IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received May 13, 2019, accepted May 31, 2019, date of publication June 10, 2019, date of current version June 26, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2921815

Dense-Residual Network With Adversarial Learning for Skin Lesion Segmentation

WENLI TU[1,2], XIAOMING LIU[1,2], WEI HU[1,2], AND ZHIFANG PAN[3,4]

[1]College of Computer Science and Technology, Wuhan University of Science and Technology, Wuhan 430065, China
[2]Hubei Province Key Laboratory of Intelligent Information Processing and Real-Time Industrial System, Wuhan 430065, China
[3]School of Information and Engineering, Wenzhou Medical University, Wenzhou 325035, China
[4]Information Technology Center, Wenzhou Medical University, Wenzhou 325035, China

Corresponding author: Xiaoming Liu (lxmspace@gmail.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 61403287, Grant 61472293, and Grant 61572381, in part by the Key Project of Hubei Provincial Department of Education under Grant D20181103, and in part by the Foundation of Wenzhou Science and Technology Bureau under Grant 2018ZG016.

ABSTRACT Skin lesion segmentation in dermoscopic images is a challenging task in the domain of medical images analysis because of the irregular and blurring edges of the lesion and the presence of various artifacts. Inspired by the successful applications of the generative adversarial network (GAN), we propose a new neural network for the segmentation of skin lesion. Different from the traditional adversarial network, the segmentation network uses encoder–decoder with the dense-residual block which enables the network to be trained more efficiently. It can also establish a direct relationship between adjacent pixels to improve segmentation accuracy. A multi-scale objective loss function is used to utilize deep supervision in multiple layers of the critic network. End point error and Jaccard distance are combined as the content loss function. It can solve the problem of boundary ambiguity and solve the lesion-background imbalance in pixel-level classification for skin lesion segmentation. We finally use a joint loss function including a multi-scale objective loss function, end point error, and Jaccard distance content loss function. The experiment results show that our algorithm is superior to other state-of-the-art algorithms on the ISBI2017 and PH2 datasets.

INDEX TERMS Adversarial learning, convolutional neural networks, dense-residual block, skin lesion, dermoscopic image.

I. INTRODUCTION

Malignant melanoma is a common and one of fastest-growing cancer in the world, and the number of deaths has increased in recent decades. Although some advanced treatment techniques are widely used, the survival rate of late melanoma within five years is less than 15% while early melanin is more than 6 times [1]. It is obvious that timely diagnosis and treatment of melanin is essential for the survival of patients. Dermatoscopy, also known as incident light microscopy, is a skin microscopy that can be magnified dozens of times. It is used to observe the skin pigmentation disease. Although the microscopy can observe more image details, it is often time consuming and complicated. To help dermatologists improve their efficiency, computerized analytical methods have been developed. The most advanced method up to 2012 is reviewed in [2].

The associate editor coordinating the review of this manuscript and approving it for publication was Kumaradevan Punithakumar.

It is an important step in computerized analysis of dermoscopic images to automatically segment melanoma from the surrounding skin [3], [4]. Clustering, thresholding, region merging and splitting, active contour models and supervised learning have been proposed by researchers in the past few years. These algorithms have many different advantages and drawbacks that have been compared and analyzed in [5], [6]. Many automatic algorithms, such as K-nearest neighbor [4], support vector machine (SVM) [3], [7], AdaBoost [8] and manifold learning methods [9], [10], have been proposed. However, the above method can't achieve very good results due to the size, shape, color, and texture of melanoma vary greatly among different types of skin. In addition, some artificial factors, such as hair on the skin, pigmentation inconsistency, blood vessels, air bubbles, etc. will increase the difficulty of segmentation task. Some examples are shown in Figure 1.

In recent years, convolutional neural networks (CNN) has made significant progress in different computer vision

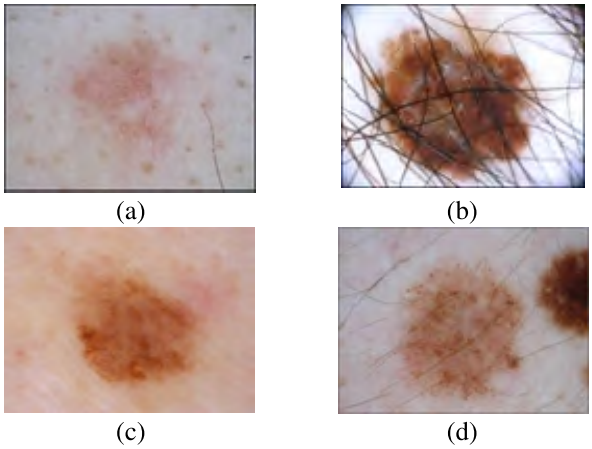

FIGURE 1. Example images from the ISBI2017 melanin detection. These samples represent the presence of typical factors such as (a) low contrast between lesion and the surrounding skin, (b) hair on the skin, (c) irregular and fuzzy borders, and (d) inconsistent pigmentation.

tasks such as image classification [11], object detection [12], semantic image segmentation [13]. In addition to the successful application on natural images, CNN algorithm is increasingly applied to the processing of medical images, such as the recognition of CT images [14], the registration of images [15], segmentation in optical coherence tomography images [7], [16], including a few works on skin lesion segmentation in dermoscopic images [17] and the classification of skin cancer [18].

Inspired by Long *et al.* [19] and Luc *et al.* [20], an adversarial learning skin lesion segmentation method is proposed in this paper. Generally, GAN produces fairly good output through mutual game learning of two modules in the framework: a generator G and a discriminator D. In the training process, the goal of G is to generate an image looks as realistic as possible to deceive the discriminator D. The goal of D is to separate the G-generated image from the real image. Thus, G and D constitute a dynamic ''gaming process''. Different from the traditional adversarial network, we propose a new adversarial network model for skin lesions segmentation.

In this paper, we propose a novel end-to-end adversarial network for skin lesion segmentation in which the segmentation network uses encoder-decoder architecture with Dense-Residual block. Densenet [21] achieves better results with fewer parameters through the features reuse. Resnet [11] accelerates network convergence and optimization. The segmentation networks(S) and critic network(C) are trained in an alternating manner that optimize the loss function. The S network is designed to generate dense feature maps and results in highly accurate lesion segmentation. In order to speed up network training and learn both global and local features, a multi-scale objective loss function is used in critic network. Furthermore, the End Point Error (EPE) [22] content loss function is utilized to precisely divide the boundaries of the melanoma region. Jaccard distance content loss function can solve lesion-background imbalance in pixel-level classification for skin lesion segmentation. We combine multi-scale loss function with EPE and Jaccard distance as new loss function. Figure2 shows the flowchart of the proposed framework.

The contributions of this paper include the following:

1. We propose a novel adversarial network for skin lesion segmentation, in which S network uses encoder-decoder architecture with our proposed Dense-Residual block. We set up multiple skip connections between encoder-decoder in segmentation network. This proposed network is designed to generate dense feature maps and results in highly accurate lesion segmentation. In addition, global convolution network [23], [24] is added to critic network to acquire large receptive field.

2. To learn both global and local features and realize deep supervision, we use a multi-scale objective loss function. Furthermore, EPE and Jaccard distance content loss functions are used to divided the boundaries of the melanoma region and solve lesion-background imbalance in pixel-level classification for skin lesion segmentation.

3. Our model is an end-to-end network architecture that handles the entire image. Experimental results show that our method achieves good segmentation results on ISBI2017 [25] and PH2 datasets [26].

The structure of this paper is organized as follows: the related work is presented in Section 2. Section 3 introduces our method in detail. The experimental results and evaluation are then presented in Section 4. Then discussions are given in Section 5. Finally, we give the conclusion in the section 6.

II. RELATED WORK

We overview some related work to our research in this section, including semantic segmentation, GANs and skin lesion segmentation techniques.

A. SEMANTIC SEGMENTATION

Semantic segmentation is regarded as a dense pixel classification task which assign a label to every pixel. Classification and localization are generally regarded as two major challenges in semantic segmentation. The model which is insensitive to location information can improve classification performance but degrade the segmentation performance. Many deep learning methods using CNNs are proposed to solve the contradiction problem in image segmentation. The classification of image patch [27] is the first application of deep learning in semantic segmentation. However, it is time consuming in inference. The Fully convolutional network (FCN) network [19] proposed by Long et al. is a typical CNN model for image segmentation, which can greatly speed up the segmentation. The FCN model replaces the fully connected layer in traditional CNN with a convolutional layer to obtain a rough label mapping, and then upsamples the label mapping with a deconvolution to obtain a high-resolution feature map. The upsampling operation is used to compensate for the reduced resolution caused by pooling but it also makes the final segmentation result fuzzy. Ronneberger *et al.*

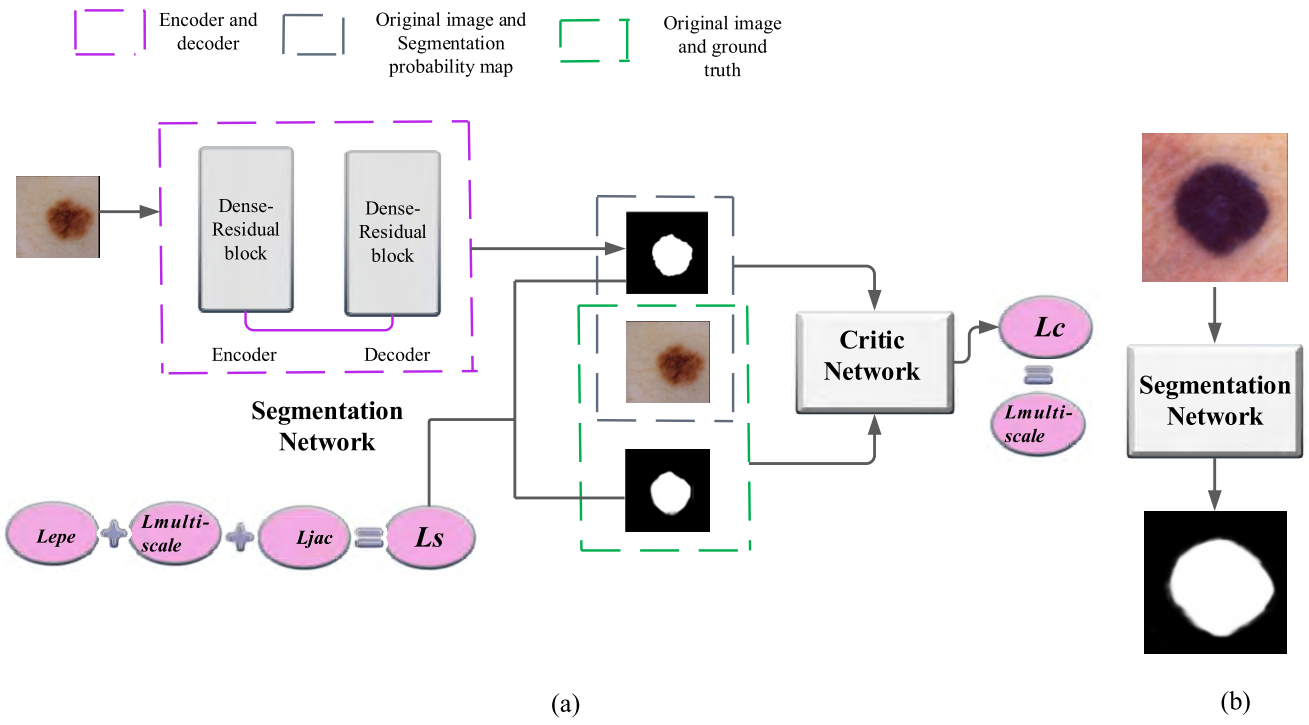

FIGURE 2. The flow chart of the network model (a), the segmentation network and critic network are optimized by minimization of *Ls* (S network loss) and *Lc* (C network loss) during the training phase. (b) shows that a test image is input and the segmentation result is output in the test phase.

proposed U-Net [28] which is an end-to-end encoder-decoder network architecture. The encoder gradually reduces the spatial dimension with the pooling layer, and the decoder gradually recover the details and spatial dimensions of the image. There are skip connections between the encoder and the decoder to help the decoder obtain image details from the encoder part. This method is widely used in medical image segmentation [29].

Recently, ResNets [11] has been incorporated into FCNs [30], [31] to simplify the training of deep (hundreds of layers) by introducing a residual block. The main feature of ResNet is to sum up the nonlinear transformation of input and its identity mapping to increases the number of connections in the network. This method is helpful for network optimization, improving segmentation accuracy and accelerating network convergence [30]. Huang *et al.* [21] proposed DenseNet that can be seen as an extension of ResNets [11]. Densenet is built by dense block and pooling operation. In the dense block, the feature maps in all the previous layers are used as input for each layer, thus it can reuse all the previous features. Through the feature reuse, it achieves better results with fewer parameters. Inspired by their success, we proposed Dense-Residual block in both the segmentation network and critic network to utilize their advantages.

B. ADVERSARIAL LEARNING

GAN was proposed by Goodfellow *et al.* [32] and it consists of a discriminator D and a generator G. The input of the generator G is a random noise vector z, following a simple distribution (such as Gaussian distribution). The output of the generator G is a synthesized image that is similar to the distribution $p_{x\sim p_{data}(x)}$ of a real sample data x. The discriminator D is a binary classifier used to estimate the probability of a sample comes from training data $p_{x\sim p_{data}(x)}$. The loss function in traditional GAN is defined as follows:

$$\min_G \max_D V(D,G) = \mathop{E}_{x\sim p_{data}(x)}[\log(D(x))] + \mathop{E}_{z\sim P_z(z)}[\log(1-D(G(z)))] \quad (1)$$

Adversarial learning was first used for semantic segmentation by Luc *et al.* [20]. In their work, an additional discriminator is used to identify the real or fake of the entire input image to improve the segmentation results. Adversarial learning has also been used in medical image segmentation [7]. A novel multi-scale loss function was proposed for both G and D by Xue *et al.* [33] to directly and effectively enforce the learning of hierarchical features. Different from [20] while similar to [33], we also used a encoder-decoder architecture and multi-scale loss function. Besides, we have incorporated Dense-Residual block in encoder-decoder architecture. For the loss function, we have also incorporated End Point Error and Jaccard distance loss function.

C. SKIN LESION SEGMENTATION IN DERMOSCOPIC IMAGES

In recent decades, skin lesion segmentation has been studied and several methods have been proposed [34]–[38]. Here we will briefly introduce some closely related CNN methods.

Yuan and Lo [17] introduce an automatic method of skin lesion segmentation using 19 layers deep FCN and is trained in end-to-end. In order to simultaneously produce the segmentation and the coarse classification result, two fully convolutional residual networks is proposed in [10]. A deep ResNet was also utilized in [39] to enhance robust visual features learning and representations. The deep ResNet has 50 layers for segmentation of skin lesions to obtain better segmentation results and improve accuracy. Enhanced Convolutional-Deconvolutional Networks [17] were used for automatic segmentation of skin lesions. However, the proposed methods can't realize deep supervision, which lead to loss of information details and can't further enhance the performance of segmentation on skin lesion. A novel multi-scale loss function is used in [33] to enforce the learning of hierarchical feature maps.

III. THE PROPOSED METHOD

Our goal is to separate the lesion in the skin image from the surrounding normal tissue without manual intervention. The proposed network architecture consists of two parts: a segmentation network (S) and a critic network (C). Our approach consists of a training and a testing phase, as shown in Figure 2.

In the training stage (Figure 2(a)), the purpose of the S network is to generate a posterior probability map similar to ground truth label from the original input image. Multi-scale objective loss function, EPE and Jaccard distance are calculated with posterior probability map and corresponding ground truth. The S network is a fully convolutional network with encoder-decoder Dense-Residual block architecture which is based on U-Net [28]. The C network has two types of inputs: One is the pixel-level multiplication of the predicted label map generated by the S network and the original image, and the other is the pixel-level multiplication of the ground truth label map and the original image. There is a problem with adversarial segmentation method that discrete label masks are contained in reference segmentations, while a continuous probability value is produced by segmentator for each class in each pixel. It may learn to discriminate between discrete and continuous values to distinguish reference label and outputs of segmentator. In order to solve the problem, we used pixel-level multiplication of the segmentation mask and the original image, ground truth label and the original image [40] as the input of the critic network. By introducing feedback from critics into the segmentation network, the segmentation network can produce more accurate skin lesion segmentation.

In the testing stage, we only use the segmentation network to generate posterior probability map for the test image (Figure 2(b)). Due to the network is fully convolutional, images of any size can be used as input and it can generate segmentation results with the same size.

In the adversarial network, the S and C networks are trained alternately by backpropagation in an adversarial manner. We first fix the C network and use the gradient calculation in the loss function of the S network to train the S network. Then, we fix the S network and use the gradient calculation in the loss function of the C network to train it. After the training is completed, both the S and C networks become very powerful. The results prove that the final S network can produce a more accurate segmentation probability map.

A. NETWORK ARCHITECTURE

1) SEGMENTATION NETWORK

We use encoder-decoder with Dense-Residual block which is based on U-Net as our segmentation network, as shown in Figure 3. The network framework is inspired by [41]. In order to obtain a deeper feature map of the input image and ensure maximum information transmission between the layers in the network, we use the Dense block [21]. Dense block leverages the potency of the network through feature reuse, resulting in a streamlined model that is easy to train and is efficiently parameterized. In order to alleviate the difficulty of training deep networks, we connected a residual block network after each Dense block. In particular, the S network consists of a compression path with five encoder blocks, a bottom block and an extension path corresponding to five decoder blocks, as listed in Table 1. In the encoder stage, skip connection exists around dense blocks 1 and 2. It leads to the linear growth of feature maps and enhance the reuse of features. In the decoder stage, the upsampling path increases the spatial resolution of feature maps, it will require too much memory if skip connection is used [41]. In order to overcome this limitation, the input of dense blocks is not connected with its output. There is no skip connection skipping dense blocks 4 and 5 in the decoder stage, as shown in Figure 3.

In the encoder, we use 3×3 convolution firstly. Then, five Dense-Residual blocks are used. Each of Dense-Residual block contains Dense block, Concatenation, Residual block and 2×2 max pooling as downsampling which reduce the dimension of the image. Figure 4(a) illustrates the layout of the Dense block, in which connect each layer directly to all subsequent layers. It is used to enhance the flow of information between layers. The l^{th} layer receives the feature-maps of all preceding layers. The formula is as follows:

$$x_l = H_l([x_0, x_1, \ldots, x_{l-1}]) \tag{2}$$

$H_l(\cdot)$ can represent Pooling, Convolution(conv), rectified linear units (ReLU) or batch standardization (BN) composite functions. The output of the l^{th} layer is expressed as $x_l.[x_0, x_1, \ldots, x_{l-1}]$ represents the connection of the feature map $x_0, x_1, \ldots, x_{l-1}$ in layers. The number of feature map outputted by each layer in each dense block is the same, and we denote it with k. In order to limit the requirement of memory, we use a small k value with 16. The k determines the contribution of information in each new layer to the global state. Once the global state is written, it can be accessed from anywhere in the network. When the number of network layers reaches a certain number, the performance of the network may be saturated. This is due to the reason when the network becomes deep, the network becomes difficult to train. It has

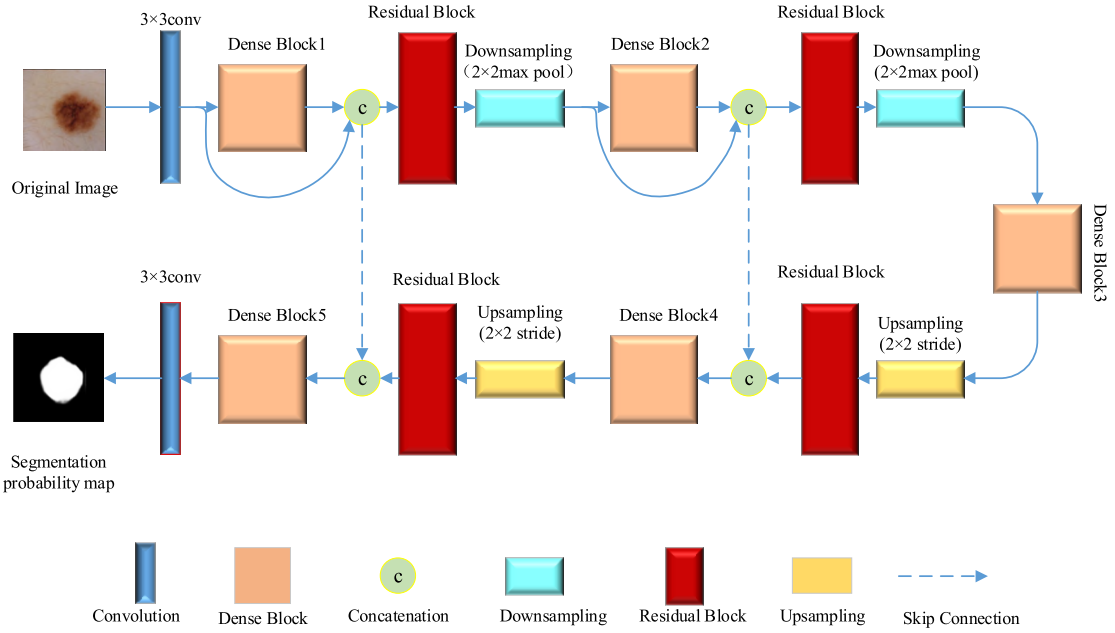

FIGURE 3. The segmentation model only two Dense-residual-blocks are shown. The final segmentation probability map is obtained from the original image.

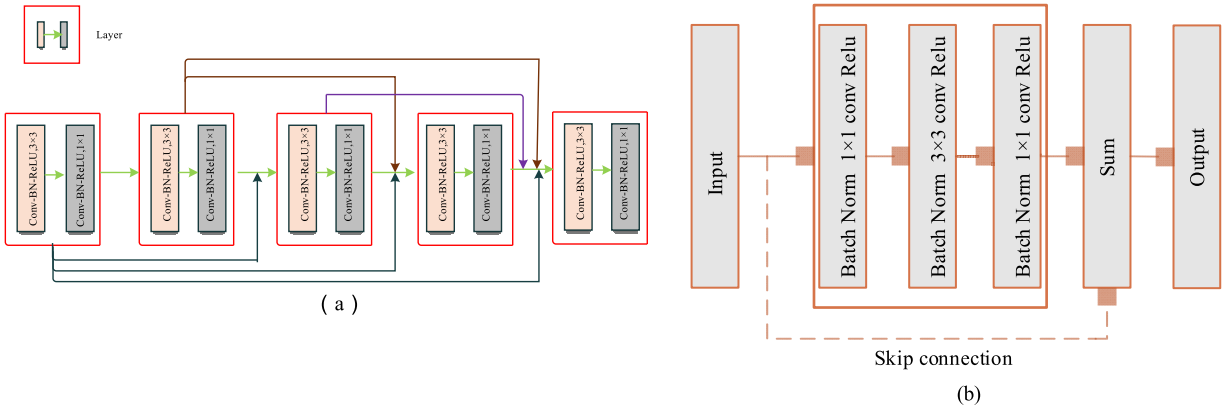

FIGURE 4. The detail of (a) demonstrate a model of the five-layer Dense block that every layer consists of 3 × 3 conv,1 × 1 conv. The model of the Residual Block network is shown (b).

been observed that residual block can well deal with the difficulty of training deep networks. The traditional convolutional feed-forward neural network connects the output of layer l as input to layer $l+1$. The Residual block add a skip connection and bypass the nonlinear transformation through an identical function:

$$x_l = H_l(x_{l-1}) + x_{l-1} \tag{3}$$

Figure 4(b) illustrates the layout of the Residual network. Our segmentation network has five Dense-Residual blocks, and the architecture detail is listed in Table 1.

In the decoding phase, we use a bilinear interpolation upsampling, Dense-Residual block and Skip Connection to recover full image resolution. The Residual block we added consists of a 1×1 convolution, a 3×3 convolution and a 1×1 convolution, as shown in Figure 4(b). A skip connection is added between the layers in the encoder and the corresponding layers in the decoder. The last layer uses the softmax activation function to output the segmentation probability map. The architecture of the segmentation network S is shown in Figure 3. More details about the segmentation network is listed in supplementary table S1.

2) CRITIC NETWORK

In the critic network, we use pixel-level multiplication of the segmentation mask and the original image, multiplication of ground truth and the original image as the input of the critic network. We used convolution and global convolutional interaction substitution to obtain a larger perceptual field. The kernels of convolution and global convolution [10] are 7×7 and 13×13, 5×5 and 11×11, 4×4 and 9×9, 4×4 and 7×7, 4×4 and 5×5 respectively, as shown in Figure 5.

TABLE 1. The structure details of the segmentation network.

3×3 Convolution			
	first step	second step	third step
Encoder 1	Dense block(layers=4), Concatenation	Residual block (1×1,3×3,1×1conv)	Downsampling (2×2 Max pool)
Encoder 2	Dense block(layers=5), Concatenation	Residual block (1×1,3×3,1×1conv)	Downsampling (2×2 Max pool)
Encoder 3	Dense block(layers=7), Concatenation	Residual block (1×1,3×3,1×1conv)	Downsampling (2×2 Max pool)
Encoder 4	Dense block(layers=10), Concatenation	Residual block (1×1,3×3,1×1conv)	Downsampling (2×2 Max pool)
Encoder 5	Dense block(layers=12), Concatenation	Residual block (1×1,3×3,1×1conv)	Downsampling (2×2 Max pool)
Bottleneck	Dense block(layers=15)		
Decoder 1	Upsampling (2×2 stride)	Residual block (1×1,3×3,1×1conv), Concatenation	Skip Connection Dense block (layers=12)
Decoder 2	Upsampling (2×2 stride)	Residual block (1×1,3×3,1×1conv), Concatenation	Skip Connection Dense block (layers=10)
Decoder 3	Upsampling (2×2 stride)	Residual block (1×1,3×3,1×1conv), Concatenation	Skip Connection Dense block (layers=7)
Decoder 4	Upsampling (2×2 stride)	Residual block (1×1,3×3,1×1conv), Concatenation	Skip Connection Dense block (layers=5)
Decoder 5	Upsampling (2×2 stride)	Residual block (1×1,3×3,1×1conv), Concatenation	Skip Connection Dense bloc k(layers=4)
3×3 Convolution			
Softmax			

The convolution blocks in our network architecture use batch normalization and ReLU activation.

It is worth noting that whether it is segmentation or critic network, all blocks after convolution use batch normalization and RELU activation functions.

B. LOSS FUNCTION

We use multiple levels of loss functions to capture the long-term and short-term spatial relationships of pixels. A loss value is calculated for each downsampling, and finally the average of all the loss values will be used as the multi-layer loss. More details can be found in Figure 5. In our proposed method, a dataset containing N training images x and corresponding ground truth value T is given. The multi-scale objective loss function is defined as [33].

$$L_{multi-scale} = \min_{S}\max_{C}\frac{1}{N}\frac{1}{L}\sum_{n=1}^{N}\sum_{i=1}^{L}\|C_i(x_n * T_n) - C_i(x_n * S(x_n)\|_1 \quad (4)$$

where x and T represent original image and ground truth label respectively, S(x) is the segmentation probability map of the input image x generated by S network and ground truth T multiply at the pixel level. L is the number of downsamplings in critic network.

The end-point error is defined as [22].

$$L_{epe} = \sqrt{(S(x)_x - T_x)^2 + (S(x)_y - T_y)^2} \quad (5)$$

It is used for preserving the melanoma boundaries to maximize Peak Signal-to-Noise Ratio. In EPE, the size and orientation of the predicted mask edges is compared with those of ground truth label. $(S(x)_x, S(x)_y)$ and (T_x, T_y) are the first derivatives of S(x) and T in x and y directions, respectively.

Skin tumors typically only account for a small fraction of the entire dermoscopic image in dermoscopy. Since pixel-level classification tends to favor the background, this increases the probability that the tumor will be partially segmented or even missed. So we use the Jaccard distance loss function which is defined as follows [17]:

$$L_{jac} = 1 - \frac{\sum_{i=1}^{N} S(x_i)T_i}{\sum_{i=1}^{N} S(x_i) + \sum_{i=1}^{N} T_i - \sum_{i=1}^{N} S(x_i)T_i} \quad (6)$$

The final loss functions of the S model and C model are as follows:

$$L_s = L_{multi-scale} + \lambda L_{epe} + \beta L_{jac} \quad (7)$$

$$L_c = L_{multi-scale} \quad (8)$$

where lambda and beta are user specified weight parameters.

It is noted that both Segmentation and Critic networks use $L_{multi-scale}$ objective loss component, which is also done by SEGAN [33]. The $L_{multi-scale}$ is used in Segmentation network for deep supervision, which enforces the segmentation result to be similar to the ground truth. The final loss functions of the Segmentation model and Critic model are shown in equation (7) and (8). We first fix the Critic network and use the gradient information of the loss function (7) of the Segmentation network to train the Segmentation network. Then, we fix the Segmentation network and use the gradient information of loss function (8) of the Critic network to train the Critic network. After the training is completed, both the Segmentation and Critic networks will perform well.

IV. EXPERIMENTATION AND RESULTS

In this Section, we investigate the performance of the proposed segmented network. The detailed design of the experiment and the results of the experiment will be presented. The proposed method will be compared with several state-of-the-art segmentation methods. The code is available at https://github.com/tuwenli/skin-lesion.

A. DATASET

We evaluated the performance of our network on a dataset published by ISBI 2017 [25]. The dataset contains 2000 training dermoscopic images and their corresponding lesion masks, validation set contains 300 skin mirror images and

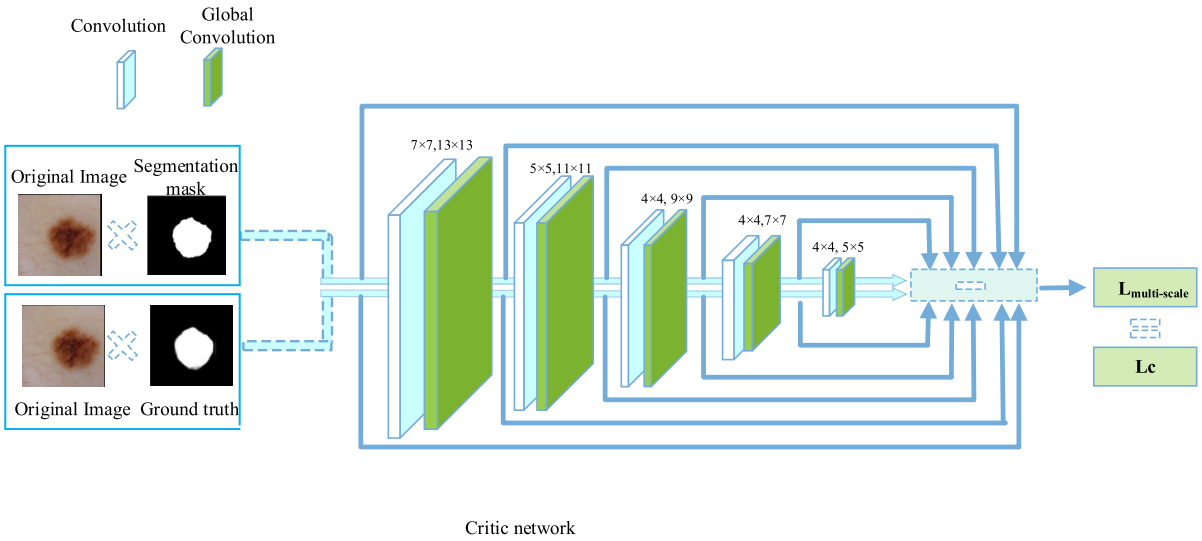

FIGURE 5. The structure of critic network shown above. We use interchangeably convolution and global convolutional five times. Every time calculates a loss value, and the average value is finally taken. The multi-scale loss $L_{multi-scale}$ is used as the loss function of the critic network.

their corresponding segmentation masks and the test set contains 600 skin mirror images and their corresponding segmentation masks. The sizes of these images are different, but they have a height to width ratio of 4/3. We scale them to a length in the range of 180 to 135, and randomly cut them into 128 × 128 squares during the training. The reason for this size is to take into account the memory usage and the speed of training. In order to augment the data, we also flip the training image horizontally and vertically. We have also evaluated our method on PH2 dataset [26] that comes from Dermatology Service of Hospital Pedro Hispano, Matosinhos, Portugal Mendonç a et al. with Tuebinger Mole Analyzer system. It contains 200 skin lesion dermoscopic images with the resolution of 768 × 560 pixels.

B. PERFORMANCE EVALUATION

The output of trained Dense-Residual segmentation network model is binarized to a lesion mask to compare with the ground truth provided by clinicians. We use Jaccard Coefficient (JAC), Dice Similarity Coefficient (DSC), accuracy (ACC), sensitivity (SE), specificity (SP) to evaluate the performance of the network. The evaluation metrics are defined as follows:

$$\mathrm{JAC} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN} + \mathrm{FP}} \tag{9}$$

$$DSC = \frac{2 \times TP}{2 \times TP + FN + FP} \tag{10}$$

$$\mathrm{ACC} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{FP} + \mathrm{TN} + \mathrm{FN}} \tag{11}$$

$$\mathrm{SE} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{12}$$

$$\mathrm{SP} = \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FP}} \tag{13}$$

where TP, FP, FN denote the number of true positives, false positives and false negatives respectively.

C. IMPLEMENTATION DETAILS

We use the Adam optimizer with a batch size of 10 to train the Segmentation and Critic network. The initial learning rate is set to 0.0002. The number of iterations is set to 500, and after every 25 epoch learning rates is decreased at a decay rate of 0.5 until 1e × 10^{-6} is reached. The values of λ and β in (7) are set as 0.5. We set a threshold to produce the final binary mask. We evaluated and compared our proposed approach with the 600 skin images test set. The entire network architecture is implemented with pytorch. We use Nvidia GeForce GTX 1080Ti GPU with 11 GB GDDR5X memory for the training.

D. ABLATION EXPERIMENTS

The segmentation results of our proposed method on some test images of the ISBI2017 are shown in Figure 6. These columns represent the segmentation results of skin images with different lesion conditions, such as benign and melanoma. We analyzed the effects of some key factors in the proposed the adversarial network on skin lesion segmentation performance. The factor includes Dense- Residual block encoder-decoder architecture in segmentation network. For the loss function, the factors include Multi-scale objective Loss function ($L_{multi-scale}$), EPE (L_{epe}) and Jaccard distance (L_{jac}) content Loss function. We denote the encoder-decoder architecture with only DenseNet block in segmentation network with $L_{multi-scale}$ as baseline 1. The baseline 2 is the baseline 1 integrated with L_{epe} and L_{jac}. The baseline 3 is Dense-Residual block encoder-decoder architecture in segmentation network with $L_{multi-scale}$. Our proposed method is the baseline 3 integrated with L_{epe} and L_{jac}. We replace one

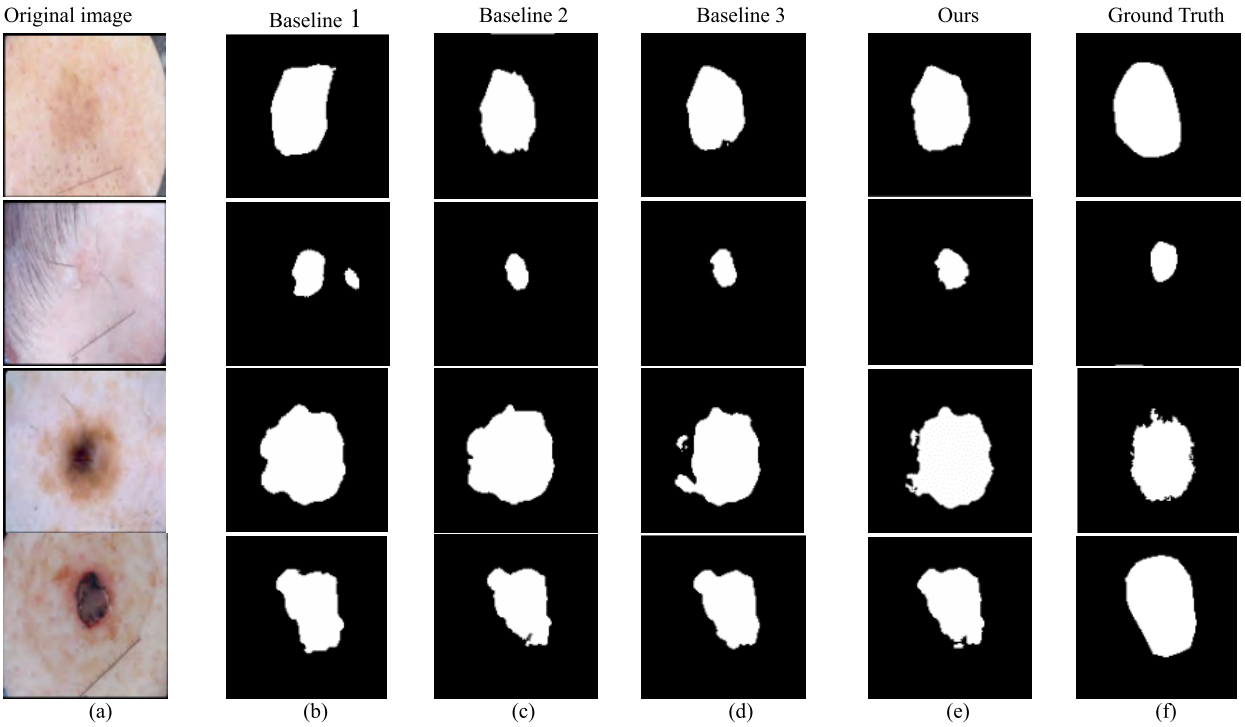

FIGURE 6. Results on ISBI 2017 (a) Original image, (b) Baseline 1, (c) Baseline 2, (d) Baseline 3, (e) Ours and (f) Ground Truth. First two rows of skin lesion images are benign while last two rows of skin lesion images are melanoma.

TABLE 2. Ablation experimentation results of the proposed method on ISBI2017 dataset.

	DenseNet block	Dense-Residual block	$L_{multi\text{-}scale}$	L_{epe} and L_{jac}
Baseline 1	Yes	No	Yes	No
Baseline 2	Yes	No	Yes	Yes
Baseline 3	No	Yes	Yes	No
(our proposed)	No	Yes	Yes	Yes

TABLE 3. The experimental results of Baseline 1, Baseline 2, Baseline 3 and Ours. The mean (standard deviation) values are reported.

Model	JAC	DSC	ACC
Baseline 1	0.726 (0.184)	0.824 (0.154)	0.910 (0.181)
Baseline 2	0.734 (0.191)	0.835 (0.152)	0.921 (0.176)
Baseline 3	0.751 (0.182)	0.846 (0.161)	0.937 (0.167)
Ours	**0.768** (0.176)	**0.862** (0.149)	**0.945** (0.178)

component with another, while leaving the others unchanged. A summary can be seen in Table 2. The experimental results are listed in Table 3, and some of the segmentation probability are shown in Figure 6.

It can be seen from the table 3 that the JAC, DSC and ACC indexes of ours improve 0.042,0.038 and 0.035 than those of Baseline 1 respectively, indicating that the added residual blocks, EPE and Jaccard distance content loss function improve the accuracy and efficiency of the model. The detailed analysis is as follows 1) and 2).

1) COMPARE WITH OR WITHOUT RESIDUAL BLOCK NETWORK

We used Dense-Residual blocks in our model, and we have also investigated the performance without Residual structure (Baseline 2). Densenet can obtain a deeper feature map of the input image and ensure maximum information transmission between the layers in the network. However, the experimental results are not good. We found that when the number of network layers reaches some level, the performance of the network will be saturated. If the number of layers of the network is increased further, its performance will begin to degenerate. This shows that when the network becomes deep, the deep network becomes difficult to train. In order to prevent the degeneration of segmentation performance, we add a residual block after each dense block. The experimental results have been greatly improved. the results on ISBI2017 are shown in Figure 7. The confidence interval for each indicator in 600 test images are also shown.

It can be seen in Table 3 and Figure 7 that after adding the residual layers, performance can be improved significantly (from 73.4% to 76.8% for JAC). In the proposed method, the number of layers and the number of features is gradually increasing in Dense block. We add residual block with identity mapping, and the skip connection of the residual network enables gradients to be propagated directly

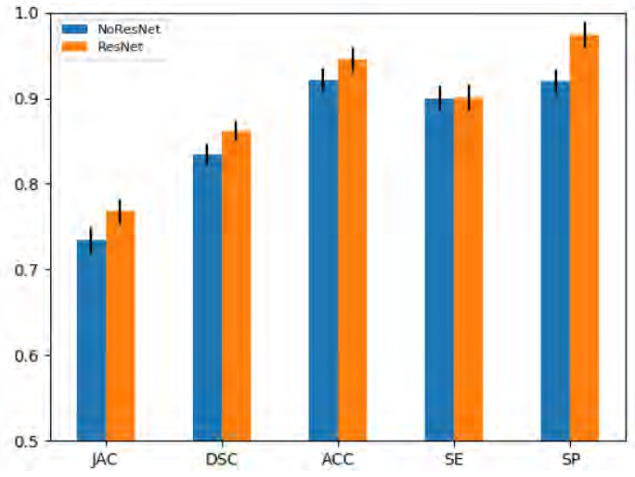

FIGURE 7. Results of the segmentation network with (ours) or without (Baseline 2) residual block on the SP, SE, ACC, JAC and DSC measurements. The confidence interval for each indicator is also shown.

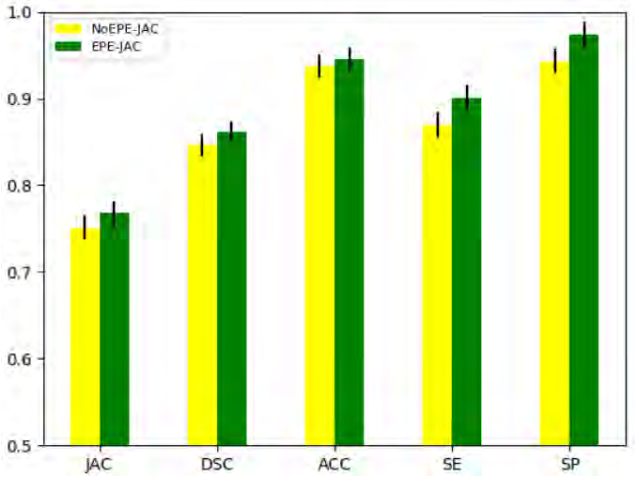

FIGURE 8. Results with (Ours) and without (Baseline 3) content loss function L_{epe} and L_{jac} on the SP, SE, ACC, JAC, and DSC evaluation metrics. The confidence interval for each indicator is also shown.

in the network. It avoids exploding gradient problem or the vanishing gradient problem [11]. In addition, we have also appropriately reduced the number of feature maps through the residual block. It greatly reduces the difficulty of training. Please refer to the supplementary table S1. This shows that when the network layer is deeper and the network training is difficult, adding the residual network can alleviate this problem and improve the accuracy of skin lesion segmentation.

2) COMPARE WITH OR WITHOUT CONTENT LOSS FUNCTION L_{epe} and L_{jac}

In order to investigate the effectiveness of our loss function and solve the various challenges in skin lesion segmentation, we compared with or without L_{epe} and L_{jac} used in Generative Adversarial Network. One has only objective loss function $L_{multi-scale}$(Baseline 3), while the other has $L_{multi-scale,}$ L_{epe} and L_{jac}(Ours).

The results are shown in Figure 8 and Table 3. By integrating the L_{epe} and L_{jac} content loss function, the results are improved on DSC, JAC, ACC, SE and SP respectively. The confidence interval for each evaluation metric in 600 test images are also shown in the Figure 8. The main reason is that our loss function not only implements deep supervision in multiple layers, but also uses L_{epe} function, which is conducive to establishing a direct relationship between adjacent pixels and solving the problem of boundary ambiguity. Moreover, most of our segmented skin lesion only account for a portion of the entire picture, and the Jaccard distance is more relevant to the segmentation task of the image. The traditional cross-entropy loss function cannot deal with the imbalance problem of foreground and background pixel, while the Jaccard distance is primarily concerned with the foreground pixels. The segmentation results of skin lesion show that we can obtain more promising semantics and more accurate segmentation results by integrating content loss function L_{epe} and L_{jac}.

TABLE 4. Averaged *JAC* results of the proposed Method with different weights for segmentation network loss function.

b \ λ	0.1	0.5	1	10
0.1	0.732	0.741	0.734	0.726
0.5	0.756	0.770	0.753	0.737
1	0.747	0.762	0.752	0.729
10	o.738	0.745	0.734	0.716

3) EFFECTS OF WEIGHT PARAMETERS ON LOSS FUNCTION

The loss function of the segmentation network in Equation (7) consists of three components. The final loss function is obtained by integrating the three components with the corresponding weight parameters. Although in theory, they have different value ranges ($L_{multi-scale,}$ is in $[0, +\infty)$. L_{jac} is in [0, 1], L_{epe} is in $[0, 2\sqrt{2}]$, we find their values are in similar scales. We get the experimental results on the validation set. The investigated value of weights λ and β for L_{epe} and L_{jac} are [0.1,0.5,1,10], and the results are listed in Table 4. From the table, we can see that when their values are in a reasonable range, the influence on the results is small. The best result is obtained when λ and β are set as 0.5, 0.5.

4) COMPARE DIFFERENT NUMBERS OF DENSE-RESIDUAL BLOCK

In order to investigate the influence of the number of layers on the performance, we conducted experiments to compare the different number of cascade Dense-Residual blocks. Table 5 summarizes several configurations, we have conducted experiments with 2-6 Dense-residual-blocks, respectively. As can be seen from Table 5, the segmentation results have improved as increasing the number of layers in the Segmentation network at the beginning. But when we increase

TABLE 5. Result of different numbers of Dense-Residual block. The mean (standard deviation) values are reported.

Model	JAC	DSC	ACC	SE	SP
Two Dense-residual-blocks	0.704 (0,196)	0.810 (0.152)	0.901 (0.181)	0.920 (0.194)	0.912 (0.181)
three Dense-residual-blocks	0.731 (0.198)	0.821 (0.169)	0.910 (0.187)	0.931 (0.173)	0.923 (0.187)
four Dense-residual-blocks	0.742 (0.182)	0.846 (0.158)	0.923 (0.167)	0.932 (0.190)	0.934 (0.175)
five Dense-residual-blocks	**0.768** (0.176)	**0.862** (0.149)	**0.945** (0.178)	**0.901** (0.181)	**0.974** (0.184)
six Dense-residual-blocks	0.751 (0.202)	0.840 (0.179)	0.930 (0.186)	0.927 (0.196)	0.943 (0.197)

TABLE 6. Performance comparison between proposed segmentation methods and other state-of-the-art methods in the ISBI 2017 Challenge. Note that SMCP AND DDCA are not deep learning based methods.

Method	JAC	DSC	ACC	no. parameter
DDCA[43]	0.718	0.810	0.922	NA
FCN[19]	0.721	0.827	0.923	18.7M
UNET[28]	0.731	0.834	0.928	9.9M
SMCP[42]	0.749	0.839	0.930	NA
RESNET[39]	0.758	0.842	0.934	17.8M
SEGAN[33]	0.762	0.851	0.941	103.2M
ECDN[17]	0.765	0.849	0.934	23.8M
Our proposed method	**0.768**	**0.862**	**0.945**	**69.8M**

the number of Dense-Residual block up to 6, the performance starts to decreases. This may be because too many layers lead to overfitting. We found that Dense-Residual block has the best result when its number is 5. According to the experimental results, we set the number of Dense-Residual blocks to 5.

E. COMPARISON WITH OTHER METHODS

The proposed skin lesion segmentation method was compared to other methods on the ISIB2017 dataset. We report three related indicators JAC, DSC and ACC. Table 6 and Figure 9 show the segmentation results of different methods. The proposed segmentation method obtained 76.8% in JAC and 86.2% in DSC. It is superior to other existing segmentation methods, and the Jaccard score is increased from 76.2% to 76.8%. This is mainly due to the proposed Dense-Residual blocks in segmentation network and the novel loss function. Its feed-forward method connects each layer with each subsequent layer, and allows subsequent layers to bypass the feature layer that enhance the propagation of features. Our loss function can strengthen the deep supervision, make the fuzzy boundary clearer and focus more on foreground pixel. Therefore, the proposed method can obtain high precision in pixel classification and accurate segmentation results, and is better than other segmentation methods. The segmentation results of our method and other methods on several typical images are shown in Figure 10.

FIGURE 9. It compares our method result with others method on the ACC, JAC, and DSC evaluation indicators.

F. EVALUATION ON PH2 DATABASE

To further evaluate the performance of our proposed model, we performed the following experiment on the separate dermoscopic image PH2 database [26], which is widely used in algorithm validation and benchmarking. We finetuned the weights of our model by further training the network with PH2 data. The setting of experiments is similar to [44], the PH2 database was randomly divided into two groups (A and B). Each group includes 100 images. In group A, we trained our method with 50 images and then validated with the rest 50 images. We set the total number of epochs to 200. The model with the best performance of the validation image in group A is selected and used to perform image segmentation in group B. The segmentation results in group B as shown in Figure 11.

We compared the performance of our proposed method with other state-of-the-art methods on the PH2 dataset. The results are listed in Table 7. It can be seen from the experimental results that our method achieved best performance in dataset PH2 and is superior to the existing segmentation methods in the literature. Our JAC index is also higher than the results on ISBI2017. One reason is that the PH2 dataset is not as complex as the ISBI2017 dataset. It has not so much size, shape, color, and texture of melanoma vary greatly among different types of skin. Another reason is that in our method, different from the traditional adversarial network model, we not only used Dense-Residual block Network in the encoder-decoder segmentation model, but also used a novel loss function which can strengthen the deep supervision, make the fuzzy boundary clearer and focus more on foreground pixel. Figure 11 shows some typical segmentation results on PH2.

G. STATISTICAL SIGNIFICANCE ANALYSIS OF EXPERIMENTAL RESULTS

Statistical significance of the differences was determined using 2-tailed paired t-test for which p value of 0.05 is considered significant. On the ISBI2017 dataset, the mean (standard deviation) of JAC and DSC of our method is 0.768 (0.176) and 0.862 (0.149), respectively, while SEGAN is 0.762 (0.203)

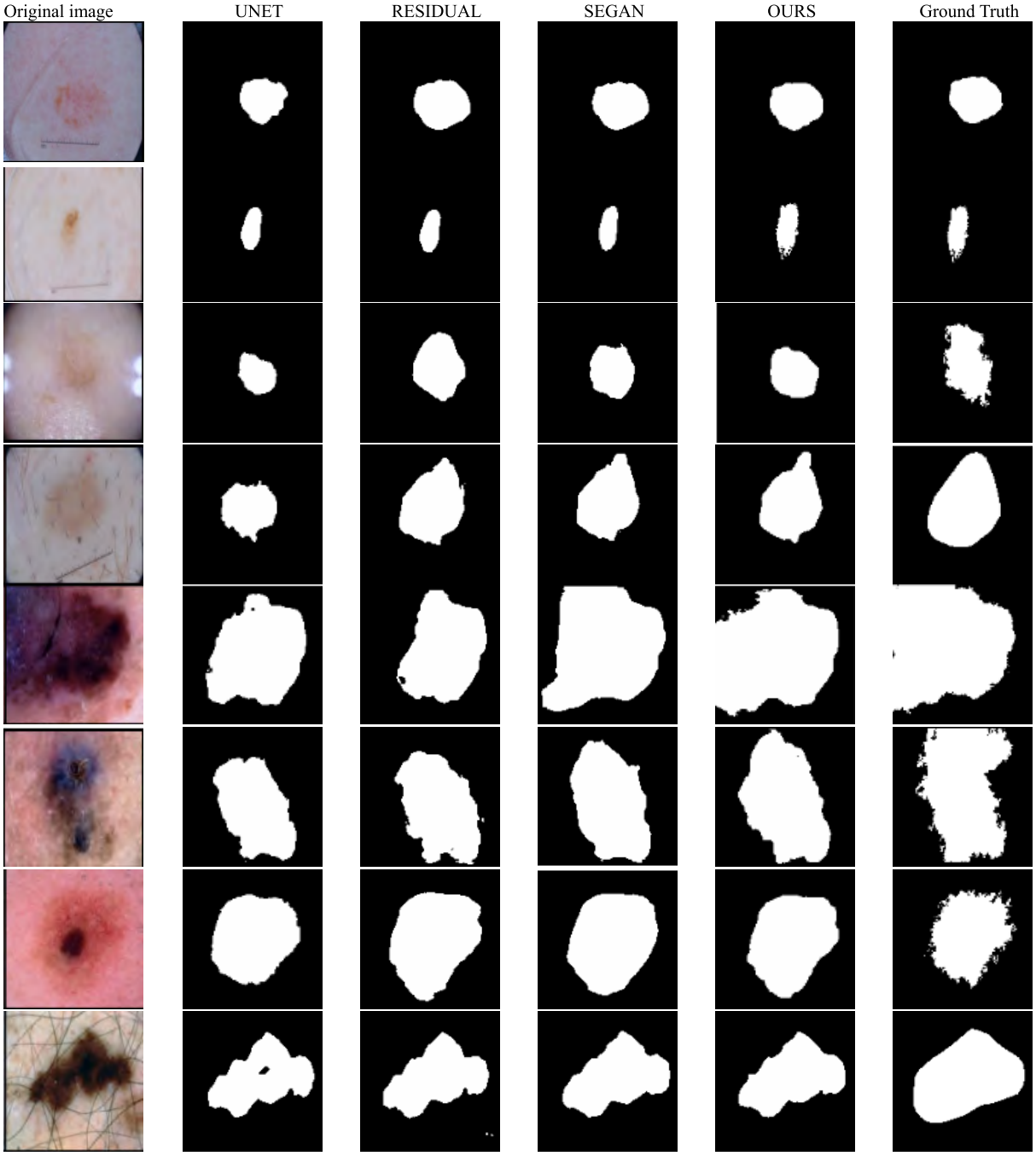

FIGURE 10. Segmentation results by Original image, UNET, RESIDUAL, SEGAN, proposed method, and Ground Truth. First four rows of skin lesion images are benign while last four rows of skin lesion images are melanoma.

and 0.851 (0.176). Due to the relatively large standard deviation, paired t-test analysis shows that the differences between our proposed method and the SEGAN for JAC and DSC are not statistically significant ($p > 0.05$).

The main reason for the large standard deviation in our analysis is the particularity of the dataset. In the ISBI2017 dataset, there are some images with very low contrast between lesion and the surrounding skin, some with irregular and fuzzy borders, and some with great hairs artifact. The segmentation results on some of these images are not optimal, resulting in lower JAC and DSC values. On other images, the proposed method can achieve great results. Figure 12 shows the results on several typical images. The standard deviations on the JAC and DSC of both the proposed our method and SEGAN are relatively large. It is very challenging to segment accurately all the difficult images.

Meanwhile, we also performed the statistical analysis on the PH2 dataset. The mean (standard deviation) of JAC

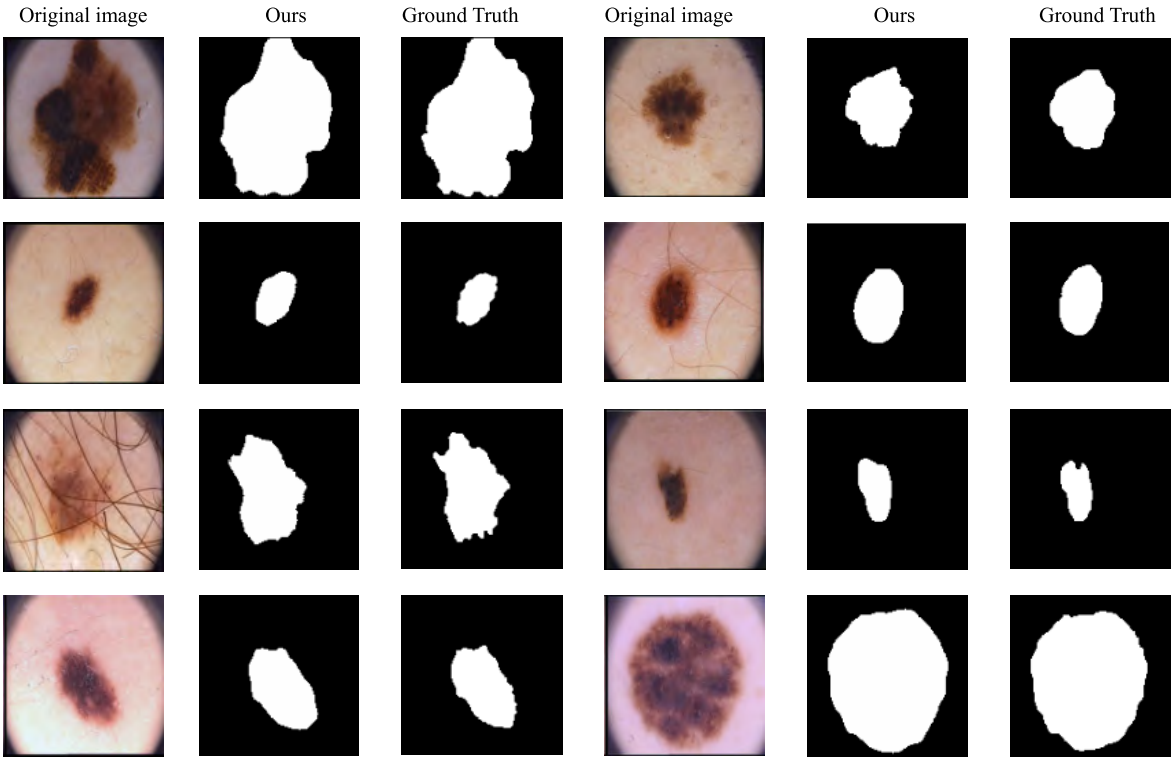

FIGURE 11. Original image, segmentation results our method and Ground Truth on PH2 dataset from the left column to the right column. Each row has two sample case.

TABLE 7. Performance comparison between the proposed segmentation and other state-of-the-art methods on PH2 dataset.

Method	JAC	DSC
Adaptive Thresholding in Silveira et al.[5]	72.4%	80.4%
Level Set Active Contours in Li et al. [45]	76.3%	83.5%
Multi-Scale Segmentation in Bozorgtabar et al. [46]	76.0%	86.1%
Deep FCN in Yuan et al. [44]	81.5%	91.5%
Densely Linked CNN in Bagher et al. [47]	85.3%	91.5%
Step-wise integration in Bi et al.[48]	85.9%	92.1%
SEGAN [33]	85.1%	92.4%
Ours	**86.3%**	**93.2%**

and DSC of SEGAN on the PH2 dataset is 0.851(0.096) and 0.924(0.082), respectively. The mean (standard deviation) of JAC and DSC of our method is 0.863(0.079) and 0.932(0.054), respectively. Paired t-test analysis shows that the differences between our proposed method and the SEGAN are statistically significant for JAC and DSC ($p < 0.05$) on PH2 dataset. The reason is that the PH2 dataset is not as complex as the ISBI2017 dataset. It has not so much size, shape, color, or texture variation among different types of skin.

V. DISCUSSION

One of the key steps in computerized recognition of melanoma on dermoscopic images is to depict lesions from surrounding skin areas. However, this step is very challenging because the size, shape, color, and texture of melanoma vary greatly among different types of skin. Some artificial factors, such as hair on the skin, pigmentation inconsistency, blood vessels, air bubbles, etc., will further increase the difficulty of segmentation tasks.

In this work, we proposed a novel adversarial network architecture for the segmentation of skin lesions, which performed better than other methods on both the ISBI2017 and the PH2 dataset. Although traditional GANs have been successfully applied in unsupervised and semi-supervised learning tasks, the application of semantic segmentation is few. We proposed Dense-Residual block which is a key component in our method for segmentation tasks. When information about an input or gradient pass through many network layers, problems like the vanishing gradient and exploding may become more prominent. To solve the problem, the segmentation network uses Dense-Residual block which can strengthen feature propagation, encourage feature reuse and help multi-scale feature mapping from different layers. GAN is generally difficult to train and training is unstable. Although Wasserstein GAN (WGAN) [49] has made progress by clipping weight in this regard. But the generated samples are poor or not suitable for segmentation. We use multi-scale loss to solve the problem. EPE is used to establish a direct relationship between adjacent pixels and solving the problem of boundary ambiguity. In addition, we add the Jaccard distance into the loss function of the segmentation network to distinguish foreground and background pixel. Compared

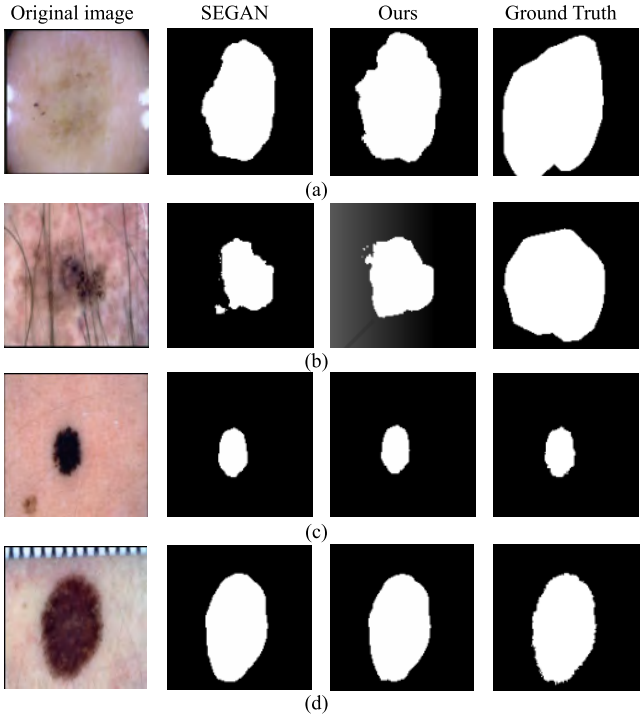

FIGURE 12. Segmentation results by Original image, SEGAN, proposed method and Ground Truth. First two rows of skin lesion images are (a) low contrast between lesion and the surrounding skin and (b) hair on the skin. They get failure segmentation results. Last two rows (c) and (d) of skin lesion images do not have these conditions and get better segmentation results.

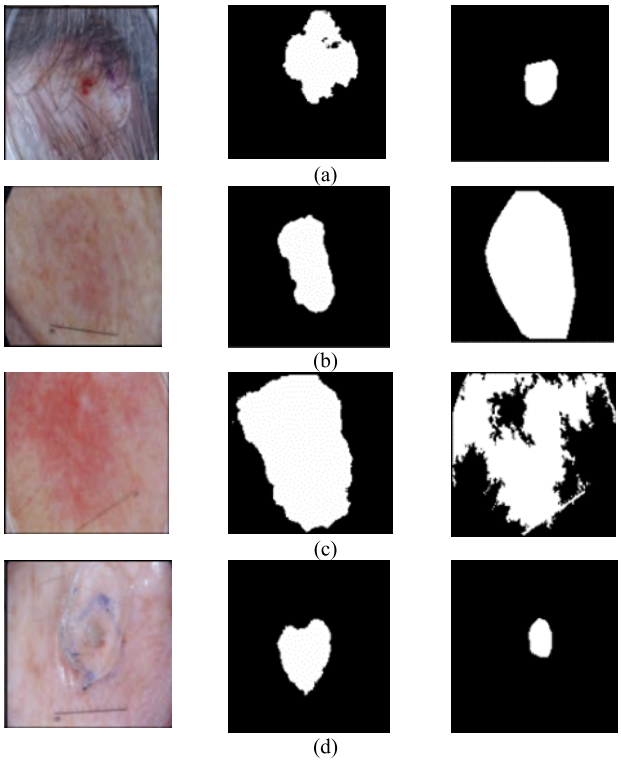

FIGURE 13. A few failure cases of our proposed method. The left column is the original image, the middle column is the segmentation probability image of our proposed method and the rightmost column is ground truth. (a) the skin lesion is surrounded by a lot of hair, (b) the contrast between the lesion area and surrounding skin is too low, (c) the skin boundary of the lesion is very blurred, and (d) the surrounding skin lesion is full of air bubbles.

with the cross entropy, it is more related to the task of image segmentation and improves the segmentation performance. To further verify the generalization ability of our model, we test it in the PH2 database which only has 200 skin lesion images. The performance of our proposed method is also better than other state-of-the-art methods. Our method uses the original image as input, the neural network can directly learn the features. Due to the affection of several images with very low contrast and great artifact in ISBI2017 dataset, the standard deviation of performance criterion is large, and the difference between results of our method and SEGAN are not statistically significant. The difference between results of our method and SEGAN are statistically significant on the PH2 dataset.

In most of the two independent databases, our model achieved good segmentation accuracy. But under some conditions of image acquisition, our model needs further improvement. Figure 13 shows some failure cases of the proposed method. the segmentation performance of our proposed method needs further improvement. Better network architecture and more effective training strategies require further research in future work. Moreover, the deep learning model is combined with traditional image segmentation methods active contours, and the recently hot Recurrent Neural Networks (RNNs) model is applied to capture context information.

VI. CONCLUSIONS

We propose a Dense-Residual block and adversarial learning based on deep convolutional neural network for skin lesion segmentation on dermoscopic images in this paper. Dense block obtains a deeper feature map of the input image and ensure maximum information transmission between the layers in the network. When increasing the number of layers of the network, its performance will begin to degenerate and lead to overfitting. Residual block is used to solve the problem. We add a multi-scale loss and EPE loss for deep supervision and make the fuzzy boundary clearer to further boost the segmentation performance and stability. At the same time, we added a Jaccard distance to the loss function of segmentation network. Our loss function directly maximizes the overlap between the foreground of ground truth and the foreground of the prediction partition mask, and it can eliminate the need for background rebalancing when the number of foregrounds is not well balanced. Compared with the traditional cross entropy loss, our loss function not only improves segmentation accuracy, but improves the stability of network. The experiment results on ISBI2017 and PH2 datasets show that the proposed method outperforms the state-of-the-art skin lesion on dermoscopic images segmentation methods. The differences between results of our method and SEGAN

are not statistically significant on the ISBI2017 dataset, while they are statistically significant on the PH2 dataset.

In our future work, we try to apply other advanced segmentation network architectures [50], [51] for skin lesion segmentation. We also try to investigate the extension of proposed method to other medical image segmentation tasks.

REFERENCES

[1] C. M. Balch, J. E. Gershenwald, S.-J. Soong, J. F. Thompson, M. B. Atkins, D. R. Byrd, A. C. Buzaid, A. J. Cochran, D. G. Coit, S. Ding, A. M. Eggermont, K. T. Flaherty, P. A. Gimotty, J. M. Kirkwood, K. M. McMasters, M. C. Mihm, Jr., D. L. Morton, M. I. Ross, A. J. Sober, and V. K. Sondak, "Final version of 2009 AJCC melanoma staging and classification," *J. Clin. Oncol.*, vol. 27, no. 36, pp. 6199–6206, Dec. 2009.

[2] K. Korotkov and R. Garcia, "Computerized analysis of pigmented skin lesions: A review," *Artif. Intell. Med.*, vol. 56, no. 2, pp. 69–90, Oct. 2012.

[3] M. E. Celebi, H. A. Kingravi, B. Uddin, H. Iyatomi, Y. A. Aslandogan, W. V. Stoecker, and R. H. Moss, "A methodological approach to the classification of dermoscopy images," *Computerized Med. Imag. Graph.*, vol. 31, no. 6, pp. 362–373, Sep. 2007.

[4] H. Ganster, P. Pinz, R. Rohrer, E. Wildling, M. Binder, and H. Kittler, "Automated melanoma recognition," *IEEE Trans. Med. Imag.*, vol. 20, no. 3, pp. 233–239, Mar. 2001.

[5] M. Silveira, J. C. Nascimento, J. S. Marques, A. R. S. Marcal, T. Mendonca, S. Yamauchi, J. Maeda, and J. Rozeira, "Comparison of segmentation methods for melanoma diagnosis in dermoscopy images," *IEEE J. Sel. Topics Signal Process.*, vol. 3, no. 1, pp. 35–45, Feb. 2009.

[6] M. E. Celebi, H. Iyatomi, G. Schaefer, and W. V. Stoecker, "Lesion border detection in dermoscopy images," *Computerized Med. Imag. Graph.*, vol. 33, no. 2, pp. 148–153, Mar. 2009.

[7] X. Liu, T. Fu, Z. Pan, D. Liu, W. Hu, and B. Li, "Semi-supervised automatic layer and fluid region segmentation of retinal optical coherence tomography images using adversarial learning," in *Proc. 25th IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2018, pp. 2780–2784.

[8] G. Capdehourat, A. Corez, A. Bazzano, R. Alonso, and P. Musé, "Toward a combined tool to assist dermatologists in melanoma detection from dermoscopic images of pigmented skin lesions," *Pattern Recognit. Lett.*, vol. 32, no. 16, pp. 2187–2196, Dec. 2011.

[9] R. Hettiarachchi and J. F. Peters, "Multi-manifold-based skin classifier on feature space Voronoï regions for skin segmentation," *J. Vis. Commun. Image Represent.*, vol. 41, pp. 123–139, Nov. 2016.

[10] Y. Li and L. Shen, "Skin lesion analysis towards melanoma detection using deep learning network," *Sensors*, vol. 18, no. 2, p. 556, 2018.

[11] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 770–778.

[12] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 779–788.

[13] H. Noh, S. Hong, and B. Han, "Learning deconvolution network for semantic segmentation," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2015, pp. 1520–1528.

[14] Z. Yan, Y. Zhan, Z. Peng, S. Liao, Y. Shinagawa, S. Zhang, D. N. Metaxas, and X. S. Zhou, "Multi-instance deep learning: Discover discriminative local anatomies for bodypart recognition," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1332–1343, May 2016.

[15] S. Miao, Z. J. Wang, and R. Liao, "A CNN regression approach for real-time 2D/3D registration," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1352–1363, May 2016.

[16] X. Liu, T. Fu, Z. Pan, D. Liu, W. Hu, J. Liu, and K. Zhang, "Automated layer segmentation of retinal optical coherence tomography images using a deep feature enhanced structured random forests classifier," *IEEE J. Biomed. Health Inform.*, to be published.

[17] Y. Yuan and Y.-C. Lo, "Improving dermoscopic image segmentation with enhanced convolutional-deconvolutional networks," *IEEE J. Biomed. Health Inform.*, vol. 23, no. 2, pp. 519–526, Mar. 2019.

[18] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, pp. 115–118, Feb. 2017.

[19] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 3431–3440.

[20] P. Luc, C. Couprie, S. Chintala, and J. Verbeek, "Semantic segmentation using adversarial networks," Nov. 2016, *arXiv:1611.08408*. [Online]. Available: https://arxiv.org/abs/1611.08408

[21] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. CVPR*, Jul. 2017, pp. 2261–2269.

[22] M. M. K. Sarker, H. A. Rashwan, F. Akram, S. F. Banu, A. Saleh, V. K. Singh, F. U. H. Chowdhury, S. Abdulwahab, S. Romani, P. Radeva, and D. Puig, "SLSDeep: Skin lesion segmentation based on dilated residual and pyramid pooling networks," May 2018, *arXiv:1805.10241*. [Online]. Available:https://arxiv.org/abs/1805.10241

[23] C. Peng, X. Zhang, G. Yu, G. Luo, and J. Sun, "Large kernel matters—Improve semantic segmentation by global convolutional network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1743–1751.

[24] X. Liu, J. Cao, T. Fu, Z. Pan, W. Hu, K. Zhang, and J. Liu, "Semi-supervised automatic segmentation of layer and fluid region in retinal optical coherence tomography images using adversarial learning," *IEEE Access*, vol. 7, pp. 3046–3061, 2018.

[25] N. C. Codella, D. Gutman, M. E. Celebi, B. Helba, M. A. Marchetti, S. W. Dusza, A. Kalloo, K. Liopyris, N. Mishra, H. Kittler, and A. Halpern, "Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (ISBI), hosted by the international skin imaging collaboration (ISIC)," in *Proc. IEEE 15th Int. Symp. Biomed. Imag.*, Apr. 2018, pp. 168–172.

[26] T. Mendonça, P. M. Ferreira, J. S. Marques, A. R. S. Marcal, and J. Rozeira, "PH^2-a dermoscopic image database for research and benchmarking," in *Proc. 35th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2013, pp. 5437–5440.

[27] C. Farabet, C. Couprie, L. Najman, and Y. LeCun, "Learning hierarchical features for scene labeling," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 1915–1929, Aug. 2013.

[28] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, Nov. 2015, pp. 234–241.

[29] J. Fan, X. Cao, P.-T. Yap, and D. Shen, "BIRNet: Brain image registration using dual-supervised fully convolutional networks," Feb. 2018, *arXiv:1802.04692*. [Online]. Available: https://arxiv.org/abs/1802.04692

[30] M. Drozdzal, E. Vorontsov, G. Chartrand, S. Kadoury, and C. Pal, "The importance of skip connections in biomedical image segmentation," in *Deep Learning and Data Labeling for Medical Applications*. Cham, Switzerland: Springer, 2016, pp. 179–187.

[31] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834–848, Apr. 2018.

[32] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 2672–2680.

[33] Y. Xue, T. Xu, H. Zhang, L. R. Long, and X. Huang, "SegAN: Adversarial network with multi-scale L_1 loss for medical image segmentation," *Neuroinformatics*, vol. 16, nos. 3–4, pp. 383–392, Oct. 2018.

[34] F. Peruch, F. Bogo, M. Bonazza, V.-M. Cappelleri, and E. Peserico, "Simpler, faster, more accurate melanocytic lesion segmentation through MEDS," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 2, pp. 557–565, Feb. 2014.

[35] H. Zhou, X. Li, G. Schaefer, M. E. Celebi, and P. Miller, "Mean shift based gradient vector flow for image segmentation," *Comput. Vis. Image Understand.*, vol. 117, no. 9, pp. 1004–1016, Sep. 2013.

[36] A. R. Sadri, M. Zekri, S. Sadri, N. Gheissari, M. Mokhtari, and F. Kolahdouzan, "Segmentation of dermoscopy images using wavelet networks," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 4, pp. 1134–1141, Apr. 2013.

[37] F. Xie and A. C. Bovik, "Automatic segmentation of dermoscopy images using self-generating neural networks seeded by genetic algorithm," *Pattern Recognit.*, vol. 46, no. 3, pp. 1012–1019, Mar. 2013.

[38] M. E. Celebi, Q. Wen, H. Iyatomi, K. Shimizu, H. Zhou, and G. Schaefer, "A state-of-the-art survey on lesion border detection in dermoscopy images," in *Dermoscopy Image Analysis*. Boca Raton, FL, USA: CRC Press, 2015, pp. 97–129.

[39] L. Bi, J. Kim, E. Ahn, and D. Feng, "Automatic skin lesion analysis using large-scale dermoscopy images and deep residual networks," Mar. 2017, *arXiv:1703.04197*. [Online]. Available: https://arxiv.org/abs/1703.04197
[40] J. M. Wolterink, K. Kamnitsas, C. Ledig, and I. Išgum, "Generative adversarial networks and adversarial methods in biomedical image analysis," Oct. 2018, *arXiv:1810.10352*. [Online]. Available: https://arxiv.org/abs/1810.10352
[41] S. Jégou, M. Drozdzal, D. Vazquez, A. Romero, and Y. Bengio, "The one hundred layers Tiramisu: Fully convolutional densenets for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jul. 2017, pp. 1175–1183.
[42] M. Jahanifar, N. Z. Tajeddin, B. M. Asl, and A. Gooya, "Supervised saliency map driven segmentation of lesions in dermoscopic images," *IEEE J. Biomed. Health Inform.*, vol. 23, no. 2, pp. 509–518, 2019.
[43] A. Galdran, A. Alvarez-Gila, M. L. Meyer, C. L. Saratxaga, T. Araújo, E. Garrote, G. Aresta, P. Costa, A. M. Mendonça, and A. C. Campilho, "Data-driven color augmentation techniques for deep skin image analysis," 2017, *arXiv:1703.03702*. [Online]. Available: https://arxiv.org/abs/1703.03702
[44] Y. Yuan, M. Chao, and Y.-C. Lo, "Automatic skin lesion segmentation using deep fully convolutional networks with Jaccard distance," *IEEE Trans. Med. Imag.*, vol. 36, no. 9, pp. 1876–1886, Sep. 2017.
[45] C. Li, C.-Y. Kao, J. C. Gore, and Z. Ding, "Minimization of region-scalable fitting energy for image segmentation," *IEEE Trans. Image Process.*, vol. 17, no. 10, pp. 1940–1949, Oct. 2008.
[46] B. Bozorgtabar, M. Abedini, and R. Garnavi, "Sparse coding based skin lesion segmentation using dynamic rule-based refinement," in *Proc. Int. Workshop Mach. Learn. Med. Imag.* Cham, Switzerland: Springer, Oct. 2016, pp. 254–261.
[47] S. B. Salimi, S. Bozorgtabar, P. Schmid-Saugeon, H. K. Ekenel, M. S. Rad, and J.-P. Thiran, "DermoNet: Densely linked convolutional neural networkfor efficient skin lesion segmentation," Ecole Polytechnique Federale Lausanne, Lausanne, Switzerland, Tech. Rep., Nov. 2018.
[48] L. Bi, J. Kim, E. Ahn, A. Kumar, D. Feng, and M. Fulham, "Step-wise integration of deep class-specific learning for dermoscopic image segmentation," *Pattern Recognit.*, vol. 85, pp. 78–89, Jan. 2019.
[49] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein GAN," Jan. 2017, *arXiv:1701.07875*. [Online]. Available: https://arxiv.org/abs/1701.07875
[50] A. Graves, A.-R. Mohamed, and G. Hinton, "Speech recognition with deep recurrent neural networks," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, May 2013, pp. 6645–6649.
[51] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, Sep. 2014, pp. 346–361.

WENLI TU is currently pursuing the master degree with the College of Computer Science and Technology, Wuhan University of Science and Technology, Wuhan, China. Her research interests include medical image processing and machine learning.

XIAOMING LIU received the Ph.D. degree from Zhejiang University, China, in 2007. From 2014 to 2015, he was a Visiting Scholar with the University of North Carolina at Chapel Hill, NC, USA. He is currently a Full Professor with the College of Computer Science and Technology, Wuhan University of Science and Technology, Wuhan, China. His research interests include medical image processing, pattern recognition, and machine learning.

WEI HU received the Ph.D. degree from Zhejiang University, China, in 2008. He is currently a Full Professor with the College of Computer Science and Technology, Wuhan University of Science and Technology, Wuhan, China. His research interests include system on chip and computer architecture.

ZHIFANG PAN received the Ph.D. degree from the Department of Computer Science and Engineering, Shanghai Jiao Tong University, China, in 2013. In 1999, he joined Wenzhou Medical University, where he is currently an Associate Professor and the Deputy Director of the Information Technology Center. His research interest includes the analysis of medical image and big data.