Copyright ⓒ 2024 KSII

Multi-scale context fusion network for melanoma segmentation

Zhenhua Li, and Lei Zhang*
School of Electrical & Information Engineering, Jiangsu University of Technology
Changzhou, Jiangsu 213001 China
[e-mail: zhlei@jsut.edu.cn]
*Corresponding author: Lei Zhang

Received November 16, 2022; revised October 1, 2023; revised April 13, 2024; accepted June 24, 2024; published July 31, 2024

Abstract

Aiming at the problems that the edge of melanoma image is fuzzy, the contrast with the background is low, and the hair occlusion makes it difficult to segment accurately, this paper proposes a model MSCNet for melanoma segmentation based on U-net frame. Firstly, a multi-scale pyramid fusion module is designed to reconstruct the skip connection and transmit global information to the decoder. Secondly, the contextural information conduction module is innovatively added to the top of the encoder. The module provides different receptive fields for the segmented target by using the hole convolution with different expansion rates, so as to better fuse multi-scale contextural information. In addition, in order to suppress redundant information in the input image and pay more attention to melanoma feature information, global channel attention mechanism is introduced into the decoder. Finally, In order to solve the problem of lesion class imbalance, this paper uses a combined loss function. The algorithm of this paper is verified on ISIC 2017 and ISIC 2018 public datasets. The experimental results indicate that the proposed algorithm has better accuracy for melanoma segmentation compared with other CNN-based image segmentation algorithms.

Keywords: Melanoma segmentation, Receptive field, Attention mechanism, multi-scale contextural information.

1. Introduction

The growth rate and mortality of patients with skin cancer are high [1]. Melanoma is the leading cause of death in most skin cancer patients. However, studies have shown that for most patients, if melanoma is diagnosed early and timely, resection can be used to remove it, thereby improving the survival rate of patients [2]. Although melanoma can be identified visually, even experienced dermatologists may be misdiagnosed because of the pigmentation of skin surface lesions caused by melanoma. In addition to this traditional method, dermatoscopy is a general method in the early diagnosis of melanoma. Dermatoscope can enhance the visual effect of deep skin, enabling dermatologists to diagnose melanoma that is invisible to the naked eye. However, due to the complexity of skin lesions and the sheer volume of images, manually checking these dermoscopic melanoma images is a laborious exercise [3]. Thereby, the use of computer-aided technology to achieve accurate melanoma image segmentation has great practical significance in pathological diagnosis.

At present, the methods of melanoma image segmentation are mainly divided into two categories: traditional image segmentation methods and image segmentation algorithms based on deep learning. The traditional segmentation methods contain threshold segmentation [4], edge detection [5], pixel clustering segmentation [6] and so on. The threshold segmentation method sets the threshold value to divide the pixel information of melanoma image, so as to segment the lesions. Garnavi in [7] first selected the appropriate color channel, and then used the threshold method to segment the lesion boundary. Sforza in [8] used adaptive threshold method to segment melanoma images. Although the threshold method is simple to operate, it ignores the spatial position relationship of pixels, which makes some information of melanoma images lost. The edge detection method uses the first order or second order differential operator to obtain the pixel value mutation points on the image, and then connect these pixel points to obtain the boundary of the region. In [9], the authors used Canny algorithm combined with partial differential equation segmentation method to segment the boundary of skin lesion area. This kind of algorithm is greatly affected by noise, resulting in wrong edge segmentation results. Pixel clustering segmentation method gathers the pixels with high similarity in the image into a region, thus forming the target segmentation region. Zhou in [10] used mean shift clustering algorithm to segment skin lesions. This algorithm requires manual initialization of pre parameters, and the segmentation speed is slow. Melanoma has problems such as different sizes and shapes of lesions, blurred boundaries, and unclear feature textures. In the field of melanoma image segmentation, traditional segmentation methods have not been widely used.

In recent years, convolutional neural networks have some achievements in the domain of medical image segmentation. The authors of [11] designed a U-shaped network, which can fuse deep semantic information with shallow fine-grained information, and has a good effect in biomedical image segmentation. To capture multi-scale spatial information, Xue in [12] proposed a multi-scale context fusion module between the U-Net network encoder and decoder. Feng in [13] used CswinUnet instead of U-Net 's original encoder to effectively improve the network's ability to model contextual information. In [14], the authors proposed self-adaption feature learning network to learn the image features of skin lesions, and then adopted a step-by-step training strategy to work out the problem of sample imbalance. The authors of [15] used the gate attention mechanism and feature fusion module to increase the receptive field of the network model, effectively improving the positioning precision of segmentation.

Although the existing U-Net model has achieved good results, this codec structure has insufficient ability to extract contextual information at each stage, continuous pooling operations have lost a lot of spatial information, and jump connection cannot explore multi-

scale information. In addition, melanoma segmentation often faces the challenges of low contrast, blurred boundaries, and hair occlusion, which makes it difficult for a simple U-Net network to extract advanced feature information, resulting in unsatisfactory segmentation results. Therefore, A new multi-scale context fusion network (MSCNet) is proposed to fuse global context information. First, to provide richer global context information to the decoder, four identical multi scale pyramid fusion modules (MSPF) are proposed to reconstruct the skip connections. Then, a context information conduction module (CIC) is innovatively designed, which is embedded at the top of the encoder. By using the dilated convolution of three cascaded branches, different sizes of receptive fields are provided for the segmentation target to better integrate multi-scale background information. In addition, adding global channel attention (GCA) to the decoder enables the network to raise more concern about the learning of melanoma features and suppress interference from irrelevant information. In the end, combined loss function is applied to alleviate the class imbalance problem.
The main contributions of this paper are summarized in five aspects:

(1) The MSPF module is innovatively designed to reconstruct the skip connection. The module is improved on the basis of ASPP. It uses dense connections to enhance feature propagation and shallow feature reuse, and adaptively transforms the features extracted by the encoder to provide global information for the decoder.
(2) Based on the idea of RFB and SPP module, this paper proposes a DAC module and a MSP module. The proposed DAC block and MSP block are combined into a CIC module, which is embedded at the top of the encoder to obtain rich global feature information and provide multi-scale receptive fields for segmenting lesions.
(3) This paper proposes GCA attention, which introduces spatial attention on the basis of ECA channel attention. It can not only highlight the features of important channels, but also explore the spatial relationship between features, improve the attention to the edge contour of melanoma, so as to achieve accurate segmentation of melanoma.
(4) The combined loss used in this paper is the weighted sum of BCE loss and Dice loss, so as to solve the problem of category imbalance.
(5) A large number of experiments have been carried out on the public skin lesion datasets ISIC 2017 and ISIC 2018. Compared with other advanced segmentation algorithms, the superiority of this method is proved.

2. Proposed Method

2.1 Network model

Although the U-Net model has achieved excellent segmentation results for most image segmentation tasks, it does not perform well in the segmentation task of melanoma. Due to melanoma there will be blurred boundaries, resulting in melanoma and the surrounding class is difficult to distinguish. In addition, the size and noise of each lesion area are different, which makes it difficult to segment melanoma. If U-Net is used directly to segment melanoma, the segmentation effect is not good. The main reason is that the encoder in U-Net uses pooling operations many times, which could lead to much spatial information loss. meanwhile, the U-Net encoder 's ability to obtain context information is insufficient. Jump connections between encoder and decoder can only pass the same level features to the decoder for fusion. It ignores the global context information and does not explore more information from a full-scale perspective. In order to effectively segment melanoma in medical images, based on the structure of U-Net coding-decoding, we present a multi-scale context fusion network MSCNet.

The MSCNet network structure is shown in **Fig. 1**.

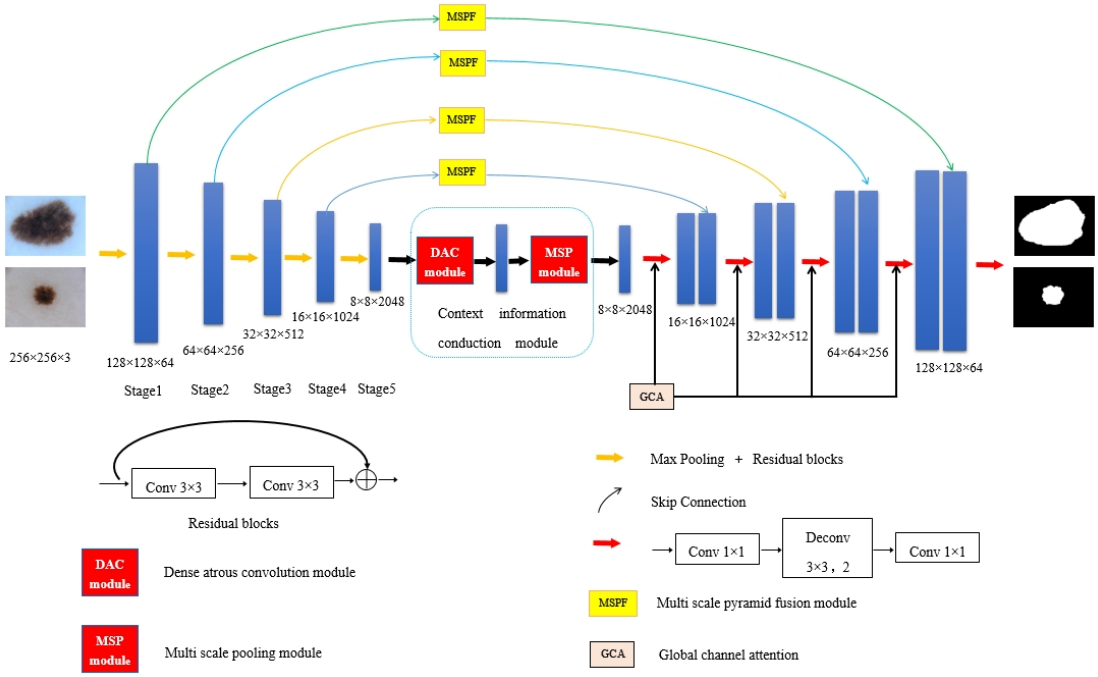

Fig. 1. Overall structure of MSCNet

The network MSCNet proposed in this paper mainly includes five parts: encoder, context information conduction module, multi-scale pyramid fusion module, global channel attention and decoder. Aiming at the problem that U-Net has insufficient ability to extract context information, this paper first innovatively designs multiple same multi-scale pyramid fusion modules. The module reconstructs the skip connection and adaptively transforms the features extracted by the encoder, so that the feature information in the decoder merges the global information. To capture multi-scale context semantic information, a context information conduction module is designed and embedded at the top of the encoder. To make the model concern the small characteristics of melanoma, this paper further innovatively designs the global channel attention, which is based on the efficient channel attention (ECA) module [16]. Unlike the ECA module, GCA adds a global maximum pooling to extract the global information of the channel. Furthermore, it can be seen from the literature [17] that spatial attention plays a key role in determining where the network pays attention, while the ECA module also ignores spatial information. In order to make the model pay attention to the spatial location information of the lesion at the same time, the spatial attention is embedded in the global channel attention to achieve accurate segmentation of melanoma.

In addition, the U-Net original encoder is replaced by the pre-trained ResNet50, so that the detailed information of melanoma can be learned more fully. While distinguishing melanoma from background better, it can also avoid the disappearance of gradient in the training procedure of the network, and the precision of melanoma segmentation can be improved.

2.2 Context information conduction module

Since the boundary of melanoma is fuzzy and the similarity between melanoma and background is high, when U-Net is used to segment melanoma, it is prone to over-segmentation and under-segmentation. Rich contextual global information is critical when dealing with complex medical images. Although shallow low-level features can be fused with high-level features in the decoder through skip connections, the continuous down-sampling operation in the encoder results in loss of global information. Therefore, this paper innovatively designs a context information conduction module, which is embedded at the top of the encoder to obtain rich global feature information and provide multi-scale receptive fields for segmentation targets. The module is composed of DAC module and MSP module. The U-Net encoder uses 3×3 convolution and pooling operations, which can only capture feature information in a limited range. Inspired by Receptive field block [18] and atrous convolution, as shown in **Fig. 2**, this paper proposes a DAC module to encode advanced features. The module embeds different scales of dilated convolutions into three cascaded branches to effectively capture deeper semantic features. In the DAC module, a residual connection is used to avoid gradient disappearance. In addition, the size of the lesion area in a medical image can change. Aiming at the problem that the lesion area of advanced skin cancer is larger than that of early stage, this paper innovatively designs a multi-scale pooling (MSP) module, the module mainly segments targets of different sizes according to different receptive fields. The size of the receptive field directly affects how much context information the model extracts. Typically, the model uses only the maximum pooling of a single pooling kernel, such as 3×3. The MSP module uses the maximum pooling of three different pooling kernels to encode the global context information, which is shown in **Fig. 3**. And then splices with the input feature map. Furthermore, to decrease the computation of the model, after the splicing operation, 1×1 convolution is used to decrease the dimension. In summary, the DAC module uses multi-scale atrous convolution to obtain rich feature information, and then the MSP module uses the maximum pooling operation of different pooling kernels to extract depth context information.

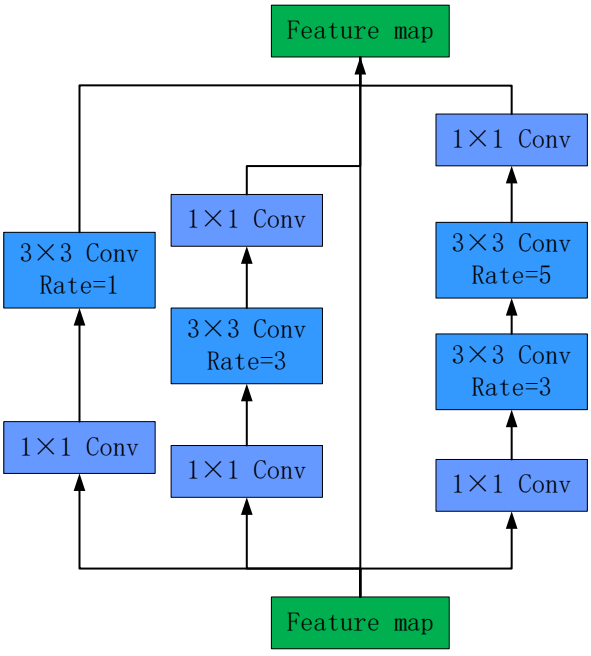

Fig. 2. Dense atrous convolution module

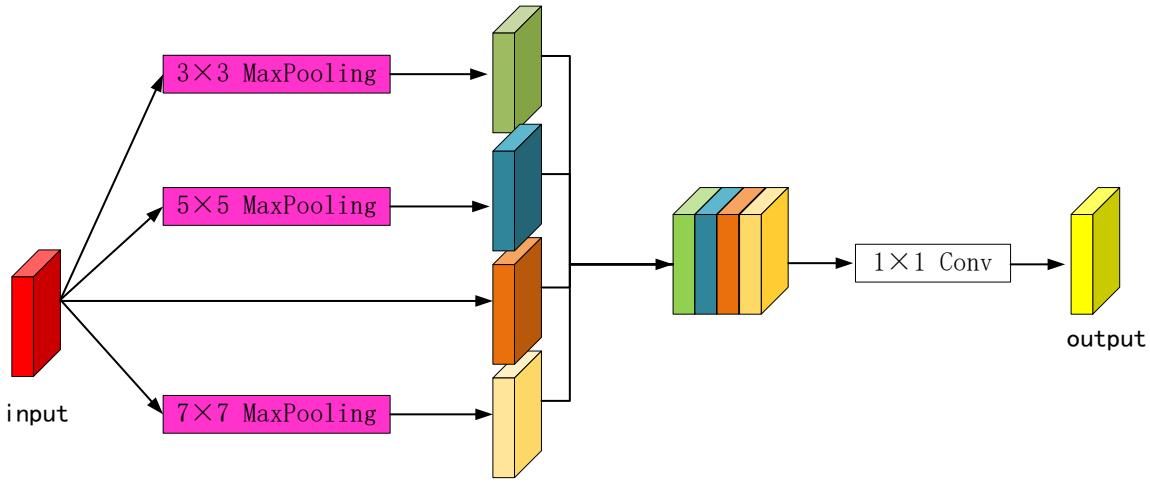

Fig. 3. Multi scale pooling module

2.3 Multi scale pyramid fusion module

The encoder of U-Net can learn its feature information from the input melanoma image, including the surrounding environment features and categories, and obtain the corresponding location information at the same time. For a single stage of the U-shaped network, the ability to extract context information is relatively weak, which may cause a lot of global information to be lost when it is transmitted to the shallow layer. Moreover, the skip connection in U-Net is only a simple superposition. The downsampling used in each stage of the encoder ignores the context information, and also introduces noise and irrelevant clutter. Therefore, this paper designs a new multi-scale pyramid fusion module, through which the skip connection is reconstructed, so that the information of the current stage can be transmitted to the decoder through the MSPF module, which brings advanced features to the decoder. In addition, the MSPF module can also suppress the background noise caused by low-level features to a certain extent, avoiding useless noise in the segmentation results. **Fig. 4** below is a schematic of the MSPF module. The module is a dual-branch network. The main branch uses two 3×3 convolutions and uses dense connections, which can enhance feature propagation and shallow feature reuse. The side branch draws on the idea of ASPP [19]. Firstly, the channel is reduced by convolution of 3×3, and then the obtained feature is subjected to 1×1 convolution, 3×3 atrous convolution with expansion rate of 6, 3×3 atrous convolution with expansion rate of 12 and pooling operation. In order to encode the feature map, the feature maps obtained by the four branch operations are spliced along the channel dimension, and 1×1 convolution is used for channel dimension reduction. In addition, in order to provide multi-scale global information for the decoder, the main branch and the side branch are spliced and 1×1 convolution is used to reduce the dimension of the channel. Considering the computational cost of the model, this paper uses four designed MSPF modules to reconstruct the jump connection. Through this module, the global context information of the current stage and the higher stage can be transmitted to the decoder, which improves the segmentation accuracy of melanoma.

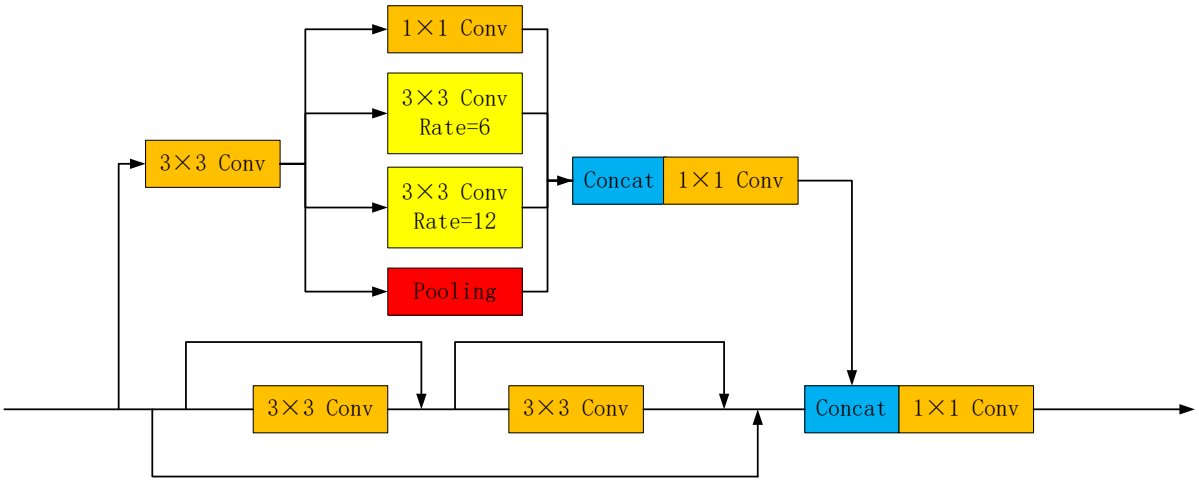

Fig. 4. Multi scale pyramid fusion module

2.4 Global Channel Attention

U-Net uses skip connections to splice the high-resolution shallow features extracted by the encoder with the deep features extracted by the decoder, although the high-resolution shallow features can make up for the lost spatial information in the coding phase. However, the features extracted from the coding part are also transmitted to the decoder by encoding-decoding, so there is information redundancy. The attention mechanism can suppress irrelevant redundant information and enrich context information [20]. In order to focus on the feature information related to melanoma, this paper further adds Global Channel Attention (GCA) to the decoder. **Fig. 5** shows the GCA module structure. The module adds parallel global maximum pooling on the basis of ECA channel attention to retain more spatial information, and also connects the spatial attention module in series. It can not only study the relationship between channels and highlight important feature channels, but also explore the spatial relationship between features, which improves the attention to the edge contour of melanoma, so as to achieve accurate segmentation of melanoma. In general, given the input feature layer $F \in R^{H\times W\times C}$, in order to compress the spatial information of the feature map , global average pooling and the global maximum pooling is used, so as to obtain two different spatial information descriptors : $F_{avg}^{c} \in R^{1\times 1\times C}$ and $F_{\max}^{c} \in R^{1\times 1\times C}$. After using the one-dimensional convolution of the shared weight for the feature descriptor, we use the element-by-element summation and the sigmoid function. Finally, the channel attention weight vector $\alpha \in R^{1\times 1\times C}$ is obtained. The average pooling and maximum pooling calculation formulas are shown in equations 1 and 2, and the channel attention module CA calculation formula is shown in equation 3. Among them, H, W and C are high, wide and channel, respectively. σ is the sigmoid function, and Con1D is the one-dimensional convolution.

$$F_{avg}^{c} = \frac{1}{H\times W}\sum_{i=1}^{H}\sum_{j=1}^{W}F(i,j) \tag{1}$$

$$F_{\max}^{c} = Max(F(i,j)) \tag{2}$$

$$\alpha = \sigma(Con1D(F_{avg}^{c}) + Con1D(F_{\max}^{c})) \tag{3}$$

In order to allocate spatial attention weight on each pixel, this paper additionally uses spatial attention block SA, which uses $F \cdot \alpha$ as the input feature map to generate spatial attention weight $\beta \in R^{H \times W \times 1}$. The SA module consists of a 3×3 and a 1×1 convolutional layer. The final output of our GCA module is

$$F' = F \times \alpha + F \times \alpha \times \beta + F \tag{4}$$

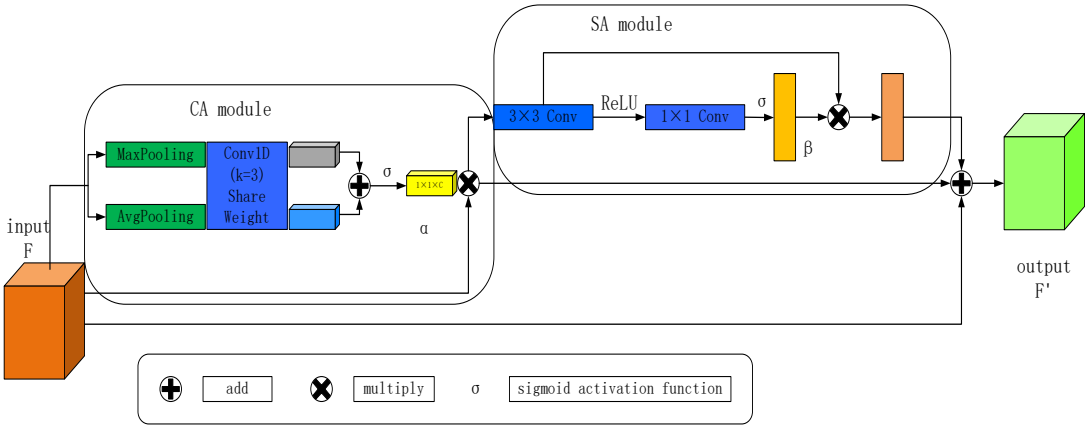

Fig. 5. Global Channel Attention module

2.5 Loss function

The segmentation task of melanoma is a binary classification problem from the perspective of pixels. The image is mainly composed of melanoma and background. Therefore, the binary cross entropy (BCE) function is usually used as a loss function. The calculation formula is as follows

$$L_{BCE} = -\sum_{i}[(1-y_i)\log(1-x_i) + y_i \log x_i] \tag{5}$$

Among them, y_i and x_i are the label value and the predicted value at the *i*-th pixel position, respectively.

However, in the segmentation task of melanoma, the background part is larger than the lesion part, which will lead to the imbalance of categories. The model will learn more background features, which will affect the feature learning of melanoma. At this time, the effect of single BCE loss function is not very ideal. The Dice loss function calculation formula is

$$L_{Dice} = 1 - \frac{2 \times \sum_i y_i x_i + \gamma}{\sum_i y_i + \sum_i x_i + \gamma} \tag{6}$$

The Dice loss function can evaluate the classification accuracy of melanoma and background pixels in melanoma images, thus solving the problem of class imbalance to a certain extent. where γ is a smoothing parameter to prevent the loss function from having a denominator of 0. Combining the characteristics of Dice loss function and BCE loss function, this paper proposes a joint loss function as follows

$$L = L_{BCE} + L_{Dice} \tag{7}$$

The joint loss function makes the network more efficient and stable in training, effectively alleviates the imbalance between melanoma and background categories, and improves the segmentation accuracy of melanoma.

3. experiments and analysis

3.1 Experimental environment

The experiment in this paper is based on the Pytorch deep learning framework. The Pycharm compiler is used. The CPU is Intel Core i5-11400. The GPU is configured as NVIDIA GeForce GTX 3060 with 12 GB memory. CUDA uses 11.1 version. The operating system is Windows 10 and the programming language is Python 3.7 of Anconda3.

3.2 Experimental dataset

In this paper, the segmentation performance of the proposed network model for melanoma is evaluated on two public skin lesion datasets. The first dataset is the skin lesion segmentation datase from the Kaggle competition platform of International Skin Imaging Cooperation Organization (ISIC) 2017 [21]. The dataset contains 2000 original images of different types of skin lesions such as melanoma, nevus and seborrheic keratosis. The original size of each image is 576 x 767 and all have a label map marked by a professional doctor. Using the data set division method in [22]. in this paper,1250 images are used for training, 150 images are used for verification and 600 images as a test set. The other is the ISIC 2018 melanoma segmentation data set released by [23]. In this dataset, there are 2594 melanoma images, and the resolution of each image is 2016x3024. The 2594 pictures are divided into three parts: 1815 for training, 259 for verification, and 520 for testing. The images and corresponding labels of the experimental dataset are shown in **Fig. 6** and **Fig. 7**.

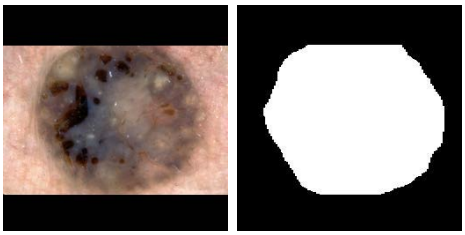

Fig. 6. ISIC2017 dataset and its annotation image

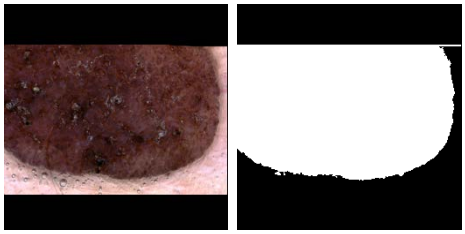

Fig. 7. ISIC2018 dataset and its annotation image

3.3 Data preprocessing and related settings

In the domain of deep learning, a large number of datasets is vital for the network models to train, considering the small number of datasets in this experiment, the model will have over-fitting. Therefore, to heighten the generalization capability of the model, the data needs to be expanded. The training set image and the corresponding label image are subjected to 0 to 360 degrees of random rotation, contrast enhancement and other data enhancement. As shown in **Fig. 8**, the first column is the original image, the second column is a randomly rotated image from 0 to 360 degrees, the third column is a brightness-enhanced image, and the fourth column is a contrast-enhanced image. Aiming at the problem that the original image is large, to save computing resources, the algorithm adjusts the image size to a resolution of 256×256.
In the experiment, the images and real labels of the preprocessed data set are sent to the network for training. In the training process of the model, the stochastic gradient descent (SGD) optimization algorithm is used to iteratively update the parameters. The learning rate is set to 0.001, and the learning rate is dynamically adjusted during the training process. The weight decay and the batch size is set to 0.0001 and 16 respectively, the momentum parameter and the epoch is set to 0.9 and 100 respectively. When the loss function decreases and tends to be stable, the training is stopped, and the weight with the smallest loss is selected for testing to obtain the best model performance data.

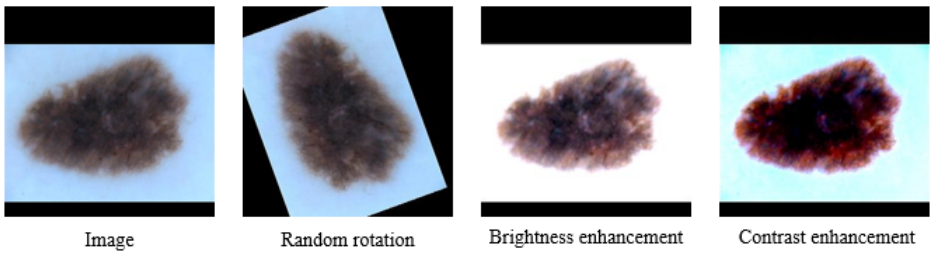

Fig. 8. Data enhancement

3.4 Contrast experiment

In order to evaluate the segmentation performance of different algorithms for melanoma, this paper uses common metrics such as accuracy, specificity, IoU, dice, precision and recall. In addition, this paper uses the receiver operating characteristic (ROC) and precision-recall (P-R) curve to further evaluate the performance of the algorithm.The results of melanoma detection on the ISIC 2017 dataset are shown in **Table 1**. The MSCNet is compared with the six image segmentation algorithms of U-net [11], Attention U-net [24], U-net ++ [25], CA-net [26], R2U-net [27] and PraNet [28] under the same conditions. The experimental results of CPFNet [29], SESV [30], MB-DCNN [31] and DAGAN [32] are from FAT-Net [33], and the Precision, PR and ROC of these methods are not given in FAT-Net [33]. The experimental results show that the MSCNet is the best in Accuracy (0.9784), Recall (0.9111), IoU (0.8347), Dice (0.9099), area under the ROC curve (0.9529) and area under the PR curve (0.9134). The ROC curve and P-R curve are shown in **Fig. 9**. The most obvious improvement is that the area under the ROC curve is 0.024 higher than that of U-net [11]. Although R2U-net [27] is superior to MSCNet in specificity and accuracy, the other six evaluation indicators are significantly lower than MSCNet, and the recall rate is the most obvious, 0.0619 lower than MSCNet in this paper.

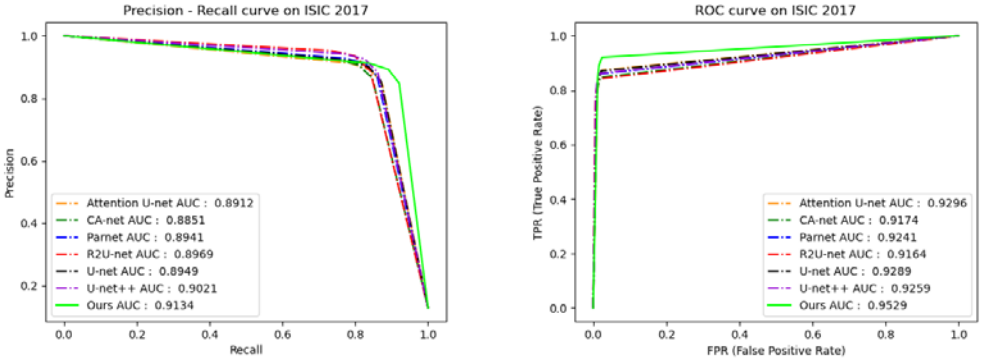

Fig. 9. ROC and PR curves on ISIC 2017 dataset

Table 1. ISIC 2017 dataset detection results and comparative experiment

Method	Year	Accuracy	Recall	Specificity	Precision	Dice	IoU	PR	ROC
U-net[11]	2015	0.9696	0.8831	0.9813	0.8655	0.8742	0.7765	0.8949	0.9289
Attention U-net[24]	2018	0.9694	0.8860	0.9807	0.8619	0.8738	0.7759	0.8912	0.9296
R2U-net[27]	2018	0.9721	0.8492	**0.9887**	**0.9110**	0.8790	0.7842	0.8969	0.9164
U-net++[25]	2019	0.9731	0.8776	0.9860	0.8952	0.8863	0.7959	0.9021	0.9259
CA-net[26]	2020	0.9676	0.8545	0.9829	0.8718	0.8630	0.7591	0.8851	0.9174
PraNet[28]	2020	0.9707	0.8758	0.9835	0.8784	0.8771	0.7823	0.8941	0.9241
CPFNet[29]	2020	0.9215	0.8344	0.9645	-	0.8403	0.7546	-	-
SESV[30]	2020	0.9223	0.8326	0.9668	-	0.8392	0.7531	-	-
MB-DCNN[31]	2020	0.9311	0.8325	0.9684	-	0.8427	0.7603	-	-
DAGAN[32]	2020	0.9304	0.8363	0.9716	-	0.8425	0.7594	-	-
FAT-Net[33]	2021	0.9326	0.8392	0.9725	-	0.8500	0.7653	-	-
Ours	2022	**0.9784**	**0.9111**	0.9875	0.9087	**0.9099**	**0.8347**	**0.9134**	**0.9529**

Table 2 shows the experimental results on ISIC 2018 dataset. MSCNet was compared with 11 other image segmentation algorithms. The experimental results of CPFNet [29], DAGAN [32], Resunet + + [34], CKDNet [35] are from FAT-Net [33], and the Precision, PR and ROC of these methods are not given in FAT-Net [33]. From the experimental results, the MSCNet is the best in accuracy (0.9658), intersection ratio (0.8267), Dice (0.9051), ROC AUC (0.9362) and P-R AUC (0.9294). Among them, compared with FAT-Net [33], Dice increased by 0.0148. Similarly, this paper gives the ROC-PR curve. As shown in **Fig.10**, the area under the ROC curve and the area under the PR curve of the algorithm are 0.9362 and 0.9294, respectively.

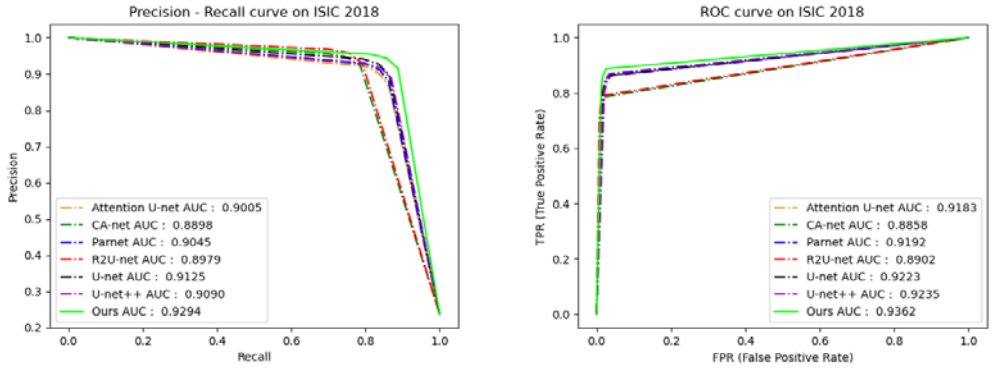

Fig. 10. ROC and PR curves on ISIC 2018 dataset

Table 2. ISIC 2018 dataset detection results and comparative experiment

Method	Year	Accuracy	Recall	Specificity	Precision	Dice	IoU	PR	ROC
U-net[11]	2015	0.9569	0.8874	0.9724	0.8775	0.8824	0.7896	0.9125	0.9223
Attention U-net[24]	2018	0.9497	0.8680	0.9679	0.8578	0.8629	0.7588	0.9005	0.9183
R2U-net[27]	2018	0.9528	0.7959	**0.9877**	**0.9355**	0.8601	0.7545	0.8979	0.8902
U-net++[25]	2019	0.9568	0.8811	0.9737	0.8819	0.8815	0.7881	0.9090	0.9235
Resunet++[34]	2019	0.9382	0.8735	0.9721	-	0.8536	0.7721	-	-
CA-net[26]	2020	0.9507	0.7876	0.9871	0.9314	0.8535	0.7444	0.8898	0.8858
PraNet[28]	2020	0.9544	0.8664	0.9739	0.8811	0.8737	0.7757	0.9045	0.9192
CPFNet[29]	2020	0.9496	0.8953	0.9655	-	0.8769	0.7988	-	-
DAGAN[32]	2020	0.9324	0.9072	0.9588	-	0.8807	0.8113	-	-
CKDNet[35]	2021	0.9492	0.9055	0.9701	-	0.8779	0.8041	-	-
FAT-Net[33]	2021	0.9578	**0.9100**	0.9699	-	0.8903	0.8202	-	-
Ours	2022	**0.9658**	0.8961	0.9813	0.9184	**0.9051**	**0.8267**	**0.9294**	**0.9362**

3.5 Ablation experiment

For the sake of verifying the effectiveness of each module and fusion strategy in the proposed algorithm, this experiment takes U-Net as the Baseline, and conducts ablation experiments on the MSPF module that fuses skip connections, the CIC module between codecs, and the GCA module embedded in the decoder on the ISIC 2017 dataset and the ISIC 2018 dataset. From the **Table 3** and **Table 4**, it can be seen the segmentation results of different fusion strategies, and the bold text is the optimal value of each column index.

Table 3. Ablation experiments of MSCNet on ISIC 2017 datasets

Method	Accuracy	Recall	Specificity	Precision	Dice	IoU	PR	ROC
Baseline	0.9596	0.9280	0.9639	0.7772	0.8459	0.7330	0.8569	0.9459
Baseline+CIC	0.9736	0.8975	0.9839	0.8837	0.8905	0.8027	0.8967	0.9407

Baseline+MSPF	0.9755	0.9164	0.9835	0.8831	0.8994	0.8173	0.9047	0.9499
Baseline+GCA	0.9739	0.9369	0.9790	0.8583	0.8958	0.8114	0.9013	0.9493
Baseline+CIC+MSPF	0.9756	0.9081	0.9847	0.8900	0.8990	0.8165	0.9045	0.9464
Baseline+CIC+GCA	0.9754	0.9027	0.9852	0.8925	0.8976	0.8142	0.9034	0.9439
Baseline+GCA+MSPF	0.9741	**0.9376**	0.9791	0.8588	0.8965	0.8124	0.9019	0.9483
Ours	**0.9784**	0.9111	**0.9875**	**0.9087**	**0.9099**	**0.8347**	**0.9134**	**0.9529**

Table 4. Ablation experiments of MSCNet on ISIC 2018 datasets

Method	Accuracy	Recall	Specificity	Precision	Dice	IoU	PR	ROC
Baseline	0.9538	0.8190	**0.9838**	0.9143	0.8659	0.7635	0.8852	0.9014
Baseline+CIC	0.9587	0.8782	0.9766	0.8934	0.8857	0.7949	0.8969	0.9274
Baseline+MSPF	0.9545	0.9063	0.9652	0.8529	0.8788	0.7838	0.8881	0.9357
Baseline+GCA	0.9596	0.8627	0.9812	0.9111	0.8862	0.7957	0.8994	0.9219
Baseline+CIC+MSPF	0.9627	0.9034	0.9759	0.8930	0.8982	0.8152	0.9070	0.9296
Baseline+CIC+GCA	0.9610	0.9068	0.9731	0.8826	0.8945	0.8092	0.9032	0.9300
Baseline+GCA+MSPF	0.9583	**0.9116**	0.9687	0.8665	0.8885	0.7994	0.8971	0.9302
Ours	**0.9658**	0.8961	0.9813	**0.9184**	**0.9051**	**0.8267**	**0.9294**	**0.9362**

From the data in the table, it can be found that on the basis of Baseline, the effect of adding any designed module is better than baseline.Meanwhile, the effect of the baseline with two modules is also better than that of adding only a single module. The performance of the MSCNet exceeds other image segmentation algorithms. On the basis of adding CIC and MSPF modules, the GCA attention module is embedded, which makes the model take notice of the edge contour information of melanoma, strengthen the attention to useful information, and achieve the best segmentation effect. On the ISIC 2017 dataset, the Dice value reached 90.99% and the accuracy rate reached 97.84 %. On the ISIC 2018 dataset, the Dice value reached 90.51 % and the accuracy rate reached 96.58 %. This also indirectly proves that the CIC, MSPF and GCA modules designed in this paper are effective in improving the segmentation performance of melanoma. Although it can be seen from **Table 4** that the specificity is lower than the baseline, which means that the misdiagnosis rate of the MSCNet will be slightly higher than the baseline. The accuracy, IOU, Dice, PR and ROC of the MSCNet are far higher than the baseline, and the segmentation results are also far better than the baseline, which is enough to prove the superiority of the MSCNet. In subsequent studies, we will focus on the improvement of specificity to reduce the rate of misdiagnosis. For example, the introduction of more negative samples in the training data can help the model learn more background information, thereby improving specificity. Or adjust the network, increase the depth of the network, and introduce more nonlinear transformations to enhance the specificity of the network.

In order to compare the effects of different loss functions, ablation experiments were performed on the ISIC 2017 and ISIC 2018 datasets for BCE loss, Dice loss, and BCE+Dice loss, respectively. The results of different loss functions on ISIC 2017 datasets are shown in **Table 5**. It can be seen that the performance indicators obtained by BCE+Dice loss used in this paper are the highest. Among them, Dice is 0.019 higher than Dice loss, and IOU is 0.0397 higher than BCE loss. From the data in **Table 6**, it can be seen that the indicators have

improved. The experimental results show that the use of BCE+Dice loss is superior to the use of the other two loss functions on the ISIC 2017 and 2018 datasets.

Table 5. Results of segmentation with different loss functions on ISIC 2017 datasets

Method	Loss	Accuracy	Recall	Specificity	Precision	Dice	IoU	PR	ROC
Ours	BCE	0.9709	0.9049	0.9744	0.8336	0.8858	0.7950	0.8926	0.9396
Ours	Dice	0.9735	0.9060	0.9826	0.8763	0.8909	0.8033	0.8967	0.9443
Ours	BCE+Dice	**0.9784**	**0.9111**	**0.9875**	**0.9087**	**0.9099**	**0.8347**	**0.9134**	**0.9529**

Table 6. Results of segmentation with different loss functions on ISIC 2018 datasets

Method	Loss	Accuracy	Recall	Specificity	Precision	Dice	IoU	PR	ROC
Ours	BCE	0.9613	0.8726	0.9699	0.8724	0.8968	0.8129	0.9045	0.9263
Ours	Dice	0.9640	0.8828	0.9721	0.9065	0.8993	0.8171	0.9103	0.9324
Ours	BCE+Dice	**0.9658**	**0.8961**	**0.9813**	**0.9184**	**0.9051**	**0.8267**	**0.9294**	**0.9362**

3.6 Visual analysis

In order to illustrate the superiority of the proposed algorithm in melanoma segmentation, five images were randomly selected in the ISIC 2017 testset and the segmentation effect maps of different algorithms were given, **Fig.11** shows the results. Obviously, melanoma image irregular shape, incomplete edges, features similar to the surrounding background, so melanoma detection is a challenging problem. It can be seen from the first row diagram that only the algorithm in this paper and CA-net [27] have better detection results, but CA-net [27] is affected by blood on the skin surface, resulting in a subtle gap in the lower edge detection. The algorithm in this paper adds attention, and the network can suppress the interference of irrelevant factors, so as to notice the edge contour information of melanoma. The boundary of the second and fifth lines of melanoma images is blurred, which causes great interference to the detection. From the detection results, only MSCNet can robustly segment melanoma, and the detection results of the other five algorithms are incomplete.

The detection results of melanoma on the ISIC 2018 dataset are visualized as shown in **Fig.12**. The melanoma image in the ISIC 2018 dataset has the interference of multiple body hairs and skin colors, and the background is more complex. Therefore, melanoma detection in this dataset is also a serious challenge. From the first row diagram, it can be seen that although the six algorithms can detect a good result, the U-net ++ [26] algorithm has a false detection, and the remaining five algorithms are not detailed enough for the edge contour detection of melanoma. Only the MSCNet in this paper can detect the detailed edge contour. Obviously, from the second row, when the melanoma area is large, most of the missed detection occurs in each algorithm. The algorithm in this paper introduces the context information transmission module, which can effectively capture more advanced information and achieve the best segmentation effect. From the third, fourth and fifth row graphs, it can be seen that all six algorithms have false detection. The algorithm in this paper designs a multi-scale pyramid fusion module, which can effectively extract and utilize multi-scale context information. This information is necessary for accurate segmentation of targets and can avoid fuzzy decisions.

In order to further verify the validity of each module in the model, this paper visualizes the heat map of the model of the ablation experiment, so as to understand the relevant features of

the segmentation that the model focuses on. The heat map is shown in **Fig.13**. It can be clearly seen from the diagram that the baseline only focuses on a small part of the lesion, while adding any module of CIC, MSPF and GCA can increase the attention to the lesion, and the pairwise combination of modules is also better than a single module. Finally, the model in this paper pays the most comprehensive attention to the lesion and is also optimal.

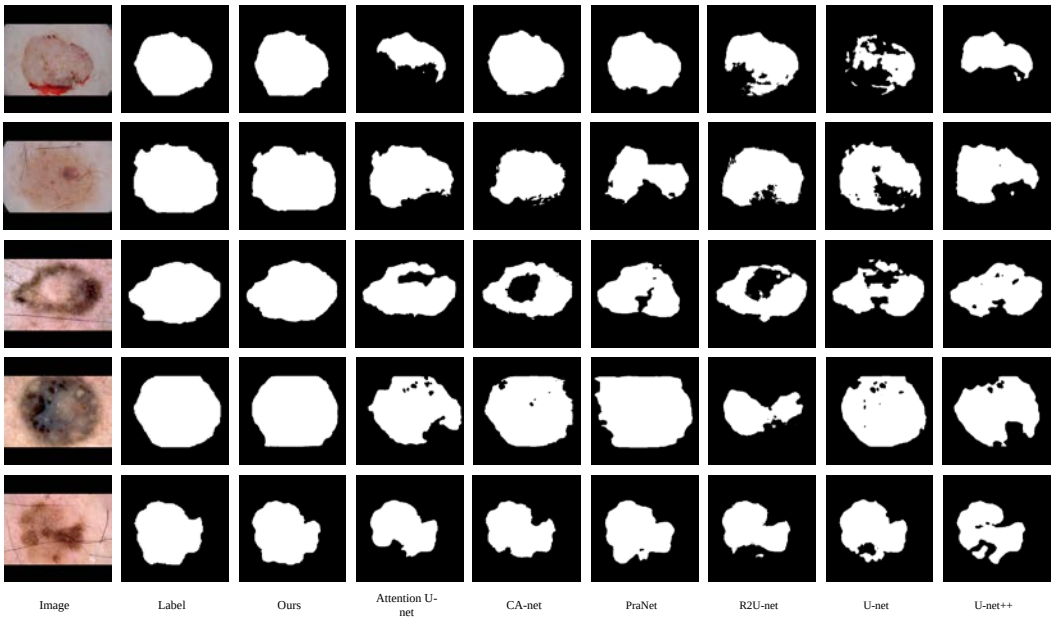

Fig. 11. Segmentation results of different algorithms on ISIC 2017 dataset

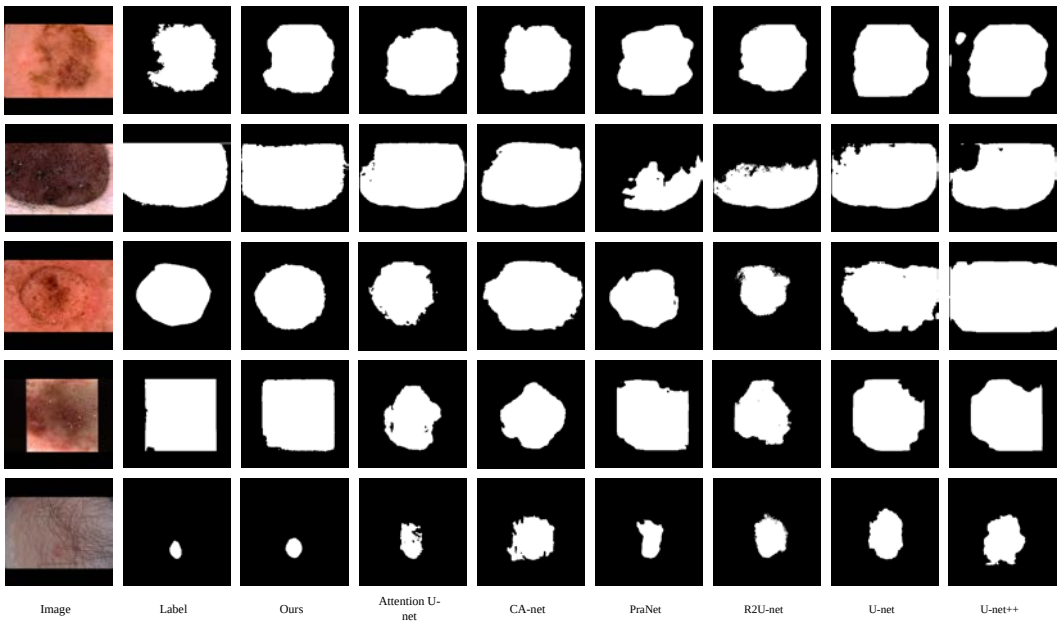

Fig. 12. Segmentation results of different algorithms on ISIC 2018 dataset

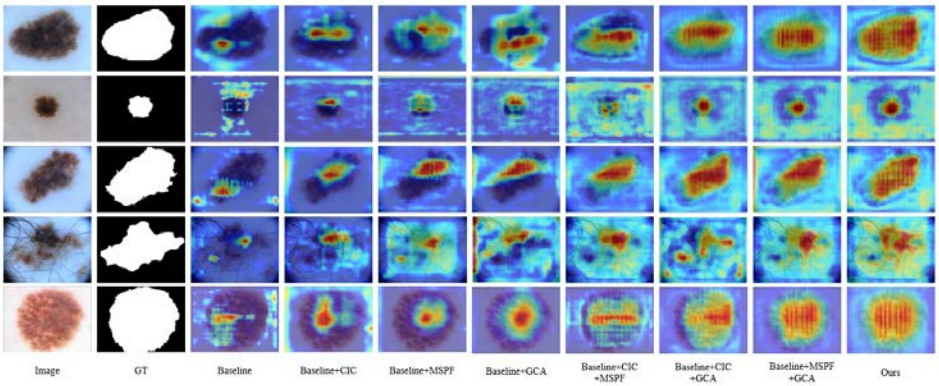

Fig. 13. Attention heatmaps of ablation experimental model

4. Conclusion

In clinical medical applications, melanoma segmentation is the basis of automatic detection of medical images and an important prerequisite for computer-aided doctor diagnosis. It is of great significance for promoting precision medical diagnosis. Aiming at the problem that the traditional U-shaped neural network has insufficient ability to extract context information, A new multi-scale context fusion network (MSCNet) is proposed to fuse global context information. Firstly, the MSPF module is innovatively designed, which is added to the skip connection to fuse the high-stage information and provide global information guidance for the decoder. Secondly, the CIC module is innovatively added to the top of the encoder, which can provide different scales of receptive fields for melanoma segmentation by using dilated convolution with different expansion rates, so as to effectively fuse Contextual feature information and enhance the accuracy of segmentation. In order to make the model take notice of the feature information related to the segmentation target, the GCA module is embedded in the decoder to enhance the attention to the edge contour and tiny features of melanoma. Finally, the combination of binary cross entropy and Dice loss function is used to solve the problem of target class imbalance. Experimental results show that MSCNet is better than other segmentation algorithms, and can accurately segment melanoma with good robustness and generalization.

Although this algorithm has achieved some results, but there are still some problems need further study. For example, the dataset of melanoma images can be augmented by adversarial networks. Moreover, the model ignores the long-range dependence of image information and can be integrated into the transformer method for future research. In the subsequent clinical application, the development of melanoma image segmentation system for clinical diagnosis is also the direction of future efforts.

References

[1] R. L. Siegel, K. D. Miller and A. Jemal, "Cancer statistics," *CA: A Cancer Journal for Clinicians*, vol.69, no.1, pp.7-34, 2019. Article (CrossRef Link)

[2] C. M. Balch, J. E. Gershenwald, S. Soong et al., "Final version of 2009 AJCC melanoma staging and classification," *Journal of clinical oncology*, vol.27, no.36, pp.6199-6206, 2009. Article (CrossRef Link)

[3] H. C. Engasser and E. M. Warshaw, "Dermatoscopy use by US dermatologists: a cross-sectional survey," *Journal of the American Academy of Dermatology*, vol.63, no.3, pp.412-419, Sep. 2010. Article (CrossRef Link)

[4] M. Abd Elaziz, S. Lu and S. He, "A multi-leader whale optimization algorithm for global optimization and image segmentation," *Expert Systems with Applications*, vol.175, Aug. 2021. Article (CrossRef Link)

[5] A. Rampun, K. López-Linares and P. J. Morrow et al., "Breast pectoral muscle segmentation in mammograms using a modified holistically-nested edge detection network," *Medical image analysis*, vol.57, pp.1-17, Oct. 2019. Article (CrossRef Link)

[6] Z. Li, X. M. Wu and S. F. Chang, "Segmentation using superpixels: A bipartite graph partitioning approach," in *Proc. of 2012 IEEE Conference on Computer Vision and Pattern Recognition*, pp.789-796, 2012. Article (CrossRef Link)

[7] R. Garnavi, M. Aldeen and M. E. Celebi et al., "Border detection in dermoscopy images using hybrid thresholding on optimized color channels," *Computerized Medical Imaging and Graphics*, vol.35, no.2, pp.105-115, Mar. 2011. Article (CrossRef Link)

[8] G. Sforza, G. Castellano and S. K. Arika et al., "Using Adaptive Thresholding and Skewness Correction to Detect Gray Areas in Melanoma in Situ Images," IEEE Transactions on Instrumentation and Measurement, vol.61, no.7, pp.1839-1847, Jul. 2012. Article (CrossRef Link)

[9] V. B. Pires and C. A. Z. Barcelos, "Edge Detection of Skin Lesions Using Anisotropic Diffusion," in *Proc. of Seventh International Conference on Intelligent Systems Design and Applications (ISDA 2007)*, pp.363-370, 2007. Article (CrossRef Link)

[10] G. E. Hinton and R. R. Salakhutdinov, "Reducing the Dimensionality of Data with Neural Networks," *Science*, vol.313, no.5786, pp.504-507, Jul. 2006. Article (CrossRef Link)

[11] O. Ronneberger, P. Fischer and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *Proc. of Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pp.234-241, Nov. 2015. Article (CrossRef Link)

[12] X. Wang, Z. Li and Y. Huang, Y. Jiao, "Multimodal medical image segmentation using multi-scale context-aware network," *Neurocomputing*, vol.486, pp.135-146, May. 2022. Article (CrossRef Link)

[13] K. Feng, L. Ren, G. Wang et al., "SLT-Net: A codec network for skin lesion segmentation," *Computers in Biology and Medicine*, vol.148, Sep. 2022. Article (CrossRef Link)

[14] P. Tang, X. Yan and Q. Liang, D. Zhang, "AFLN-DGCL: Adaptive Feature Learning Network with Difficulty-Guided Curriculum Learning for skin lesion segmentation," *Applied Soft Computing*, vol.110, Oct. 2021. Article (CrossRef Link)

[15] A. Iqbal, M. Sharif, M. A Khan et al., "FF-UNet: a U-Shaped Deep Convolutional Neural Network for Multimodal Biomedical Image Segmentation," *Cognitive Computation*, vol.14, no.4, pp.1287-1302, 2022. Article (CrossRef Link)

[16] Q. Wang, B. Wu, P. Zhu et al., "Supplementary material for 'ECA-Net: Efficient channel attention for deep convolutional neural networks," in *Proc. of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp.11534-11542, 2020. Article (CrossRef Link)

[17] S. Woo, J. Park, J. Y. Lee and I. S. Kweon, "CBAM: Convolutional Block Attention Module," in *Proc. of the European conference on computer vision (ECCV)*, pp.3-19, 2018. Article (CrossRef Link)

[18] S. Liu, D. Huang, and Y. Wang, "Receptive Field Block Net for Accurate and Fast Object Detection," in *Proc. of the European conference on computer vision (ECCV)*, pp.385-400, 2018. Article (CrossRef Link)
[19] L. C. Chen, G. Papandreou, I. Kokkinos et al., "DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs," *IEEE transactions on pattern analysis and machine intelligence*, vol.40, no.4, pp.834-848, 2018. Article (CrossRef Link)
[20] W. Zhang, Z. Zhu, Y. Zhang et al., "Cell Image Segmentation Method Based on Residual Block and Attention Mechanism," *Acta Optica Sinica*, no.17, pp.76-83, 2020. Article (CrossRef Link)
[21] N. C. F. Codella, D. Gutman, M. E. Celebi et al., "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. of 2018 IEEE 15th international symposium on biomedical imaging (ISBI 2018)*, pp.168-172, 2018. Article (CrossRef Link)
[22] R. Azad, A. Bozorgpour, M. Asadi-Aghbolaghi et al., "Deep Frequency Re-Calibration U-Net for Medical Image Segmentation," in *Proc. of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops*, pp.3274-3283, 2021. Article (CrossRef Link)
[23] N. Codella, V. Rotemberg, P. Tschandl et al., "Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the international skin imaging collaboration (ISIC)," [Online]. arXiv preprint arXiv:1902.03368, 2019. Article (CrossRef Link)
[24] O. Oktay, J. Schlemper, L. L. Folgoc et al., "Attention U-Net: Learning Where to Look for the Pancreas," in *Proc. of 1st Conference on Medical Imaging with Deep Learning (MIDL 2018)*, 2018. Article (CrossRef Link)
[25] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh et al., "UNet++: Redesigning Skip Connections to Exploit Multiscale Features in Image Segmentation," *IEEE transactions on medical imaging*, vol.39, no.6, pp.1856-1867, June. 2020. Article (CrossRef Link)
[26] R. Gu, G. Wang, T. Song et al., "CA-Net: Comprehensive Attention Convolutional Neural Networks for Explainable Medical Image Segmentation," *IEEE transactions on medical imaging*, vol.40, no.2, pp.699-711, Feb. 2021. Article (CrossRef Link)
[27] M. Z. Alom, M. Hasan, C. Yakopcic et al., "Recurrent Residual Convolutional Neural Network based on U-Net (R2U-Net) for Medical Image Segmentation," arXiv preprint arXiv:1802.06955, 2018. Article (CrossRef Link)
[28] D. P. Fan, G. P. Ji, T. Zhou et al., "PraNet: Parallel Reverse Attention Network for Polyp Segmentation," in *Proc. of Medical Image Computing and Computer Assisted Intervention – MICCAI 2020*, pp.263-273, 2020. Article (CrossRef Link)
[29] S. Feng, H. Zhao, F. Shi et al., "CPFNet: Context Pyramid Fusion Network for Medical Image Segmentation," *IEEE transactions on medical imaging*, vol.39, no.10, pp.3008-3018, Oct. 2020. Article (CrossRef Link)
[30] Y. Xie, J. Zhang, H. Lu et al., "SESV: Accurate Medical Image Segmentation by Predicting and Correcting Errors," *IEEE Transactions on Medical Imaging*, vol.40, no.1, pp.286-296, Jan. 2021. Article (CrossRef Link)
[31] Y. Xie, J. Zhang, Y. Xia, C. Shen, "A Mutual Bootstrapping Model for Automated Skin Lesion Segmentation and Classification," *IEEE transactions on medical imaging*, vol.39, no.7, pp.2482-2493, Jul. 2020. Article (CrossRef Link)
[32] G. Yang, S. Yu, H. Dong et al., "DAGAN: Deep De-Aliasing Generative Adversarial Networks for Fast Compressed Sensing MRI Reconstruction," *IEEE transactions on medical imaging*, vol.37, no.6, pp.1310-1321, Jun. 2018. Article (CrossRef Link)
[33] H. Wu, S. Chen, G. Chen et al., "FAT-Net: Feature adaptive transformers for automated skin lesion segmentation," *Medical Image Analysis*, vol.76, Feb. 2022. Article (CrossRef Link)
[34] D. Jha, P. H. Smedsrud, M. A. Riegler et al., "ResUNet++: An Advanced Architecture for Medical Image Segmentation," in *Proc. of 2019 IEEE International Symposium on Multimedia (ISM)*, pp.225-2255, 2019. Article (CrossRef Link)
[35] Q. Jin, H. Cui, C. Sun et al., "Cascade knowledge diffusion network for skin lesion diagnosis and segmentation," *Applied Soft Computing*, vol.999, Feb. 2021. Article (CrossRef Link)

Zhenhua Li received the M.S. degree in School of Electrical and Information Engineering from Jiangsu University of Technology, Changzhou, China, in 2023. His research interests include deep learning and image processing.

Lei Zhang (Member, IEEE) received the Ph.D. degree in information and communication engineering from Southeast University, Nanjing, China, in 2016. He is currently an Associate Professor with the School of Electrical and Information Engineering, Jiangsu University of Technology, Changzhou, China. From 2019 to 2020, he has been a Visiting Scholar at Queen Mary University of London, U.K. His research interests include deep learning, machine vision, wireless communication and intelligent reflecting surface.