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

Medical Image Segmentation Based on 3D U-net

1st Silu Chen
School of Internet of things engineering
Jiangnan University
Wuxi,China
794258229@qq.com

2st Guanghao Hu
School of Internet of things engineering
Jiangnan University
Wuxi,China
huguanghao520@gmail.com

3st Jun Sun
School of Internet of things engineering
Jiangnan University
Wuxi,China
sunjun_wx@hotmail.com

Abstract—For medical image processing, as the target area of the tumor lesions is small, and the boundaries of the organs are blurred, so the segmentation of the medical images is difficult. In the original 3D U-net model, feature extraction is performed on the interest image region by increasing the channel attention mechanism, so that the model keep a watchful eye on key region before segmentation. Test results indicate that the improved model has significantly improved segmentation accuracy relative to the original 3D U-net model and is a valid image segmentation model.(*Abstract*)

Keywords—Medical image segmentation; 3D U-net; channel attention mechanism(*Keywords*)

I. INTRODUCTION

Malignant tumors are one of the main causes of human death at this stage, which seriously threatens human life and health. Every year, there are a large number of new cases of cancer deaths at home and abroad, and about 60% of cancer patients receive radiotherapy during the treatment process.

In traditional medical diagnosis, the diagnostic results mainly depend on the judgment of experts. Generally speaking, medical images require very high segmentation accuracy. Still, because of the low SNR (Signal to Noise Ratio) of medical images, even doctors with long-term professional training may be affected by some external factors lead to inaccurate diagnosis by the doctor. Therefore, as deep learning has realized excellent performance in all fields of machine vision today, applying it to medical images to achieve automatic segmentation of tumor lesions is necessary. It can improve the accuracy of segmentation as well as reduce doctor errors result from fatigue.

Many scholars have proposed various medical image segmentation methods before. Still, in the tumor segmentation task, due to the large differences in shape and texture between different tumors, coupled with medical images involving patient privacy and other issues, the amount of data is small, these problems have increased the difficulty of medical image processing[1]. So as to solve the problem of medical image segmentation, the predecessors proposed many segmentation methods, mainly divided into threshold processing, region growth, learning methods, deformable models, and neural networks.

The neural network is a practical artificial intelligence technology since Hinton G E et al.[2] proposed deep learning, it has entered a rapid development stage. Due to the problem of the low computational efficiency of traditional neural networks, Long J et al.[3] proposed FCN (Fully Convolutional Networks). FCN solves the problem of image segmentation at the semantic level by classifying images at the pixel level. In 2015, Ronneberger O et al. [3] proposed a U-shaped structure and named it U-Net by improving the FCN network, and the network solves the problem of pixel positioning through shallower layers and the classification of pixels through deeper layers to better achieve image segmentation. After the U-net network was proposed, the segmentation accuracy of medical images has been improved to a certain extent. Many scholars did medical image segmentation based on the U-net network. In 2017, Cicek et al.[4] realized the three-dimensional segmentation of medical images through the dimensional upgrade operation of the U-net network.

3D data is redundant for biomedical data analysis, for the following reasons: (1) Since computer screens can only display 2D slices, it is challenging to segment labels on a three-dimensional level. (2) The information of adjacent

2473-3636/20/$31.00 ©2020 IEEE
DOI 10.1109/DCABES50732.2020.00042

layers is almost the same , and taking slices labeled layer by layer is tedious and redundant. Therefore, fully annotating 3D data is not an effective way to create large and rich training data sets, especially for learning algorithms that require large amounts of labeled data[6].3D U-net is a network that can perform 3D data segmentation only, by training with 2D labeled data[7].

II. RELATE THEORIES

A. 3D U-net

Network structure:

1) There is an encoding path and a decoding path, and each has four resolution levels.
2) Each layer of the encoding path contains two 3 × 3 × 3 convolutions, each followed by a ReLU layer, and then a 2 × 2 × 2 maximum pooling layer with a step size of 2 in each direction.
3) In the decoding path, each layer contains a 2 × 2 × 2 deconvolution layer with a step size of 2, followed by two 3 × 3 × 3 convolution layers, each followed by a ReLU layer.
4) Pass the same resolution layer in the encoding path to the decoding path through a shortcut to provide it with the original high-resolution features.
5) The last layer is a 1 × 1 × 1 convolution layer, which can reduce the number of output channels, and the final number of output channels is the number of label categories.

The input of the network is a set of 132 × 132 × 116 pixels with three channels. The size of the output is 44 × 44 × 28, and the size of each pixel is 1.76 × 1.76 × 2.04 μm^3. Since each predicted segmented pixel has an approximate acceptance range of 155 × 155 × 180 μm^3, each output pixel has enough space for training and learning.

We set the weight of unlabeled pixels to zero, then the network learns only from labeled pixels and generalizes to the whole stereo data.

B. Channel attention mechanism

The visual attention mechanism allows humans to quickly select valueable information from plentiful data with limited attention resources. When we promptly scan an image, there is usually a target area that needs our focus, and this is what we call center of attention. Subsequent observations require more attention resources for this vital area to get more detailed information about the target you need to focus on while suppressing other useless information [8].

The attention mechanism can obtain the key information required by the target from a large amount of information. The core idea of channel attention is that the network learns feature weights through loss to increase the weight of usful feature maps, and reduce feature maps that are ineffective or less effective. So that training results will be better and further improving network performance.

The channel attention mechanism shows in Figure 1:

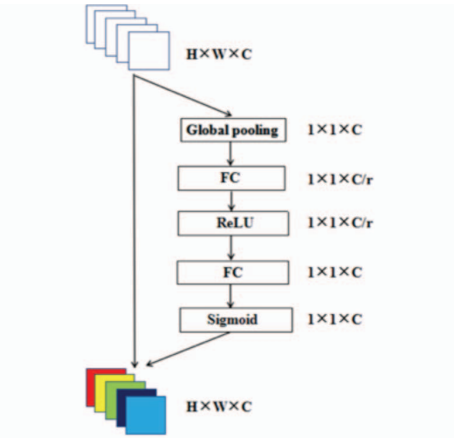

Figure. 1: Channel Attention Mechanism Block

By the above can be known, the input and output feature maps have latitudes of H × W × C, that is, C feature maps of size H × W. First, perform global pooling on the input feature map. The formula is as follows:

$$z_c = F_{GP}(X_c) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} X_c(i,j) \quad (1)$$

Where Xc is the channel c of the input feature map, and Xc (i, j) is the image pixel value in the channel c at position (i, j). $F_{GP}(\cdot)$ is global pooling. The input of $H \times W \times C$ converts into the output of $1 \times 1 \times C$ by Formula 1. Through his process, the numerical distribution of the C feature maps of the layer is obtained, that is, the global information.

After the global pooling operation, the following operations performs, and the formula is as follows:

$$s = \sigma(W_2\delta(W_1 z)) \quad (2)$$

Step:

1) Firstly , use $W_1 \times z$, which is a fully-connected operation, including a dimension reduction operation with a factor of r.
2) The output dimension of the first fully connected operation is $1\times1\times c/r$. Through ReLU layer, the activation function is δ (·), and the output dimension is unchanged through the activation function.
3) Enter the second fully-connected operation, which includes the dimension increasing operation with a factor r, and the output dimension is 1 × 1 × c.
4) Finally, σ (·) is the Sigmoid activation function to obtain s.

The above operation results in an s dimension of 1 × 1 × c (c is the number of channels), which is the core of the channel attention mechanism and is used to represent the weight of each feature map. This weight learns through the previous fully connected layer and the non-linear layer.

After getting s, multiply it by the original channel Xc, which is the corresponding weight. The final channel attention expression is as follows:

$$\widetilde{X_c} = F(X_c, s_c) = s_c X_c \quad (3)$$

III. U-NET NETWORK BASED ON 3D CHANNEL ATTENTION

A. 3D channel attention mechanism

The network structure of the 3D channel attention proposed in this paper shows in Figure 2

The main body of the 3D channel attention mechanism is the same as the 2D channel attention mechanism above. After global pooling, the FC layer, the ReLU layer, the FC layer, and the Sigmoid activation layer[9].

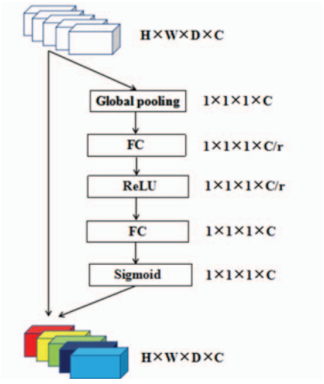

Figure. 2 3D Channel Attention Mechanism Block

Based on the 2D channel attention mechanism[10], 3D channel attention mechanism converts from 2D pixels to 3D voxels, and inputs convert from H × W × C to H × W × D × C, the corresponding global pooling formula is as follows:

$$z_c = F_{GP}(X_c) = \frac{1}{H\times W\times D}\sum_{i=1}^{H}\sum_{j=1}^{w}\sum_{m=1}^{D} X_C(i,j,m) \quad (4)$$

B. Network structure

we insert the channel attention mechanism into the 3D U-net network and add the channel attention mechanism to each layer in the 3D U-net decoding path. The entire network is shown in Figure 3 below:

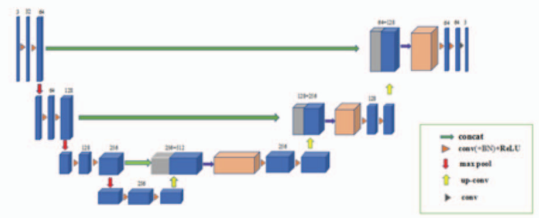

Figure. 3 Network structure

IV.EXPERIMENTAL SIMULATION AND RESULT ANALYSIS

A. Data

The data set used in this experiment is the Brats 2018 data set, which is a data set for multimodal brain tumor segmentation. There are 285 cases in the data set, and each case has four modalities, which are t1, t1ce, t2, and flair. Three parts need to be divided: whole tumor, enhance tumor, and tumor core.

B. Evaluation standard

The Dice coefficient names according to Lee Raymond Dice. It is a kind of similarity measure function. It is usually used to calculate the degree of overlap between the two samples to determine the similarity between the two. It is defined as follows:

$$s = \frac{2|X \cap Y|}{|X| + |Y|} \quad (5)$$

Among them, X is the algorithm segmentation result, Y is the actual segmentation image in this paper,. The dice

value ranges from [0, 1], and the larger the dice value, the higher the experimental segmentation accuracy.

C. *Experimental results*

A total of 500 epochs set in the experimental code and training complets to 197 epochs. Through the box diagram of Figure 4, we can see the experimental results clearly.

The experimental segmentation results are as follows:

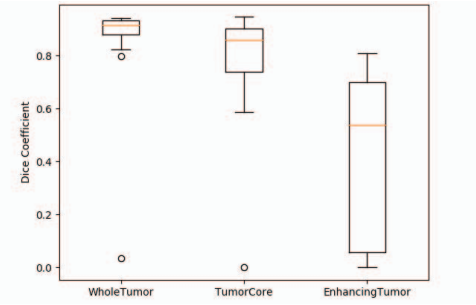

Figure. 4 The box diagram of experimental results

D. *Overview*

The result of the experiment is loU (intersection over union) to measure, that is, compare the overlap between the generated image and the correct labeled part of the image. LoU is defined as follows:

$$\text{loU} = \frac{TP}{TP + FP + FN} \tag{6}$$

Compare the model in this paper with 2D U-net and 3D U-net. The results show in Table 1 below:

Table 1 Comparison of the split loU of each model

Test volume	2D U-net	3D U-net	Text model
1	0.603	0.645	0.906
2	0.678	0.786	0.889
3	0.422	0.779	0.908
Average	0.567	0.736	0.901

It can be seen from Table 1 that the image loU segmented by the model in this paper is better than the segmentation results of the 2D U-net and 3D U-net models. The segmentation accuracy is significantly better than the other two models and can achieve better segmentation results.

V. CONCLUTION

As a classic model of medical image segmentation, 3D U-net mostly solves the embarrassing situation of directly converting 3D images into 2D slices and then feeding them into the model for training. 3D U-net retains the excellent features of 2D U-net while using 3D images can directly improve the segmentation accuracy. In this paper, we add the channel attention mechanism to the original 3DUnet model. In the decoding path of 3D U-net, extracting the target area of the image, and then performing other operations corresponding to the 3D U-net, can enhance the segmentation effect of the network to a certain extent, while reducing the computational pressure.

REFERENCES

[1] Ritter F , Boskamp T , Homeyer A , et al. Medical image analysis[J]. Computer Physics Communications, 2013, 2(6):60-70.

[2] Hinton G E, Osindero S, Teh Y W. A fast learn ing algorithm for deep belief nets[J]. Neural computation, 2006, 18(7): 1527-1554.

[3] Long J , Shelhamer E , Darrell T . Fully Convolutional Networks for Semantic Segmentation[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2014, 39(4):640-651.

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

[5] Çiçek, Özgün, Abdulkadir A , Lienkamp S S , et al. 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation[J]. 2016.

[6] Sebastian Neppl, Guillaume Landry, Christopher Kurz,等. Evaluation of proton and photon dose distributions recalculated on 2D and 3D Unet-generated pseudoCTs from T1-weighted MR head scans[J]. Acta Oncologica, 2019:1-6.

[7] Pizer S M , Fletcher P T , Joshi S , et al. DeformableM-Repsfor 3D Medical Image Segmentation[J]. International Journal of Computer Vision, 2003, 55(2-3):85-106.

[8] Lu, Yue, Zhou, Yun, Jiang, Zhuqing, Guo, Xiaoqiang, & Yang, Zixuan. . Channel attention and multi-level features fusion for single image super-resolution.

[9] Nie, Weizhi, Wang, Kun, Liang, Qi, & He, Roubing. . Panorama based on multi-channel-attention cnn for 3d model recognition. *Multimedia Systems*.

[10] Hu, Jie, Shen, Li, Albanie, Samuel,等. Squeeze-and-Excitation Networks[J].