# Research on a Medical Image Semantic Segmentation algorithm based on Deep Learning

Ai Chen[1], Song Yifan[2], Lv Qi[3]
*1 College of Science, Nanchang University, Nanchang, Jiangxi, 330000*
*2 Jiluan Academy, Nanchang University, Nanchang, Jiangxi, 330000*
*3 Northeast Agricultural University, Harbin, Heilongjiang, 150030*

**Background:** Brain tumor is a kind of tumor with high fatality rate, which accounts for 2.4% of human tumors. Glioma is the primary tumor with the highest incidence and the worst prognosis. Gliomas include glioblastoma (high-grade glioblastoma, HGG) and low-grade cell tumor (LGG)), which have different degrees of aggressiveness and prognosis. Generally speaking, gliomas include edema areas, necrotic core and non-enhanced tumor core areas, and enhanced tumor core areas. Accurate identification of the location and contour information of the tumor and its subregions is a prerequisite for doctors to make accurate diagnosis and treatment plans (such as predicting patient survival, radiotherapy, etc.). Clinically, the accumulation of doctors' personal medical knowledge, differences in the level of experience and visual fatigue and other uncertain factors will affect the analysis of image results. Therefore, it is extremely important to construct a system model for fine segmentation of brain tumors.
**Methods**: A 2.5-dimensional (2.5D) convolution neural network based on cascade is proposed. The task is divided into three sub-tasks: global segmentation of brain tumor, segmentation of tumor nucleus and segmentation of enhanced tumor, and the three results are combined to produce the final result. In each sub-task, the three-dimensional (3D) images are cut in axial, vector and crown directions to generate 2.5D images; 2.5D images are input into the proposed 2.5DV-Net for training; 2.5D segmentation results are spliced into 3D results to generate segmentation results of different sub-tasks.
**Results**:The experimental results show that the average Dice values of the proposed method for tumor whole, tumor nucleus and enhanced tumor segmentation are 0.9071, 0.8542 and 0.8140, respectively, which basically meet the clinical needs.
**Conclusions**: In this paper, a multimodal brain tumor fine segmentation based on cascaded convolution neural network is proposed. through experimental comparison, it is proved that the cascade second-class segmentation network structure is more accurate than the end-to-end convolution neural network structure. In addition, the advantages and disadvantages of 3D, 2D and 2.5D networks are compared and analyzed under the premise of the same convolution neural network structure.

## 1. Introduction

Usually, doctors diagnose patients through multimodal three-dimensional (3D) magnetic resonance imaging (MRI) of the brain. MRI is a medical focus imaging technology realized by generating and collecting magnetic resonance signals, spatial coding and Fourier sampling. 3DMRI technology is the most important imaging

method in the task of brain tumor detection, which is to reconstruct the sequential two-dimensional MRI image sequence into 3D image. Compared with X-ray plain film, angiography scanning and other two-dimensional medical images, 3DMRI can provide doctors with image information of any coordinate points in space, which is convenient for staff to carry out quantitative and qualitative mathematical and medical analysis. In the process of 3DMRI, MRI images of different modes can be scanned due to the difference of auxiliary conditions such as contrast medium. For brain MRI, there are four common modes: magnetic resonance imaging fluid attenuated inversion recovery sequence (FLAIR) image, T1-weighted image (T1 weighted image), T1ce image (post-contrast T1-weighted image) and T2-weighted image (T2 weighted image). MRI images of different modes have different imaging effects on the tumor or its internal subregions. For example, T1-weighted images can accurately identify the location and outline of the tumor core, T2-weighted images can accurately identify the edema area, T1ce can accurately detect whether there is an enhanced tumor in the tumor core, and FLAIR can scan the overall outline of the tumor more clearly. It can be seen that according to multimodal 3D MRI, doctors can not only observe brain tumors from multiple angles, but also obtain more sufficient brain information, which avoids the omission of information and improves the accuracy of diagnosis.

Because of the heterogeneity of tumor microenvironment and the different appearance of different sub-regions of brain tumor, fine segmentation from multimodal MRI images is very complex. Traditional segmentation methods such as graph cut, grab cut, dense CRF and other machine learning algorithms based on graph theory and maximum flow can segment a variety of foreground and background, and achieve more accurate segmentation results on natural images[2-3]. Because most of these methods achieve segmentation by using the color information and boundary contrast information of the image, the feature of the medical image is not as clear as the natural image, and the outline is not distinct, so the segmentation effect is not ideal [4-8].

Since 2012, with the popularity of deep learning and the proposal of full convolution network (FCN), the image segmentation method based on deep convolution network has been developed and applied, and achieved excellent results. Deep Lab network expands the receptive field to obtain global information and improves the segmentation accuracy by introducing hole convolution without lowering the sampling layer. Pyramid scene decomposition network (PSP Net)based on the idea of spatial pyramid pooled (SPP), the features of down sampling at different scales are fused to obtain more comprehensive deep information. Similarly, the segmentation algorithm based on convolutional neural network has achieved good segmentation results on two-dimensional natural image.

For medical images, the proposal of V-Net solves the problem of 3D medical image segmentation and improves the accuracy of 3D image segmentation. However, 2D convolution kernel can not extract spatial features, simply splicing 2D segmentation results into 3D segmentation results, there may be aliasing, faults and other problems, affecting the accuracy of segmentation.

In order to solve the above problems, this paper proposes a cascaded multi-frame two-dimensional convolution neural network, that is, cascaded 2.5D (2.5D) convolution network, and builds a multi-modal brain tumor fine segmentation system. The experimental results show that the network structure can achieve high segmentation accuracy in all parts of the brain tumor. In clinical diagnosis, the proposed brain tumor fine segmentation system can assist doctors in diagnosis. For example, to provide auxiliary information for doctors, doctors can further judge the nature of the tumor and the prognosis of patients according to the location marked by the computer, or assist doctors in making surgical plans and determining a more accurate surgical location according to the location of the tumor.

## 2. Algorithm principle

### 2.1 Algorithm framework

In the task of brain tumor segmentation, if the 3DMRI image of the brain is segmented directly, it will cause misjudgment and blurred boundaries between different sub-regions, and it is easy to misdetect tumors in non-tumor areas such as the back of the brain and brainstem. In order to improve the segmentation accuracy of brain tumor, tumor nucleus and enhance tumor, and reduce false positive, a cascade-based multi-frame two-dimensional convolution neural network, namely cascaded 2.5D convolution network, is proposed. As shown in figure 1, the cascade 2.5D convolution network subdivides the segmentation task of the tumor and its sub-regions into the segmentation of the whole tumor (WT, including all parts of the tumor), the segmentation of the tumor nucleus (TC, including enhanced tumor, non-enhanced tumor and tumor necrosis) and the segmentation of enhanced tumor (ET). This cascade network structure improves the accuracy of fine segmentation of brain tumors.

As shown in figure 2, the cascaded 2.5D convolution network includes image cutting module, tumor global segmentation module, tumor nucleus segmentation module and enhanced tumor segmentation module.

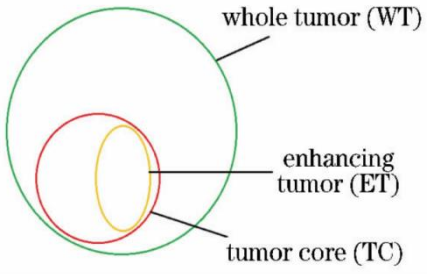


Figure 1 Map of the relationship between the whole brain tumor, tumor nucleus and enhanced tumor location

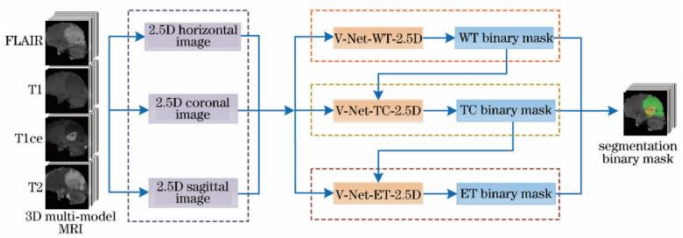


Figure 2 Segmentation system based on cascaded 2.5D convolution neural network

The main results are as follows: 1) 2.5D image is obtained by image clipping of four-mode 3DMRI. 2) the 2.5D image is input into the whole tumor segmentation module for processing, and the whole tumor segmentation image is obtained. 3) input the segmented image of the whole tumor and the 2.5D image into the tumor nucleus segmentation module to get the segmented image of the tumor nucleus. 4) input the segmented image of the tumor nucleus and the 2.5D image into the enhanced tumor segmentation module to get the segmented image of the enhanced tumor. 5) the whole tumor segmentation image, tumor nucleus segmentation image and enhanced tumor segmentation image are combined to get the final tumor segmentation image.

**2.2 Image clipping module**

In the implementation of convolution neural network, because 3D images occupy too much video memory, a multi-frame two-dimensional image format is proposed. In order to distinguish from the two-dimensional image of a single frame, it is called 2.5D image. Assuming that the size of a four-mode 3D image is 4 × 128pixel × 128pixel × 128pixel (4 represents the four modes of the image), it can be regarded as a sequence of 128 four-mode 2D images with the size of 128pixel × 128pixel, in which any continuous N images can be used to form a 2.5D image. Assuming that the size of Numbai 11 128pixel 2.5D image is 4 × 11 × pixel × 128 pixel, it can be understood as an 11-layer continuous four-mode 2D image. In the cropping process, any one of 128 sequence images can be used as the center of 11 layers of continuous 2D images, so. Each 4 × 128pixel × 128pixel × 128pixel 3D image can generate 1284 × 11 × 128pixel × 128pixel 2.5D images (11 layers are filled outward if the central layer is at both ends of the 3D image). In addition, due to the different cutting angles, the above cutting process can be repeated according to the direction of different axes in space, which is called axial cutting, sagittal cutting and crown cutting. Therefore, through this process, each 3D image with the size of 4 × 128pixel × 128pixel × 128pixel can be cut into 384 2.5D images for network training.

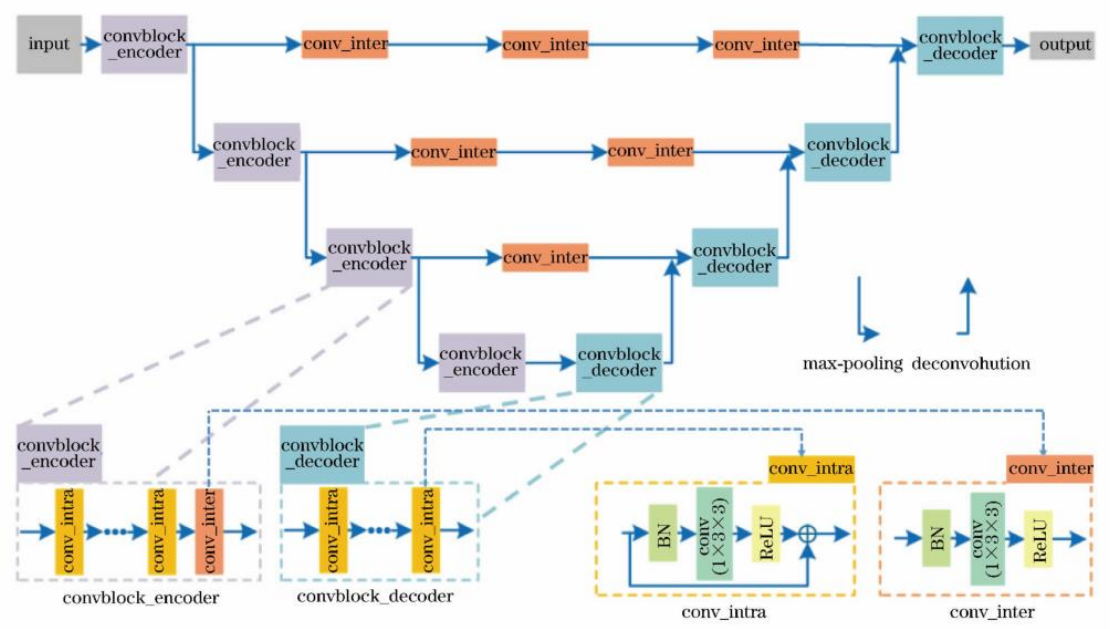


Figure 3 The proposed 2.5DV-Net network structure diagram

In the process of training, 2.5D image clipping includes two parts: image clipping and label clipping. The image clipping process is described above, and the label cropping process refers to cutting out the label map of the three layers in the center of the 2.5D image. For example, a 2.5D image with a size of 4 × 11 × 128pixel × 128pixel is obtained, and the segmented tag map of the middle three layers, that is, the image with the size of 4 × 3pixel × 128pixel × 128pixel, is called 2.5D tag.

Generally speaking, the size of the image and the label is the same in the training process of the convolutional network-based segmentation algorithm, but as mentioned above, the size of the two is not the same in the proposed algorithm, this is because the designed 2.5DV-Net has different inputs and outputs. In this paper, the mechanism of 2.5DV-Net network is discussed in detail in section 2.3.

**2.3 2.5D convolution network**

The classical network structure V-Net is improved and the 2.5DV-Net network structure is built. As shown in figure 3, the network structure is mainly realized by two kinds of convolution operations: intra-layer convolution and inter-layer convolution, which is composed of encoder, decoder and hopping connection.

Both intra-layer convolution and interlayer convolution are realized by changing the convolution kernel size of 3D convolution layer. In the intra-layer convolution layer, the convolution kernel size of the 3D convolution layer is set to 1 × 3 × 3, so that the convolution layer only extracts the intra-layer information, but not the inter-layer information. therefore. The essence of intra-layer convolution layer is 2D convolution of multiple two-dimensional images at the same time, that is, it can be used to extract the features of each layer plane image in 2.5D data. The intra-layer convolution layer keeps the same size of data input and output by filling zeros on the input data. It is assumed that an intra-layer convolution layer contains 64 convolution cores if the input size is 16 × 11 × 128pixel × 128pixel. The output of the convolution layer is 64 × 11 ×

128pixel × 128pixel. By setting the convolution kernel of 3D convolution layer to 3 × 1 × 1, the interlayer convolution layer only extracts interlayer information, but not intra-layer information. In addition, the interlayer convolution layer uses a non-zero-filling convolution operation, so the convolution itself will change the size of the data. For example, if an interlayer convolution layer contains 64 convolution cores and its input data size is 64 × 11 × 128pixel × 128 pixels, its output is 64 × 9 × 128pixel × 128pixel. As shown in figure 3, the encoder consists of four coded convolution blocks for feature extraction and feature dimensionality reduction. Among them, each encoder convolution block is realized by a cascade of intra-layer convolution, an inter-layer convolution and a reduced sampling layer. Here, the down sampling layer adopts the maximum pool layer with a sliding window size of 1 × 2 × 2, which is similar to the intra-layer convolution layer, and the down sampling layer only pools the features within the layer. for example, if the input size of a down sampling layer is 64 × 9 × 128pixel × 128 pixel, its output is 64 × 9 × 64pixel × 64pixel.

The decoder consists of four decoded convolution blocks for image reconstruction and scale restoration. The decoding convolution block is realized by the cascade of a deconvolution layer and a plurality of intra-layer convolution layers. The convolution kernel size of deconvolution is 1 × 2 × 2, which can be regarded as the inverse transformation of down sampling layer. For example, if the input size of a deconvolution layer is 256 × 3 × 32pixel × 32 pixel, its output is 256 × 3 × 64pixel × 64pixel.

The jump connection part is used for the fusion of deep features and shallow features. Because the shallow features are not equal to the deep features before each jump layer connection, this paper adds a different number of interlayer convolution layers to different jump layers, so that the two groups of features have the same size in space, which is convenient for fusion.

Through the improvement on the basis of V-Net, the 2.5D version is realized. As shown in figure 3, in the 2.5DV-Net structure, the features on each branch are convoluted four times. Therefore, assuming that the input size is 4 × 11 × 128pixel × 128 pixel, the network output size is c × 3 × 128pixel × 128 pixel, that is, the segmentation result of the middle part of the input 2.5D image. Where c represents the number of categories divided. For example, if the type of segmentation includes both tumor and background, then cym2.

**2.4 Whole tumor segmentation module**

In order to reduce the false positive in the back of the brain and brainstem and improve the overall segmentation accuracy of the tumor, a whole tumor segmentation module is designed to get the segmented image of the whole tumor, as shown in figure 4.

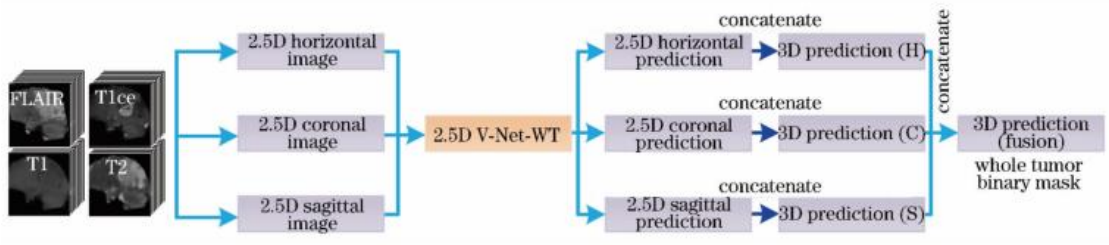


Figure 4 Whole tumor Segmentation Module structure Diagram

**2.5 Tumor nuclear segmentation module**

In the process of tumor nuclear segmentation, some samples are easy to cause inaccurate boundary prediction and significant false positive. In order to obtain accurate tumor nucleus boundary, a tumor core segmentation module is used to use the whole tumor segmentation image as auxiliary information, together with the 2.5D image of four modes as the input of the segmentation network, so as to modify the segmentation boundary of the tumor nucleus. To improve the segmentation accuracy, the segmentation flow chart is shown in figure 5.

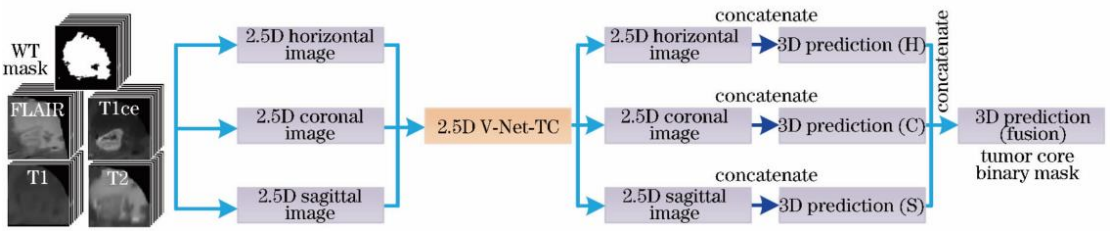


Figure 5 Structure diagram of tumor nuclear segmentation module

**2.6 Enhanced tumor segmentation module**

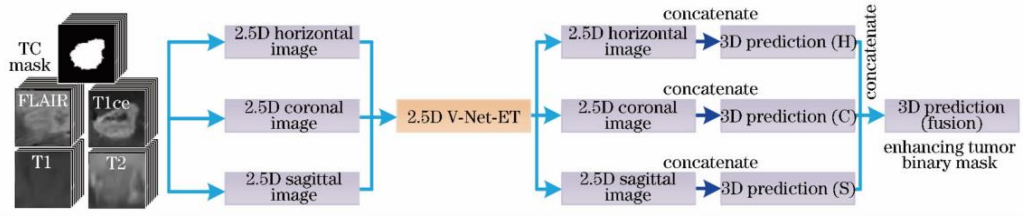


Figure 6 Enhanced tumor segmentation module structure diagram

Using an enhanced tumor segmentation module, the tumor nucleus segmentation image is used as auxiliary information, and the 2.5D images of four modes are used as the input of the segmentation network, so as to obtain more accurate enhanced tumor segmentation results. The segmentation flow chart is shown in figure 6.

**3.Experiment and results**

**3.1Data set and evaluation index**

The data set used in this paper comes from BraTS2018. The data includes 210cases of HGG data and 75 cases of LGG samples. Each case includes registered 3DMRI images of FLAIR, T1, T1ce and T2. The same coordinate voxel values of different modal images correspond to the same location in the patient's brain. The size of the image is 155pixel × 240pixel × 240pixel. As shown in figure 7, gliomas can be divided into a variety of heterogeneous subregions, such as edema (green), necrotic core and non-

enhanced tumor core (red), and enhanced tumor core (yellow). In this paper, 285 samples were randomly divided into 10 samples in order to carry out 10% discount cross-validation.

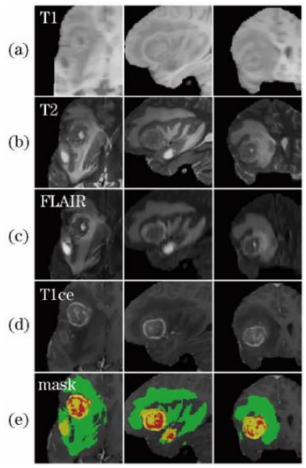


Figure 7 The images of different modes of MRI on three sections. (a) T1 image; (b) T2 image; (c) FLAIR image; (d) T1ce image; (e) fine segmentation label map

### 3.2Result analysis

The effect of cascading V-Net (Cascaded V-Net) and traditional end-to-end segmented convolution neural network is analyzed and discussed. As shown in Table 1, the effects of traditional multi-class segmentation V-Net and 3D, 2D, 2.5D cascade segmentation V-Net are compared. The data in the table are the cross-verification results of each algorithm on the training set.

| Method | End-to-end-V-Net | Cascaded V-Net-3D | Cascaded V-Net-2D | Cascaded V-Net-2.5D |
|---|---|---|---|---|
| Dice WT | 0.8714 | 0.8832 | 0.8986 | 0.9071 |
| Dice TC | 0.7562 | 0.7924 | 0.8413 | 0.8542 |
| Dice ET | 0.6524 | 0.7122 | 0.7919 | 0.8140 |

Table 1 Comparison of various algorithms for 10% discount Cross Verification

In Table 1, Dice WT, Dice TC and Dice ET represent the segmentation similarity coefficients of the whole tumor (including all parts of the tumor), the tumor nucleus (including the necrotic core region and the enhanced / non-enhanced tumor core region) and the enhanced tumor, respectively. The network structure represented by End-to-end-V-Net is a 3D convolution segmentation network similar to V-Net-WT network structure. In the process of training, the network output is a 4-channel thermal map, which represents the probability that each voxel point belongs to edema, enhanced tumor, non-enhanced tumor, necrosis and background, respectively.

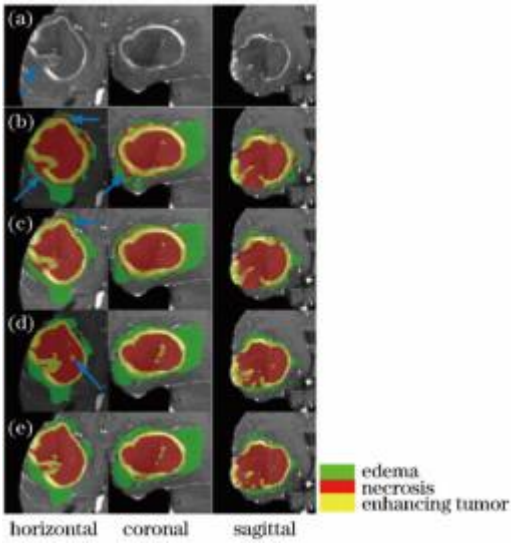


Figure 8 The segmentation results of sample 1 by different network structures in three directions. (a) T1ce image; (b) end-to-end VMuNet; (c) cascadedV-Net-3D; (d) cascadedV-Net- 2D; (e) cascadedV-Net- 2.5D

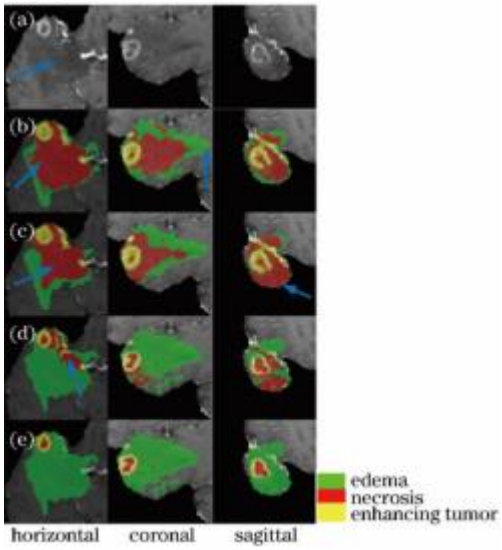


Figure 9 The segmentation results of sample 2 by different network structures in three directions. (a) T1ce image; (b) end-to-end VMuNet; (c) cascadedV-Net-3D; (d) cascadedV-Net- 2D; (e) cascadedV-Net- 2.5D

It can be seen from Table 1 that the Cascaded V-Net-3D network shows more outstanding network performance than the End-to-end-V-Net network of the same scale, which shows that for the task of fine segmentation of brain tumors, the cascade of multi-class segmentation into multiple second-class segmentation networks is helpful to reduce the difficulty of network training and improve the segmentation accuracy of brain tumors.

The effects of 3D, 2D and 2.5D networks based on V-Net on the segmentation results are also compared in Table 1. It is found that 2D network has better network performance than 3D network. A reasonable explanation is that the 2D network has a more sufficient amount of data, assuming that the input sample size of the z × h × wth 2D network is z+h+w times that of the 3D network, so the segmentation result of the 2D network is more accurate. For the enhanced tumor area, the Dice-ET of 2D network was 0.0797 higher than that of 3D network, and the number of false positive samples decreased significantly. This phenomenon shows that the 3D network structure is easy to predict more false positive sample points because of the scarcity of samples, and for enhanced tumor segmentation, some LGG samples do not exist or only a small amount of this region, so the false positive defect of 3D network is more obvious. In addition, compared with the 2D network, the 2.5D network further improves the segmentation accuracy. Because the 2.5D network extracts inter-layer information, more abundant features can be extracted for each voxel point. Moreover, the 2.5D network has the same number of training samples as the 2D network, which can integrate the advantages of 2D and 3D networks and obtain higher performance. The segmentation effect of each

network in Table 1 is shown in figure 8.

The sources of samples 1 and 2 used in figures 8 and 9 are the verification set part of the BraTS2018 data set, and the gold standard of the sample shown is not listed because the data label of this part is not disclosed. 2.5DV-Net can effectively reduce the false positive voxel points in each segmentation module, and can predict a more accurate contour.

| Direction | Horizontal prediction | Coronal prediction | Sagittal prediction | Fused prediction |
|---|---|---|---|---|
| Dice WT | 0.8989 | 0.8808 | 0.8912 | 0.9071 |
| Dice TC | 0.8457 | 0.8367 | 0.8401 | 0.8542 |
| Dice ET | 0.8081 | 0.7922 | 0.8001 | 0.8140 |

Table 2 Comparison of 2.5DV-Net Segmentation effects when inputting pictures in different directions

As shown in Table 2, the segmentation effects of 2.5DV-Net in axial, vector and coronal directions and after fusion are analyzed experimentally, and it is proved that the segmentation accuracy of multi-direction fusion is higher than that of single direction. It can be seen from Table 2 that the segmentation results of the three directions are different. A reasonable explanation is that because there are advantages and disadvantages of image labeling in different directions, and this is inevitable in the work of image labeling, therefore, the trained network model will reflect the prediction differences in different directions. As shown in Table 2, the prediction result of the crown direction is usually lower than that of the other two directions. The prediction results before and after axial, sagittal and coronal fusion are shown in figure 10. As can be seen in figure 10, fusion can solve the false positive problem of the network and improve the boundary accuracy of segmentation.

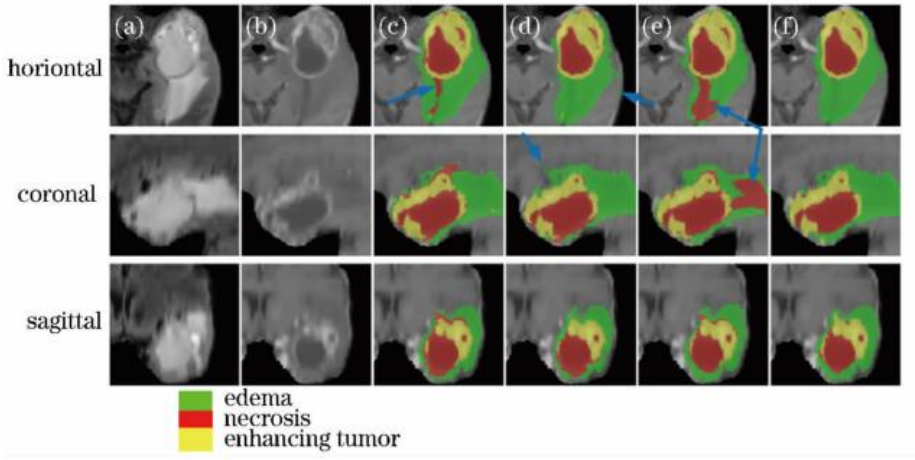


Figure 10 Qualitative comparison of segmentation results before and after fusion. Fusion result of (a) FLAIR image; (b) T1ce image; (c) horizontal to; (d) crown; (e) sagittal;

## 4.Conclusion

In this paper, a multimodal brain tumor fine segmentation based on cascaded convolution neural network is proposed. through experimental comparison, it is proved that the cascade second-class segmentation network structure is more accurate than the end-to-end convolution neural network structure. In addition, the advantages and

disadvantages of 3D, 2D and 2.5D networks are compared and analyzed under the premise of the same convolution neural network structure. It is proved by experiments that 2.5D network has the most advantage in 3D image fine segmentation. Because compared with 3D network, 2.5D network has a larger sample size, and can improve the segmentation accuracy through multi-directional fusion; compared with 2D network, 2.5D network can extract more abundant image features for prediction. Finally, the weighted loss functions Dice Loss and Jaccard Loss, are proposed to further improve the effect of network segmentation. Finally, the cascade convolution network can achieve high-precision and fine segmentation of brain tumors (including high-grade cell tumors and low-grade gliomas), and has high robustness.

## Reference


[1]Mayer G S , Vrscay E R . Self-similarity of Fourier domain MRI data[J]. Nonlinear Analysis Theory Methods & Applications, 2009, 71(12):e855-e864.

[2] Stanescu A , Fleck P , Schmalstieg D , et al. Semantic Segmentation of Geometric Primitives in Dense 3D Point Clouds[C]// IEEE International Symposium on Mixed and Augmented Reality Adjunct. 0.

[3] Freedman D , Zhang T . Interactive graph cut based segmentation with shape priors[C]// Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on. IEEE, 2005.

[4] Ren Lu, Li Qiang, Guan Xin, et al. Three-dimensional MRI segmentation of brain tumors based on improved continuous maximum flow algorithm [J]. Progress in laser and optoelectronics

[5] Li Renzhong, Liu Yangyang, Yang Man, et al. 3D point cloud segmentation based on improved region growth [J]. Advances in Laser and Optoelectronics, 2018 55 (5): 051502.

[6] Xie Zhinan, Zheng Dong, Chen Jiayao, et al. Mass segmentation method of liver cancer ablation CT image based on improved Chan-Vese model [J]. Advances in Laser and Optoelectronics, 2017, 54 (2): 021702.

[7] Chu Jinghui, Wang Xingyu, Lu Wei. Three-dimensional breast MRI segmentation based on inter-frame correlation [J]. Journal of Tianjin University, 2017 50 (8): 835 Murray 8

[8] Yao Hongbing, Bian Jinwen, Cong Jiawei, et al. Medical image segmentation model based on local sparse shape representation [J]. Advances in Laser and Optoelectronics, 2018 55 (5): 051011