

SE-UXNet: A Medical Imaging Segmentation Network Based on Attention Mechanism

Zhongyu Huang¹, Wenxin Hu¹, Wei Zheng²⁺

¹ School of Data Science and Engineering, East China Normal University, China

² Information Technology services, East China Normal University, China

Abstract. Automatic medical image segmentation is an important part of medical image analysis, and plays an indispensable role in computer-aided diagnosis. Recently, FCN (Fully Convolutional Network) and U-Net have become the main frameworks for image segmentation task. Based on U-Net, the U-Net++ is proposed to predict robust segmentation maps. However, it is difficult for both U-Net and U-Net++ to obtain better results as the model depth grows. In this paper we propose a new medical image segmentation network SE-UXNet using attention mechanism. This model bridges the encoder-decoder models of two different structures to increase model capacity, and recalibrates the weight of the features in the bridge process by SE blocks. Using the Dice coefficient and mIoU as evaluation criteria, we evaluate our model by performing nuclei, liver, prostate, polyps and pulmonary nodule segmentation tasks. The results show that, the Dice coefficients of the above segmentation tasks using SE-UXNet were 92.75%, 90.49%, 90.34%, 95.33% and 73.44%, respectively, all higher compared to results of the U-Net and U-Net++. The experimental results also show that the method we proposed improves the accuracy of medical image segmentation.

Keywords: Medical image segmentation, SE-block, encoder-decoder, U-Net, U-Net++

1. Introduction

With the development of medical imaging technology, image diagnosis plays an increasingly important role in disease diagnosis including ultrasound (X-RAY), computed tomography (CT) and magnetic resonance imaging (MRI). To divide the image into several areas and extract the area of interest, image segmentation technique is used. At present, it is still mainly performed by doctors, which relies heavily on the professional experience of the doctor and is somewhat subjective. In some cases, doctors may make mistakes because of stress and fatigue. Therefore, deep learning-based image segmentation algorithms are introduced to relieve doctors' working stress and improve the accuracy of medical image segmentation.

The end-to-end medical image segmentation model based on deep learning has grown rapidly over the past few years. FCN (Fully Convolutional Network) [1], U-Net [2] and U-Net++ [3] are all popular structures. This kind of network structure can output the segmentation results intuitively and has a good effect on the medical image segmentation data sets. However, due to the defects of the network structure itself, it is difficult for U-Net and U-Net++ to construct a well-behaved deep network. Compared to a single U-Net, Wanli Chen et al. [4] proposed that the use of a bridge structure helps to solve this problem by expanding model capacity. As an important attention structure, SE block [5] can be easily embedded in skip connections. We believe that the SE blocks help to pass the important features between different branches of the bridge model more effectively. In this paper, we propose a novel structure as SE-UXNet, which is based on the U-Net and U-Net++, to improve the accuracy of medical image segmentation.

The main contributions are summarized as follows:

- We propose a composite structure as SE-UXNet by combining the SE blocks with the bridging model of U-Net and U-Net++. We utilize the bridging structure to improve model capacity and further improve the model accuracy by adjusting the weight of the model feature channel.

⁺ Corresponding author. Tel.: + 18116316193; fax: +62542986.
E-mail address: wzhenh@admin.ecnu.edu.cn

- Our method achieved very competitive performance on four public data sets (Data Science Bowl 2018, MICCAI PROMISE12, Liver Tumor Segmentation Challenge and CVC-ClinicDB) and the lung nodule data set.

The rest of this paper is mainly as follows: we first introduce the related work regarding image segmentation in Section 2. We then describe the details of our proposed model and the feature generation in Section 3. In Sections 4, we introduce the experimental settings and results. Finally, we conclude our work and indicate future directions in Section 5.

2. Related Work

A variety of traditional medical image segmentation algorithms have been proposed over the past few decades. Otsu[6] was a classic method based on dynamic thresholding, which classified foreground and background according to whether the intensity value is above a threshold. But this method focused on the histogram distribution of the image with ignoring the spatial information. To solve this problem, region growing[7] was proposed, which performed region base growing based on the initial seed point but often encountered problems of over-segmentation. Edge detection and grouping [8] was an efficient segmentation method, but it was very susceptible to noise. Compared with ordinary images, the difference in foreground and background is not obvious, which limits the performance of this method. To obtain ideal results, active contour models such as gradient vector flow (GVF) [9] has also been widely used on CT and MR images.

In addition, many machine learning methods are applied to image segmentation tasks. Tara Saikumar et.al.[10] proposed a method which combined the watershed transform, Improved FCM and level set method and performed well on MR images. In 2012, Reza Kharghanian et al.[11] proved combination of Gabor features and line features provided a good performance for vessel segmentation. Multiple features can effectively improve the segmentation effect. In 2014, Chi-Man Pun and Pan Ng [12] proposed a skin color segmentation system which combined color and textures feature as segmentation feature, which performed better than GMM algorithm.

In recent years, CNN (Convolutional Neural Network) has shown great potential in image deep learning. While in terms of image segmentation, the FCN has become one of the main frames. It accepts images with any size, recovers the category of each pixel from abstract features and classifies the images in pixel-level. The U-Net model proposed by Olaf Ronneberger et al.in 2015 was a classic FCN model. It used the encoder-decoder structure to merge features, which had high accuracy and robustness in the task of medical image segmentation with small datasets. To further utilize the input image features, Zongwei Zhou et al. proposed the U-Net++ in 2018. Compared to the U-Net, the U-Net++ used dense nested skip connections instead of ordinary skip connections in the encoder-decoder structure to facilitate the simultaneous utilization of features from different levels. The results show that U-Net++ perform well in Data Science Bowl 2018, ASU-Mayo, MICCAI 2018, LiTS Challenge and other data sets. However, the main disadvantage of U-Net and U-Net++ was the difficulty to construct a deep network, and the deep network structure can not necessarily bring better results. To solve this problem, Wanli Chen et al. proposed the W-net model in 2018 by bridging two U-Net models. As the champion model of the 2017 ImageNet competition, SE-net had a great effect in image classification tasks by adjusting the feature channel weights through feature recalibration. Because the SE block can be effectively combined with skip connection, Chaitanya Kaul et al. proposed FocusNet [13] in 2019, which incorporated SE blocks within bridged U-Net and performed well on lung segmentation data set.

3. Model

This section describes our architecture. We first describe the overall structure and logical relationship of the SE-UxNet. After that, the SE block is explained in detail, because of its contribution to the feature recalibration of U-Net and U-Net ++ bridging model.

4.1. SE-UxNet

Our model is based on the U-Net and the U-Net++.U-Net is widely used in the field of image segmentation as a classic medical model. The U-Net++ makes improvement network performance optimization and precision by replacing simple skip connections with dense skip connections. UXNet obtains preliminary

feature extraction by U-Net, and then obtains the final segmentation result through U-Net++.

We refer to these two parts as feature extraction branch and segmentation branch. In order to achieve better segmentation result, we further propose the SE-UXnet using attention structure to recalibrate feature layers. Both branches of our network model are encoder-decoder structures, and in the feature extraction branch, the decoder output of each layer will pass through an SE block. It will then be connected with the output of the SE block, which belongs to the upper layer of the segmentation branch encoder. The structure of our model is shown in the Figure 1.

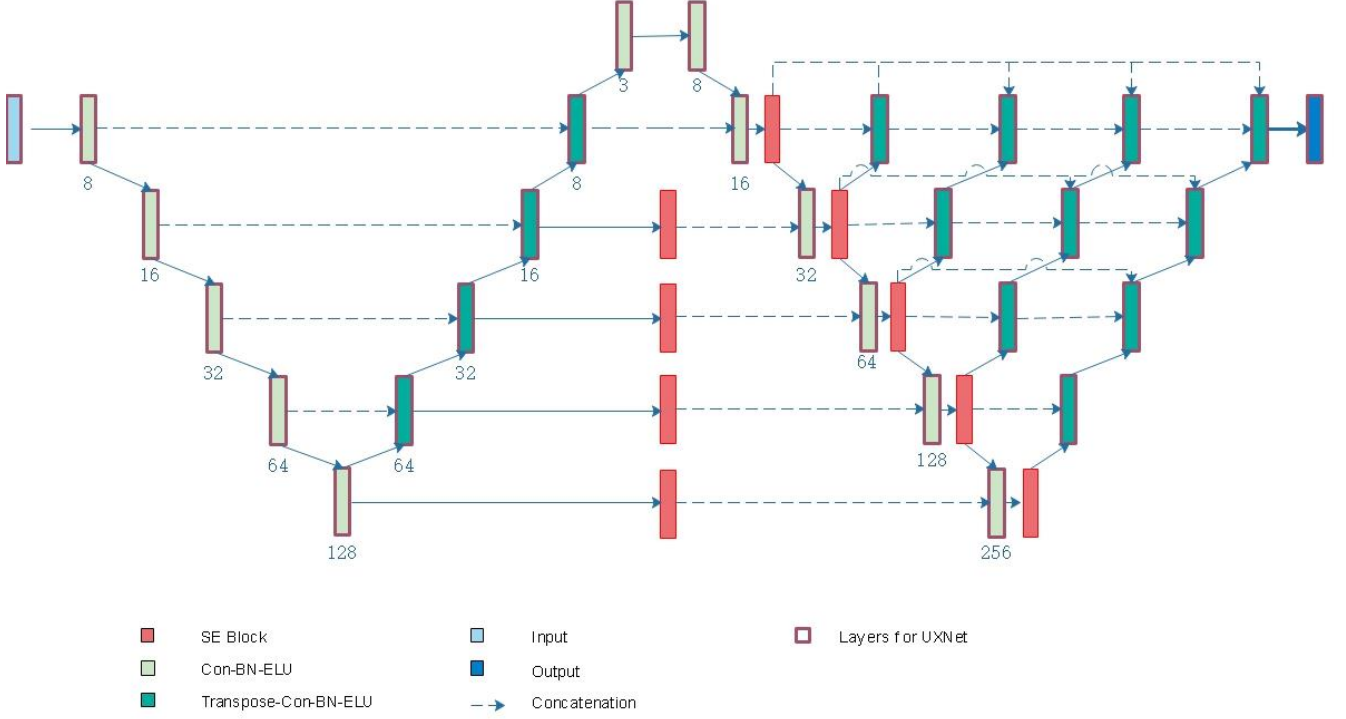


Fig. 1: Overview of the SE-UXNet and UXNet. The arrows indicate the direction of the information flow in the network.

In the feature extraction branch, we enter a picture $x \in X$, X represents the mini batch of pictures each time. Each layer of the encoder will learn a set of feature maps G based on x , given by:

$$E_i^a = G_i(x, W_i) \quad (1)$$

W represents the weights of the standard convolutional layer. The decoder gets the output of the same layer encoder and describes the feature as follows:

$$D_i^a = [E_i^a; H_i(D_{i+1}^a(x, W_{i+1}))] \quad (2)$$

The $[;]$ denotes connection via skip connection, D_{i+1}^a represents the output of the previous decoder layer and H denotes the output from the up-sampling layer.

In the segmentation branch, the output of the outermost encoder structure can be expressed as:

$$E_i^b = \sigma([E_{i-1}^b; \sigma(D_i^a)]) \quad (3)$$

Where E_{i-1}^b represents the output of the upper layer encoder structure in the segmentation branch, $\sigma()$ represents the recalibration of the input features through the SE block. Through the dense skip connection, the output features of the outermost encoder are recalibrated layer by layer, and the output of the segmented branch can be expressed as:

$$y^{i,j} = \begin{cases} E_i^b & j = 0 \\ ([[y^{i,k}]_{k=0}^{j-1}; U(y^{i+1,j-1})]) & j > 0 \end{cases} \quad (4)$$

U denotes a transposed convolution layer, and (i, j) denotes the location of the node in the segmentation branch, for example, the output of the lowest node is $y^{4,0}$. The output of the fourth layer is $y^{3,0}$ and $y^{3,1}$, from left to right. When $j=0$, there are two inputs to the segmentation branch encoder node, respectively the output of the upper encoder node and the output of the same layer feature extraction branch decoder node. At this point, the output $y^{i,0}$ of the encoder node is equivalent to E_i^b . When $j = 1$, there are also two inputs to the encoder node, which are the output of the node at the same layer with $j = 0$, and the transposed convolution output of the node at the lower layer with $j = 0$. For example, node $N^{0,1}$, which outputs $y^{0,1}$, accepts inputs $y^{0,0}$ and $y^{1,0}$. When $j > 1$, the encoder node has $j + 1$ inputs, where j inputs are the outputs of the previous j nodes on the same skip path, and the last input is the output of the transposed convolution layer on the nodes of the lower skip path. All previous feature maps will accumulate and reach the current node. Finally, we apply a 1×1 convolution with sigmoid activation to generate prediction results for each pixel. Upsampling is done using transposed convolution layer rather than up-sampling layer. The other convolutional layers use 3×3 filters and always use dropout with a fixed rate of 0.2.

4.2. SE-Block

The skip connections in the encoder-decoder structure helps provide better gradient flow, which makes it easier to train deeper networks. Similarly, it's also effective on using skip connections between different branches. The structure of the SE module is shown in Figure 2:

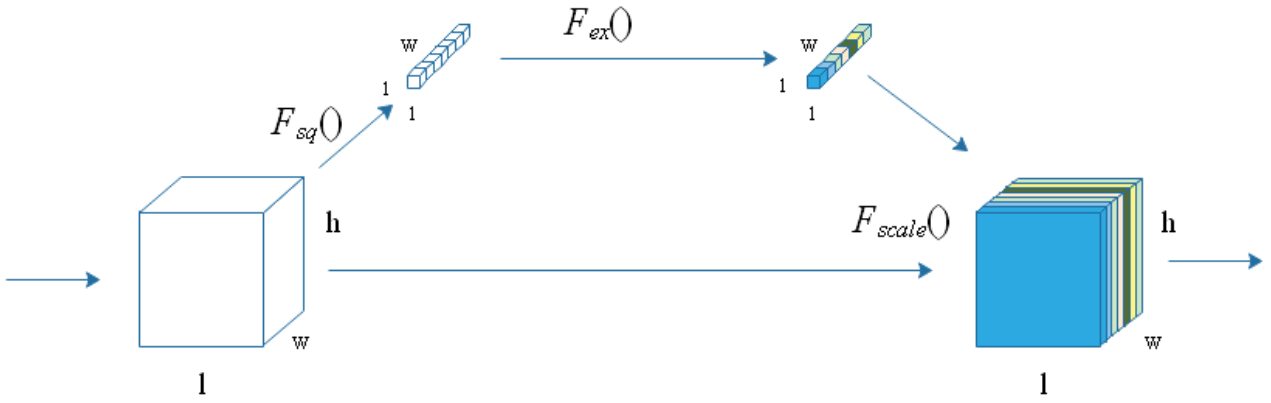


Fig. 2: Recalibration of weights input to SE module features

After the SE module obtains the input, it transforms the input at the feature layer level through a series of convolution and pooling operations. Unlike the traditional CNN, the SE block performs the following three steps. The first one is the Squeeze operation. We perform feature compression along the spatial dimension and turn \mathbf{W} two-dimensional feature channels of $1 \times h$ into \mathbf{W} real numbers. These real numbers have a global receptive field to some extent, and the dimensions of the output match the number of feature channels of the input. Then it uses the Excitation operation to generate weights for each feature channel. The final operation is feature recalibration. We consider the output value of Excitation as the importance of each feature channel after feature selection, and then multiply it with the input features to complete the recalibration of the original features in the channel dimension.

In the SE-UXNet, we use global average pooling for Squeeze operations. Then we use two fully connected layers to model the correlation between the channels and output the same number of weights as the input features. We first use a fully connected layer with a ReLU activation function to reduce the feature dimension to $1/8$ of the input, then rise back to the original dimension through a fully connected layer with a sigmoid activation function to obtain a normalized weight between 0 and 1. Finally, this weight is multiplied with the input features to achieve recalibration, as shown in Figure 3. The advantages of using two fully-connected layers are twofold. They can reduce parameters and have more nonlinearities, furthermore the complex correlations can be fitted better between channels.

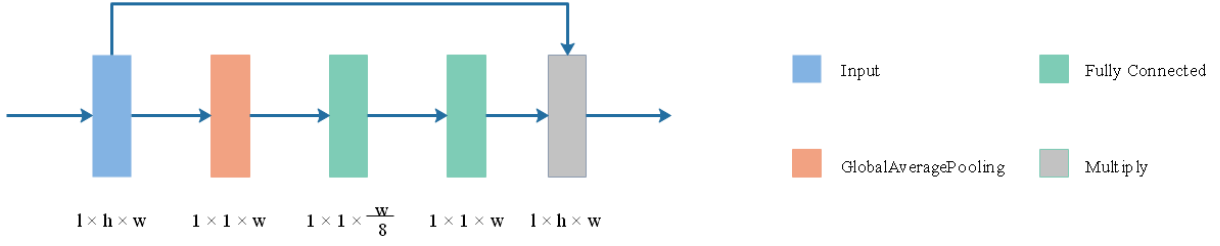


Fig. 3: structure of SE module

4. Experiments

Our experiment is designed to validate whether the SE-UXNet and UXNet can perform better than the simple U-Net and U-Net++. In the following, we first introduce the statistics and evaluation indicators of the experimental data sets, and then present the setting of the experimental parameters. Regarding the experimental results, we compare our methods with several other advanced methods. In addition, we also compare the effects of the SE blocks on the model.

4.1. Data Sets

DSB2018(Data Science Bowl 2018), LiTS (Liver Tumor Segmentation Challenge), MICCAI PROMISE12 and CVC-ClinicDB[14]databases are public, common medical image data sets.

DSB2018 provides 670 sets of training data that are different in terms of cell type, magnification and imaging mode. Each set of data contains an image and multiple labels. Each label contains only one nucleus, and the labels will not overlap each other. LiTS provided 130 CTs of the patient's liver and 64 slices were cut on each CT in 2 mm steps to build a data set. The PROMISE12 data set contains MRI images of 50 prostates obtained from clinical settings and provided by different professional institutions with different prostate sizes and appearances to a certain extent. CVC-ClinicDB is a data set that extracts frames from a colonoscopy video. It provides a total of 612 sets of data, each of which contains image data and label data of intestinal polyps.

Each data set was divided into training set, validation set and test set by a ratio of 8:1:1. In the DSB data set, we use 536 sets of images for training, while using 67 sets of images as the validation set and 67 sets of images as the test set. In the LiTS data set, in order to reduce the duplication of training data, we randomly shattered 8320 slices and selected one every 10 slices to join the new data set. The new data set contains 666 sets of images for training ,83 sets of images in its validation set and 83 sets of test images. In PROMISE12, slices containing prostate parts were cut out, and 259 sets of images for the training set data were obtained, and 32 sets of images for each of the test data set and the validation data set. For the CVC-ClinicDB, we use 490 sets of images as the training data set, and then take 61 sets of images as the validation and test sets respectively.

4.2. Evaluation Metrics

Segmenting a medical image actually divides each pixel of the image into two categories: target pixels and background pixels. In order to quantitatively compare the standard results of expert manual segmentation with the segmentation results of our model, we use the mIoU and the Dice coefficient to measure the performance of the segmentation models. The specific calculation methods are:

$$mIoU = \frac{TP}{TP + FP + FN} \quad (5)$$

$$DSC = \frac{2TP}{2TP + FP + FN} \quad (6)$$

The formulas above contain four statistical indicators. True positive (TP) indicates that the target pixel is correctly segmented. False positive (FP) indicates that the target pixel is incorrectly segmented. True negative (TN) indicates that the background pixel is correctly segmented. False negative (FN) indicates a background pixel point with incorrect segmentation. The higher the degree of overlap between images, the closer the mIoU and the Dice coefficient are to 1, and the closer our segmentation network results are to the standard results of expert manual segmentation.

4.3. Experiment Details

In this section, we present more details regarding the implementation environment and data augmentation strategies. All experiments were done in python, using Keras [15] with a Tensorflow [16] backend. The Stochastic Gradient Descent optimization function is used to optimize the network parameters, and the cosine loss [4] is used to measure the classification error of the network. The initial learning rate is set to 0.001, and the batch size for training and testing was kept at 8. All experiments were run for a maximum of 300 epochs.

4.4. Experiment Results and Analysis

4.4.1. Comparative Analysis of Image Processing Results

The segmentation experiment results obtained on the four public data sets are shown in Figure 4:

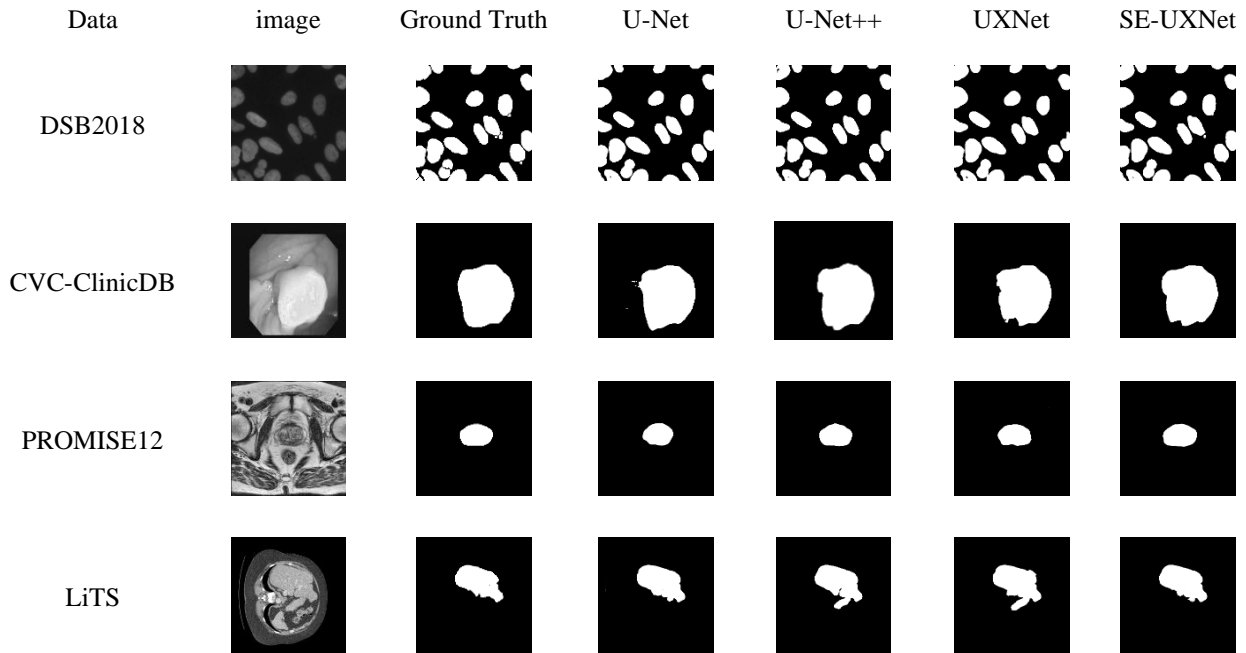


Fig. 4: Qualitative comparison between U-Net, U-Net++, UXNet and SE-UXNet showing segmentation results for DSB2018, CVC-ClinicDB, PROMISE12 and LiTS data sets. From the left:(1) image, (2) Ground truth, (3) U-Net, (4) U-Net++, (5) UXNet and (6) SE-UXNet.

In this study, the experimental results of the SE-UXNet are compared with the U-Net and the U-Net++, and also compared with the UXNet without the SE module, to verify the role of the SE module in skipping connections.

The images in lines 1 to 4 are taken from the DSB2018, CVC-ClinicDB, PROMISE12 and LiTS data sets, respectively. Columns 1 to 5 represent the original image, the label image manually labeled by experts, the segmentation result image of the U-Net, U-Net++, UXNet and the SE-UXNet. By comparison, it can be seen that the SE-UXNet can accurately segment organs and tissue structures with more distinct contours (such as images of liver and prostate) and better divide the fine gaps between images of different cells.

In addition, we used the lung nodule data set to validate the effectiveness of our model. The data set contains 19 patient data, divided into training set, validation set and test set by a ratio of 8:1:1. The results are shown in Figure 5.

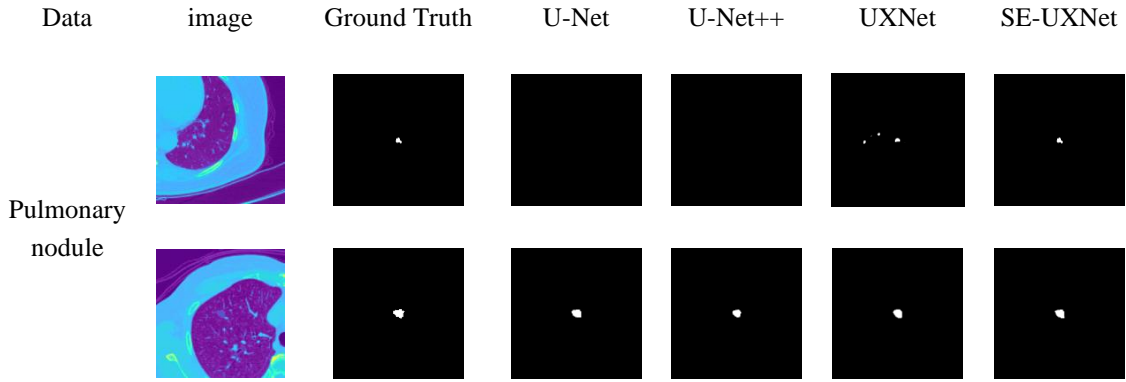


Fig. 5: U-Net, U-Net++, UXNet and SE-UXNet experimental results on lung nodule data set

Experiments show that compared with other structures, our network structure is easier to find and segment smaller nodules, which helps to find and treat the disease earlier.

4.4.2. Comparative Analysis of Results Indicators

To validate the effectiveness and robustness of our method, we analyze the two indicators of mIoU and Dice coefficients, and perform experiments on four public datasets to compare the results with the literatures. Table 1 lists the mIoU values and Dice coefficients of U-Net, U-Net++, UXNet, SE-UXNet, and other advanced models:

- FCN-8S [17], a typical FCN model, up sampling with scale of 8.
- E-D Network [18], propose a deep encoder-decoder network with a novel data augmentation method.
- V-Net [19], present a volumetric convolutional neural network.
- UdeM 2D [20], combine fully convolutional networks with fully convolutional ResNet.
- 3D GCA-Net [21], a 3D encoder-decoder segmentation network.
- Wnet [4], a stacked U-net suitable for small image data sets.
- 2D-Densely Connected CNN [22], using densely connected fully convolutional neural network.
- ResU-Net [23], incorporation of ResNet and U-Net.
- MIMO-FAN (DPS) [24], propose a network architecture for unified multi-scale feature abstraction.

Except for the cell nuclear dataset, the segmentation effect of SE-UXNet is better than the original U-Net, U-Net++, and UXNet. Compared with U-Net and U-Net++, SE-UXNet has a significant improvement in prostate segmentation results, with mIoU increasing by an average of 3.96% and 0.84%, and Dice coefficients increasing by an average of 3.74% and 1.33%.

The UXNet network performs poorly on the polyp segmentation and liver segmentation datasets. On the liver segmentation dataset, the mIoU values of the UXNet network are lower than U-Net++ and SE-UXNet by 0.37% and 0.67%.

However, the UXNet achieves the best segmentation effect on the cell nuclear segmentation data set, indicating that the SE module plays a better role in the data set where only a single segmentation instance exists in each slice. But it does not perform well in segmentation tasks where there are multiple instances in a slice. We think the reason for this is that the SE block has a global field of view, which makes it possible to treat different nuclei in a slice as one. This makes SE-UXNet blur the gaps between nuclei compared to UXNet.

Table 1: Evaluation of SE-UXNet on four public datasets

Dataset	Model	mIoU	DSC
DSB2018	U-Net [2]	90.90	91.02
	U-Net++ [3]	92.11	92.65
	UXNet	93.23	93.37
	SE-UXNet	92.66	92.75
CVC-ClinicDB	FCN-8S [17]	81.0	-
	U-Net	83.14	84.29
	U-Net++	85.24	86.44
	UXNet	86.12	88.58
	E-D Network [18]	88.9	89.35
	SE-UXNet	89.17	90.49
PROMISE12	U-Net	86.42	86.6
	V-Net [19]	-	86.9
	UdeM 2D [20]	-	87.4
	3D GCA-Net [21]	-	88.9
	U-Net++	89.54	89.01
	UXNet	89.94	89.7
	Wnet[4]	-	89.56
	SE-UXNet	90.38	90.34
LiTS	2D-Densely Connected CNN [22]	-	92.3
	U-Net	93.08	93.15
	U-Net++	93.48	94.03
	ResU-Net [23]	-	94.1
	UXNet	93.11	93.86
	MIMO-FAN (DPS) [24]	-	95.1
	SE-UXNet	93.78	95.33

In addition, we evaluated the performance of SE-UXNet on the lung nodule dataset. Unlike medical images of organs such as the liver or prostate, the number of background pixels in a lung nodule image is much larger than the number of pixels in the target image, which makes segmenting a lung nodule more difficult than segmentation tasks such as the liver. Experiments have shown that our model can also perform well in lung nodule segmentation tasks, as shown in Table 2.

Table 2: Evaluation of SE-UXNet on the lung nodule dataset

Dataset	Model	mIoU	DSC
pulmonary nodule	U-Net	66.2	71.23
	U-Net++	69.77	71.35
	UXNet	70.43	71.58
	SE-UXNet	72.45	73.44

As shown in Table 3, SE-UXNet was compared to the bridging U-Net with SE modules (SE-UUNet) and bridging U-Net++ with SE modules (SE-XXNet). We can see that our proposed network performs better than the bridging U-Net and the bridging U-Net++ which has more parameters. Therefore, we choose SE-UXNet as our experimental model.

Table 3: The comparison of SE-UXNet, SE-UUNet and SE-XXNet

Dataset	Model	mIoU	DSC	Paras
CVC-ClinicDB	SE-UUNet	83.82	86.11	1.13M
	SE-XXNet	86.94	87.75	1.30M
	SE-UXNet	89.17	90.49	1.21M

5. Conclusion

In this paper, we propose the innovation structure SE-UXNet, which is a deep learning framework using the attention module. This framework is suitable for image segmentation tasks of small image data sets such as medical image data sets. We propose a new bridging model and discuss the role of the SE module in skip connections. Data Science Bowl 2018, MICCAI PROMISE12, Liver Tumor Segmentation Challenge, CVC-ClinicDB data sets and lung nodule data set were used to evaluate our network model. The results show that our network model is better than U-Net and U-Net++. At the same time, it has certain competitiveness with other advanced models. Future works will aim at further segmenting tumors or nodules on segmented organs and segmenting slices containing multiple instances by adjusting structure of SE block for multiple field of view.

6. Acknowledgements

We thank all viewers who provided the thoughtful and constructive comments on this paper. The third author is the corresponding author. This research is funded by Shanghai Key Laboratory of Multidimensional Information Processing, East China Normal University, Shanghai 200241, China. The computation is supported by ECNU Multifunctional Platform for Innovation (001).

7. References

- [1] Long J, Shelhamer E and Darrell T, "Fully convolutional networks for semantic segmentation," Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 3431-3440, 2015
- [2] Ronneberger, Olaf, Philipp Fischer, and Thomas Brox, "U-Net: Convolutional networks for biomedical image segmentation," International Conference on Medical image computing and computer-assisted intervention. Springer, Cham, pp. 234-241, 2015
- [3] Zhou, Z., Siddiquee, M. M. R., Tajbakhsh, N., and Liang, J, "Unet++: A nested U-Net architecture for medical image segmentation," Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. Springer, Cham, pp. 3-11, 2018
- [4] Chen, Wanli, et al. "Prostate segmentation using 2D bridged U-Net," 2019 International Joint Conference on Neural Networks (IJCNN). IEEE, pp.1-7, 2019
- [5] Hu, Jie, Li Shen, and Gang Sun, "Squeeze-and-excitation networks," Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 7132-7141, 2018
- [6] Otsu, Nobuyuki. "A threshold selection method from gray-level histograms," IEEE transactions on systems, man, and cybernetics 9.1, pp. 62-66, 1979
- [7] Adams, Rolf, and Leanne Bischof, "Seeded region growing," IEEE Transactions on pattern analysis and machine intelligence 16.6, pp. 641-647, 1994
- [8] Canny, John. "A computational approach to edge detection," IEEE Transactions on pattern analysis and machine intelligence 6, pp.679-698, 1986
- [9] Xu, Chenyang, and Jerry L. Prince, "Snakes, shapes, and gradient vector flow," IEEE Transactions on image processing 7.3, pp. 359-369, 1998
- [10] Tara. Saikumar, P. Yugander, P.S. Murthy, and B. Smitha, "Improved Fuzzy C-Means Clustering Algorithm Using Watershed Transform on Level Set Method for Image Segmentation," International Journal of Machine Learning and Computing vol. 2, no. 1, pp. 19-23, 2012
- [11] Reza Kharghanian and Alireza Ahmadyfard, "Retinal Blood Vessel Segmentation Using Gabor Wavelet and Line Operator," International Journal of Machine Learning and Computing vol.2, no. 5, pp. 593-597, 2012
- [12] Chi-Man Pun and Pan Ng, "Skin Segmentation Using GMM Classifier and Texture Feature Extraction," International Journal of Machine Learning and Computing vol.4, no. 1, pp. 57-62, 2014
- [13] Kaul, Chaitanya, Suresh Manandhar, and Nick Pears, "FocusNet: An attention-based Fully Convolutional Network for Medical Image Segmentation," 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), pp. 455-458, 2019

- [14] Bernal, Jorge, et al. "WM-DOVA maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians," *Computerized Medical Imaging and Graphics* 43, pp. 99-111, 2015
- [15] François Chollet et al. "keras," 2015
- [16] Abadi, Martín, et al. "Tensorflow: Large-scale machine learning on heterogeneous distributed systems," *arXiv preprint arXiv:1603.04467*, 2016
- [17] Akbari, Mojtaba, et al. "Polyp segmentation in colonoscopy images using fully convolutional network," 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, pp. 69-72, 2018
- [18] Nguyen, Quang, and Sang-Woong Lee. "Colorectal segmentation using multiple encoder-decoder network in colonoscopy images," 2018 IEEE First International Conference on Artificial Intelligence and Knowledge Engineering (AIKE). IEEE, pp. 208-211, 2018.
- [19] Milletari, Fausto, Nassir Navab, and Seyed-Ahmad Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," 2016 Fourth International Conference on 3D Vision (3DV). IEEE, pp. 565-571, 2016.
- [20] Drozdal, Michal, et al. "Learning normalized inputs for iterative estimation in medical image segmentation," *Medical image analysis*,44, pp. 1-13, 2018
- [21] Jia, Haozhe, et al. "3D Global Convolutional Adversarial Network for Prostate MR Volume Segmentation," *arXiv preprint arXiv:1807.06742*, 2018
- [22] Kaluva, Krishna Chaitanya, et al. "2d-densely connected convolution neural networks for automatic liver and tumor segmentation," *arXiv preprint arXiv:1802.02182*, 2018
- [23] Han, Xiao. "Automatic liver lesion segmentation using a deep convolutional neural network method," *arXiv preprint arXiv:1704.07239*, 2017
- [24] Fang, Xi, et al. "Unified Multi-scale Feature Abstraction for Medical Image Segmentation." *arXiv preprint arXiv:1910.11456*, 2019