2020 19th International Symposium on Distributed Computing and Applications for Business Engineering and Science (DCABES)

# Semantic Segmentation of Brain MRI Based on U-net Network and Edge Loss

1st Zude Wang
*IoT school*
*Jiangnan University*
Wuxi, China
wzd1003@163.com

2nd Leixin Zhang
*IoT school*
*Jiangnan University*
Wuxi, China
1137804516@qq.com

***Abstract*—Brain MRI analysis is of great significance for extracting clinical information from patients and providing diagnostic recommendations for doctors. However, brain MRI is difficult to detect and segment because of its complex structure and great difference. To deal with these problems, in the field of medical image semantic segmentation, the model based on U-net structure in the depth neural network model has excellent performance. In this work, in order to increased the influence of image edge pixels on segmentation results and improved the accuracy of medical image segmentation, we provided a more optimized U-net model that can be applied to medical image semantic segmentation. The model integrated the convolution block attention module and after the feature extraction, combined with the edge detection network, the segmentation effect of edge pixels was enhanced. At the same time, by improving the residual convolution block, the amount of parameters was greatly reduced. On the BraTS-2017 data set, we had good experiment results.**

***Keywords—semantic segmentation, attention module, edge detection, BraTS-2017***

## I. Introduction

In recent years, deep convolutional neural networks have achieved good results in various fields such as natural image classification and segmentation, and they have good migration capabilities, making deep learning research and applications extremely extensive. Also in the field of medical imaging, the research of deep convolutional neural network is increasing day by day. The research focus of this paper is the semantic segmentation of MRI brain.

With the rapid progress of medical imaging technology, most of the disease detection depends on the analysis of medical imaging, and the final diagnosis of the disease is still dependent on the pathologist's subjective experience. This process is too dependent on the ability of the doctors, and in the actual diagnosis process, not only consumes a lot of time, but also easy to misdiagnose, so the use of computer-assisted doctors to analyze medical images can greatly reduce the work intensity and improve the diagnosis efficiency.

Magnetic resonance imaging (MRI) is the most common in medical imaging. Because its lines are relatively clear and the diagnosis process is relatively harmless to the human body, it is often used for pathological detection of various tissues such as the brain and spine. In the actual diagnosis, the MRI image structure of the human brain is extremely complicated, so different sequences will be used to scan and image different lesions. The diagnosis usually refers to four produce different responses to different tumor tissues. Semantic segmentation can use multiple modes to automatically distinguish tumor tissues to assist doctors in diagnosis and treatment.

## II. Model Building

In this paper, based on the deep residual network proposed by He et al. [1], combined with the convolutional attention module in [2], it is improved on the basis of U-net, and designed to be suitable for MRI semantic segmentation. In the task residual attention U-net network, in the process of feature extraction and upsampling, we redesign-ed the residual block. Introduce convolutional attention module in layer jump connection. In the process of training parameters, an edge detection network is introduced and gradient descent algorithm is used for training.

### A. Convolution Block Attention Module

Convolutional Block Attention Module (CBAM) is a high efficiency and low memory consumption attention module for convolutional neural networks. This module has two consecutive sub-modules: channel attention module and spatial attention module. The all feature map is flexibly redefined in each convolution block of the deep network by the convolutional attention module. It is precisely because the convolution operation is to pick up information features by mixing channels and spatial information together, and the convolution attention module is to decompose the meaningful features of the two main dimensions of channels and spaces. The convolutional attention module effectively helps the information flow in the network by understanding the information to be strengthened or suppressed. The model of the CBAM network is shown in Figure 1.

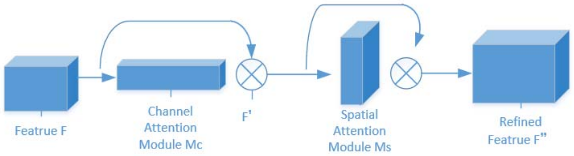

Figure 1. Convolution attention module.

Given the convolutional feature map F as input, the channel attention $M_c$ and spatial attention $M_s$ are calculated through CBAM. The result $F''$ and the process of the entire convolution attention module are as follows:

$$F' = M_c(F) \otimes F \tag{1}$$

$$F'' = M_s(F') \otimes F' \tag{2}$$

2473-3636/20/$31.00 ©2020 IEEE
DOI 10.1109/DCABES50732.2020.00048

Among them: $\otimes$ means element-by-element multiplication, the channel attention $M_c$ and spatial attention $M_s$ processes are defined as follows, and $\sigma$ represents the sigmoid function:

$$M_c(F) = \sigma\big(MLP\big(Avgpool(F)\big) + MLP\big(Maxpool(F)\big)\big) \quad (1)$$

(3)

$$M_s(F') = \sigma\big(f^{7\times7}\big([Avgpool(F')); (Maxpool(F')]\big)\big) \quad (2)$$

(4)

The input feature map F is respectively subjected to the maximum pooling operation and the average pooling operation, and then to the two-layer shared neural network layer (MLP). The result is added as the output feature, and then activated by the activation function to generate the final channel attention feature map. The generated channel attention feature map F' and the input feature map F element are multiplied correspondingly to generate the input feature F' in the spatial attention module. Take the output of the channel attention module as a channel-based maximum pooling and average pooling operation, and then connect the two results based on the channel dimension. Then, after a convolution operation, the dimension is reduced to a channel. Then through the activation function to generate spatial attention feature map. Finally, the feature map and the input feature of the spatial attention module are multiplied to obtain the final feature F".

Since CBAM is an easy to build generic module, it can be seamlessly integrated into any convolutional neural network. The consumption of architecture can be ignored, and can also achieve good results.

### *B. Edge detection network*

The loss function commonly used in the conventional medical image semantic segmentation network is the cross-entropy loss of global pixels. The loss function checks each pixel separately and compares the class prediction (pixel vector in the depth direction) with our thermally encoded target vector. Because the distinction between the edge pixels and the surrounding pixels of the brain MRI image is not obvious, the conventional method has certain difficulties in the training process, so we propose an edge detection network and increase the loss of edge pixels. This paper uses gradient descent algorithm for training.

S is the input image, and $f_\theta$ is the U-net feature extraction network:

$$\hat{S} = f_\theta(S) \in [0,1]^{H*W*C} \quad (5)$$

The semantic segmentation result obtained by the network is $\hat{S} \in [0,1]^{H*W*C}$, where H represents the height of the input image, W represents the width of the input image, and C represents the number of channels.

$g_\varphi$ is the edge feature extraction network, and get the edge feature map $\hat{E}$ :

$$\hat{E} = g_\varphi(\hat{S}) \in [0,1]^{H*W*2} \quad (6)$$

### *C. Network model*

In order to alleviate the problem of gradient dis-appearance caused by the increase in network depth, the network in this paper designs each basic convolutional block as a residual convolutional block. Automatically find important features in the entire channel distribution and spatial distribution. Since most of the current high-accuracy literatures are trained on GPU clusters, which leads to low practicality, in order to enable programs to be trained and run on a single GPU, this paper uses the shallowest network possible to achieve better effect.

The network can be divided into two parts, one is the feature extraction network, and the other is the edge detection network[3]. Perform simple data preprocessing on the data before training. The network model diagram is shown in Figure2.

The encoder part first uses a convolution block of size 32 for convolution operation, and then uses the residual convolution module for feature extraction. Each operation is composed of convolution with Batch Normalization (BN) and Relu activation function, and then the layer-by-layer connection combined with the convolution attention module, the filter size gradually increases to 32, 64, 128, 256 . The decoder part uses the same residual convolution module as the encoder, combines the channel attention and spatial attention features collected by feature extraction, and performs upsampling operations. The final result is a convolution operation, and it is output through the sigmoid function. The detailed network parameters are shown in TABLE Ⅰ.

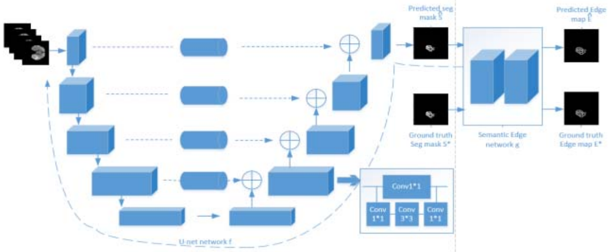

Figure 2. Network structure diagram.

TABLE I. NETWORK PARAMETER TABLE

| | ***Input*** | ***Filter*** | ***Output*** |
|---|---|---|---|
| Input | 176*176*4 | - | - |
| Conv | 176*176*4 | 32 | 176*176*32 |
| ResBlock1 | 88*88*32 | 32 | 88*88*32 |
| ResBlock2 | 44*44*32 | 64 | 44*44*64 |
| ResBlock3 | 22*22*64 | 128 | 22*22*128 |
| ResBlock4 | 11*11*128 | 256 | 11*11*256 |
| ResBlock5 | 11*11*256 | 256 | 11*11*256 |
| ResBlock6 | 22*22*384 | 256 | 22*22*256 |
| ResBlock7 | 44*44*320 | 128 | 44*44*64 |
| ResBlock8 | 88*88*160 | 64 | 88*88*64 |
| Conv | 176*176*96 | 32 | 176*176*32 |
| Output | - | 1 | 176*176*1 |

Our edge detection network performs Gaussian smoothing on the predicted semantic segmentation results and real labels output by the network, and then uses the sobel operator($G_x, G_y$)as a convolution kernel to convolve, and the output loss is calculated as the mean square error, which is backpropagated, training network parameters.

The typical loss of deep U-net network training for semantic segmentation is Per-Pixel Cross Entropy (PPCE) loss $L^{PPCE}$ . When edge detection networks are not considered, PPCE will treat each pixel equally. The loss of the prediction result $\hat{S}$ and the true marker $S^*$ calculated by the loss function is:

$$\mathcal{L}^{ppce}(\hat{S}, S^*) = -\sum_{i=1}^{H}\sum_{j=1}^{W}\sum_{c=1}^{C} S_{i,j}^{c*} \log(\hat{S}_{i,j}^{c}) \tag{7}$$

Now we can get the edge $E^*$ of the real mark through the true mark $E^*$ and directly through the edge detection network$g_\varphi$:

$$E^* = g_\varphi(S^*) \tag{8}$$

Get the proper network parameters of edge detection through back propagation training:

$$\varphi^* = argmin_\Psi \mathcal{L}^{ppce}(\hat{E}, E^*) \tag{9}$$

Through global pixel information and edge pixel information, optimize the parameters through the loss function value:

$$\theta^* = argmin_\theta \mathcal{L}^{ppce}(\hat{S}, S^*) + \lambda_1 \mathcal{L}^{ppce}(\hat{E}, E^*) \tag{10}$$

When we superimpose convolutional layers and non-linear layers, deep network embedding captures more abstract information. Therefore, the edge-based PPCE is not the best method to use the edge detection network to match the two segmentation templates. The multitasking formula optimizes the same loss, but the gradient of the back propagation through the segmentation head does not consider edge constraints, which is why the architecture does not enforce structural information in prediction.

In the framework of image synthesis, better results are obtained by defining losses that contain fixed pre-trained networks. The idea of this method is to measure the semantic difference between the two images as the difference expressed by the feature calculated by the fixed network. Therefore, we define the edge loss as:

$$\mathcal{L}^{g}(\hat{S}, S^*) = \sum_{i=1}^{H}\sum_{j=1}^{W}\sum_{c=1}^{C} \lambda_c |\Psi_{i,j}^{c}(\hat{S}) - \Psi_{i,j}^{c}(S^*)| \tag{11}$$

Where $\Psi_{i,j}^{c}$ represents the pixel embedded at the depth c of the semantic edge detection network $g_\varphi$ ($\Psi_{i,j}^{c}$ is a c-dimensional vector, where c is the number of channels in layer 1), $\lambda_c$ Is a hyperparameter that represents importance. The final loss of the entire network consists of two parts: global pixel loss and edge pixel loss:

$$\mathcal{L}(\hat{S}, S^*) = \mathcal{L}^{ppce}(\hat{S}, S^*) + \mathcal{L}^{g}(\hat{S}, S^*) \tag{12}$$

## III. EXPERIMENTAL RESULTS AND ANALYSIS

### *A. Experimental data set and evaluation index*

Because different brain tumors have different degrees and large deformations, we use the BraTS-2017 data set [4,5,6] multi-modal brain tumor image segmentation bench-mark, which contains 210 high-grade glioma cases (HGG) and 75 low-grade glioma cases (LGG). Because there are many useless pixels on the edge of the image, the middle 176*176*128 is taken as the input of the experiment. When training, first select 0.15 as the verification set of cross-validation. The total number of parameters for the entire training process is about 570,000. The result of the labeling of this data set is shown in Figure 3 below. Figure A shows the tumor edema area (yellow), Figure B shows the tumor core area (red), Figure C shows the tumor necrosis area and focus area (dark green and green), and Figure D shows the comprehensive marker of the entire tumor structure.

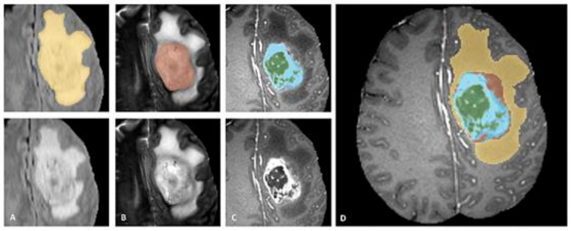

Figure 3. Annotation of the data set.

In order to verify the effectiveness of the model, this paper uses the experimental model of the graphics card GeForce GTX 1080 Ti, the system is Ubuntu16.04 system, using Python language based on TensorFlow 1.8.0 and Keras 2.2.4 experiments, experimental results and comparison of mainstream models The results are shown in the table II below, where WT is the accuracy rate of the entire tumor area, TC is the accuracy rate of the tumor core area, and ET is the accuracy rate of the HGG tumor enhancement area. The number of training cycles epoch is 20, the batch size is 32, the activation function of the convolutional layer is Relu, the initial learning rate is set to 0.005, when the accuracy of the verification set is stagnant, iterative adjustment reduces the learning rate. Compare the experimental results of multiple models in the following TABLE II

TABLE II. EXPERIMENTAL RESULTS

| | ***WT*** | ***TC*** | ***ET*** |
|---|---|---|---|
| *EMMA* | 90.10 | 79.70 | 73.80 |
| *MC* | 90.41 | 78.48 | 72.91 |
| *OM-Net* | **91.28** | 82.50 | **80.84** |
| *CU-Net* | 88.80 | 83.10 | 78.40 |
| ***Ours*** | 90.59 | **86.64** | 77.50 |

Analysis of experimental results: Since this model improves the residual convolution module and adds the convolution attention module, it uses fewer parameters than other models. Under the same configuration, the training network takes the shortest time and is segmented in areas with obvious edge features. The effect is obvious and the accuracy is high.

EMMA[7] is a combination of multiple models, so there is no targeted solution to the problem, but only a large number of operations to reduce errors and consume a lot of system resources. Although MC[8] has good results, it still has the following disadvantages. First of all, it usually needs to train multiple depth models, which greatly increases the model complexity and storage memory space consumption. Second, each model uses its own training data for individual training, while ignoring the correlation between deep models. Third, the MC runs the depth model one-to-one, which leads to alternate calculations of GPU-CPU and lack of online interaction between tasks. In order to achieve a single brain tumor segmentation and better solve the imbalance problem. The result of OM-Net[9] experiment is preferably the increase of the number of models built in MC. Although it can be proved that the effectiveness of the models in MC and the superposition of models can achieve good results, a large number of models will cause the system to be complicated and larger in storage. Consumption, followed by blindly increasing the number of models until the final segmentation effect is not obvious, and even the accuracy rate may decline, so multiple experiments are required to select the appropriate model, and the success of OM-Net also stems from his introduction of cross Cross-task Guided Attention (CGA) is a CGA that can guide the subsequent task through the prediction result of the previous task, and obtain the category-specific statistical information of each channel in advance. The category-specific information also enables the CGA to separately predict the channel-by-channel correlation of voxels for a particular category, so the OM-Net+CGA experiment has high accuracy. CU-Net[11] uses two cascaded U-Net networks, which deepens the network depth, exacerbates the problem of gradient disappearance, and consumes more time and memory.

### B. *Ablation experiment*

TABLE III. ABLATION EXPERIMENTS

| | ***No edge detection network*** | | | ***With edge detection network*** | | |
|---|---|---|---|---|---|---|
| | WT | TC | ET | WT | TC | ET |
| **Ours** | 88.44 | 86.30 | 76.15 | **90.59** | **86.64** | **78.17** |

The ablation experiment results show that the use of edge loss function can improve the segmentation effect of medical images and improve the segmentation accuracy. Compared with the model that does not use the edge detection network only, the segmentation effect of the overall tumor (WT) added to the edge detection network is increased by 2.15%, the segmentation effect of the core area of the tumor is increased by 0.34%, and the accuracy of the HGG tumor enhancement area is increased by 2.02%.

## IV. CONCLUSION

This paper proposes a network structure based on the residual convolution module and the convolution attention module, combined with the improvement of the loss function, applied to the semantic segmentation of medical images of the brain MRI, which improves the accuracy of pixel segmentation and reduces the parameters. The amount of use, through the experimental test of the data set BraTS-2017, shows that the model has a high accuracy rate for the screening of various types of brain lesions, and can assist doctors in the actual situation in screening the lesions. Experimental expansion: related articles prove that dilated convolution is effective for brain tumor segmentation. Introducing context without losing output spatial resolution is a direction for future exploration.

## REFERENCES

[1] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition [C] .Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV: IEEE, 2016:770-778.

[2] Sanghyun Woo, Jongchan Park,et al. CBAM: Convoluional Block Attention Module.In CVPR,2018.

[3] Yifu Chen, Arnaud Dapogny. SEMEDA: Enhancing Segmentation Precision with Semantic Edge Aware Loss.In CVPR,2019.

[4] I. Kokkinos. Ubernet: Training a universal convolutional neural network for low-, mid-, and high-level vision using diverse datasets and limited memory. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 6129–6138, 2017.

[5] J. Johnson, A. Alahi, and L. Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In European conference on computer vision, pages 694–711. Springer, 2016.

[6] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. In International Conference on Learning Representations, 2015.

[7] Konstantinos Kamnitsas, Wenjia Bai.15-Ensembles of Multiple Models and Architectures for Robust Brain Tumour Segmentation.In CVPR,2017.

[8] Xinchao Wang, Dacheng Tao.One-pass multi-task convolutional neural networks for efficient brain tumor segmentation. in Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent, 2018

[9] Guotai Wang, Wenqi Li.Automatic Brain Tumor Segmentation using Cascaded Anisotropic Convolutional Neural Networks. In CVPR,2017.

[10] Chenhong Zhou, Changxing Ding.One-pass Multi-task Networks with Cross-task Guided Attention for Brain Tumor Segmentation. In CVPR,2019.

[11] Hongying Liu, Xiongjie Shen.Cascaded U-Net with Loss Weighted Sampling for Brain Tumor Segmentation. Image and Video Processing,2019.