenate G_f , G_b and the RGB image to get a $5 \times H \times W$ network input. As the initial segmentation gives important information of the wound location, the corresponding G_c is passed through three 3×3 convolutional layers, and the resulting feature maps are concatenated with the output of the encoder and then fed into the SFA module. This information helps constrain the updated segmentation results not to deviate too much from the initial result, and facilitates training of the model.

With the new defined inputs, the whole network is retrained to give the updated segmentation.

D. Loss Functions

For both FANet and IFANet, a joint loss L_{total} consisting of binary cross-entropy loss L_{BCE} , Dice loss L_{Dice} and weighted cross-entropy edge loss L_{edge} is used to train the network, detailed as follows:

$$L_{total} = L_{BCE} + L_{Dice} + L_{edge} \quad (3)$$

$$L_{BCE} = -\frac{1}{n} \sum_i^n t_i \log(o_i) + (1 - t_i) \log(1 - o_i) \quad (4)$$

$$L_{Dice} = 1 - \frac{1}{n} \sum_i^n \frac{2t_i o_i}{t_i^2 + o_i^2} \quad (5)$$

$$L_{edge} = -\frac{1}{n} \sum_i^n \alpha \text{Canny}(t_i) \log(o'_i) + (1 - \alpha) [1 - \text{Canny}(t_i)] \log(1 - o'_i) \quad (6)$$

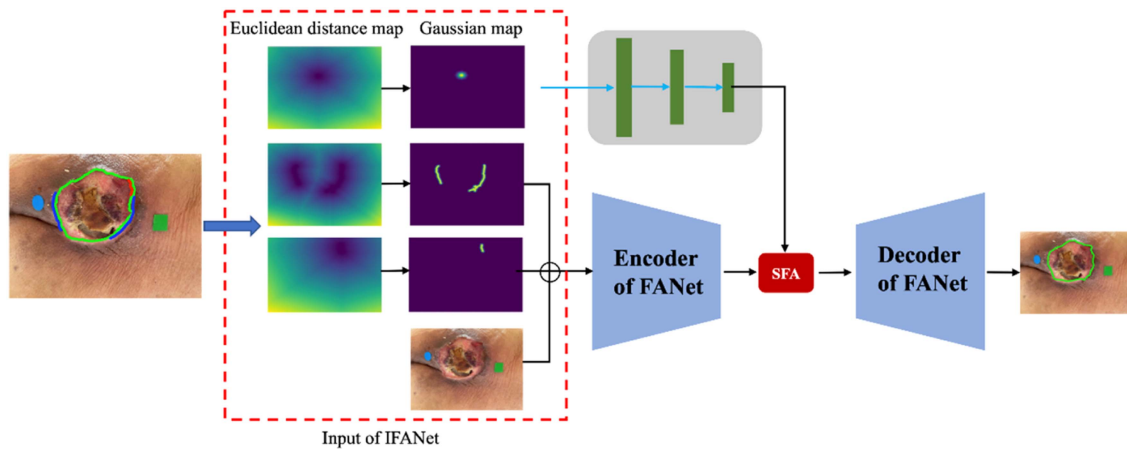


Fig. 4. An overview of the proposed IFANet. In the input, red scribbles refer to false positive and blue scribbles refer to false negative.

where t_i represents the pixel value in the ground truth, o_i represents the pixel value of the prediction map of the network, o'_i represents the pixel value of the edge prediction map output by the edge extraction branch, n represents the number of all pixels in the image, i represents the i th pixel, and α represents the weighting coefficient to deal with the unbalanced number of edge and non-edge pixels. Here α is empirically set as 0.75.

IV. EXPERIMENTAL SETTINGS

A. Datasets

Two datasets were used for training and testing of the proposed networks. The first dataset (MISC) contains 552 skin wound images which includes miscellaneous wound types such as burns, diabetic foot ulcers, venous or arterial varicose ulcers, orthopedic wounds, pressure ulcers, etc. Among them, 118 images were provided by Suzhou Science and Technology Town Hospital and Fariver Innovation Technologies Co., Ltd., Suzhou. The collection and analysis of image data were approved by the Institutional Review Board of Suzhou Science and Technology Town Hospital and the Institutional Review Board of Soochow University, and adhered to the tenets of the Declaration of Helsinki. The remaining 434 images were from the Medetec Wound Database [38]. The images were taken with a front view of the wound, and most of the wounds were in the center of the image. The image size ranged from 560×384 to 3968×2976 , all resized to 512×512 for network input. Ground truth for segmentation was obtained by manual delineation (P.Z.) trained by a wound care specialist (Y.T.) and further corrected by this specialist, using the Adobe Photoshop software. Five-fold cross-validation were performed for this dataset.

The second dataset (FUSeg) was from the MICCAI 2021 Foot Ulcer Segmentation Challenge [9], [39]. 1010 images with segmentation masks were available. The images were obtained through collaboration between the University of Wisconsin-Milwaukee and Advancing the Zenith of Healthcare Wound and Vascular Center. The IRB review and approval were waived by the Institutional Review Board of University of Wisconsin-Milwaukee. The images were taken as close-ups with the camera

facing the wound surface. All images were zero-padded to 512×512 . Segmentation masks were manually obtained and further reviewed and verified by wound care specialists and nurses. In our experiments, 610 images were used for training, 200 for validation and the remaining 200 for testing.

B. IFANet With Simulated Interactions

For training of the IFANet, interaction for each of the image is needed, which is laborious if done manually. Instead, we adopted simulated interaction with automatic generated scribbles while keeping some of the random nature of human interactions. First, the automatic segmentation results were compared with the ground truth to obtain false positive and false negative regions. Secondly, the skeletons of each connected components were extracted. Regions whose skeleton contains less than 40 pixels were ignored to simulate that human often overlook small segmentation errors. Lastly, a random number of skeletons were selected as interactive scribbles for IFANet, imitating human behaviors.

For ablation test and comparative test of IFANet, simulated interactions were also used. As some randomness were involved in generating the scribbles, the average indices of five tests were recorded. Tests with real human interactions were also performed on a subset of the data.

C. Implementation Details

The encoder of the proposed networks was based on the ResNet34 pretrained on ImageNet. The implementation is based on the public Pytorch platform and NVIDIA GeForce RTX 2080Ti GPU with 11GB memory. In the training process, the image was resized to 512×512 , and online data augmentation including random left-to-right flipping, up-and-down flipping, -30° to 30° rotation, and additive Gaussian noise were applied. The batch size was 4. The stochastic gradient descent (SGD) algorithm with initial learning rate of 0.01, momentum of 0.9 and weight decay rate of 0.0001 was used to optimize the network. The FANet was trained for 120 iterations, and the IFANet was trained for 100 iterations. The code and data will be

TABLE I
RESULTS OF ABLATION EXPERIMENTS FOR EFA AND SFA MODULES IN FANET

Methods	MISC Dataset				FUSeG Dataset				Inference time(s)
	DSC(%)	IoU(%)	Sen(%)	Spe(%)	DSC(%)	IoU(%)	Sen(%)	Spe(%)	
Baseline	87.31±1.11	79.85±1.25	88.74±1.36	98.96±0.25	83.10±1.21	75.36±1.23	89.13±1.82	99.85±0.28	0.1477
Baseline +EFA	88.86±1.21	82.00±1.47	90.43±0.87	98.97±0.33	83.97±1.21	76.45±1.22	88.78±1.79	99.88±0.25	0.1772
Baseline +SFA	88.33±1.16	81.27±1.27	90.49±1.49	98.87±0.38	84.00±1.13	76.58±1.29	87.51±1.84	99.89±0.20	0.1420
FANet	89.16±0.79	82.16±1.06	90.76±0.84	98.94±0.15	84.36±1.20	76.87±1.24	90.73±1.55	99.87±0.23	0.1721

Bold font indicates the best performance.

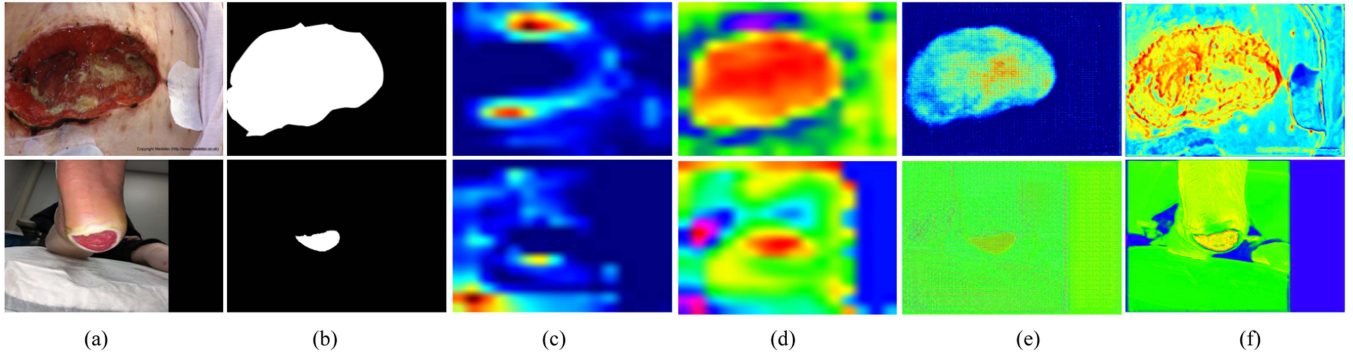


Fig. 5. Feature maps from different layers of the FANet. (a) original image (b) ground truth (c) input feature of SFA (d) output feature of SFA (e) regional features from the decoder (f) edge features from the edge extraction branch.

released at <https://github.com/yinzi/Interactive-Skin-Wound-Segmentation-Based-on-Feature-Augment-Networks>.

D. Evaluation Metrics

Four indices were used for quantitative evaluation, including the Dice similarity coefficient (DSC), intersection-over-union (IoU), sensitivity (Sen), and specificity (Spe), calculated as follows:

$$DSC = \frac{2TP}{2TP + FP + FN} \quad (7)$$

$$IoU = \frac{TP}{TP + FP + FN} \quad (8)$$

$$Sen = \frac{TP}{TP + FN} \quad (9)$$

$$Spe = \frac{TN}{TN + FP} \quad (10)$$

where TP , TN , FP and FN are true positive, true negative, false positive and false negative for segmentation, respectively.

V. RESULTS

A. Results of Automatic Segmentation

1) *Ablation Study*: In order to verify the effectiveness of the proposed EFA and SFA module in FANet, corresponding ablation experiments were carried out on both datasets. It can be seen from Table I that compared with the baseline, adding the EFA module or adding the SFA module can improve the DSC and IoU for both datasets. Adding both modules improves the DSC

by 1.85% and 1.26%, improves the IoU by 2.31% and 1.51%, and improves the Sen by 2.02% and 1.60% for the two datasets, respectively, while Spe is kept high. Compared with the baseline, adding the two modules slightly increases the inference time.

Fig. 5 shows some feature maps obtained by the FANet. It can be seen that after the SFA module, the features become much more focused on the wound region. The decoder outputs feature maps coarsely corresponding to the wound region and the edge extraction branch outputs feature maps with more edge information. These feature maps from two branches can complement each other to give more accurate segmentation results.

2) *Comparisons to other Networks*: We compared the proposed FANet with some state-of-the-art segmentation algorithms: U-Net [40], Attention U-Net [41], PSPNet [42], CE-Net [43], CPFNet [44], CS²-Net [45], and DANet [46], among which, Attention U-Net, CS²-Net, and DANet also employ attention modules, while PSPNet, CE-Net, CPFNet emphasize the global context and multi-scale features. We also compared the results obtained by three models proposed for wound segmentation, namely, cGAN [25], MobilenetV2 [9], and Mask R-CNN [24]. The performance comparison is summarized in Table II. From the table, we can see that the performance of our proposed network is better than all the above networks, giving DSC of $89.16 \pm 0.79\%$ and $84.36 \pm 1.20\%$ for the two datasets, respectively. The wound areas are generally smaller in the FUSeG dataset, which leads to lower segmentation indices with all models. Five sample results from each dataset are also shown in Figs. 6 and 7 to visually compare our method with these methods. With the enhanced spatial relationship and edge information, for both large and small wounds, the proposed

TABLE II
RESULTS OF FANET COMPARED WITH OTHER METHODS

Methods	MISC Dataset				FUSeG Dataset				Inference time(s)
	DSC(%)	IoU(%)	Sen(%)	Spe(%)	DSC(%)	IoU(%)	Sen(%)	Spe(%)	
U-Net [40]	75.58±1.73	65.47±1.94	81.33±2.45	97.70±0.44	81.94±1.22	73.92±1.24	85.13±1.21	99.88±0.23	0.1337
Att_U-Net [41]	81.65±1.68	73.19±1.95	81.97±1.80	99.01±0.09	80.85±1.24	72.75±1.25	84.20±1.22	99.88±0.29	0.1701
PSPNet [42]	83.91±0.99	75.50±1.33	86.94±1.64	98.59±0.36	79.11±1.37	70.23±1.47	83.33±1.25	99.85±0.38	0.1589
CE-Net [43]	88.07±0.72	80.99±0.84	89.57±1.30	98.97±0.21	83.85±1.05	76.10±1.24	88.27±1.71	99.89±0.21	0.1426
CPFNet [44]	87.77±1.00	80.56±1.19	89.76±1.65	98.97±0.23	83.62±1.18	76.14±1.32	87.25±1.93	99.88±0.19	0.1398
CS ² -Net [45]	84.21±1.33	75.86±1.72	86.22±1.38	98.62±0.28	81.01±1.35	72.98±1.49	86.09±2.12	99.85±0.33	0.1424
DANet [46]	88.77±0.70	81.88±0.95	90.33±1.35	98.94±0.27	83.16±1.20	75.54±1.34	87.03±1.89	99.89±0.19	0.1807
cGAN [25]	66.76±2.50	54.76±2.59	69.53±2.62	98.34±0.23	52.62±3.35	42.75±3.14	60.35±3.36	99.58±0.61	0.1320
MobileNetV2 [9]	81.66±1.22	72.37±1.61	84.42±2.37	98.48±0.21	77.43±2.39	67.59±2.51	83.02±2.22	99.82±0.31	0.1828
Mask R-CNN [24]	86.88±1.17	81.06±1.23	88.77±1.23	98.56±0.44	82.69±1.20	74.48±1.22	88.00±1.21	99.70±0.16	0.1736
FANet	89.16±0.79	82.16±1.06	90.76±0.84	98.94±0.15	84.36±1.20	76.87±1.24	90.73±1.55	99.87±0.23	0.1721

Bold font indicates the best performance.

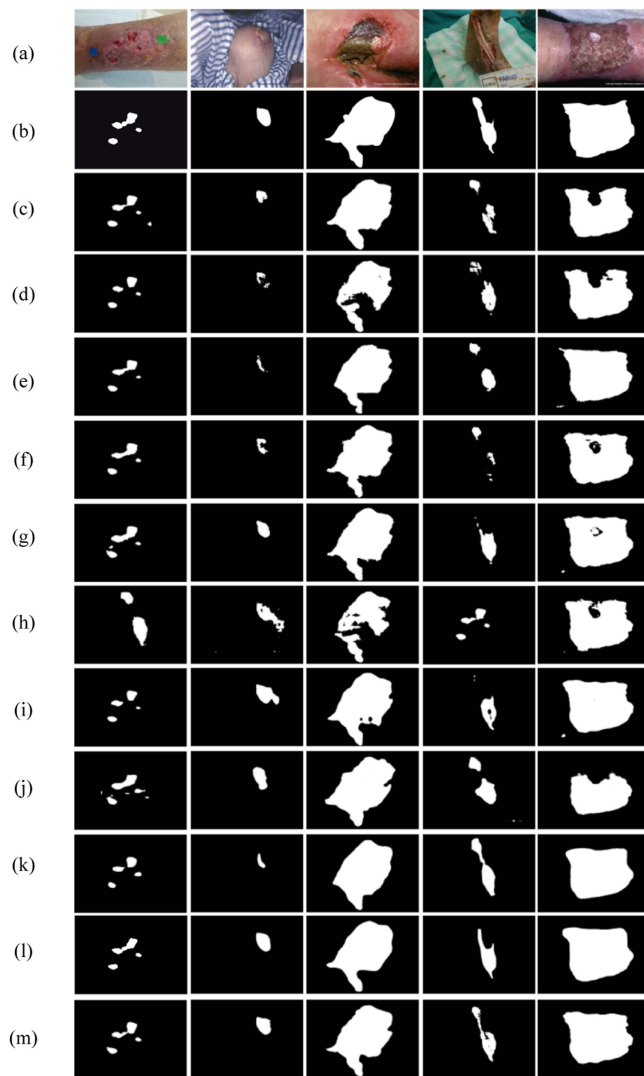


Fig. 6. Visual comparison of results from the MISC dataset. (a) original image (b) ground truth (c) U-Net (d) Attention U-Net (e) PSPNet (f) CE-Net (g) CPFNet (h) CS²-Net (i) DANet (j) cGAN (k) MobileNetV2 (l) Mask R-CNN (m) FANet.

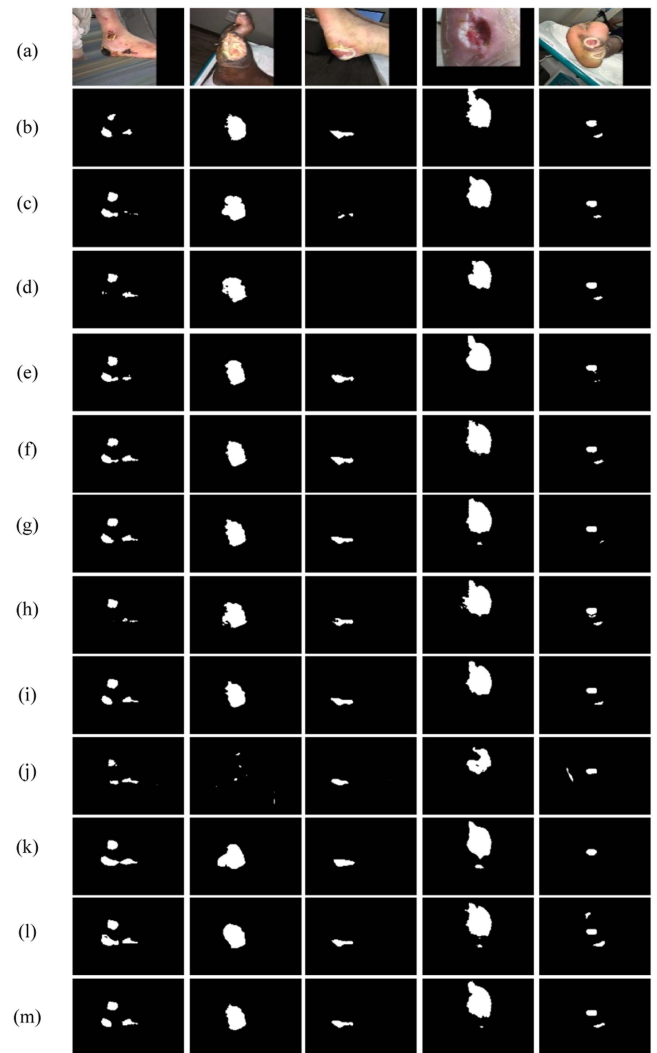


Fig. 7. Visual comparison of results from the FUSeG dataset. (a) original image (b) ground truth (c) U-Net (d) Attention U-Net (e) PSPNet (f) CE-Net (g) CPFNet (h) CS²-Net (i) DANet (j) cGAN (k) MobileNetV2 (l) Mask R-CNN (m) FANet.

TABLE III
RESULTS OF ABLATION EXPERIMENTS FOR EFA AND SFA MODULES IN IFANET

Methods	MISC Dataset				FUSeG Dataset			
	DSC(%)	IoU(%)	Sen(%)	Spe(%)	DSC(%)	IoU(%)	Sen(%)	Spe(%)
Baseline	93.26±0.70	88.70±1.02	93.80±0.81	99.49±0.03	86.42±0.97	79.70±1.10	90.03±1.53	99.91±0.14
Baseline +EFA	93.59±0.79	89.09±1.01	94.20±0.58	99.44±0.11	87.38±0.88	80.92±1.02	90.56±1.60	99.92±0.11
Baseline +SFA	93.85±1.22	89.36±1.52	93.69±1.24	99.55±0.05	87.15±0.85	80.44±0.99	91.21±1.33	99.89±0.25
IFANet	94.03±0.87	89.65±1.14	94.28±0.56	99.51±0.06	87.88±0.77	81.37±0.92	91.92±1.21	99.91±0.15

Bold font indicates the best performance.

FANet is able to segment the wound more accurately, even if there are different colors or illumination in the wound area.

The inference time of different networks is also shown in Table II. The time cost for the proposed model is in the same order as other existing models, and can fulfill the real-time requirement of clinical application.

B. Results of Interactive Segmentation

1) *Ablation Study*: In IFANet, we also did ablation experiments to verify the effectiveness of the EFA module and SFA module. From Table III, we can see that after adding either or both the proposed feature augment modules, most indices are improved. Compared with the baseline, with both modules, the DSC, IoU and Sen of the proposed IFANet are improved by 0.77%, 0.95% and 0.48% for the MISC dataset, and are improved by 1.46%, 1.67% and 1.89% for the FUSeG Dataset, respectively, while Spe is kept high. The improvement is less for the MISC dataset because the performance indices are already on a high level. These results indicate that spatial relationship and edge information is also useful for segmentation refinement.

2) *Improvement with Interaction*: Comparing results in Tables II and III, we can see that the results of IFANet show substantial improvement over FANet. The DSC, IoU and Sen are improved by 4.87%, 7.49%, 3.52% for the MISC dataset, and are improved by 3.52%, 4.50%, 1.19% for the FUSeG dataset, respectively. Increasing in all indices proves that the true positives are increased while false positives are reduced, with guidance of simple scribble interactions. Fig. 8 shows the comparison of the segmentation results before and after interactive refinement. The segmentation results are effectively corrected as expected by human users.

3) *Interaction with other Networks*: In order to prove that the proposed FANet as backbone of the interactive network is better than other networks, we applied the proposed interactive method using other networks as backbone and compared the results. The structure of network was not changed, while the initial segmentation and interactions were input in the same way as in IFANet. Specifically, G_b , G_f and the original image were concatenated as the input, and G_c went through convolution layers and was then inserted after the last layer of the encoder. In order to focus on the performance of the interactive segmentation stage, the result of FANet was used as the initial segmentation for all networks. The performance indices are shown in Table IV, and results of three sample pictures are given in Fig. 9. It can be seen that the interactive method works with all these U-shaped

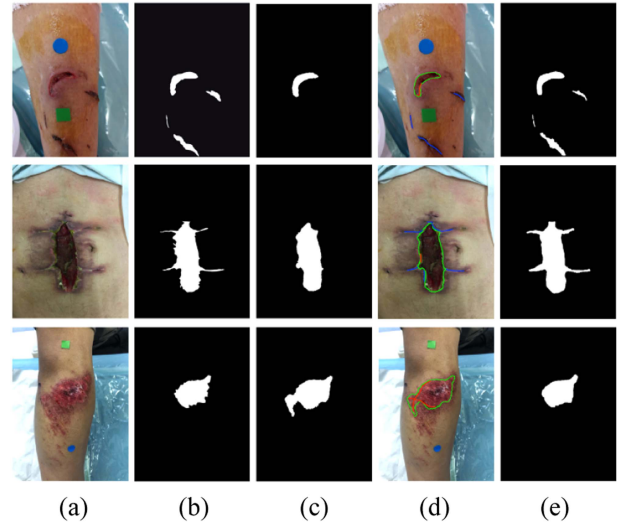


Fig. 8. Visual comparison of FANet and IFANet. (a) original image (b) ground truth (c) result of FANet (d) Manual interaction (Green: automatic segmentation, Red: false positive, Blue: false negative) (e) result of IFANet.

networks and can improve the segmentation (compared with results of FANet in Table II). When the proposed FANet is used as the backbone network, it shows more excellent performance with the highest indices. Visually compared, the IFANet gives boundaries closer to the ground truth.

4) *Comparison with Existing Interactive Segmentation Methods*: We compared the results of IFANet with two classical interactive segmentation methods, Graph Cut [47] and GrabCut [48], and two deep learning based interactive segmentation methods, DEXTR [28] and f-BRS [31]. Note that these methods all require interactive inputs before segmentation starts, and the results can be refined with more inputs. Graph Cut takes scribble interactions, GrabCut needs a bounding box first and scribbles for refinement. DEXTR takes four extreme points first and a fifth point in the boundary can be added to reduce errors. F-BRS uses click interactions. For Graph cut, GrabCut and f-BRS, the maximum number of user refinement steps was set as 10, while for IFANet, only one round of refinement was performed. For fair comparison, all interactions were simulated based on ground truth masks.

The performance indices after interactive refinement are shown in Table IV. It can be seen that the IFANet results in the highest Dice and IoU while the Sen and Spe are relatively high and balanced. Graph Cut has high Sen and low Spe, indicating

TABLE IV
RESULTS OF IFANET COMPARED WITH OTHER METHODS

Methods	MISC Dataset				FUSeG Dataset			
	DSC(%)	IoU(%)	Sen(%)	Spe(%)	DSC(%)	IoU(%)	Sen(%)	Spe(%)
U-Net	90.87±1.18	85.10±1.55	92.01±1.13	99.20±0.10	87.02±0.82	80.12±0.98	90.41±1.28	99.90±0.18
+interactive Att_U-Net	90.89±0.85	85.15±1.24	90.55±0.60	99.40±0.13	87.36±0.85	80.12±1.04	90.41±1.35	99.90±0.15
+interactive PSPNet	91.53±1.07	85.99±1.29	92.03±0.91	99.27±0.10	84.41±0.95	76.57±1.13	88.76±1.46	99.89±0.16
+interactive CE-Net	93.76±0.89	89.23±1.24	94.12±0.84	99.48±0.07	87.44±0.80	80.77±0.95	92.08±1.15	99.90±0.16
+interactive CPFNet	93.74±1.10	89.26±1.36	94.15±1.03	99.49±0.07	87.12±0.87	80.42±0.99	91.38±1.34	99.89±0.17
+interactive CS ² -Net	91.27±1.04	85.66±1.35	92.09±0.79	99.23±0.13	87.17±0.86	80.55±1.03	91.34±1.27	99.90±0.19
+interactive DANet	93.62±0.63	89.19±0.79	93.51±0.76	99.50±0.07	86.80±0.81	79.77±0.96	90.86±1.24	99.90±0.15
+interactive Graph Cut [47]	90.97±8.90	84.44±12.50	94.56±6.88	98.75±1.68	87.17±10.78	78.68±15.00	90.03±10.31	99.85±0.23
GrabCut [48]	91.05±11.33	84.97±13.81	89.70±12.84	99.44±0.57	87.25±13.23	79.18±15.79	85.87±14.05	99.94±0.08
DEXTR [28]	83.01±1.38	73.08±1.72	83.57±1.83	98.31±0.33	87.06±0.93	78.21±1.35	91.90±1.25	99.88±0.11
f-BRS [31]	88.20±0.61	81.48±1.83	91.02±1.53	98.19±0.40	84.20±1.03	76.45±1.17	89.61±1.01	99.88±0.21
IFANet	94.03±0.87	89.65±1.14	94.28±0.56	99.51±0.06	87.88±0.77	81.37±0.92	91.92±1.21	99.91±0.15

Bold font indicates the best performance.

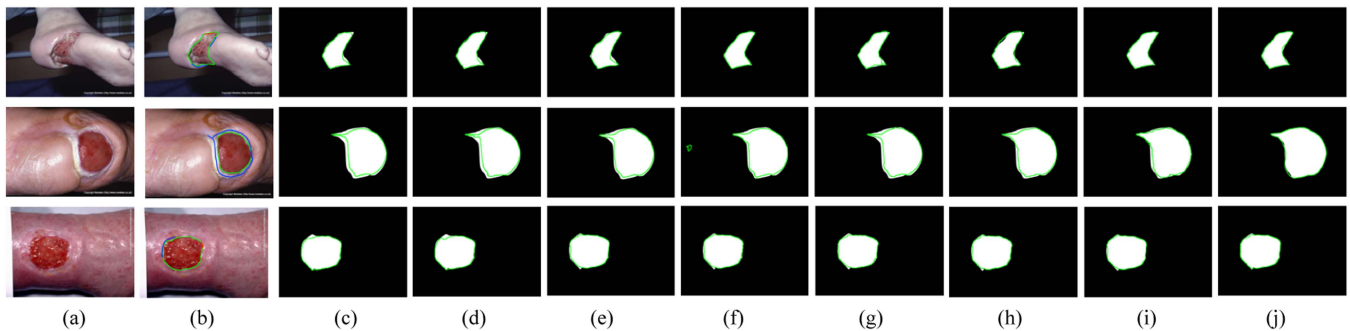


Fig. 9. Visual comparison of interactive segmentation results with other networks as backbone. (a) original image (b) initial segmentation and interactions (Green: automatic segmentation boundary, Red: marks for background, Blue: marks for foreground) (c) U-Net (d) attention U-Net (e) PSPNet (f) CE-Net (g) CPFNet (h) CS²-Net (i) DANet (j) IFANet (Green: interactive segmentation results, white: ground truth mask).

that it tends to over-segment the wound. On the contrary, GrabCut has low Sen and high Spe, indicating under-segmentation. DEXTR has higher performance on the FUSeG dataset with smaller wound, as it crops and resizes the ROI to form the input. Some examples are shown in Fig. 10. It can be seen that with IFANet, only some manual corrections are needed near the boundaries, because the initial automatic segmentation is already quite accurate. Other methods, especially the traditional ones, need more interactions to reach satisfactory results.

5) *Test with Real Human Interaction*: Simulated interactions were designed to avoid human labor in training the interactive network and to facilitate performance comparison. However, in order to verify the performance of IFANet in practical scenarios, we tested its performance when the inputs were real human interactions, on a subset of our dataset. Two medical workers independently performed manual interaction on one-fold of

TABLE V
RESULTS OF REAL HUMAN INTERACTION COMPARED WITH SIMULATED INTERACTION ON A SUBSET OF 110 IMAGES

Interaction	DSC(%)
simulated	94.15
medical staff (a)	93.18
medical staff (b)	93.37

MISC dataset with 110 images using in-house built software, where the initial segmentation boundaries were displayed on the image, and the scribbles were drawn with mouse. The results obtained by the two sets of interactions respectively are compared with those obtained by simulated interaction inputs, as shown in Table V. It is found that using real interactions, results are slightly worse than that using the simulated interaction. The difference is less than 1.0%. The reason maybe as follows. First,

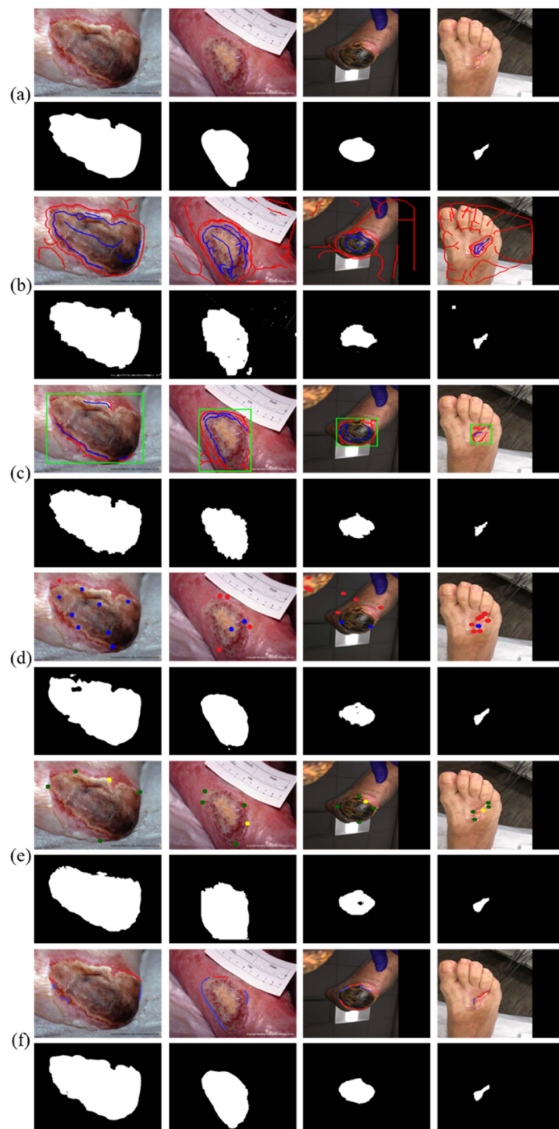


Fig. 10. Visual comparison with results of existing interactive segmentation methods. (a) Original image and ground truth (b) Graph Cut (c) GrabCut (d) f-BRS (e) DEXTR (f) IFANet. For Graph Cut, GrabCut, f-BRS and IFANet, red refers to background marks and blue refers to foreground marks. For GrabCut, green refers to user-provided bounding box. For DEXTR, green refers to extreme points, and yellow refers to the extra boundary point for refinement.

in simulation, the center pixels of the incorrect segmented area are selected as the seed points for calculating the distance map. While for human users, the scribbles deviate from the center line most of the times. Secondly, if the error region has a thin but elongated shape along the edges, it can easily be overlooked by human users, but will be picked up in simulation. However, the experiment still proves the effectiveness of our proposed interactive network in refining segmentation results. The inter-user variation is also small, with 0.19% difference in DSC.

VI. DISCUSSION AND CONCLUSION

In this work, we propose a two-stage skin wound segmentation framework based on convolutional neural networks. In the first stage, taking the characteristics of the skin wound

image into consideration, the FANet is proposed for initial automatic segmentation. In FANet, an edge feature extraction branch, composed of EFA modules and supervised by edge loss is designed. The edge features are augmented by the multi-scale features from the encoder and effectively supplements the features for generating the final segmentation map. Furthermore, an SFA module emphasizing global context is designed to provide enhanced features with spatial relationships. In general, the proposed network structure is suitable for segmentation of regions with complex color and textures, especially near the edges.

The second stage is for interactive refinement aiming at the need in clinical practice. Based on the same structure of FANet, the IFANet is retrained with combined input of the original image, the initial segmentation, and the information from the scribble-based interaction. Both the initial segmentation and the scribbles are transformed into Gaussian distance maps but some difference exists. A flatter Gaussian kernel, with tripled standard deviation, is used for the initial segmentation, because the total wound area is usually larger than the incorrect segmentation areas. Furthermore, as the scribbles represent areas that are difficult to segment, their corresponding Gaussian maps go through more convolution layers for better feature learning.

Experiments were performed on two datasets, one with various types of skin wounds, and another with small foot ulcer wounds which are more difficult to segment. Ablation tests show that the proposed EFA and SFA modules both contribute to the automatic segmentation and segmentation refinement results. Comparative tests show that the proposed network structure outperforms other related networks in both stages. Comparison between results of FANet and IFANet shows that significant improvement is achieved with simple interactions. Unlike traditional interactive segmentation methods such as Graph Cut [47], GrabCut [48], or deep learning based method such as in [26], [27], [28], [29], [30], [31], [32], [33], [34], the proposed framework does not need interactive inputs in the beginning, making it more convenient in practical use. In fact, the interactive refinement may not be necessary when the user considers the initial results accurate enough. Comparative experiments with some existing interactive methods prove that the IFANet requires fewer interactions and achieves better refinement results. The interaction using scribbles is also easy to handle. In short, the proposed framework is designed to minimize human operations while obtaining high segmentation accuracy.

The proposed method is the core component for photo-graph-based non-contact wound measurement. With a scale or ruler placed beside the wound, the actual area and diameter of the wound can be calculated. It also has the potential to be applied to telemedicine. In the future, we will further study the accuracy of wound measurement in practical settings. In practice, in order to reduce the error of 2D measurement of the 3D space, the shape of the scale can be detected in real-time and used to enforce frontal imaging with respect to the wound plane, or to perform geometric corrections to some extent. Furthermore, to overcome the projection errors in 2D area calculation, we will explore models that can segment multi-view images, and perform more accurate measurement in the reconstructed 3D space. On the other hand, in tracking of the wound healing process, it is also important to measure different tissue types of

the wound separately, such as granulation, slough and necrosis. We will further explore extension of the proposed model for the task of wound tissue segmentation.

REFERENCES

- [1] M. Zhu et al., "Mesenchymal stromal cells pretreated with pro-inflammatory cytokines promote skin wound healing through VEGFC mediated angiogenesis," *Stem Cells Transl. Med.*, vol. 9, no. 10, pp. 1218–1232, 2020.
- [2] C. Meads, E. Lovato, and L. Longworth, "The debrisoft monofilament debridement pad for use in acute or chronic wounds: A NICE medical technology guidance," *Appl. Health Econ. Health Policy*, vol. 13, no. 6, pp. 583–594, 2015.
- [3] D. T. Ubbink, R. Lindeboom, A. M. Eskes, H. Brull, D. A. Legemate, and H. Vermeulen, "Predicting complex acute wound healing in patients from a wound expertise centre registry: A prognostic study," *Int. Wound J.*, vol. 12, no. 5, pp. 531–536, 2015.
- [4] M. Marian and J. Ziegler, "Venous ulcers of the lower extremities," *Topics Clin. Nutr.*, vol. 31, no. 1, pp. 3–23, 2016.
- [5] C. Lachenbruch, D. Ribble, K. Emmons, and C. VanGilder, "Pressure ulcer risk in the incontinent patient analysis of incontinence and hospital-acquired pressure ulcers from the international pressure ulcer prevalence™ survey," *J. Wound, Ostomy, Continence Nurs.*, vol. 43, no. 3, pp. 235–241, 2016.
- [6] R. S. Kirsner and A. C. Vivas, "Lower-extremity ulcers: Diagnosis and management," *Brit. J. Dermatol.*, vol. 173, no. 2, pp. 379–390, 2015.
- [7] L. I. F. Moura, A. M. Dias, E. Carvalho, and H. C. de Sousa, "Recent advances on the development of wound dressings for diabetic foot ulcer treatment-A review," *Acta Biomaterialia*, vol. 9, no. 7, pp. 7093–7114, 2013.
- [8] R. G. Frykberg and J. Banks., "Challenges in the treatment of chronic wounds," *Adv. Wound Care*, vol. 4, no. 9, pp. 560–582, 2015.
- [9] C. Wang et al., "Fully automatic wound segmentation with deep convolutional neural networks," *Sci. Rep.*, vol. 10, no. 1, 2020, Art. no. 21897.
- [10] L. K. Branski, G. G. Gauglitz, D. N. Herndon, and M. G. Jeschke, "A review of gene and stem cell therapy in cutaneous wound healing," *Burns*, vol. 35, no. 2, pp. 171–180, 2009.
- [11] M. Romanelli, "Technological advances in wound bed measurements," *Wounds*, vol. 14, no. 2, pp. 58–66, 2002.
- [12] O. P. David, D. Sierra-Sosa, and B. G. Zapirain, "Pressure ulcer image segmentation technique through synthetic frequencies generation and contrast variation using toroidal geometry," *Biomed. Eng. Online*, vol. 16, 2017, Art. no. 4.
- [13] S. Zahia, M. B. Garcia Zapirain, X. Sevillano, A. González, P. J. Kim, and A. Elmaghraby, "Pressure injury image analysis with machine learning techniques: A systematic review on previous and possible future methods," *Artif. Intell. Med.*, vol. 102, 2020, Art. no. 101742.
- [14] P. Unger, C. Fife, and D. Weir, "Capturing the essence of the wound evaluation," *Today's Wound Clinic*, vol. 2, no. 2, pp. 38–42, 2008.
- [15] T. A. Krouskop, R. Baker, and M. S. Wilson, "A noncontact wound measurement system," *J. Rehabil. Res. Develop.*, vol. 39, no. 3, pp. 337–346, 2002.
- [16] T. Gilman, "Wound outcomes: The utility of surface measures," *Int. J. Lower Extremity Wounds*, vol. 3, no. 3, pp. 125–132, 2004.
- [17] C. Ahn and R. S. Salcido, "Advances in wound photography and assessment methods," *Adv. Skin Wound Care*, vol. 21, no. 2, pp. 85–93, 2008.
- [18] M. F. A. Fauzi, I. Khansa, K. Catignani, G. Gordillo, C. K. Sen, and M. N. Gurcan, "Computerized segmentation and measurement of chronic wound images," *Comput. Biol. Med.*, vol. 60, pp. 74–85, 2015.
- [19] M. Dhane et al., "Fuzzy spectral clustering for automated delineation of chronic wound region using digital images," *Comput. Biol. Med.*, vol. 89, pp. 551–560, 2017.
- [20] L. Wang et al., "An automatic assessment system of diabetic foot ulcers based on wound area determination, color segmentation, and healing score evaluation," *J. Diabetes Sci. Technol.*, vol. 10, no. 2, pp. 421–428, 2015.
- [21] H. Wannous, S. Treuillet, and Y. Lucas, "Robust tissue classification for reproducible wound assessment in telemedicine environments," *J. Electron. Imag.*, vol. 19, no. 2, 2010, Art. no. 023002.
- [22] A. F. M. Hani, L. Arsha, A. S. Malik, A. Jamil, and F. Y. B. Bin, "Haemoglobin distribution in ulcers for healing assessment," in *Proc. 4th Int. Conf. Intell. Adv. Syst.*, 2012, pp. 362–367.
- [23] H. Lu et al., "Wound intensity correction and segmentation with convolutional neural networks," *Concurrency Computation: Pract. Experience*, vol. 29, no. 6, 2017, Art. no. e3927.
- [24] C. Jiao, K. Su, W. Xie, and Z. Ye, "Burn image segmentation based on mask regions with convolutional neural network deep learning framework: More accurate and more convenient," *Burns Trauma*, vol. 7, 2019, Art. no. 6.
- [25] S. Sarp, M. Kuzlul, M. Pipattanasomporn, and O. Guler, "Simultaneous wound border segmentation and tissue classification using a conditional generative adversarial network," *J. Eng.*, vol. 2021, no. 3, pp. 125–134, 2021.
- [26] N. Xu, B. Price, S. Cohen, J. Yang, and T. Huang, "Deep interactive object selection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 373–381.
- [27] J. H. Liew, Y. Wei, X. Wei, S. H. Ong, and J. Feng, "Regional interactive image segmentation networks," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 2746–2754.
- [28] K. Maninis, S. Caelles, J. Pont-Tuset, and L. V. Gool, "Deep extreme cut: From extreme points to object segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 616–625.
- [29] N. Xu, B. Price, S. Cohen, J. Yang, and T. Huang, "Deep grabcut for object selection," in *Proc. 2017 Brit. Mach. Vis. Conf.*, 2017, pp. 1–12.
- [30] W. Jang and C.-S. Kim, "Interactive image segmentation via backpropagating refinement scheme," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 5297–5306.
- [31] K. Sofiuk, I. Petrov, O. Barinova, and A. Konushin, "F-BRS: Rethinking backpropagating refinement for interactive segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 8623–8632.
- [32] X. Chen, Z. Zhao, F. Yu, Y. Zhang, and M. Duan, "Conditional diffusion for interactive segmentation," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2021, pp. 7325–7334.
- [33] X. Chen, Z. Zhao, Y. Zhang, M. Duan, D. Qi, and H. Zhao, "FocalClick: Towards practical interactive image segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 1290–1299.
- [34] G. Wang et al., "Interactive medical image segmentation using deep learning with image-specific fine tuning," *IEEE Trans. Med. Imag.*, vol. 37, no. 7, pp. 1562–1573, Jul. 2018.
- [35] G. Wang et al., "DeepGeoS: A deep interactive geodesic framework for medical image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 7, pp. 1559–1572, Jul. 2019.
- [36] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.
- [37] X. Wang, R. Girshick, A. Gupta, and K. He, "Non-local neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 7794–7803.
- [38] Medetec Wound Database, 2017. [Online]. Available: <http://www.medetec.co.uk/files/medetec-image-databases.html>
- [39] Foot Ulcer Segmentation Challenge, 2021, [Online]. Available: <https://github.com/uwm-bigdata/wound-segmentation>
- [40] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Intervention*, 2015, pp. 234–241.
- [41] J. Schlemper et al., "Attention gated networks: Learning to leverage salient regions in medical images," *Med. Imag. Anal.*, vol. 53, pp. 197–207, 2019.
- [42] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 6230–6239.
- [43] Z. Gu et al., "CE-Net: Context encoder network for 2D medical image segmentation," *IEEE Trans. Med. Imag.*, vol. 38, no. 10, pp. 2281–2292, Oct. 2019.
- [44] S. Feng et al., "CPFNet: Context pyramid fusion network for medical image segmentation," *IEEE Trans. Med. Imag.*, vol. 39, no. 10, pp. 3008–3018, Oct. 2020.
- [45] L. Mou et al., "CS²-net: Deep learning segmentation of curvilinear structures in medical imaging," *Med. Image Anal.*, vol. 67, 2021, Art. no. 101874.
- [46] J. Fu et al., "Dual attention network for scene segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 3141–3149.
- [47] Y. Y. Boykov and M. -P. Jolly, "Interactive graph cuts for optimal boundary & region segmentation of objects in N-D images," in *Proc. IEEE 8th Int. Conf. Comput. Vis.*, 2001, pp. 105–112.
- [48] C. Rother, V. Kolmogorov, and A. Blake, "Interactive foreground extraction using iterated graph cuts," *ACM Trans. Graph.*, vol. 23, pp. 309–314, 2004.