IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received January 12, 2020, accepted January 31, 2020, date of publication February 24, 2020, date of current version March 4, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2974512

iMSCGnet: Iterative Multi-Scale Context-Guided Segmentation of Skin Lesion in Dermoscopic Images

YUJIAO TANG[1,2], ZHIWEN FANG[1,2], SHAOFENG YUAN[3], CHANG'AN ZHAN[1,2], YANYAN XING[1,2], JOEY TIANYI ZHOU[4], AND FENG YANG[1,2]

[1]School of Biomedical Engineering, Southern Medical University, Guangzhou 510515, China
[2]Guangdong Provincial Key Laboratory of Medical Image Processing, Southern Medical University, Guangzhou 510515, China
[3]Shanghai United Imaging Healthcare Co., Ltd., Shanghai 201807, China
[4]IHPC, A*STAR, Singapore 138632

Corresponding authors: Zhiwen Fang (fzw310@gmail.com) and Feng Yang (yangf@smu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61771233 and Grant 61702182, in part by the Hunan Provincial Natural Science Foundation of China under Grant 2018JJ3254, and in part by the Key Project of Special Development Foundation of Shanghai Zhangjiang National Innovation Demonstration Zone under Grant 1701-JD-D1112-030.

ABSTRACT Despite much effort has been devoted to skin lesion segmentation, the performance of existing methods is still not satisfactory enough for practical applications. The challenges may include fuzzy lesion boundary, uneven and low contrast, and variation of colors across space, which often lead to fragmentary segmentation and inaccurate boundary. To alleviate this problem, we propose a multi-scale context-guided network named as MSCGnet to segment the skin lesions accurately. In MSCGnet, the context information is utilized to guide the feature encoding procedure. Moreover, because of the information loss in spatial down-sampling, a context-based attention structure (CAs) is designed to select effective context features in the decoding path. Furthermore, we boost the performance of MSCGnet with iterations and term this upgraded version as iterative MSCGnet, denoted as iMSCGnet. To supervise the training of iMSCGnet in an end-to-end fashion, a novel objective function of deep supervision, which consists of the terms of each encoding layers and the terms from each MSCGnet output of iMSCGnet, is employed. Our method is evaluated extensively on the four publicly available datasets, including ISBI2016 [1], ISBI2017 [2], ISIC2018 [3] and PH2 [4] datasets. The experimental results prove the effectiveness of proposed components and show that our method generally outperforms the state-of-the-art methods.

INDEX TERMS Skin lesion segmentation, multi-scale context, attention, deep supervision.

I. INTRODUCTION

Melanoma has been considered as the most dangerous type of skin cancer. It's less than 6.5% of all skin cancer but 75% of (skin cancer) deaths are related to melanoma [5]–[7]. In order to analyze the skin lesion, which might be skin cancer, dermoscopy is commonly used in imaging skin because it is safe, non-invasive and effective [8], [9]. However, based on dermoscopic images, accurate skin lesion location by experts for melanoma diagnosises is high time-consuming, subjective and labor-intensive [9], [10]. Therefore, a smart automatic skin lesion segmentation method is highly desired in the computer-aided diagnosis (CAD) systems [11]–[13], which can enhance the diagnosis capability of experts. Nevertheless, because of the complex issues (e.g., fuzzy lesion boundary, various lesion appearance, and low contrast between lesions and their surrounding normal skin), skin lesion segmentation is still a challenging task.

Currently, aiming to achieve accurate segmentation, most researchers utilize convolutional neural networks (CNNs) to design deep-learning-based methods [14]–[18]. As a classical framework of biomedical image segmentation, UNet [14] achieves high performance by adding shortcuts between the layers of its encoder and decoder. Inspired from UNet, a rewiring method for shortcuts is proposed to link subnetworks densely by UNet++ [19]. However, due to the low contrast of some skin lesions, it is hard to segment the lesion accurately. The main reason is that

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

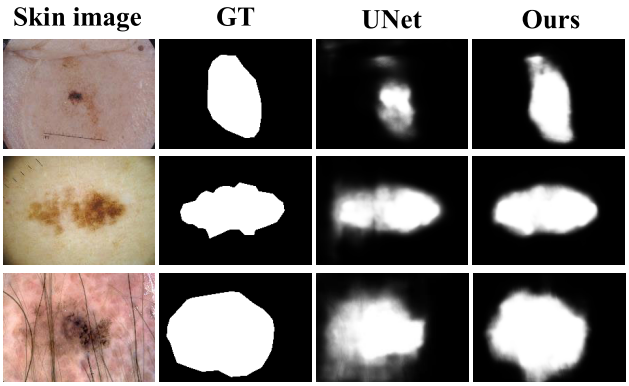


FIGURE 1. Skin Lesion segmentation based on the multi-scale context information. Column 1: three dermoscopic images. Column 2: their corresponding ground truths (GT). Column 3 and 4: the segmentation results of UNet [14] and our method, respectively. Compared with UNet, our method can achieves better performance.

deep-learning-based methods without context information would be confused while meeting the skin lesion whose appearance is similar to that of its surrounding normal skin. This problem will lead to fragmentary segmentation and fuzzy boundary. Some examples are given in Fig. 1. The third column is the results of UNet, and it can be observed that extra context information makes network more discriminative to detect the low contrast lesion shown in the forth column. In order to overcome this problem, several methods [20], [21] adopt the context information to improve the segmentation performance. In [20], Mirikharaji *et al*. refines the skin segmentation by combining the skin image and its context information in an auto-context scheme. Bi *et al*. [21] design a multi-stage architecture including context information as an extra input to segment the skin lesion. However, the context information is only used as the input information. It can not guide the feature extraction well at different scales, which contain different-level semantic information.

In this paper, we propose a multi-scale context-guided network for skin lesion segmentation, which is denoted as MSCGnet. Different from traditional context-based methods taking only one scale into consideration, MSCGnet utilize context information from multi-scale feature layers. Inspired by [22], [23], which shows that coarse segmentation results can infer the context information, we resize a coarse segmentation result to different scales and fuse them into all layers of UNet. Due to the information loss in the procedure of down-sampling [24], [25], a context-based attention structure is introduced in the decoding path to obtain effective context information. Shown in the last column of Fig. 1, our MSCGnet can produce better results than UNet due to the multi-scale context information. Furthermore, inspired by [21] introducing a multi-stage segmentation to refine the results step by step, MSCGnet can be updated to an iterative MSCGnet, named as iMSCGnet. In iMSCGnet, except the first context information from the result of UNet, the successive context information comes from the result of the previous MSCGnet. In training phase, a weighted loss function, which comprises of the terms from each encoding layers of an MSCGnet and the terms from each MSCGnet output of iMSCGnet, is designed to train the whole model end-to-end.

The proposed method is evaluated on four public datasets (ISBI 2016 [1], ISBI 2017 [2], ISIC 2018 [3] and PH2 [4]). The experimental results demonstrate the effectiveness of iMSCGnet, and the main contributions of this paper include:

- A Multi-Scale Context-Guided network denoted as MSCGnet is proposed for skin lesion segmentation by using the context information to guide the feature extraction in the encoding and decoding layers.
- Aiming to improve the performance step by step, an end-to-end iterative MSCGnet named as iMSCGnet is introduced. In iMSCGnet, the initial context information is from the result of UNet, and the result of each MSCGnet is treated as the context information of the next MSCGnet.
- A novel objective function of deep supervision, which consists of the terms from each encoding layers of an MSCGnet and the terms from each MSCGnet output of iMSCGnet, is designed to train the whole model end-to-end.

The preliminary conference version of this work was presented in ISBI 2019 [26]. This paper is a substantial extension, including:

- A context-based attention structure is designed in the decoding path.
- Deep supervision of the encoder is introduced to supervise the training of iMSCGnet.
- The experimental results on new datasets are presented for demonstrating the effectiveness of our iMSCGnet.

The remainder of this paper is organized as follows: the related work is reviewed in Sec. II. Then the overview of MSCGnet is illustrated and introduced in Sec. III. Sec. IV introduces the proposed MSCGnet. An iterative version of MSCGnet and the objective function are detailed in Sec. V and VI, respectively. The experimental results and discussions are conducted in Sec. VII. Finally, Sec. VIII concludes the whole paper.

II. RELATED WORK

A. SKIN LESION SEGMENTATION BASED ON HAND-CRAFTED FEATURES

In the past decades, many classical algorithms [27]–[32] have been developed to segment skin lesions. In [28], lesion segmentation is achieved using a histogram analyzing method. Wong *et al*. [29] adopt a iterative stochastic region merging method to segment lesions. According to the techniques for tracking curve movement, the deformable models may be divided into two categories, including parametric and geometric models. The active contour model is a part of the parametric model. In [32], the skin lesion segmentation is performed by a biologically inspired geodesic active contour technique. However, the above methods rely on the hand-crafted features, which can not capture the high-level semantic information.

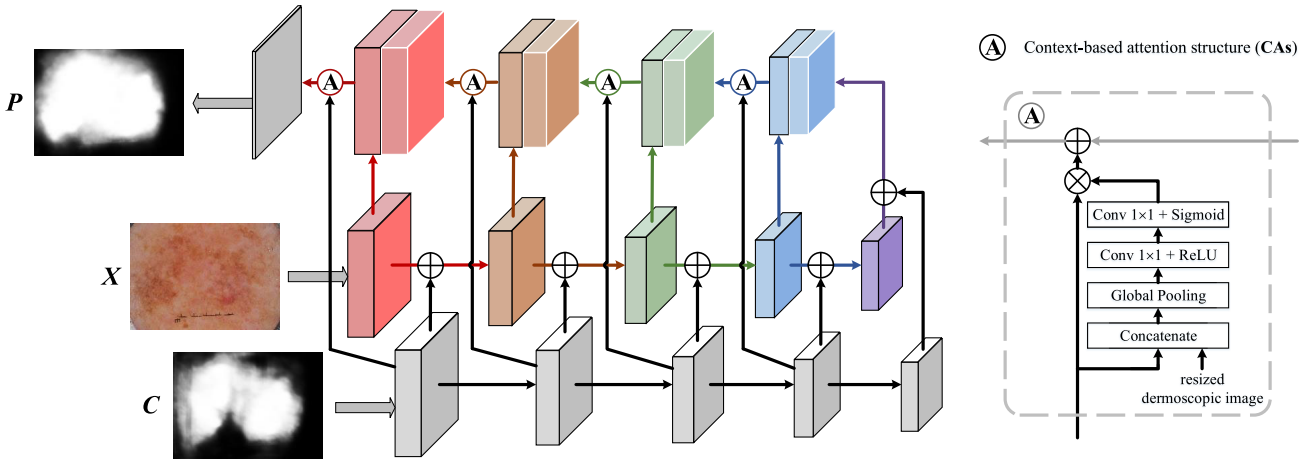


FIGURE 2. The architecture of MSCGnet. MSCGnet consists of two encoders and a decoder. The context information from an context encoder is embedded into multi-scale layers of MSCGnet. Thus, the inputs of MSCGnet are a dermoscopic image *X* and a context image *C*. Considering the information loss of the auto-encoder [24], [25], a context-based attention structure (CAs) is designed to select context features in the decoding path. The structure of CAs is shown on the right side. Different from the traditional channel attention mechanism, a resized demoscopic image is also an input of the attention structure. More details about CAs are presented in Sec. IV-B. This figure is best viewed in color.

B. SKIN LESION SEGMENTATION BASED ON DEEP LEARNING

Recently, deep learning methods have won a great success in skin lesion segmentation [20], [21], [33]–[45]. Yuan *et al*. [33] train the traditional fully convolutional networks (FCN) [18] using a jaccard distance loss to segment the skin lesions. The SkinNet [34] replaces the conventional convolution layers with dense convolution blocks in encoding and decoding of UNet [14] for skin lesion detection. Some studies also improve segmentation accuracy by employing context information in CNN models. Mirikharaji *et al*. [20] combine image appearance information and contextual information in an auto-context scheme to detect the lesion boundaries. Bi *et al*. [21] design a multi-stage FCN (mFCN-PI) architecture which employ context information to detect lesion region repeatedly. In mFCN-PI, the context from the previous probability map is fed into FCN along with the dermoscopic image to obtain segmentation results. However, the above methods only make use of the single-scale context information.

Some researchers [35], [36] also explore the potential of CNN by using multi-scale context in skin lesion segmentation task. In [35], Yu *et al*. propose a fully convolutional residual network (FCRN), which applies the multi-scale context information by the skip connection for skin lesion segmentation. However, FCRN has difficulty in restoring the shape of skin leisons when faced with challenging cases. To automatically delineating skin lesion, Wang *et al*. [36] suggest the pyramid attention network (PA-Net) which is based on an encoder-decoder architecture. In PA-Net, the proposed pyramid attention module (PAM) jointly uses pyramid pooling [46] and attention structure to integrate the multi-scale context information at the beginning of the decoder.

Different form above context-based methods [20], [36], which pay more attention to some special layers, our proposed iMSCGnet implements the context fusion in all layers. Moreover, under a novel deep supervision, multi-scale context information can be used to effectively guide the feature extraction in the whole model.

III. OVERVIEW

A good skin lesion segmentation method should be able to analyze the lesions according to their context information due to the nonuniformity of contrast and/or color. It means that, in a skin lesion region, some lesions are salient but others are ambiguous. Obviously, the segmentation of ambiguous skin lesions will benefit from that of their surrounding salient lesions. According to this observation, some existing methods [20], [21] adopt the context information as an input of their models. However, the single-scale context information from the input can not guide feature extraction or feature selection of different scales effectively because the size and the shape of skin lesions are various.

In this work, we propose a multi-scale context-guided network, denoted as MSCGnet, which incorporates the context information into different layers of an encoder and a decoder. It can guide the multi-scale feature extraction in the encoding and decoding paths. The architecture is shown in Fig. 2. Inspired by [22], [23], which shows that coarse segmentation results can infer the context information, the coarse segmentation is used to obtain multi-scale context features through successive convolution operations. In the decoding path, a context-based attention structure (CAs) is designed to select the discriminative context features. More details will be introduced in Sec. IV. In order to refine the results step by step, we design an iterative mechanism including multiple MSCGnet, and the iterative version is named as iMSCGnet. The details of iMSCGnet will be discussed in Sec. V, and its corresponding objective function will be presented in Sec. VI. This function, which comprises of the terms from each encod-

ing layers of an MSCGnet and the terms from each MSCGnet output of iMSCGnet, is designed to train iMSCGnet end-to-end.

IV. MSCGNET FOR SKIN LESION SEGMENTATION

In this section, we will introduce the details of the proposed multi-scale context-guided network. The architecture of MSCGnet will be present in Sec.IV-A, and its context-based attention structure (CAs) will be illustrated in Sec. IV-B.

A. MSCGnet

Generally, the context information is useful for improving the performance in many visual applications [47], [48]. As mentioned above, the information of salient lesions is also useful for segmenting the ambiguous lesions in the surrounding region of salient lesions. Thus, we hope to enhance the ability of skin lesion segmentation by introducing multi-scale context information into the encoding and decoding layers. The architecture of MSCGnet is shown in Fig. 2. In our MSCGnet, the multi-scale context information is used to guided the feature learning in both encoding and decoding paths, which is helpful for accurate semantic segmentation due to the provision of local and global information [49]. In MSCGnet, the inputs are the dermoscopic image X and the context image C, and the output is the result of skin lesion segmentation P.

In the encoding path, the context information is incorporated via the addition operation $\oplus$ of features shown in Fig. 2. Aiming to ensure the scale consistency, the feature map of the context image goes though the same encoding processing as that of the dermoscopic image. The encoders of the context image and the dermoscopic image consist of multiple repetitive blocks including two 3×3 convolutions followed by a rectified linear unit (ReLU) and a 2×2 max pooling. The multi-scale features of the context image and the dermoscopic image are added element by element (i.e., $\oplus$ in Fig. 2) before max pooling operation.

In order to alleviate the influence of information loss caused by consecutive pooling operations in the encoding path, the skip connections between the same resolution layers of the encoder and the decoder proposed in UNet [14] are also utilized in our MSCGnet. From the experimental results, it is observed that, except the information symmetry via the skip connections, extra information obtained from the context image and the dermoscopic image can effectively improve the performance of skin lesion segmentation. Thus, in the decoding path, a context-based attention structure (CAs) is designed to achieve accurate segmentation results. The details of CAs will be presented in the next section (i.e., Sec. IV-B).

B. CONTEXT-BASED ATTENTION STRUCTURE

Because of the information loss in the auto-encoder [24], [25], we design a context-based attention structure (CAs) to select context features in the decoding path. It can help each encoding layer obtain effective context information according to its previous feature layer. Some visualization examples are given in Fig. 3. Ⓐ and $\oplus$ are the context-based attention structure CAs and the element-wise addition, respectively. From column (a) and column (c), we can see that the features obtained from Ⓐ are more effective than that from $\oplus$, especially for ambiguous lesions. It means that the feature selection of the context information is necessary.

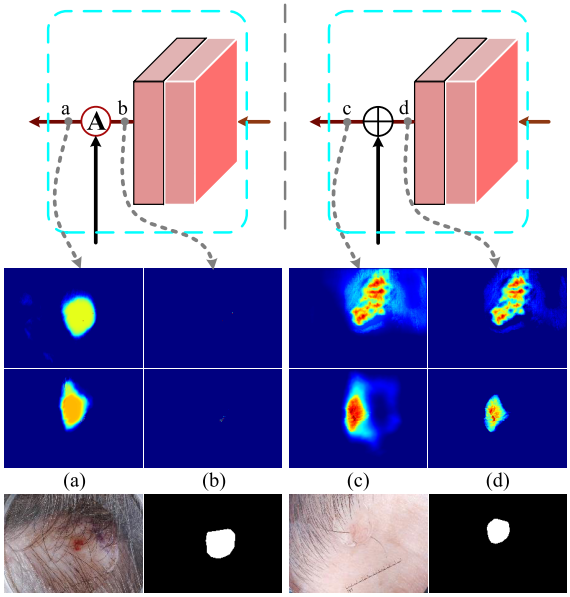


FIGURE 3. Comparison between CAs and the traditional element-wise addition. Ⓐ and $\oplus$ denote our context-based attention structure CAs and the element-wise addition, respectively. Two samples and their corresponding ground truths are listed at the bottom of this figure. Based on Ⓐ and $\oplus$, the feature maps of the samples are shown in the middle of this figure. (a–d) are visualizations of feature maps at different places. The feature maps in the first row are extracted from the first sample and the second row is for the second one. Obviously, our context-based attention structure can achieve more effective features.

Shown in Fig. 2, the context-based attention structure (i.e., Ⓐ in Fig. 2) is implemented after each up-sampling operation. The details of CAs is illustrated on the right of Fig. 2. Different from the traditional channel attention mechanism [50], the resized demoscopic RGB image is also an input of the attention structure. The main reason is that demoscopic images can provide more details than context images. Some visualization examples are listed in Fig. 4. Ⓐ and $Ⓐ_{w/o}$ represent the attention structure with and without the demoscopic image, respectively. From column (a) and column (c), it can be observed that, for ambiguous lesions, the attention structure including the demoscopic image can provide more robust features than that without the demoscopic image.

In CAs, after concatenating the context feature and the resized demoscopic image, a global average pooling, a convolution followed by ReLU and a convolution followed by Sigmoid will be implemented to output a weight vector. Then, a multiply operation is utilized between the context feature and the weight vector to select effective context information. Finally, the selected context features and the corresponding features of the decoding path will be added element by element.

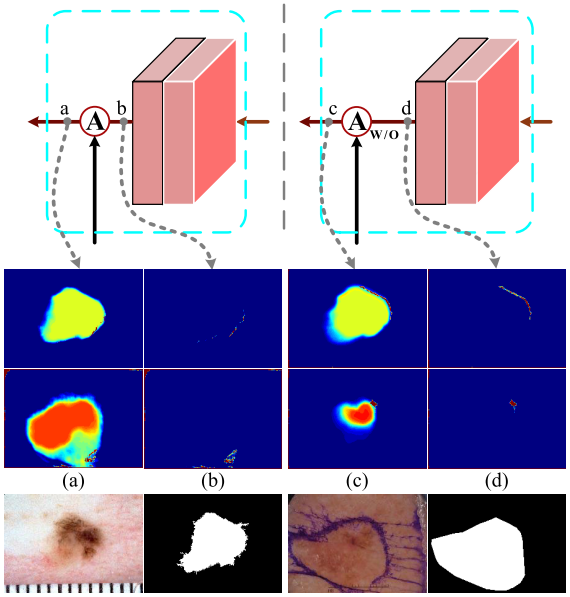


FIGURE 4. Comparison between CAs with and without the dermoscopic image. Ⓐ and $Ⓐ_{w/o}$ represent CAs with and without the dermoscopic image, respectively. Two samples and their corresponding ground truths are listed at the bottom of this figure. Based on Ⓐ and $Ⓐ_{w/o}$, the feature maps of the samples are shown in the middle of this figure. (a–d) are visualizations of feature maps at different places. The feature maps in the first row are extracted from the first sample and the second row is for the second one. It can be observed that more details can be preserved by the context information of the dermoscopic images.

V. ITERATIVE MSCGnet

Inspired by [51], [52], the iterative mechanism can effectively improve the segmentation performance. Thus, we update the proposed MSCGnet to an iterative version denoted as iMCSGnet. The diagram of iMSCGnet is illustrated in Fig. 5. Given the stage $s \in \{1, 2, \ldots S\}$, the input of the s^{th} MSCGnet is the dermoscopic image X and the context image C_s, and its output is P_s. In iMSCGnet, C_1 is provided by UNet, and $C_{s+1}, s \in \{1, \ldots S-1\}$ comes from the result P_s of the s^{th} MSCGnet.

VI. OBJECTIVE FUNCTION

Inspired by the deep supervision introduced in [53]–[55] which can help deep networks achieve high performance benefiting from deep semantic information, we train iMSCGnet by a novel objective function $\mathbb{L}$. Shown in Fig. 6, the introduced $\mathbb{L}$ comprises of the terms $\mathbb{L}_F^{s,l}$ from the l^{th} encoding layers of the s^{th} MSCGnet and the terms $\mathbb{L}_P^s$ from the s^{th} MSCGnet output of iMSCGnet. It is defined as

$$\mathbb{L} = \sum_{s=1}^{S} \alpha^s \sum_{l=1}^{L} \left(\mathbb{L}_P^s + \beta^{s,l} \mathbb{L}_F^{s,l} \right), \tag{1}$$

where α^s and $\beta^{s,l}$ are the weights of the s^{th} MSCGnet and the l^{th} feature layer of the encoding path in the s^{th} MSCGnet, respectively; L is the number of the encoding layers; S is the iterative number of iMSCGnet. The deep supervision

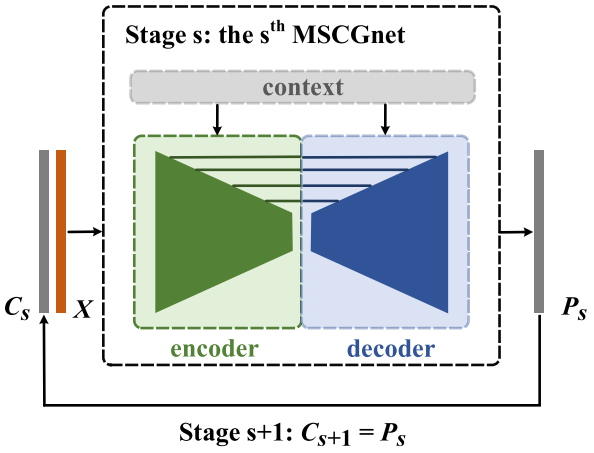


FIGURE 5. The diagram of the iterative MSCGnet. At the s^{th} stage, the input of the s^{th} MSCGnet is the dermoscopic image X and the context image C_s, and its output is P_s. C_1 is provided by UNet when $s = 1$, and $C_{s+1} = P_s, s > 1$.

is shown in Fig. 6. According to the definition of jaccard distance loss [33], $\mathbb{L}_P^s$ and $\mathbb{L}_F^{s,l}$ are defined as

$$\mathbb{L}_P^s = 1 - \frac{\sum_{(i,j)} G(i,j) P_s(i,j)}{\sum_{(i,j)} G(i,j)^2 + \sum_{(i,j)} P_s(i,j)^2 - \sum_{(i,j)} G(i,j) P_s(i,j)}, \tag{2}$$

and

$$\mathbb{L}_F^{s,l} = 1 - \frac{\sum_{(i,j)} G(i,j) F_{s,l}(i,j)}{\sum_{(i,j)} G(i,j)^2 + \sum_{(i,j)} F_{s,l}(i,j)^2 - \sum_{(i,j)} G(i,j) F_{s,l}(i,j)}, \tag{3}$$

where (i, j) represents the spatial index; P_s is the predicted result of the s^{th} MSCGnet; $F_{s,l}$ denotes the predicted result provided from the l^{th} feature layer of the encoding path in the s^{th} MSCGnet; $G(i,j)$, $P_s(i,j)$ and $F_{s,l}(i,j)$ are the $(i,j)^{th}$ values of the ground truth, P_s and $F_{s,l}$, respectively. The objective function $\mathbb{L}$ can supervise the training of iMSCGnet in an end-to-end fashion.

VII. EXPERIMENTS

In experiments, we evaluate the performance of iMSCGnet on four publicly available datasets. The datasets are from ISBI2016 [1], ISBI2017 [2] and ISIC2018 [3] challenge named as "Skin Lesion Analysis Towards Melanoma Detection" and PH2 [4]. All datasets provide RGB dermoscopic images and their corresponding ground truths. In the ISBI2016 dataset [1], there are 900 training images and 379 testing images, which varies from 542 × 718 to 2848 × 4288. The ISBI2017 challenge dataset [2] contains 2000 images for training, 150 images for validation and 600 images for testing. The image size of ISBI2017 varies from 540 × 722 to 4499 × 6748 pixels. In ISIC2018, the training set consists of 2594 RGB dermoscopic images with spatial resolutions ranging from 540 × 722 to 4499 × 6748. The ground truths of ISIC 2018 [3] validation and

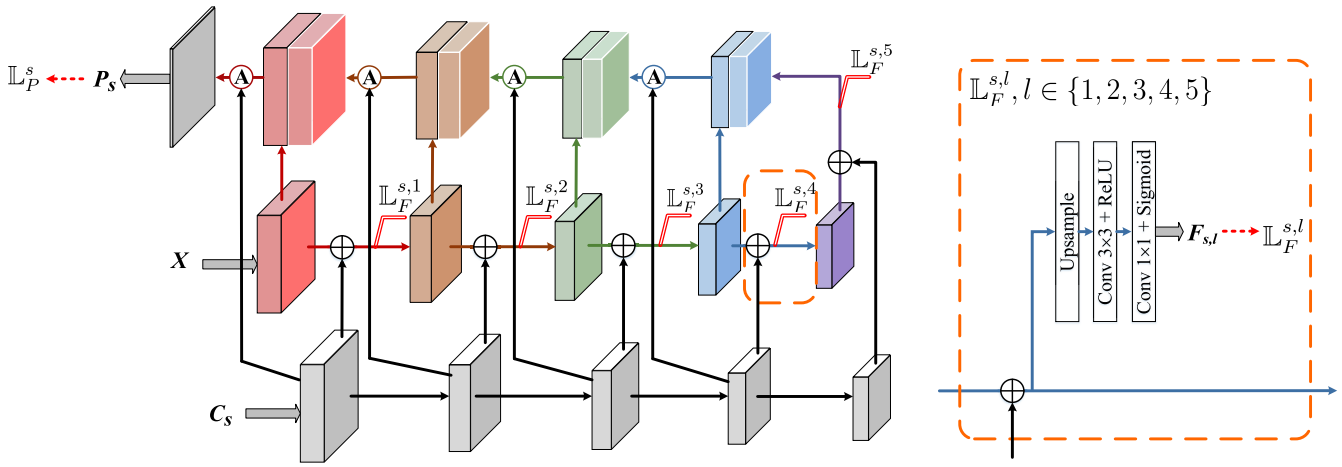


FIGURE 6. Deep supervision for iMSCGnet. The objective function of iMSCGnet includes $\mathbb{L}_P^s$ and $\mathbb{L}_F^{s,l}$, where $s \in \{1, 2, \ldots S\}$ and $l \in \{1, 2, \ldots L\}$. An example of $\mathbb{L}_F^{s,l}$ is detailed on the right side and the up-sampling operation is used to resize the size of feature maps to the size of the ground truth. The whole function is presented in Eqn. 1, and (S, L) is set to $(4, 5)$ in the experiments.

test data have not yet been released. Thus, we divide the training data into 80% training (2076 images) and 20% validation set (518 images), which are used to evaluate the performance of our approach. In order to guarantee the robustness of the model against different parts of the available data, five-fold cross-validation is used. In PH2 [4] public dataset, there are 160 non-melanoma cases and 40 melanoma cases with the size varying from 553 × 763 to 577 × 769.

To evaluate the segmentation capability of the proposed iMSCGnet on ISBI2016 [1], ISBI2017 [2] and PH2 [4], we choose three metrics [1], [2] including Jaccard index (JA), Dice coefficient (DI) and accuracy (AC). They are defined as:

$$JA = \frac{|TP|}{|FP| + |FN| + |TP|}, \tag{4}$$

$$DI = \frac{2 \times |TP|}{|FP| + |FN| + 2 \times |TP|}, \tag{5}$$

$$AC = \frac{|TP| + |TN|}{|FP| + |FN| + |TP| + |TN|}, \tag{6}$$

where TP, TN, FP and FN are the number of true positives, true negatives, false positives and false negatives, respectively. For ISIC2018, except JA and DI, we also use the threshold Jaccard index JA_{th} [3] to evaluate the performance. JA_{th} is defined as

$$JA_{th} = \begin{cases} 0 & JA < 0.65 \\ JA & JA \geq 0.65. \end{cases} \tag{7}$$

Because the height to width ratio of most images is close to 3:4, we resize all images to 160 × 224 and normalize them in the image-wise level. To alleviate the overfitting of iMSCGnet, the data augmentation, including rotation in the range of (-25°, 25°) and randomly horizontal and vertical flipping, is employed to enlarge the training dataset. The channel number of 5 layers are set as 16, 32, 64, 128 and 256, respectively. The iteration number $S = 4$ in iMSCGnet. In Eqn. 1, α^s and $\beta^{s,l}$ are two tunable parameters. In all experiments, $\alpha^s, s \in \{1, 2, 3, 4\}$ are set to $\{0.7, 0.8, 0.9, 1\}$, and $\beta^{s,l} = 0.01, s \in \{1, 2, 3, 4\}$ and $l \in \{1, 2, 3, 4, 5\}$.

In the training phase, Adam optimizer [56] with an initial learning rate 0.0001 is used to minimize the objective function (i.e., Eqn. 1) of iMSCGnet and the mini-batch size is 16. All experiments are implemented on a computer with an i7-5930K CPU and NVIDIA GTX1080Ti GPUs.

This section is organized as follows: in Sec. VII-A, VII-B and VII-C, the ablation studies on four datasets are performed to analyze the effectiveness of the proposed components. Then, in Sec. VII-D, we conduct the comparison experiments of the context information embedded in different layers, e.g., shallow feature layers, deep feature layers and all feature layers. In Sec. VII-E, we do the parameter analysis about the iteration number S. In order to prove the effectiveness of increasing weights in the iterative scheme, the comparison experiments are carried out in Sec. VII-F. Finally, the performance of iMSCGnet is compared with other state-of-the-art methods in Sec. VII-G.

In the ablation experiments, several abbreviations are defined for clarity. Between the features of the dermoscopic image and the context image, the element-wise addition structure of fusing the context (i.e., ⊕ in the encoding path) is named as *CFs*. The concatenation operation labeled as *CON* is used for the comparison with *CFs*. In the decoding path, the context-based attention structure (i.e., Ⓐ) is denoted as *CAs*. About the analyses of the objective function (i.e., Eqn. 1), the whole deep supervision is marked as *DS*. For comparison, a simplified version only including the multiple stage supervision, which is $\sum_{s=1}^{S} \alpha^s \mathbb{L}_p^s$, is called as *MS*. Further-

TABLE 1. Ablation studies on the PH2, ISBI2016, ISBI2017 and ISIC2018 datasets. The bold number indicates the best performance in each column. iMSCGnet is the iterated multi-scale context-guided network. *CON* denotes the concatenation operation. *CFs* means the context fusion structure in the encoder. *CAs* represents the context-based attention structure in the decoder. *SS* and *MS* are the single stage supervision and the multiple stage supervision, respectively. *DS* denotes the whole objective function defined in Eqn. 1. iMSCGnet + Δ represents iMSCGnet with the Δ module.

Methods	PH2			ISBI 2016			ISBI 2017			ISIC 2018		
	JA	DI	AC	JA	DI	AC	JA	DI	AC	JA	DI	JA_{th}
iMSCGnet+CON+SS	84.70	91.07	93.69	83.21	89.90	95.34	75.22	84.06	92.90	83.45	89.68	78.69
iMSCGnet+CON+MS	85.24	91.11	93.77	84.96	91.12	95.72	76.07	84.62	93.02	83.97	90.05	79.83
iMSCGnet+CFs+SS	85.35	91.52	94.02	83.76	90.42	95.43	75.82	84.33	93.04	84.12	90.14	80.04
iMSCGnet+CFs+MS	87.28	92.70	95.11	85.32	91.45	95.87	76.57	84.85	93.23	84.82	90.60	81.16
iMSCGnet+CAs+MS	86.69	92.13	94.73	85.23	91.39	95.91	76.28	84.59	93.25	84.73	90.52	81.16
iMSCGnet+CFs+CAs+MS	87.63	92.83	95.30	85.61	91.66	95.94	77.32	85.53	93.47	85.01	90.69	81.47
iMSCGnet+CFs+CAs+DS	**88.21**	**93.36**	**95.71**	**85.98**	**91.91**	**96.08**	**77.75**	**85.83**	**93.58**	**85.34**	**90.95**	**81.91**

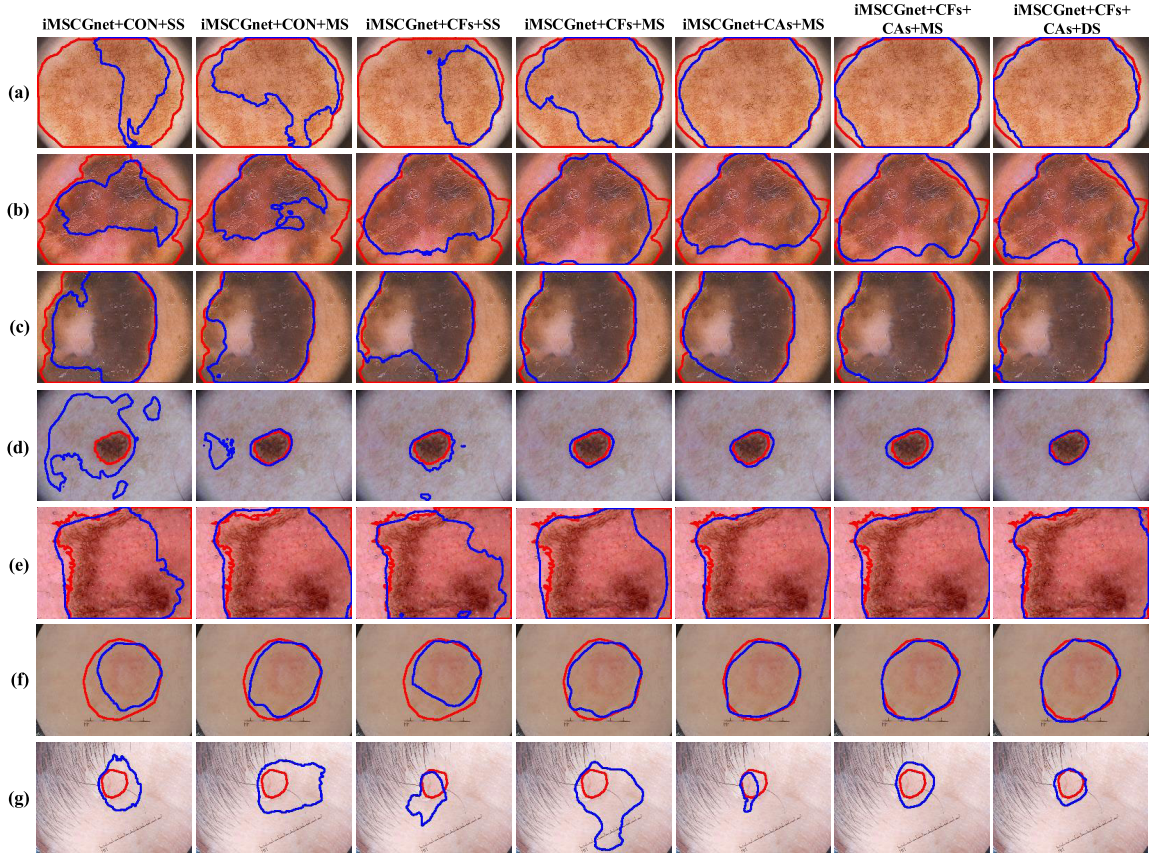


FIGURE 7. Qualitative evaluation of ablation study on three datasets. The samples in row (a-c), row (d) and row (e-g) are from PH2, ISBI2016 and ISBI2017, respectively. Red and blue contours are ground truths and segmentation results, separately.

more, single supervision of the last stage $\mathbb{L}_P^S$ is represented by *SS*.

A. ABLATION STUDIES ON PH2

In the section, 200 dermoscopic images from PH2 are tested on the model, which is trained on the ISBI2017 training data [2]. It can be used to verify the generalization ability of our method. The quantitative and qualitative results are shown in Table 1 and Fig. 7, respectively. It can be seen that:

- Compared with *CON* in Table 1, *CFs* significantly improves all metrics. For example, under the supervision of *MS*, the improvements of JA, DI and AC are (2.04%, 1.59%, 1.34%). The reason would be that the element-wise addition is better for fusing the spatial information of the context image.
- Benefit from *CAs*, the performances of JA, DI and AC are further improved in Table. 1. It means that the selected context information can help the decoder to obtain better segmentation results.
- Among *SS*, *MS* and *DS*, the supervision with more supervisors generally enhances the ability of the skin lesion segmentation. Obviously, the deep supervision can help the model encode more semantic information and make it more discriminative.
- In Fig. 7, three samples from the PH2 dataset are shown from row (*a*) to row (*c*). The multi-scale context information (*CFs* and *CAs*) and the deep supervision (*DS*) are useful for skin lesion segmentation. The contours of segmentation results show in the last column are close to that of ground truths. It means that our proposed method can achieve

accurate segmentation boundaries due to the multi-scale context information.

B. ABLATION STUDIES ON ISBI2016 AND ISBI2017

In the section, ablation studies are executed on the ISBI2016 (900 images for training and 379 images for testing) and ISBI2017 (2000 images for training, 150 images for validation and 600 images for testing) datasets. Table 1 presents the quantitative results, and the qualitative evaluation is shown in row (d-g) of Fig. 7. It can be observed that:

- On the two datasets, *CFs* and *CAs* can improve JA, DI and AC generally. Under the supervision of *MS*, iMSCGnet with *CFs* and *CAs* outperforms iMSCGnet with *CON* in terms of JA, DI and AC. The improvement of (JA, DI, AC) is (0.65%, 0.54%, 0.22%) on the ISBI2016 dataset, and (1.25%, 0.91%, 0.45%) for the ISBI2017 dataset.
- Compared with the *SS* supervision, the *MS* supervision can achieve better performances on the two datasets. Moreover, the *DS* supervision can further improve the performances of iMSCGnet, because the deep supervision can lead to effective pattern encoding in multi-scale layers.
- In row (d-g) of Fig. 7, the segmentation results in the last column are close to the ground truths due to the supports of *CFs*, *CAs* and *DS*. Benefiting from the context information, the segmentation results of low-contrast skin lesions are notable. The main reason is that these ambiguous skin lesions can be mined according to the salient lesions around them.

C. ABLATION STUDIES ON ISIC2018

In this section, we conduct the ablation experiment on ISIC2018 training data, which divided into 80% for training and 20% for validation [57]. To fully evaluate the segmentation performance of iMSCGnet, five-fold cross-validation is performed and we report the mean of the five results in Table 1. We can see that:

- The employments of *CFs* and *CAs* both raise the value of all metrics including JA, DI and JA_{th} in Table 1. Under the supervision of *MS*, *CFs* and *CAs* get improvements of (0.85%, 0.55%, 1.33%) and (0.76%, 0.47%, 1.33%) in terms of (JA, DI, JA_{th}), respectively.
- iMSCGnet with the *DS* supervision gains an improvement of (JA, DI, JA_{th}), because the multiple supervisors make the model encode more discriminative features. With the assistance of *DS*, the segmentation accuracy of the iMSCGnet with CAs and CFs is higher than iMSCGnet with *CON* with a large margin: (1.89%, 1.27%, 3.22%) in (JA, DI, JA_{th}).

D. STRUCTURE ANALYSIS OF MULTI-SCALE CONTEXT GUIDANCE

To further analyze the structure of iMSCGnet, we conduct experiments on ISBI2017 to explore the effectiveness of context information at different layers. Based on this motivation, the context information is embedded in the first layer, the fifth layer and all layers, respectively. Four samples are listed in Fig. 8. The segmentation results are shown in columns 3, 4 and 5, separately. We can observe that:

- Based on the context information embedded in the first layer, the iMSCGnet pays more attention to salient skin lesions while being poor at the segmentation of low-contrast lesions. Because of the low-level semantic information and the limited receptive field, it is hard for iMSCGnet with low-level context information to mine ambiguous lesions in a large range.
- Different from the low-level context information, the context information used in the fifth layer includes more semantic patterns. From the results shown in the fourth column, it is found that some ambiguous lesions can be mined successfully. However, due to lacking details contained in the low-level context information, the results of some ambiguous lesions are unclear, especially the result of the first sample.
- Thanks to multi-scale context information, our iMSCGnet achieves the best segmentation results in both salient and ambiguous lesion regions. Shallow context features contain local patterns and deep context features are rich in the global context information. Thus, the combination of multi-scale context information can enhance the discriminative ability of iMSCGnet.

E. PARAMETER SENSITIVITY INVESTIGATION ON ITERATION NUMBER S

In order to explore the performance of iMSCGnet with different iteration number S, we conduct the comparison experiments on the ISBI2017 and PH2 datasets. JA and DI are applied as the evaluation indicator and the result is demonstrated in Fig. 9. It is observed that iMSCGnet obtains the highest JA and DI when $S = 4$. The iterative scheme can be regarded as a process of finding the optimal value. It means that the iteration number, which is larger than the optimal value, might lead to worse segmentation results. Thus, in all experiments, we set S to 4.

F. PARAMETER SENSITIVITY INVESTIGATION ON ITERATION WEIGHT α^s

Aiming to investigate the parameter sensitivity of the weights α^s in Eqn. 1, we design a comparison experiment between iMSCGnet with different combinations of the weights α^s. Because the iteration number S is set as 4, we choose $\{1, 1, 1, 1\}$ and $\{0.7, 0.8, 0.9, 1\}$ for comparison. The first combination consists of equal weights and incremental weights for the second one. Some examples are shown in Fig. 10. We can see that:

- No matter what weight is given, the segmentation results can be refined stage by stage, because the iteration mechanism can help iMSCGnet refine the results of skin lesion segmentation by reducing the distance between the solution and the optimal one.
- iMSCGnet with the incremental weights achieves better segmentation accuracy than that with equal weights. The main reason is that the results of early stages are

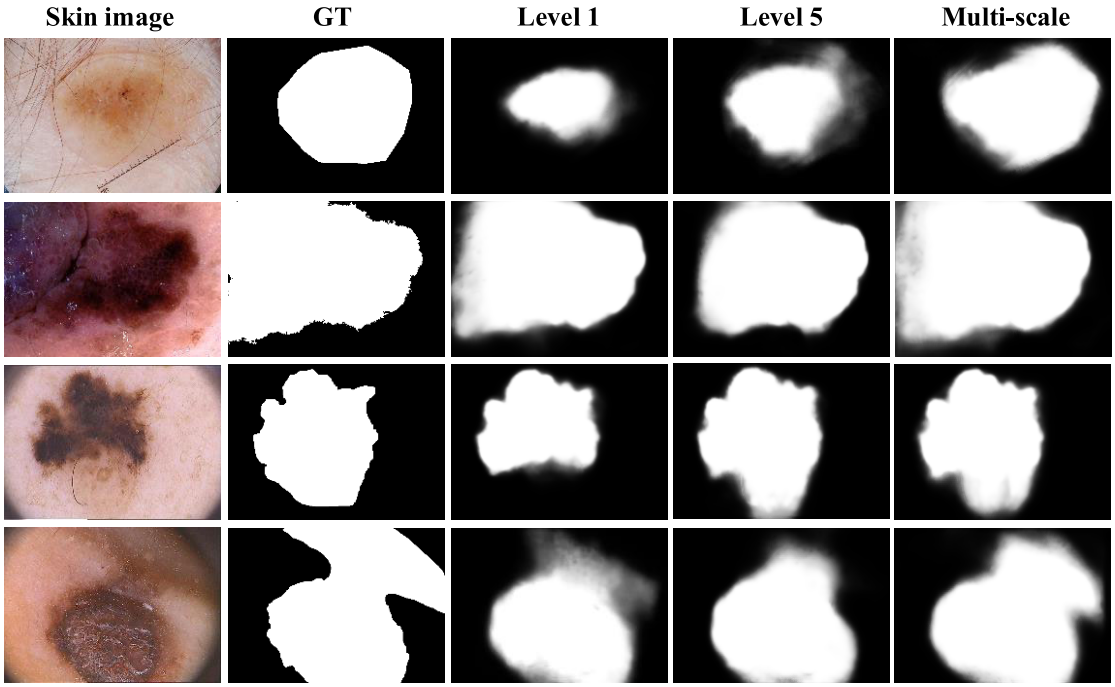
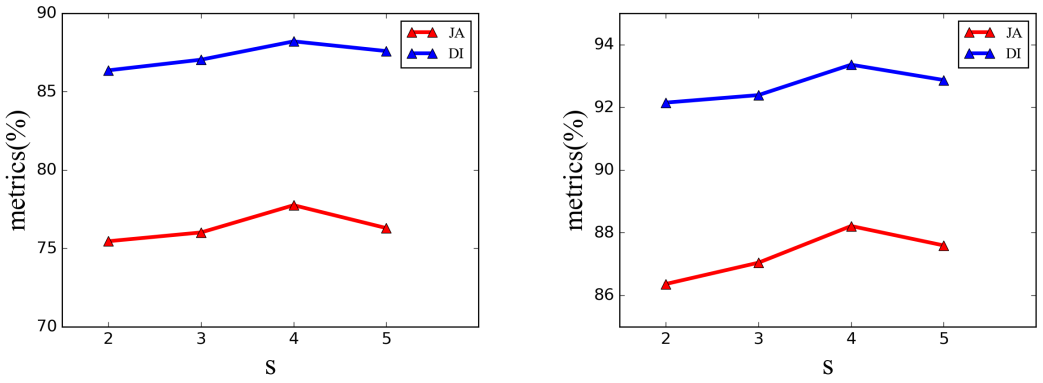


FIGURE 8. Examples of the segmentation results of iMSCGnet with different structures. In order to analyze the effectiveness of the context information embedded in different layers, the context information is fused in the shallow layer (i.e., layer 1), the deep layer (i.e., layer 5) and all layers (i.e., multi-scale), respectively. Column 1: four dermoscopic images. Column 2: ground truths (GT). Column 3-5: segmentation results.



(a) ISBI2017 (b) PH2

FIGURE 9. Parameter sensitivity investigation on iteration number S. The iteration number is varied along the x-axis and the percentage of metrics according to the iteration number is plotted on the y-axis.

not reliable. Thus, a large wight of the early stage will lead to noises in the final result. According to this observation, the weights $\{0.7, 0.8, 0.9, 1\}$, where small values are assigned to the early stages, are adopted in the experiments.

G. COMPARISON WITH THE STATE-OF-THE-ARTS

Here, we analyze iMSCGnet on the PH2, ISBI2016, ISBI2017 and ISIC2018 datasets. On the Ph2 dataset, we compare iMSCGnet with different state-of-the-art methods: mFCN-PI [21], Peng *et al.* [58], DermoNet [59], Xie *et al.* [60], DCL-PSI [37], Goyal *et al.* [41] and FrCN [40]. On the ISBI2016 dataset, we choose the state-of-the-arts as follows: Team-CUMED [35], Team-Rahman [61], DermoNet [59], Mirikharaji *et al.* [20], Deng *et al.* [62], mFCN-PI [21], Yuan *et al.* [33], Nasr-Esfahani *et al.* [63], Xie *et al.* [60] and DCL-PSI [37]. On the ISBI2017 dataset, the comparison methods are: Team-BMIT [64], Team-NLP LOGIX [65], Team-MtSinai [66], FocusNet [67], Li *et al.* [38], SkinNet [34], Tu *et al.* [39], Tschandl *et al.* [68], DCL-PSI [37], FrCN [40] and PA-Net [36]. For the ISIC2018, FrCN [40], GAN-FCN [69] and MobileGan [57] are adopted. All details are listed in Table 2, 3, 4 and 5. It can be seen that:

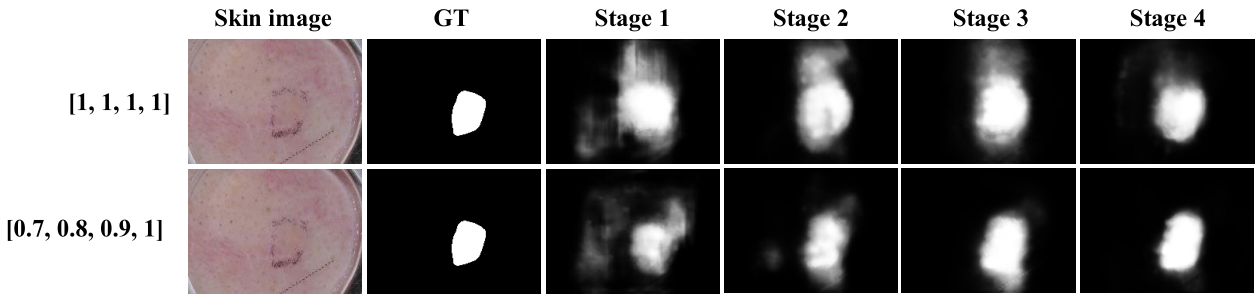


FIGURE 10. Parameter sensitivity investigation on iteration weight α^s. Column 1: dermoscopic images. Column 2: ground truths (GT). Column 3-6: results of four stages with different combinations of iteration weights.

TABLE 2. Comparison with other methods on the PH2 dataset. The bold number indicates the best performance in each column.

Methods	JA	DI	AC	Training data
mFCN-PI. [21]	83.99	90.66	94.24	ISBI2016
Peng *et al.* [58]	85.00	90.00	93.00	ISBI2016
DermoNet [59]	85.30	91.50	\	ISBI2016
Xie *et al.* [60]	85.70	91.90	94.90	ISBI2016
DCL-PSI [37]	85.90	92.10	**95.30**	ISBI2016
Ours	**86.31**	**92.27**	94.93	ISBI2016
Goyal *et al.* [41]	83.90	90.70	93.80	ISBI2017
FrCN [40]	84.79	91.77	95.08	ISBI2017
Ours	**88.21**	**93.36**	**95.71**	ISBI2017

TABLE 3. Comparison with other methods on the ISBI2016 dataset. The bold number indicates the best performance in each column.

Methods	JA	DI	AC
Team - CUMED [35]	82.90	89.70	94.90
Team – Rahman [61]	82.22	89.50	95.20
DermoNet [59]	82.50	89.40	\
Mirikharaji *et al.* [20]	83.30	90.11	95.02
Deng *et al.* [62]	84.10	90.70	95.30
mFCN-PI [21]	84.64	91.18	95.51
Yuan *et al.* [33]	84.70	91.20	95.50
Nasr-Esfahani *et al.* [63]	85.50	91.90	95.70
Xie *et al.* [60]	85.80	91.80	93.80
DCL-PSI [37]	85.92	91.77	95.78
Ours	**85.98**	**91.91**	**96.08**

TABLE 4. Comparison with other methods on the ISBI2017 dataset. The bold number indicates the best performance in each column.

Methods	JA	DI	AC
Team – BMIT [64]	76.00	84.40	93.40
Team – NLP LOGIX [65]	76.20	84.60	93.20
Team – MtSinai [66]	76.50	84.90	93.40
FocusNet [67]	75.62	83.15	92.14
Li *et al.* [38]	76.50	86.60	93.90
SkinNet [34]	76.70	85.50	93.20
Tu *et al.* [39]	76.80	86.20	**94.50**
Tschandl *et al.* [68]	77.00	85.30	\
DCL-PSI [37]	77.70	85.60	94.00
FrCN [40]	77.11	**87.08**	94.03
PA-Net [36]	77.60	85.80	93.60
Ours	**77.75**	85.83	93.58

TABLE 5. Comparison with other methods on the IBSI2018 dataset. The bold number indicates the best performance in each column.

Methods	JA_{th}
FrCN [40]	74.60
GAN-FCN [69]	77.80
MobileGan [57]	78.40
Ours	**81.91**

• Because the PH2 dataset is very small, the state-of-the-art methods usually train their models on other big datasets (e.g., ISBI2016 and ISBI2017). Then, the models are tested on the PH2 dataset for evaluating their generalization ability. In Table 2, the training data of the first part is the ISBI2016 dataset, and the ISBI2017 dataset for the second part. It can be found that our method generally outperforms other methods. Due to the large training set of ISBI2017, our iMSCGnet achieves significant improvements of JA, DI and AC. It means that our method has strong generalization ability. It is very useful for lots of applications, which are lack of enough training samples.

• On the ISBI2016 dataset, our method also obtains the best results. The main reason is that, under the deep supervision, multi-scale context information effectively guide the feature extraction in an iterative fashion.

• On the ISBI2017 dataset, the proposed iMSCGnet generally achieves better results than most of the state-of-the-art methods. In Table 4, we can see that FrCN [40] has better DI and AC than ours. However, in Table 2, compared with FrCN [40], our method gets an improvement of $(3.42\%, 1.59\%, 0.63\%)$ in terms of (JA, DI, AC). It means that FrCN [40] might tend to overfit to the training set of ISBI2017. Obviously, our iMSCGnet gains strong generalization ability on the PH2 dataset while achieving comparable performances on the ISBI2017 dataset.

• The ISIC2018 dataset is a new challenging dataset published last year. Due to inter-observer and intra-observer variability, the metrics used in the ISBI2016 and ISBI2017 datasets can not accurately measure the segmentation accuracy [3]. Thus, on the ISIC2018 dataset, a new measurement, denoted as JA_{th}, is designed to evaluate the performance of skin lesion segmentation. Shown as Table 5, our method still outperforms other state-of-the-arts.

VIII. CONCLUSION

In this work, we propose the multi-scale context-guided network, denoted as MSCGnet, which introduces the context information into multi-scale feature layers to enhance the performance of the skin lesion segmentation. In the encoding

path, multi-scale context information is extracted to guide the feature learning. In the decoding path, the context-based attention structure is proposed to effectively select the context information. Furthermore, we upgrade the MSCGnet to its iterative version, named as iMSCGnet. iMSCGnet can refine the segmentation result step by step, and be trained in an end-to-end fashion under the deep supervision of a novel objective function. The experimental results on four challenging datasets demonstrate the effectiveness of the proposed iMSCGnet.

REFERENCES

[1] D. Gutman, N. C. F. Codella, E. Celebi, B. Helba, M. Marchetti, N. Mishra, and A. Halpern, "Skin lesion analysis toward melanoma detection: A challenge at the international symposium on biomedical imaging (ISBI) 2016, hosted by the international skin imaging collaboration (ISIC)," 2016, *arXiv:1605.01397*. [Online]. Available: http://arxiv.org/abs/1605.01397

[2] N. C. F. 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. (ISBI)*, Apr. 2018, pp. 168–172.

[3] N. Codella, V. Rotemberg, P. Tschandl, M. Emre Celebi, S. Dusza, D. Gutman, B. Helba, A. Kalloo, K. Liopyris, M. Marchetti, H. Kittler, and A. Halpern, "Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the international skin imaging collaboration (ISIC)," 2019, *arXiv:1902.03368*. [Online]. Available: http://arxiv.org/abs/1902.03368

[4] T. Mendonça, P. M. Ferreira, J. S. Marques, A. R. 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.

[5] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2019," *CA, Cancer J. Clinicians*, vol. 69, no. 1, pp. 7–34, 2019.

[6] 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.

[7] C. M. Balch *et al.*, "Final Version of 2009 AJCC Melanoma Staging and Classification," *JCO*, vol. 27, no. 36, pp. 6199–6206, Dec. 2009.

[8] J. Mayer, "Systematic review of the diagnostic accuracy of dermatoscopy in detecting malignant melanoma," *Med. J. Aust.*, vol. 167, no. 4, pp. 206–210, Aug. 1997.

[9] M. E. Vestergaard, P. Macaskill, P. E. Holt, and S. W. Menzies, "Dermoscopy compared with naked eye examination for the diagnosis of primary melanoma: A meta-analysis of studies performed in a clinical setting," *Brit. J. Dermatol.*, vol. 159, no. 3, pp. 76–669, Jun. 2008.

[10] M. E. Celebi, H. Iyatomi, W. V. Stoecker, R. H. Moss, H. S. Rabinovitz, G. Argenziano, and H. P. Soyer, "Automatic detection of blue-white veil and related structures in dermoscopy images," *Computerized Med. Imag. Graph.*, vol. 32, no. 8, pp. 670–677, Dec. 2008.

[11] 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.

[12] R. B. Oliveira, M. E. Filho, Z. Ma, J. P. Papa, A. S. Pereira, and J. M. R. S. Tavares, "Computational methods for the image segmentation of pigmented skin lesions: A review," *Comput. Methods Programs Biomed.*, vol. 131, pp. 127–141, Jul. 2016.

[13] S. Pathan, K. G. Prabhu, and P. C. Siddalingaswamy, "Techniques and algorithms for computer aided diagnosis of pigmented skin lesions—A review," *Biomed. Signal Process. Control*, vol. 39, pp. 237–262, Jan. 2018.

[14] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assisted Intervent.* Cham, Switzerland: Springer, 2015, pp. 234–241.

[15] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-Net: Fully convolutional neural networks for volumetric medical image segmentation," in *Proc. 4th Int. Conf. 3D Vis. (3DV)*, 2016, pp. 565–571.

[16] O. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-Net: Learning dense volumetric segmentation from sparse annotation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assisted Intervent.* Cham, Switzerland: Springer, 2016, pp. 424–432.

[17] 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.

[18] E. Shelhamer, J. Long, and T. Darrell, "Fully convolutional networks for semantic segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 4, pp. 640–651, Apr. 2017.

[19] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: Redesigning skip connections to exploit multiscale features in image segmentation," *IEEE Trans. Med. Imag.*, to be published.

[20] Z. Mirikharaji, S. Izadi, J. Kawahara, and G. Hamarneh, "Deep auto-context fully convolutional neural network for skin lesion segmentation," in *Proc. IEEE 15th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2018, pp. 877–880.

[21] L. Bi, J. Kim, E. Ahn, A. Kumar, M. Fulham, and D. Feng, "Dermoscopic image segmentation via multistage fully convolutional networks," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 9, pp. 2065–2074, Sep. 2017.

[22] Z. Tu and X. Bai, "Auto-context and its application to high-level vision tasks and 3D brain image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 10, pp. 1744–1757, Oct. 2010.

[23] Y. Zhou, X. Sun, Z.-J. Zha, and W. Zeng, "Context-reinforced semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2019, pp. 4046–4055.

[24] S. Sabour, N. Frosst, and G. E. Hinton, "Dynamic routing between capsules," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 3856–3866.

[25] M. Yang, K. Yu, C. Zhang, Z. Li, and K. Yang, "DenseASPP for semantic segmentation in street scenes," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 3684–3692.

[26] Y. Tang, F. Yang, S. Yuan, and C. Zhan, "A multi-stage framework with context information fusion structure for skin lesion segmentation," in *Proc. IEEE 16th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2019, pp. 1407–1410.

[27] R. Garnavi, M. Aldeen, M. E. Celebi, A. Bhuiyan, C. Dolianitis, and G. Varigos, "Automatic segmentation of dermoscopy images using histogram thresholding on optimal color channels," *Int. J. Med. Med. Sci.*, vol. 1, no. 2, pp. 126–134, 2010.

[28] H. Fan, F. Xie, Y. Li, Z. Jiang, and J. Liu, "Automatic segmentation of dermoscopy images using saliency combined with Otsu threshold," *Comput. Biol. Med.*, vol. 85, pp. 75–85, Jun. 2017.

[29] A. Wong, J. Scharcanski, and P. Fieguth, "Automatic skin lesion segmentation via iterative stochastic region merging," *IEEE Trans. Inf. Technol. Biomed.*, vol. 15, no. 6, pp. 929–936, Nov. 2011.

[30] Z. Liu and J. Zerubia, "Skin image illumination modeling and chromophore identification for melanoma diagnosis," *Phys. Med. Biol.*, vol. 60, no. 9, pp. 3415–3431, May 2015.

[31] R. B. Oliveira, N. Marranghello, A. S. Pereira, and J. M. R. S. Tavares, "A computational approach for detecting pigmented skin lesions in macroscopic images," *Expert Syst. Appl.*, vol. 61, pp. 53–63, Nov. 2016.

[32] R. Kasmi, K. Mokrani, R. K. Rader, J. G. Cole, and W. V. Stoecker, "Biologically inspired skin lesion segmentation using a geodesic active contour technique," *Skin Res. Technol.*, vol. 22, no. 2, pp. 208–222, May 2016.

[33] 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.

[34] S. Vesal, N. Ravikumar, and A. Maier, "SkinNet: A deep learning framework for skin lesion segmentation," in *Proc. IEEE Nucl. Sci. Symp. Med. Imag. Conf. (NSS/MIC)*, Nov. 2018, pp. 1–3.

[35] L. Yu, H. Chen, Q. Dou, J. Qin, and P.-A. Heng, "Automated melanoma recognition in dermoscopy images via very deep residual networks," *IEEE Trans. Med. Imag.*, vol. 36, no. 4, pp. 994–1004, Dec. 2016.

[36] H. Wang, G. Wang, Z. Sheng, and S. Zhang, "Automated segmentation of skin lesion based on pyramid attention network," in *Proc. Int. Workshop Mach. Learn. Med. Imag.* Cham, Switzerland: Springer, 2019, pp. 435–443.

[37] 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.

[38] H. Li, X. He, F. Zhou, Z. Yu, D. Ni, S. Chen, T. Wang, and B. Lei, "Dense deconvolutional network for skin lesion segmentation," *IEEE J. Biomed. Health Inform.*, vol. 23, no. 2, pp. 527–537, Mar. 2019.
[39] W. Tu, X. Liu, W. Hu, and Z. Pan, "Dense-residual network with adversarial learning for skin lesion segmentation," *IEEE Access*, vol. 7, pp. 77037–77051, 2019.
[40] M. A. Al-masni, M. A. Al-antari, M.-T. Choi, S.-M. Han, and T.-S. Kim, "Skin lesion segmentation in dermoscopy images via deep full resolution convolutional networks," *Comput. Methods Programs Biomed.*, vol. 162, pp. 221–231, Aug. 2018.
[41] M. Goyal, A. Oakley, P. Bansal, D. Dancey, and M. H. Yap, "Skin lesion segmentation in dermoscopic images with ensemble deep learning methods," *IEEE Access*, vol. 8, pp. 4171–4181, 2020.
[42] M. M. K. Sarker, H. A. Rashwan, F. Akram, S. F. Banu, A. Saleh, V. K. Singh, F. U. Chowdhury, S. Abdulwahab, S. Romani, and P. Radeva, "SLSDeep: Skin lesion segmentation based on dilated residual and pyramid pooling networks," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assisted Intervent.* Springer, 2018, pp. 21–29.
[43] V. K. Singh, M. Abdel-Nasser, H. A. Rashwan, F. Akram, N. Pandey, A. Lalande, B. Presles, S. Romani, and D. Puig, "FCA-Net: Adversarial learning for skin lesion segmentation based on multi-scale features and factorized channel attention," *IEEE Access*, vol. 7, pp. 130552–130565, 2019.
[44] Z. Wei, H. Song, L. Chen, Q. Li, and G. Han, "Attention-based DenseUnet network with adversarial training for skin lesion segmentation," *IEEE Access*, vol. 7, pp. 136616–136629, 2019.
[45] G. Zhang, X. Shen, S. Chen, L. Liang, Y. Luo, J. Yu, and J. Lu, "DSM: A deep supervised multi-scale network learning for skin cancer segmentation," *IEEE Access*, vol. 7, pp. 140936–140945, 2019.
[46] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 2881–2890.
[47] W.-C. Hung, Y.-H. Tsai, X. Shen, Z. Lin, K. Sunkavalli, X. Lu, and M.-H. Yang, "Scene parsing with global context embedding," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2017, pp. 2631–2639.
[48] H. Zhang, K. Dana, J. Shi, Z. Zhang, X. Wang, A. Tyagi, and A. Agrawal, "Context encoding for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7151–7160.
[49] D. Lin, Y. Ji, D. Lischinski, D. Cohen-Or, and H. Huang, "Multi-scale context intertwining for semantic segmentation," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 603–619.
[50] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7132–7141.
[51] S. Ross, D. Munoz, M. Hebert, and J. A. Bagnell, "Learning message-passing inference machines for structured prediction," in *Proc. CVPR*, Jun. 2011, pp. 2737–2744.
[52] J. Carreira, P. Agrawal, K. Fragkiadaki, and J. Malik, "Human pose estimation with iterative error feedback," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 4733–4742.
[53] C.-Y. Lee, S. Xie, P. Gallagher, Z. Zhang, and Z. Tu, "Deeply-supervised nets," in *Proc. 18th Int. Conf. Artif. Intell. Statist.* San Diego, CA, USA: PMLR, 2015, pp. 562–570.
[54] Q. Dou, L. Yu, H. Chen, Y. Jin, X. Yang, J. Qin, and P.-A. Heng, "3D deeply supervised network for automated segmentation of volumetric medical images," *Med. Image Anal.*, vol. 41, pp. 40–54, Oct. 2017.
[55] B. Lei, S. Huang, R. Li, C. Bian, H. Li, Y.-H. Chou, and J.-Z. Cheng, "Segmentation of breast anatomy for automated whole breast ultrasound images with boundary regularized convolutional encoder–decoder network," *Neurocomputing*, vol. 321, pp. 178–186, Dec. 2018.
[56] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*. [Online]. Available: http://arxiv.org/abs/1412.6980
[57] M. M. K. Sarker, H. A. Rashwan, M. Abdel-Nasser, V. K. Singh, S. F. Banu, F. Akram, F. U. H. Chowdhury, K. A. Choudhury, S. Chambon, P. Radeva, and D. Puig, "MobileGAN: Skin lesion segmentation using a lightweight generative adversarial network," 2019, *arXiv:1907.00856*. [Online]. Available: http://arxiv.org/abs/1907.00856
[58] Y. Peng, N. Wang, Y. Wang, and M. Wang, "Segmentation of dermoscopy image using adversarial networks," *Multimedia Tools Appl.*, vol. 78, no. 8, pp. 10965–10981, Apr. 2019.
[59] S. Baghersalimi, B. Bozorgtabar, P. Schmid-Saugeon, H. K. Ekenel, and J.-P. Thiran, "DermoNet: Densely linked convolutional neural network for efficient skin lesion segmentation," *EURASIP J. Image Video Process.*, vol. 2019, no. 1, p. 71, 2019.
[60] F. Xie, J. Yang, J. Liu, Z. Jiang, Y. Zheng, and Y. Wang, "Skin lesion segmentation using high-resolution convolutional neural network," *Comput. Methods Programs Biomed.*, vol. 186, Apr. 2020, Art. no. 105241.
[61] M. Rahman, N. Alpaslan, and P. Bhattacharya, "Developing a retrieval based diagnostic aid for automated melanoma recognition of dermoscopic images," in *Proc. IEEE Appl. Imag. Pattern Recognit. Workshop (AIPR)*, Oct. 2016, pp. 1–7.
[62] Z. Deng, H. Fan, F. Xie, Y. Cui, and J. Liu, "Segmentation of dermoscopy images based on fully convolutional neural network," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2017, pp. 1732–1736.
[63] E. Nasr-Esfahani, S. Rafiei, M. H. Jafari, N. Karimi, J. S. Wrobel, S. Samavi, and S. M. Reza Soroushmehr, "Dense pooling layers in fully convolutional network for skin lesion segmentation," *Computerized Med. Imag. Graph.*, vol. 78, Dec. 2019, Art. no. 101658.
[64] L. Bi, J. Kim, E. Ahn, and D. Feng, "Automatic skin lesion analysis using large-scale dermoscopy images and deep residual networks," 2017, *arXiv:1703.04197*. [Online]. Available: http://arxiv.org/abs/1703.04197
[65] M. Berseth, "ISIC 2017-skin lesion analysis towards melanoma detection," 2017, *arXiv:1703.00523*. [Online]. Available: http://arxiv.org/abs/1703.00523
[66] 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.
[67] C. Kaul, S. Manandhar, and N. Pears, "FocusNet: An attention-based fully convolutional network for medical image segmentation," in *Proc. IEEE 16th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2019, pp. 455–458.
[68] P. Tschandl, C. Sinz, and H. Kittler, "Domain-specific classification-pretrained fully convolutional network encoders for skin lesion segmentation," *Comput. Biol. Med.*, vol. 104, pp. 111–116, Jan. 2019.
[69] L. Bi, D. Feng, and J. Kim, "Improving automatic skin lesion segmentation using adversarial learning based data augmentation," 2018, *arXiv:1807.08392*. [Online]. Available: http://arxiv.org/abs/1807.08392

YUJIAO TANG received the B.S. degree from Southern Medical University, China, in 2018, where she is currently pursuing the M.S. degree with the School of Biomedical Engineering. Her research interests include medical image analysis and machine learning.

ZHIWEN FANG received the B.S. and M.S. degrees from the Automation School, Beihang University, and the Ph.D. degree from the Huazhong University of Science and Technology, China. He was a Research Fellow with the Institute of Media Innovation, Nanyang Technological University, and a Research Scientist with the Institute of High Performance Computing, Research Agency for Science, Technology, and Research, Singapore. He is currently an Associate Professor with the School of Biomedical Engineering, Southern Medical University, China. His research interests include object detection, object tracking, anomaly detection, medical image analysis, and machine learning.

SHAOFENG YUAN received M.S. degree from Southern Medical University, China, in 2018. He is currently an Engineer with Shanghai United Imaging Healthcare Co., Ltd., China. His research interests include machine learning, medical image processing, deep learning, and computer vision.

CHANG'AN ZHAN received the Ph.D. degree in biomedical engineering from the University of Science and Technology of China, Hefei, China, in 2001. He is currently a Professor with the School of Biomedical Engineering, Southern Medical University, China. His research has been published in *The Journal of Neuroscience*, *Cerebral Cortex*, *NeuroImage*, the IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING, *Biomedical Signal Processing and Control*, and among others, in the fields of neuroscience and biomedical signal processing. His current research interests include biomedical signal processing, machine learning, and neural engineering.

YANYAN XING received the B.S. degree from Southern Medical University, China, in 2017, where she is currently pursuing the M.S. degree with the School of Biomedical Engineering. Her research interests include machine learning, computer vision, and medical image analysis.

JOEY TIANYI ZHOU received the Ph.D. degree in computer science from Nanyang Technological University, Singapore, in 2015.

He is currently a Scientist with the Institute of High Performance Computing, Research Agency for Science, Technology, and Research, Singapore.

Dr. Zhou was a recipient of the Best Poster Honorable Mention at ACML 2012, the Best Paper Award from the BeyondLabeler Workshop on IJCAI 2016, the Best Paper Nomination at ECCV 2016, and the NIPS 2017 Best Reviewer Award. He has served as an Associate Editor for IEEE ACCESS and *IET Image Processing*.

FENG YANG received the M.S. degree in biomedical signal and image processing from Sun Yat-sen University, China, in 1993, and the Ph.D. degree in communication and electronic systems from the South China University of Technology, China, in 1998. He joined the Division of Image Processing (LKEB), Leiden University Medical Center, The Netherlands, from April 2010 until April 2011, as a Visiting Scholar. He is currently with the School of Biomedical Engineering, Southern Medical University, China, as a Professor and the Director with the Department of Electronic Technology. His research interests include wavelet analysis, medical image processing, and pattern recognition.