A Medical Image Segmentation Method based on Multi-scale Features and Contour Loss Constrain

Jian Niu[1], Zhijun Fang[1,*], Chen Qian[1], Xihe Qiu[2]
[1] School of Computer Science and Technology, Donghua University, Shanghai, China
[2] School of Electrical and Electronic Engineering, Shanghai University Of Engineering Science, Shanghai, China
* Corresponding author email: zjfang@dhu.edu.cn

Abstract—**In recent years, deep learning has made breakthroughs in medical image segmentation, especially the U-Net architecture, which is becoming a benchmark for various medical image segmentation tasks due to the accuracy of its segmentation results. Although U-Net has achieved great success in many medical image segmentation tasks, it is still unsatisfactory in segmenting object boundaries and small objects. This is due to the fact that segmentation networks gradually lose information, especially edge information and small object information, during the process of convolution and downsampling of features. In order to solve the above problems, we design a novel method that uses a multi-scale module as a feature extractor in the contraction path of the U-shaped structure, which better captures the scale changes of the target object by acquiring image features at different scales; in the prediction stage, a contour prediction branch is constructed to constrain the loss of the target's contour, so that the segmentation network pays more attention to the boundaries of the target. We have validated the performance of our method on the Automated Cardiac Diagnosis Challenge (ACDC) and the spleen segmentation tasks of the Medical Segmentation Decathlon (MSD). The results show that our method obtained the best 95% Hausdorff Distance (HD) metrics on both the ACDC dataset and the Spleen dataset, as well as being quite competitive with other state-of-the-art methods in terms of Dice scores.**

Keywords-multi-scale feature; contour constrain; medical image segmentation;

I. INTRODUCTION

Medical image segmentation is an important topic in the field of medical imaging analysis, which aims to outline anatomical structures and other regions of interest (ROIs) from medical images. Accurate segmentation is crucial for modern computer-aided diagnosis (CAD) applications such as disease diagnosis, treatment planning, and disease progression monitoring. Recently, the emergence of Convolutional Neural Networks (CNNs) has greatly facilitated the development of medical image segmentation, and many high-performance models have appeared, among which the most widely used one has to be U-Net [1]. It mainly relies on a U-shaped encoder-decoder architecture and skip connections connecting the encoder and decoder, where the encoder learns global contextual representations by progressively downsampling the extracted features, while the decoder up-samples the extracted representations to the input resolution for pixel/voxel semantic prediction, and the skip connections connect the encoder's output with the decoder at different resolutions so that spatial information lost during downsampling can be recovered.

Although U-Net has achieved great success in many medical image segmentation tasks, it is still not able to segment the edges of target objects satisfactorily. In addition, due to the complexity of medical image imaging modalities, unwanted blurred images may appear during the imaging process. To address these issues, numerous studies have focused on developing contour-aware networks that consider object boundaries[2-4]. However, most of the above solutions use additional up-sampling branches to predict the contour of the segmented target which greatly increases the number of parameters of the network, or use only shallow features for contour prediction and do not incorporate deeper semantic features, resulting in poor contour prediction.

In this work, we propose a medical image segmentation method that combines multi-scale features and contour loss constraints. In order to avoid smaller target objects being ignored or larger target objects being inaccurately segmented, we use the MultiBlock module as a feature extractor in the contraction path of the U-shape structure, which better captures the scale changes of the target objects by acquiring image features of different scales. And in the prediction stage, we construct a contour prediction branch to constrain the loss of the target's contour so that the segmentation network pays more attention to the target's boundaries. We validate the effectiveness of our method on the ACDC [5] dataset and the Spleen dataset in MSD [6]. The results show that our method obtains the optimal 95% Hausdorff Distance (HD) on the ACDC dataset and Spleen dataset, and also outperforms other state-of-the-art methods on the Dice metric.

II. RELATED WORK

A. *Multi-scale Segmentation*

Multi-scale medical image segmentation is an effective method designed to overcome the problem of insufficient local information and inconsistent global information at a single scale. In past studies, many scholars have proposed various multi-scale medical image segmentation methods to improve the quality and stability of segmentation results. For example, Tian *et al.* proposed a multi-scale liver tumor segmentation method based on pyramidal convolutional neural networks, which obtained more accurate liver tumor segmentation results by integrating feature information at different scales [7]. Dou *et al.* proposed a lung nodule detection method based on multi-scale segmentation fusion, which improves the accuracy and robustness of nodule detection by integrating segmentation results at different scales [8].

DOI reference number: 10.18293/SEKE2024-040

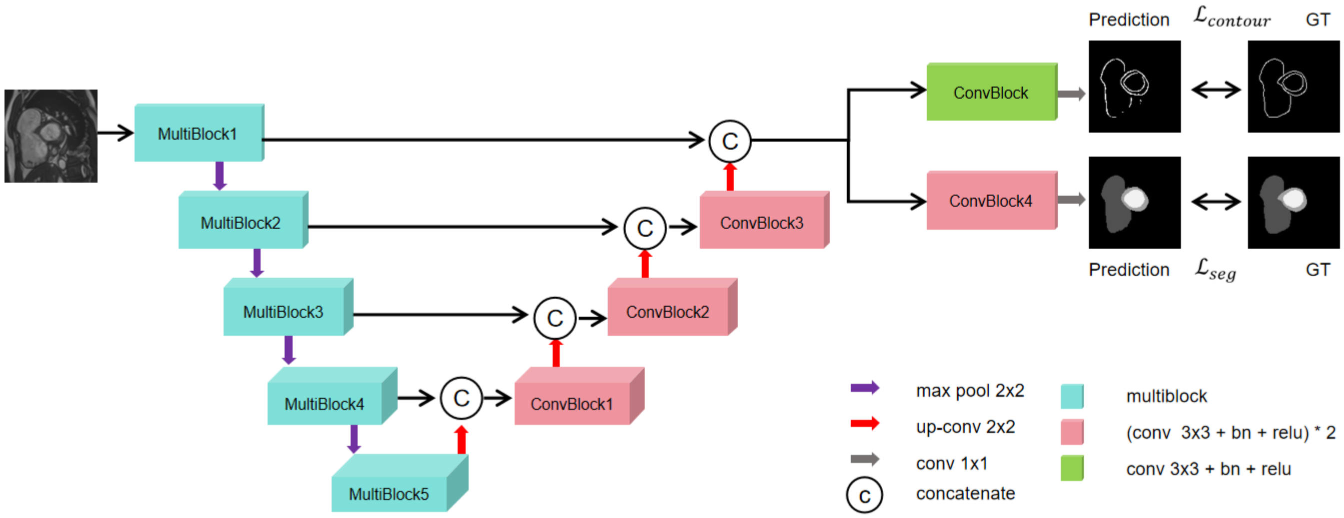

Figure 1. Structure diagram of the whole network, where $\mathcal{L}_{contour}$ is the loss of contour, $\mathcal{L}_{seg}$ is the loss of segmentation and GT is the GroundTruth.

B. *Segmentation with Contour Constraints*

To enhance medical image segmentation, researchers have focused on the relationship between target contours and segmentation outcomes, achieving satisfactory results. For example, Mirikharaji et al. proposed a star-shaped loss which preserves the segmented region with a star shape for dermoscopic skin lesion segmentation tasks and significantly improved the segmentation performance [9]. Chen et al. used a dual decoder structure to predict the segmentation result and the target's boundaries separately, which significantly enhances the boundary-awareness of the segmentation network [2]. However, in the above solutions, the loss function is difficult to design and cannot be reused, and the additional branches may significantly increase the number of parameters in the model.

III. METHOD

We outline the proposed model in Fig 1. The model has two improvements over U-Net: (1) using the MultiBlock module as a feature extractor in the downsampling stage of the U-structure to better capture the scale variation of the target object; (2) constructing a contour prediction branch in the prediction stage to better constrain the loss of the contours of the target object.

A. *MultiBlock Module*

The details of the MultiBlock module are shown in Fig 2. [10] proposed a novel network architecture called DenseUNet, which uses dense convolution instead of the normal CNN convolution in the traditional U-Net architecture, thus improving the segmentation accuracy of the model. However, since DenseNet mixes the features of all previous layers at each layer, this results in a slower model with higher memory consumption. In order to solve the above problems, a new One-Shot Aggregation (OSA) module is proposed in [11] instead of Dense Block. The OSA module avoids the problems of computational inefficiency and high energy consumption in DenseNet due to the dense connections by aggregating all the features at once.

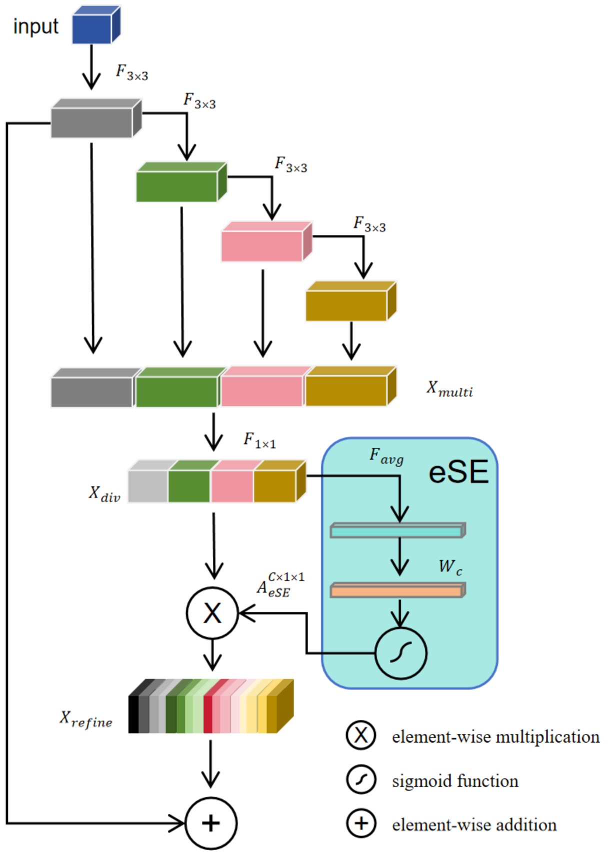

Figure 2. Structure diagram of the MultiBlock modules. $F_{1\times1}$, $F_{3\times3}$ denote 1 × 1, 3 × 3 conv layer, respectively, F_{avg} is global average pooling, $W_C$is a fully-connected layer, A_{eSE} is channel attention map.

The MultiBlock module we use in this article is a variant of the OSA module [12]. Assuming that the number of channels of the input data is k, we first use the ordinary convolutional layer to double the channels to get more features, and the number of channels of the doubled feature X_2 is 2 * k. Then, successive convolution of X_2 is performed to obtain features at different scales and these features are spliced together with X_2 to form the multi-scale feature X_{multi}, and the number of X_{multi} channels is 8 * k. X_{multi} is then downscaled using a 1×1 convolutional layer to obtain X_{div}, which has 2 * k channels. Then X_{div} is enhanced by an eSE (Effective Squeeze-Excitation) channel attention module, and the enhanced feature is X_{refine}, and the residuals of X_{refine} and X_2 are processed to get the final output. Unlike the normal SE (Squeeze-Excitation) channel attention module which uses two fully connected layers for dimensionality reduction and then dimensionality enhancement, the eSE module uses only one fully connected layer, which reduces the possibility of information loss due to dimensionality reduction. The eSE process is defined as:

$$A_{eSE}(X_{div}) = \sigma(W_C(F_{gap}(X_{div}))), \tag{1}$$

$$X_{refine} = A_{eSE}(X_{div}) \otimes X_{div} \tag{2}$$

where F_{gap} is channel-wise global average pooling, W_C is weight of a fully-connected layers, σ denotes the sigmoid function, A_{eSE} is a channel attentive feature descriptor and $\otimes$ denotes element-wise multiplication.

B. *Contour Prediction Branch*

One of the reasons limiting the performance of medical image segmentation is the inaccurate segmentation of the target's contour. There are two main methods to solve the above problem, one is to modify the loss function to make the network more concerned about the contour loss, but the design of the loss function is more difficult; the other method is to use the dual-stream decoder structure to build a separate branch of the decoder to predict the target's contour, but this greatly increases the number of parameters of the model. It is well known that the shallow features of neural networks contain rich boundary information, but also contain a lot of interference information that is not related to the contour, so it is difficult to predict the contour using only the shallow features. In order to significantly enhance the hidden details in the shallow feature maps and avoid the interference of irrelevant information, we use the shallow features before the first skip connection and the deep features after the expansion operation as the inputs to the contour prediction branch. The shallow features before the first skip connection contain a lot of boundary information and can provide a clear boundary, while the deep features after the expansion operation remove a lot of irrelevant information, which can avoid the interference of irrelevant information and let the model focus more on the contour information. To increase the number of parameters as little as possible, we use only one layer of normal CNN convolution as the contour prediction branch.

C. *Loss Function*

For the overall loss, we define it using the following equation:

$$\mathcal{L}_{total} = \mathcal{L}_{seg} \times 0.9 + \mathcal{L}_{contour} \times 0.1 \tag{3}$$

Where $\mathcal{L}_{total}$ is the loss of the whole network, $\mathcal{L}_{seg}$ is the loss of the segmentation branch, and $\mathcal{L}_{contour}$ is the loss of the contour prediction branch. The loss functions for all the above branches are a combination of soft dice loss and cross-entropy loss, and in practice we use DiceCELoss in the monai framework as the loss function of each branch.

IV. EXPERIMENTS

A. *Dataset*

ACDC (MRI): the ACDC dataset [5] is a MICCAI 2017 challenge to perform left ventricular (LV), right ventricular (RV), and myocardial (Myo) segmentation of end-diastolic (ED) and end-systolic (ES) frames from cardiac dynamic magnetic resonance imaging (cine-MRI). There are 150 cases in this dataset, 100 cases in the training set and 50 cases in the test set. Since the images in this dataset are not uniform in size, we cropped the data uniformly to (128, 128). Then 70 cases were randomly selected from the training set data for training, the remaining 30 cases were used for validation, and finally evaluated on 50 cases in the test set. For specific tests, we used only end-diastole for training and testing.

Spleen (CT): The Spleen dataset [6] is the 9th subtask of the MSD (Medical Segmentation Decathlon) and the goal is to segment the spleen from CT images. The dataset consists of 41 CT data with annotated spleen bodies. The original size of the image slices is (512, 512), and we cropped the slices to (256, 256) to minimize the background interference while preserving the target region intact. We randomly selected 30 out of 41 cases for training the model, 5 cases for validation, and 6 cases for testing.

Since the above datasets do not have the ground truth of the contours, we use Canny Edge Detection Algorithm [13] to perform edge detection on the ground truth of the segmentation results to obtain the ground truth of all the contours.

B. *Implementation Details*

We implemented our method in Python using Pytorch and trained and evaluated our method on a computer equipped with an NVIDIA RTX-2050. In the experiments, the batch size of the ACDC dataset was 4 and the batch size of the Spleen dataset was 2. The initial learning rate was 0.001 and 20 iterations were performed using the AdamW optimizer. We used Dice score and 95% Hausdorff Distance (HD) to evaluate the accuracy of the segmentation results, where Dice score describes the overall accuracy of the segmentation results and 95% HD describes the accuracy of the target's contour.

C. *Experimental Result*

We conducted experiments on the ACDC dataset and the Spleen dataset and compared the proposed model with five previous state-of-the-art techniques:1)U-Net[1]; 2) UNet++[14]; 3) AttnUnet[15]; 4) MultiResUNet[16] and 5) CUnet[3].

The specific results of the comparison experiments are shown in Table I. From Table I, we can see that our method achieves the optimal results in 95% HD metrics, which means that our method obtains the optimal segmentation boundaries, and at the same time, in Dice scores, our method is better than other advanced methods. The experimental results demonstrate that the model can obtain better boundary segmentation results after combining multi-scale features and contour loss constraints.

TABLE I. Experimental Performance Of Different Methods

Methods	**ACDC**								**Spleen**	
	RV		***Myo***		***LV***		***Avg***		***Avg***	
	Dice	***95%HD***	***Dice***	***95%HD***	***Dice***	***95%HD***	***Dice***	***95%HD***	***Dice***	***95%HD***
U-Net[1]	0.8558	4.4777	0.8535	2.9563	0.9322	2.3046	0.8805	3.2462	0.9286	7.5744
UNet++[14]	0.8607	4.3240	0.8581	2.9627	0.9302	2.1707	0.8830	3.1524	0.9342	9.5044
AttnUnet[15]	0.8626	**4.0574**	0.8598	2.8303	0.9342	2.3713	0.8855	3.0863	0.9326	6.8223
MultiResUNet[16]	0.8647	4.3447	0.8589	3.0264	0.9321	2.5795	0.8852	3.3169	0.9306	6.7520
CUnet[3]	0.8594	4.5413	**0.8713**	2.7021	0.9413	1.9828	0.8907	3.0754	0.9338	7.1515
Ours	**0.8689**	4.1603	0.8706	**2.5038**	**0.9459**	**1.8000**	**0.8952**	**2.8214**	**0.9384**	**5.8814**

TABLE II. Results Of Ablation Experiments

Models	**ACDC**		**Spleen**	
	Dice	***95%HD***	***Dice***	***95%HD***
U	0.8805	3.2462	0.9286	7.5744
U + M	0.8914	2.9132	0.9310	6.6722
U + C	0.8856	2.9965	0.9333	7.0622
U + M + C	**0.8952**	**2.8214**	**0.9384**	**5.8814**

D. Ablation

We conducted ablation studies on our proposed method on different datasets. Table II shows the results of the ablation study, where U denotes the baseline model U-Net, M denotes the MultiBlock module, and C denotes the contour prediction branch. From the results of the ablation experiments, it can be seen that our proposed model shows a considerable improvement on different datasets, with the highest improvement in Dice scores on the ACDC dataset by 1.47% (from 88.05% to 89.52%), and the highest decrease in 95HD on the Spleen dataset by 1.693 (from 7.5744 to 5.8814).

V. CONCLUSION

In this paper, we propose a new medical image segmentation method, which uses multi-block modules to acquire features at different scales to better capture the scale variation of the target object, and adds a contour prediction branch in the prediction stage to make the network focus more on the contour information of the target. We validated the effectiveness of the method in different segmentation tasks in MRI and CT modalities. The results show that the method performs significantly better than other mainstream segmentation methods on the ACDC dataset and achieves excellent results in the spleen segmentation task.

REFERENCES

[1] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference. Germany, pp. 234-241, October 2015.

[2] H. Chen, X. Qi, L. Yu, Q. Dou, J. Qin, and P.-A. Heng, “DCAN: Deep contour-aware networks for object instance segmentation from histology images,” Medical Image Analysis, pp. 135–146, Feb. 2017.

[3] Z. Cheng, A. Qu, and X. He, “Contour-aware semantic segmentation network with spatial attention mechanism for medical image,” The Visual Computer, pp. 749–762, Mar. 2022.

[4] R. Wang, S. Chen, C. Ji, J. Fan, and Y. Li, “Boundary-aware context neural network for medical image segmentation,” Medical Image Analysis, p. 102395, May 2022.

[5] O. Bernard et al., "Deep Learning Techniques for Automatic MRI Cardiac Multi-Structures Segmentation and Diagnosis: Is the Problem Solved?" in IEEE Transactions on Medical Imaging, vol. 37, no. 11, pp. 2514-2525, Nov. 2018.

[6] AmberL. Simpson et al., “A large annotated medical image dataset for the development and evaluation of segmentation algorithms,” arXiv: Computer Vision and Pattern Recognition,arXiv: Computer Vision and Pattern Recognition, Feb. 2019.

[7] H. Tian, B. Ji, W. Quan, and J. Qin, “MPA-Net: Multi-scale Pyramid Attention network for Liver Tumor Segmentation,” in 2021 International Conference on Electronic Information Engineering and Computer Science (EIECS), Changchun, China, Sep. 2021.

[8] Q. Dou, H. Chen, L. Yu, J. Qin, and P.-A. Heng, “Multilevel Contextual 3-D CNNs for False Positive Reduction in Pulmonary Nodule Detection,” IEEE Transactions on Biomedical Engineering, pp. 1558–1567, Jul. 2017.

[9] Z. Mirikharaji and G. Hamarneh, “Star Shape Prior in Fully Convolutional Networks for Skin Lesion Segmentation,” in Medical Image Computing and Computer Assisted Intervention – MICCAI 2018, Lecture Notes in Computer Science, 2018, pp. 737–745.

[10] Y. Cao, S. Liu, Y. Peng, and J. Li, “DenseUNet: densely connected UNet for electron microscopy image segmentation,” IET Image Processing, pp. 2682–2689, Oct. 2020.

[11] Y. Lee, J. Hwang, S. Lee, Y. Bae, and J. Park, “An Energy and GPU-Computation Efficient Backbone Network for Real-Time Object Detection,” in 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Long Beach, CA, USA, 2019.

[12] Y. Lee and J. Park, “CenterMask : Real-Time Anchor-Free Instance Segmentation,” in 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, Jun. 2020.

[13] J. Canny, "A Computational Approach to Edge Detection," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. PAMI-8, no. 6, pp. 679-698, Nov. 1986.

[14] Z. Zhou, M. M. Rahman Siddiquee, N. Tajbakhsh, and J. Liang, “UNet++: A Nested U-Net Architecture for Medical Image Segmentation,” in Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support,Lecture Notes in Computer Science, 2018, pp. 3–11.

[15] O. Oktay et al., “Attention U-Net: Learning Where to Look for the Pancreas,” arXiv: Computer Vision and Pattern Recognition,arXiv: Computer Vision and Pattern Recognition, Apr. 2018.

[16] N. Ibtehaz and M. S. Rahman, “MultiResUNet: Rethinking the U-Net architecture for multimodal biomedical image segmentation,” Neural Networks, pp. 74–87, Jan. 2020.