Qnet: Convolutional Network and Semantic segmentation with Attention for Computer Cells Image Processing Technology

Yinghan Guan, Shengying Dong*
Chongqing Three Gorges University, Wanzhou District, Chongqing, 404000, China

*Corresponding author: shengyingdong@sanxiau.edu.cn

Abstract. The semantic segmentation of tumor cells in glioma is for clinical diagnosis and judgment of tumor location. However, there are many problems with this task in testing, such as high cost, excessive expenditure of human resources, and insufficient detection accuracy. The data extracted in brain CT can accurately segment and often fail to meet the requirements. For further fine segmentation of tumor cell positions, we propose a new semantic segmentation method: QNet. Through network training, the location of active tumors can be segmented more accurately, which is more helpful for doctors to judge the location of tumors. Evaluate our model and classic semantic segmentation model on 2420 brain CT data sets. The experimental results show that the newly proposed QNet can realize the detailed recognition of the cavity in the tumor, the PA is as high as 0.975, and the CPA is as high as [0.986,0.507], mlou is as high as 0.698, which is better than other classic semantic segmentation models. This model can be used to segment the tumor more accurately, and meets.

Keywords: Semantic segmentation, medical image processing, deep learning, convolutional neural network, attention mechanism.

1. Introduction
Glioma is a tumor of the spinal cord and brain, occurs in glial cells. In the medical field, accurately finding the tumor location in brain CT is a fine-grained segmentation task to determine. The correctness affects the follow-up treatment plan and surgical strategy.

The classic methods of traditional semantic segmentation include OSTU [1], FCM [2], CUT [3], etc. The traditional semantic segmentation is mainly performed by extracting low-level features of the image, and the results of the segmentation are not semantically labeled. In addition, it is to represent complex classes with only level information.

With the advent of FCN [4], Unet [5], Segnet [6], FastFCN [7] and other models appearing in medical image processing are all the rage, but these models still need to improve the segmentation of details.

To make up for the lack of accuracy of existing semantic segmentation methods, we propose a CNN model named Qnet for semantic segmentation of tumors. Inspired by the structure of autoencoder structure, the backbone structure in Qnet divided into encoder (QDown) and decoder (QUp). Qnet constitutes the QDown part based on convolution and residual and the QUp part based on convolution and pixel-shuffle. The corresponding features extracted by QDown and QUp are globally
stitched through CBAM (Convolutional Block Attention Module) [8], multi-scale and multi-level features are fused. Finally, the segmentation of the tumor is achieved well.

The performance to segment tumors with cavity features. Unfixed tumor feature points can also be better located.

2. Related Work
In 2015, the FCN proposed by Jonathan Long et al. a milestone algorithm for image segmentation. FCN the information loss caused by pooling, and the results are lack details. The relation between pixels is not fully considered, and spatial consistency is lacking.

In the same year, Olaf Ronneberger et al. proposed Unet, which performed well on medical images. However, due to the semantic abstraction, too low-level information not only has little effect, but also introduces a lot of noise.

In 2016, Vijay Badrinarayanan and others proposed Segnet. Segnet has low confidence in the position of the classification boundary, and for overlapping categories, the uncertainty will increase.

In 2017, the feature map predicted by DeppLabV3[9] proposed by Liang-Chieh Chen et al. was directly bilinearly up-sampled 16 times to the desired size, which would result in insufficient detailed information.

The FastFCN proposed by Huikai Wu et al. in 2019 cannot effectively focus the network's attention on useful features.

On this basis, this paper proposes Qnet, which realizes the recognition of tumors in the internal cavity, and achieves good results for the identification of details.

3. Model
To improve the performance of the active tumor segmentation, we propose and design Qnet: a convolutional network for tumor cells image segmentation based on the attention mechanism. The model structure is shown in Figure 1. In the figure, the depth of each layer of the feature map is marked at the top, the icons in the upper frame illustrate the meaning of different symbols, and the detailed structure of the Skip connection with attention mechanism is in the upper frame. Pointed out on the right. For the feature extraction of the encoder part (Q Down), considering the high resolution of the CT image, the residual convolution module and the layer jump connection (REPC)[10] are used to effectively extract the detailed features of each level of the target. To solve the ability to express the hollow details of the tumor.

CBAM which include spatial attention block (SAB) and channel attention block (CAB) is a module of weighting and highlighting important features through features and channels. The convolutional attention module (CBAM) is placed in the layer jump connection and residual connection of backbone respectively. The module infers the attention weight based on the space branch and the channel branch, and then multiplies it with the original feature map for adaptive control.

For the super-resolution multi-scale feature fusion module (RDBC) [11], as the depth of the network increases, the features of each CONV layer will gradually be graded because of different receptive fields. Hierarchical features provide important information for image reconstruction. The incomplete block and the dense block are combined to form a super-resolution multi-scale feature fusion module dense connection layers, local feature fusion and local residual learning, and continuous memory. It is implemented by passing the state of the previous convolutional layer to the layer of the current convolutional layer. Let $F_{i_d}$ and $F_{o_d}$ be the input and output layer of the layer D respectively, and they all have feature maps G. The output of the convolutional layer C of the layer D can be expressed as:

$$ F_{o_d} = \sigma(W_{d_i}(F_{i_d}, F_{o_i}, \ldots, F_{o_{i-1}})) $$ (1)
\( \sigma \) represents the ReLU activation function. \( W_{d,c} \) is the C convolutional layer weight, and the bias term is omitted here. \([F_{d,1}, F_{d,1}, \ldots, F_{d,c-1}] \) refers to the feature map generated from the D convolutional layer to the C convolutional layer extracts features. Then introduce the \( 1 \times 1 \) convolutional layer. \( H_{d} \) represents the local feature fusion function.

\[
F_{d,2} = H_{d}^l ([F_{d,1}, F_{d,1}, \ldots, F_{d,1}, \ldots, F_{d,c}])
\]

By these, Qnet uses two convolutional layers to calculate channel features. Finally, the output of the module is used as a weighting vector. The formula is as follows:

\[
M_{d}(F) = \sigma\left(f^2([\text{AvgPool}(F), \text{MaxPool}(F)])\right) = \sigma\left(f^2([F_{\text{avg}}, F_{\text{max}}])\right)
\]

AvgPool represents average pooling, MaxPool represents maximum pooling, and represents the activation function. The output of RDBC is spliced by the output of CBAM and PixelShuffle, and finally the output is obtained.

In the decoder (QUp) process on the right, this process uses PixelShuffle to fuse the feature of the tensor and expand the resolution, and then pass through the \( 3 \times 3 \) convolutional layer and the BN layer, input to the Relu layer for activation, repeat 4 times, and finally output the image. The size is the same as the original picture.

![Model Structure](image)

**Figure 1** . Model Structure

4. Preparation and Parameters

The experiment prepared 2420 slices of brain CT, marked active tumors and necrotic tumors in the data set. This article specifically selected active tumors for semantic segmentation and divided the data set into training at a ratio of 8:1:1 set, validation set, test set. The experimental parameters are as follows:
Table 1. The hyperparameters of the experiment

| Parameter   | Value     |
|-------------|-----------|
| Learning Rate | 0.001     |
| Device       | Tesla T4  |
| Epoch        | 16        |
| Batch size   | 8         |
| Optimizer    | Adam      |

The input image is set to the size of 224*224, and the loss function chooses to use BCEWithLogitsLoss:

\[
l_{n} = -\omega [ y_{n} \cdot \log(\sigma(x_{n})) + (1 - y_{n}) \cdot \log(1 - \sigma(x_{n}))]
\]  

(4)

\(n\) is the number of classification categories, \(y_{n}\) is the probability of being predicted as a type of label, \(x_{n}\) is the result of the predicted value. The advantage of this function is that it sets a weight vector for the loss value(\(\omega\)). The increase in weight effectively increases the ratio of detected features in the binary classification problem, greatly highlights the importance of defects in the sample, and enhances the impact of defect label changes on the loss value.

5. Experiments
In the test set, there are a total of 242 brain CT, and we randomly select 3 for testing.

In the experiment, we used PA, CPA, mIou to evaluate the effect of the model. PA is the ratio of the number with correct prediction pixels category to the total number of pixels:

\[
PA = \frac{\sum_{i=0}^{n} \sum_{j=0}^{k} p_{ij}}{\sum_{i=0}^{n} \sum_{j=0}^{k} p_{ij}}
\]  

(5)

CPA is the accuracy rate of pixels that belong to category i among the predicted values of category i. In other words, the model has many predicted values of category i, among which there are right and wrong. The ratio of the correct predicted value to the total predicted value, in this paper in the two categories (including background), there are two categories of indicators that need to be calculated:

- category 1: \(P1 = TP / (TP + FP)\)
- category 2: \(P2 = TN / (TN + FN)\)  

(6)

mIou is the ratio of the intersection and union of each type of prediction result and the true value of the model, the result of summing and averaging.

Table 2. the results of PA, CPA and mIou for the six models

| Model   | PA    | CPA              | mIou         |
|---------|-------|-----------------|--------------|
| Ours    | 0.975 | [0.986, 0.507]  | 0.698        |
| Unet    | 0.972 | [0.984, 0.487]  | 0.637        |
| SegNet  | 0.968 | [0.982, 0.468]  | 0.632        |
| DeepLabV3 | 0.971 | [0.985, 0.478]  | 0.645        |
| FCN     | 0.973 | [0.986, 0.470]  | 0.645        |
| FastFCN | 0.973 | [0.986, 0.501]  | 0.670        |
\[
    m_{iou} = \frac{1}{k+1} \sum_{i=0}^{k} \frac{p_{ij}}{\sum_{j=0}^{k} p_{ij} + \sum_{j=0}^{k} p_{ji} - p_{ij}} \tag{7}
\]

In order to verify the superiority of this model, this paper prepares another five models to test the results of PA, CPA and mIou for the six models on the data sets respectively, namely UNet, SegNet, DeepLabV3, FCN and FastFCN. With PA, CPA and mIou as the evaluation criteria, Table 2 shows the results of the three network indicators on the data set.

The experimental results are shown in Figure 2.

Figure 2. Comparison of training results of six models on different types of brain tumor sample images

According to the experimental comparison, the PA of QNet is 0.975, the CPA is \([0.986, 0.507]\), and the mIou is 0.698, which is the best among the five models tested.

Figure 2 shows the results of 6 model tests, and Figure 3 shows the verification results. According to the image results, UNet, SegNet, DeepLabV3 and FCN cannot segment active tumor cells. The tumor and the cavity of the posterior choroidal artery, and although FastFCN can segment the approximate range of the active tumor, the segmentation range is accurate. Compared with FastFCN, QNet can accurately segment the active tumor position and accurately cut out the tumor cavity of the posterior choroidal artery. can display more image details.

Figure 3. Segmentation results of six models of brain tumor samples. The training results of QNet and FastFCN are excellent, but the results of the other four models are not ideal

In summary, QNet has the best segmentation effect on active tumors, and its ability to segment objects with holes is also significantly better than other models.

6. Conclusion and Discussion

This paper proposes a new semantic segmentation model based on Unet. By analyzing the results of experimental images, QNet has the best segmentation performance on objects with holes. However, Qnet needs a long time to train and cannot implement segmentation with real time. In summary, QNet solves the segmentation problem of active tumors in medical diagnosis. It has achieved the best results both in PA, CPA and mIOU indicators. We believe that Qnet will contribute to the field of tumor cells image segmentation.
Acknowledgements
This paper was supported by XiangHui Qin who provided efficient guidance for the experiment. XiangHui Qin was the co-author of this paper. Thanks to all support people.

References
[1] P. Smith, D. B. Reid, C. Environment, L. Palo, P. Alto, and P. L. Smith, “NOBUYUKI OTSU. - 1979 - A Threshold Selection Method from Gray-Level Histograms,” IEEE Trans. Syst. Man Cybern., vol. 20, no. 1, pp. 62–66, 1979.
[2] Z. Ji, Q. Sun, Y. Xia, Q. Chen, D. Xia, and D. Feng, “Generalized rough fuzzy c-means algorithm for brain MR image segmentation,” Comput. Methods Programs Biomed., vol. 108, no. 2, pp. 644–655, 2012, doi:https://doi.org/10.1016/j.cmpb.2011.10.010.
[3] T. Park, A. A. Efros, R. Zhang, and J.-Y. Zhu, “Contrastive Learning for Unpaired Image-to-Image Translation,” in Computer Vision -- ECCV 2020, 2020, pp. 319–345.
[4] J. Long, E. Shelhamer, T. Darrell, and IEEE, "Fully Convolutional Networks for Semantic Segmentation," in 2015 Ieee Conference on Computer Vision and Pattern Recognition (IEEE Conference on Computer Vision and Pattern Recognition, New York: IEEE, 2015, pp. 3431-3440.
[5] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in Medical Image Computing and Computer-Assisted Intervention, Pt Iii, vol. 9351, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds. (Lecture Notes in Computer Science, Cham: Springer International Publishing Ag, 2015, pp. 234-241.
[6] V. Badrinarayanan, A. Kendall, and R. Cipolla, “SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 39, no. 12, pp. 2481–2495, 2017, doi: 10.1109/TPAMI.2016.2644615.
[7] H. Wu, J. Zhang, K. Huang, K. Liang, and Y. Yu, “FastFCN: Rethinking Dilated Convolution in the Backbone for Semantic Segmentation,” no. 1, 2019, [Online]. Available: http://arxiv.org/abs/1903.11816.
[8] S. H. Woo, J. Park, J. Y. Lee, and I. S. Kweon, "CBAM: Convolutional Block Attention Module," in Computer Vision - Eccv 2018, Pt Vii, vol. 11211, V. Ferrari, M. Hebert, C. Sminchisescu, and Y. Weiss, Eds. (Lecture Notes in Computer Science, Cham: Springer International Publishing Ag, 2018, pp. 3-19.
[9] L. Chen, “Rethinking Atrous Convolution for Semantic Image Segmentation.”
[10] X. Ding, X. Zhang, N. Ma, J. Han, G. Ding, and J. Sun, “RepVGG: Making VGG-style ConvNets Great Again,” no. 2017, 2021, [Online]. Available: http://arxiv.org/abs/2101.03697.
[11] Y. L. Zhang, Y. P. Tian, Y. Kong, B. N. Zhong, Y. Fu, and IEEE, "Residual Dense Network for Image Super-Resolution," in 2018 IEEE/Cvf Conference on Computer Vision and Pattern Recognition(IIEEE Conference on Computer Vision and Pattern Recognition, New York: IEEE, 2018, pp.2472-2481.