2022 14th International Conference on Measuring Technology and Mechatronics Automation (ICMTMA) | 978-1-6654-9978-1/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICMTMA54903.2022.00064

A U-Net Network Model for Medical Image Segmentation Based on Improved Skip Connections

Jing Di, Shuai Ma*,Jing Lian,Guodong Wang

School of Electronics & Information Engineering, Lanzhou Jiaotong University, Lanzhou, 730070, China
*corresponding author
E-mail: 46891771@qq.com

***Abstract*— To address the loss problem introduced by downsampling in the classical U-Net architecture, this paper improves the U-Net model and uses the model for medical image segmentation. The essence of the proposed model is still the classical U-Net encoder-decoder network, in which the encoder and decoder sub-networks are connected by a skip connection. First, we improve the connection position of the skip connection, and the two ends of the connection are changed from the second convolution result of the original convolution block to the first result and the decoder convolution block for concatenation; second, the concatenation operation is added to the convolution block of the downsampling part, and the two improvements aim at retaining more image underlying information, and thus achieving more efficient fusion of high and low level image information; finally, on the public medical image segmentation dataset, the classical U-Net, FCN-8s and the improved model in this paper are comparatively evaluated for cell nucleus segmentation in microscope images and liver segmentation in abdominal CT scans. The experiments show that the improved U-Net model mIoU, Aver_dice in this paper improves by 2~3% compared with the control model.**

Keywords- Deep Learning; U-Net; Medical Image Segmentation; FCN-8s; Network Model

I. INTRODUCTION

With the development of Convolutional Neural Network (CNN) in the field of computer vision and medical image processing, deep learning has become the mainstream method in medical image segmentation tasks to gradually realize the automatic segmentation of medical images. Deep learning segmentation methods are used to achieve segmentation of medical images by classifying pixels. Unlike traditional pixel or superpixel classification methods that use hand-crafted features, deep learning methods automatically learn task-relevant features from medical images and classify pixels based on these features, which in turn enables end-to-end segmentation.

Since 2012, when AlexNet [1] won the ImageNet competition, convolutional neural networks have gradually become a widely adopted image classification architecture.In the early days when convolutional neural networks were applied to medical image segmentation tasks, many researchers used an image block-based training approach to enable the neural networks to obtain good segmentation results.The earliest deep learning-based medical image segmentation method was proposed by Cirean et al [2], using a strategy based on image blocks and sliding windows to segment neurons from microscopic images.After that, Kamnitsas et al [3] combined Fully Connected Conditional Random Field (CRF) to propose an efficient Pereira et al [4] proposed a deep learning method for automatic segmentation of brain tumors.

Most recent researchers have prioritized the use of FCNs over sliding-window-based classification with the aim of reducing redundant computations.The most famous of these novel convolutional neural network architectures is the U-Net proposed by Ronneberger et al [5] in 2015.The main innovations in the U-Net are the downsampling encoding and upsampling decoding layers and the rational design of skip connections. The skip connection is able to connect the downsampling path to the upsampling path.From a training point of view, this means that the whole image can be segmented directly by U-Net processing one pass forward, allowing U-Net to take into account the information of the whole image.U-Net has now become the baseline for most medical image semantic segmentation tasks, and has inspired a large number of researchers to think about U-type semantic segmentation networks.

Drozdzal et al [6] investigated the use of ResNet-like short-skip connections in addition to long-skip connections in regular U-Net, and experimentally demonstrated that short-skip connections accelerate the convergence of the learning process and enable very deep network architectures with relatively few parameters.Milletari et al [7] proposed a 3D variant of the U-Net architecture called V-Net.V -Net uses 3D convolutional layers and loss functions of Dice coefficients for 3D image segmentation.Nabil Ibtehaz [9] et al. argue that the direct skip connection leaves a semantic gap in coding and decoding, which is not conducive to the model's understanding of image information, and then propose to improve the skip connection to a convolutional chain of MultiResUNet, and improved the convolutional block in the downsampling part of the original model. In this paper, the U-Net network model with improved skip connections is proposed for the loss problem introduced by downsampling. Firstly, since multiple convolution will lose the underlying image information to some extent, this paper changes the first convolution result to concatenating with upsampling in the channel dimension; secondly, since concatenation is beneficial to achieve more image information preservation, this paper performs concatenation in the channel dimension after convolving the convolution block twice; finally, to solve the overfitting problem, the dropout operation is used, and the above improvements are applied to the public medical.Finally, the dropout operation is used to solve the overfitting problem.

II. CLASSICAL U-NET

The classical U-Net structure is shown in Figure 1. The key points of the simplified U-Net are only three lines: downsampling encoding, upsampling decoding, and skip connection. Among them, downsampling performs

978-1-6654-9978-1/22/$31.00 ©2022 IEEE
DOI 10.1109/ICMTMA54903.2022.00064

information compression and upsampling performs pixel recovery, which is a part of other segmentation networks, and U-Net does not skip out of this framework. As can be seen, U-Net performs 4 times of maximum pooling downsampling, uses convolution for information extraction after each sampling to get the feature map, and then goes through 4 times of upsampling to recover the input pixel size. The most critical and unique part of U-Net is the skip connection that connects the up-sampling and down-sampling parts of the graph, and each down-sampling has a skip connection with the corresponding up-sampling for concatenating. With more detailed graph features, the bottom level (deep level) has more downsampling, and the information is convolved a lot, with large spatial loss, but it helps to extract the target features, which is beneficial to the classification and judgment of the target area, and when the high level and low level features are fused, the segmentation effect is often very good.

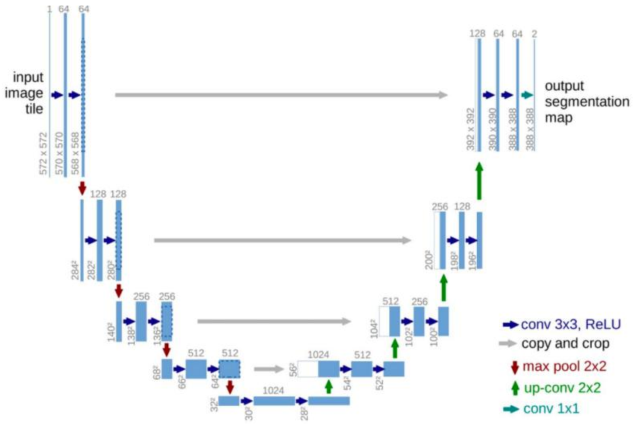


Figure 1. Classical U-Net structure

Through the study and research of U-Net and its variant structures, and inspired by UNET++ proposed by Zongwei Zhou et al [8] and MultiResUNet proposed by Nabil Ibtehaz et al [9], this paper considers the question of why the U-Net structure undergoes two convolutions before it is stitched with the upsampled part of the feature map through skip connection, and also whether the concatenating operation can be applied to other parts of the structure becomes the goal of study and research in this paper.

III. U-NET WITH IMPROVED SKIP CONNECTION

The classical U-Net structure downsampling part is mainly composed of convolutional blocks containing two convolutions and maximum pooling. Theoretically, multiple downsampling can make the model learn deeper image information, but in 2014 Geoffrey Hinton gave a talk on his capsule networks project at MIT, mentioning some problems with convolutional neural networks, including the loss of underlying features.

A. Skip connection improvement

The first improvement point takes the input and output layers as an example, where the input image is convolved once with the output layer's result from upsampling in the channel dimension by skip connection, as shown in Figure 2.

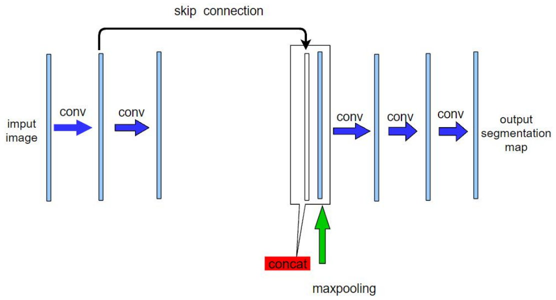


Figure 2. Skip connection improvements

Concatenating in the channel dimension requires the feature map to be of the same size and number of channels, padding='same' ensures that the input and output dimensions of the convolution are the same, while the number of channels of the subnetwork located at the same height is the same in the design of U-Net, and this design facilitates the improvement in this paper. Compared with going through two convolutions, convolution once can both extract image information and will retain more image low-resolution information, and combined with the high-resolution information passed to the same height after concatenate operation, the fusion of high- and low-level information is more obvious.

B. Convolutional layer concatenating

The concatenation in the channel dimension will form a thicker feature map and fuse more image information at different levels, while the two convolution results located at the same height happen to have the same size and number of channels, which can be concatenated, as shown in Figure 3. Since adding too much concatenation will significantly increase the number of parameters of the model, which is detrimental to the advantage of the small size of U-Net, and is likely to cause overfitting in the case of small dataset images, the concatenate operation is added only after the fifth convolution block.

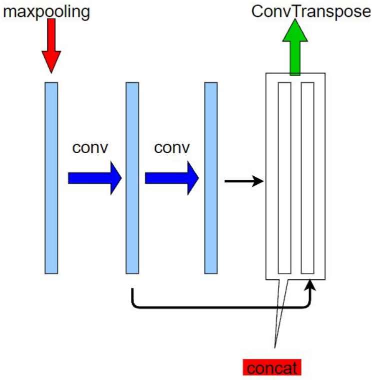


Figure 3. Concatenate two convolution results

Set padding='same' to ensure the same size of the two convolution results, concatenate the results of the two convolutions, fuse different levels of image information, so that the information complements each other, double the number of channels of the feature map obtained by concatenation, note that the number of channels after upsampling should be set to fit with the subsequent convolution.

C. Evaluation Metrics

The evaluation metrics used in the experiment are mIoU, Aver_dice, and Loss.

The mIoU is defined as the intersection of the predicted and actual regions for each class divided by the concatenated set IoU of the predicted and actual regions, summed and then averaged, and the IoU formula is defined as:

$$IoU = \frac{TP}{TP + FP + FN} \quad (1)$$

In the formula, TP is the number of true positives, i.e., the number of positives in the label and the number of positives in the predicted value. TN is the number of true negatives, i.e., the number of negatives in the label and the number of negatives in the predicted value. fp is the number of false positives, i.e., the number of negatives in the label and the number of positives in the predicted value. fn is the number of false negatives, i.e., the number of positives in the label and the number of negatives in the predicted value. fn is the number of false negatives, i.e., the number of positives in the label and the number of negatives in the predicted value. FN is the number of false negatives.

Aver_dice is defined as the intersection of twice divided by the pixel sum, summed and then averaged. The formula is defined as:

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

Loss is the cross-entropy loss function commonly used in pytorch. Assuming that the probability distribution p is the expected output, the probability distribution q is the actual output, and H(p, q) is the cross-entropy, the formula is defined as:

$$H(p,q) = -\sum_{x}(p(x)\log q(x)) \quad (3)$$

IV. EXPERIMENTAL RESULTS

The public medical image dataset used for the experiments is shown in Table 1.

Table 1 The image segmentation datasets used in experiments

Dataset	Images	Size
Liver	400	512×512
Cell	30	512×512

An example of the original image in the Liver dataset is shown in Figure 4.

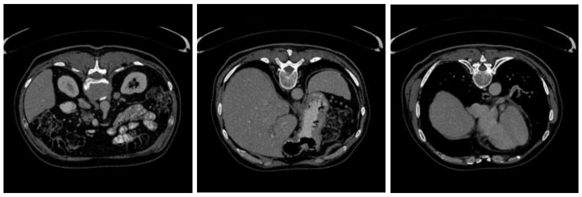


Figure 4. Example of Liver dataset images

An example of the original image in the Cell dataset is shown in Figure 5.

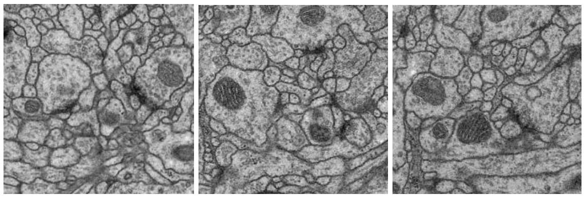


Fig.5 Example of cell dataset images

In order to verify the effectiveness of the improvements in this paper, the classical U-Net and the improved U-Net and FCN-8s [10] (VGG16) in this paper are used for a comparative study, and the number of parameters of each model is shown in Table 2.

Table 2 parameters

Architecture	parameter
U-Net	7,759,521
Ours1	8,289,697
Ours2	7,765,409
FCN-8s	143,667,240

From the table, it is known that the two models Ours1 and Ours2 improved in this paper have slightly higher number of parameters than classical U-Net and much less than FCN-8s.

The models are tested and compared on the liver dataset with the experimental parameters epoch set to 32 and batch_size set to 1. The evaluation values mIoU, Aver_dice and loss obtained on the liver dataset are shown in Table 3.

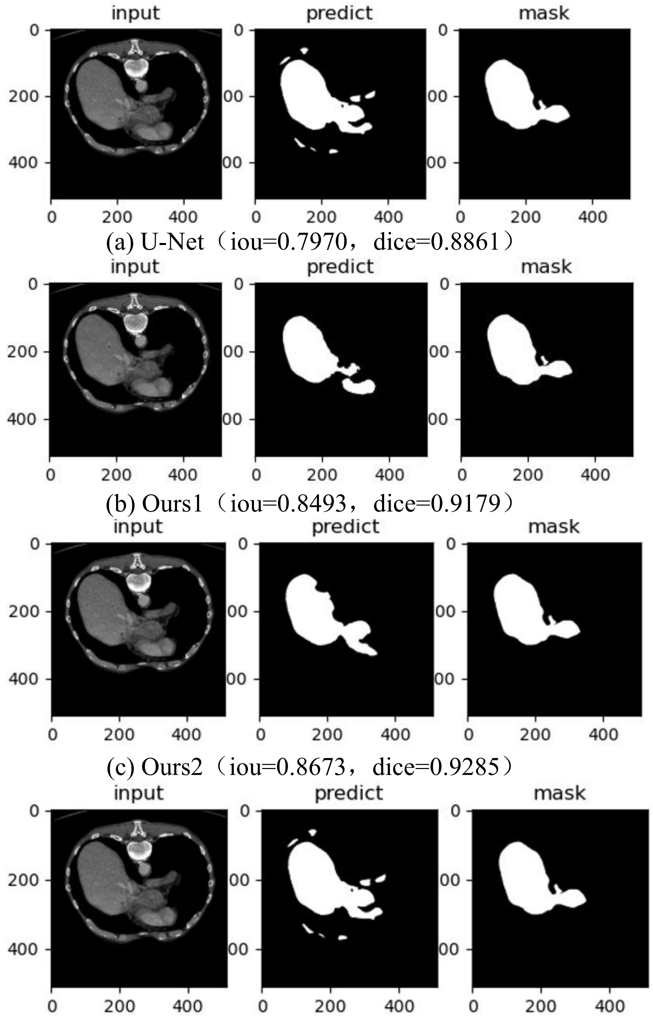
Table 3 Segmentation result on liver

Architecture	mIoU	Aver_dice	loss
U-Net	0.859039	0.923332	2.302
Ours1	0.846205	0.926563	**2.167**
Ours2	**0.880900**	**0.936283**	2.556
FCN-8s	0.802639	0.889594	2.755

(Ours1: improved skip connection and convolution block, Ours2: improved skip connection only)

From the data in the table, we can see that the improvements in this paper effectively improve the segmentation effect. Ours1, the model with improved skip connection suggested in this paper, achieves the highest mIoU and Aver_dice values, and Ours2, the model with improved skip connection and convolutional blocks, is second only to Ours1 and has the smallest loss value, which indicates that the two improvements in this paper do have an optimization effect on liver segmentation.

A randomly selected image in the liver dataset, the segmentation results of four models, U-Net, Ours1, Ours2, and FCN-8s, are shown in Figure 6.



(a) U-Net（iou=0.7970，dice=0.8861）

(b) Ours1（iou=0.8493，dice=0.9179）

(c) Ours2（iou=0.8673，dice=0.9285）

(d) FCN-8s（iou=0.7490，dice=0.8514）

Figure 6. Segmentation results of U-Net, Ours1, Ours2 and FCN-8s on the same image

From the figure, we can see that each result has over-segmentation or under-segmentation due to less training times, but relatively Ours2 can get higher segmentation scores.

In order to increase the credibility of the experimental results, this paper conducted another test on the cell dataset. When epoch was set to 32 and batch_size was set to 1, the experimental results mIoU, Aver_dice and loss values are shown in Table 4.

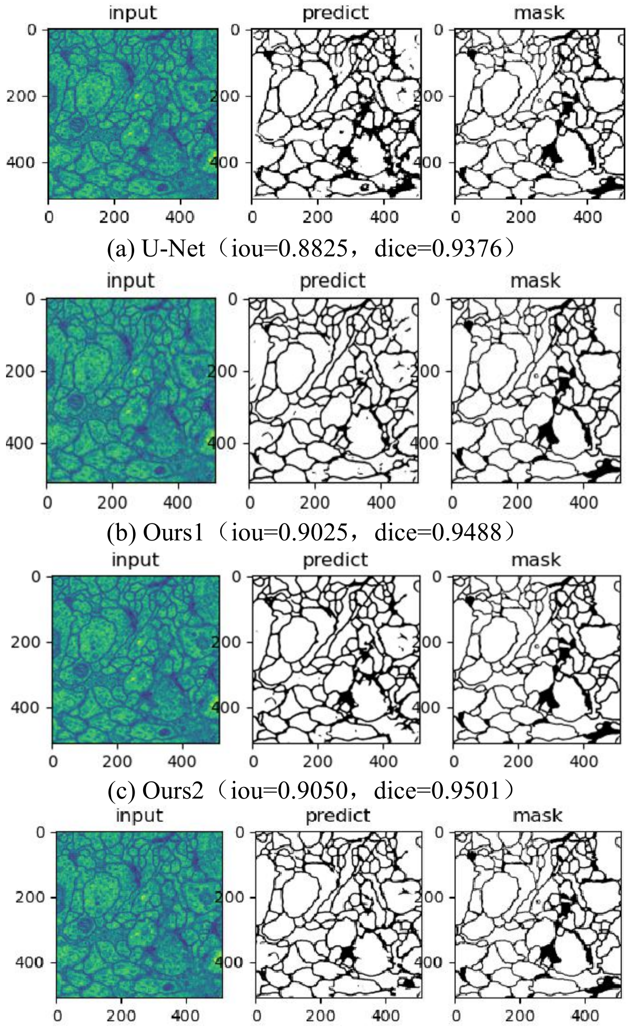
Table 4 Segmentation result on cell

Architecture	mIoU	Aver_dice	loss
U-Net	0.899314	0.946975	3.284
Ours1	0.899895	0.947300	2.829
Ours2	0.900599	0.947671	2.794
FCN-8s	0.893102	0.943514	1.764

(Ours1: improved skip connection and convolution block, Ours2: improved skip connection only)

As seen in the table, the Ours1 model achieved the highest mIoU and Aver_dice values, while the Ours2 model, although not as good as Ours1, performed better compared to the other two models, indicating that both improvements in this paper are beneficial for enhancing the segmentation effect.

The segmentation results of each model on the same image of the cell dataset are shown in Figure 7.



(a) U-Net（iou=0.8825，dice=0.9376）

(b) Ours1（iou=0.9025，dice=0.9488）

(c) Ours2（iou=0.9050，dice=0.9501）

(d) FCN-8s（iou=0.8941，dice=0.9499）

Fig.7 The result of each model segmenting the same cell image

From the figure, it can be seen that the Ours2 model proposed in this paper obtains the closest results to the labels in cell segmentation with the highest segmentation score, and the Ours1 model is the second highest, which reflects the desirability of the improvements in this paper.

Through experiments, it can be seen that the proposed models Ours1 and Ours2 consistently outperform the classical U-Net and FCN-8s in all cases and achieve the highest scores on both datasets which shows that the models are reliable and robust. Although the improvements in this paper only slightly improve the segmentation results, it still proves that the improvements do have the effect of enhancing the segmentation ability.

V. CONCLUSIONS

In order to make medical image segmentation more accurate, an improved approach to U-Net is proposed in this paper.The proposed structure redesigns the position of the skip connection, and the redesigned skip connection aims to reduce the loss introduced by the downsampling process and better achieve the fusion of high and low level feature information to complement each other. This design can combine both the local information of high resolution and the information of larger area of low resolution, which can be combined to obtain finer segmentation results, and the addition of concatenation after the convolutional block is also based on this purpose. In order to alleviate the overfitting problem that may be caused by the complexity of the model, dropout is added after the convolution, and the improved scheme is compared in two public medical image datasets, including nucleus segmentation and liver segmentation.The final experiments show that the improved model achieves more satisfactory segmentation results and achieves the highest scores compared to classical U-Net and FCN-8s. However, the model in this paper still has problems such as large number of parameters, which is contrary to the original intention of U-Net design, and the next improvement will focus on this problem.

ACKNOWLEDGEMENTS

This work was supported by the National Natural Science Foundation of China under Grant No. 62061023 and No. 61941109 and Gansu Provincial Institutions of Higher Learning Innovation Ability Promotion Project (2019B-052).

REFERENCES

[1] Technicolor T, Related S, Technicolor T, et al.ImageNet Classification with Deep Convolutional Neural Networks.

[2] Cirean D C, Giusti A, Gambardella L M, et al.Deep Neural Networks Segment Neuronal Membranes inElectron Microscopy Images[J].Advances in neural information processing systems, 2012, 25:2852--2860.

[3] Kamnitsas K, Ledig C, Newcombe V, et al.Efficient Multi-Scale 3D CNN with Fully Connected CRF for Accurate Brain Lesion Segmentation[J].Medical Image Analysis, 2016, 36:61.

[4] Pereira S, Pinto A, Alves V, et al.Brain Tumor Segmentation Using Convolutional Neural Networks in MRI Images[J].IEEE Transactions on Medical Imaging, 2016, 35(5):1240-1251.

[5] Ronneberger O, Fischer P, Brox T.U-Net: Convolutional Networks for Biomedical Image Segmentation[J].Springer International Publishing, 2015.

[6] Drozdzal M, Vorontsov E, Chartrand G, et al.The Importance of Skip Connections in Biomedical Image Segmentation[C]// Springer International Publishing.Springer International Publishing, 2016

[7] F Milletari, Navab N, Ahmadi S A.V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation[C]// 2016 Fourth International Conference on 3D Vision (3DV).IEEE, 2016.

[8] Zhou Z, Siddiquee M, Tajbakhsh N, et al.UNet++: A Nested U-Net Architecture for Medical Image Segmentation[C]// 4th Deep Learning in Medical Image Analysis (DLMIA) Workshop.2018.

[9] Ibtehaz N, Rahman M S.MultiResUNet: Rethinking the U-Net Architecture for Multimodal Biomedical Image Segmentation[J].Neural Networks, 2019.

[10] Long J, Shelhamer E, Darrell T.Fully Convolutional Networks for Semantic Segmentation[J].IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 39(4):640-651.

[11] Ronneberger O, Fischer P, Brox T.U-Net:Convolutional Networks for Biomedical Image Segmentation[J].Springer International Publishing, 2015.

[12] Dhruv P, Naskar S.Image Classification Using Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN): A Review[M].2020

[13] Lecun Y, Bottou L.Gradient-based learning applied to document recognition[J].Proceedings of the IEEE, 1998, 86(11):2278-2324.