Multi-category Graph Reasoning for Multi-modal Brain Tumor Segmentation

Dongzhe Li Baoyao Yang⋆ Weide Zhan and Xiaochen He

Guangdong University of Technology, Guangzhou

Abstract. Many multi-modal tumor segmentation methods have been proposed to localize diseased areas from the brain images, facilitating the intelligence of diagnosis. However, existing studies commonly ignore the relationship between multiple categories in brain tumor segmentation, leading to irrational tumor area distribution in the predictive results. To address this issue, this work proposes a Multi-category Region-guided Graph Reasoning Network, which models the dependency between multiple categories using a Transformer-based Multi-category Interaction Module (TMIM), thus enabling more accurate subregion localization of brain tumors. To improve the recognition of tumors' blurred boundaries, a Region-guided Reasoning Module is also incorporated into the network, which captures semantic relationships between regions and contours via graph reasoning. In addition, we introduce a shared cross-attention encoder in the feature extraction stage to facilitate the comprehensive utilization of multi-modal information. Experimental results on the BraTS2019 and BraTS2020 datasets demonstrate that our method outperforms the current state-of-the-art methods.

Keywords: Tumor Segmentation · Multi-modal Medical Images · Graph Reasoning

1 Introduction

Brain tumors are among the most common cancers worldwide, with gliomas being the most prevalent malignant brain tumors, varying in degrees of invasiveness. Precise and automatic segmentation of biomedical images would assist radiologists in reducing labor-intensive and time-consuming tasks. However, it remains challenging to achieve high-precision segmentation results[23]. To obtain comprehensive information necessary for accurate segmentation, multi-modal MRI scans with different imaging parameters are typically required in brain tumor segmentation. Common modalities include Fluid Attenuated Inversion Recovery (FLAIR), T1-weighted (T1), contrast-enhanced T1-weighted (T1ce), and T2-weighted (T2) imaging. Images from different modalities capture different pathological information and can effectively complement each other. They play a crucial role in segmenting various types of brain tumor regions such as

---

⋆ Corresponding author.

edema (ED), necrosis and non-enhancing tumor (NCR/NET), and enhancing tumor (ET).

In recent years, deep learning has largely advanced the field of computer-aided diagnosis (CAD), especially medical image segmentation. Methods based on deep learning have become mainstream in this field. The most popular approaches utilize convolutional neural networks (CNNs) geared towards semantic segmentation, such as Fully Convolutional Networks (FCN)[22], U-Net[9], and V-Net[3] to segment brain tumors. The major issue in applying these methods in multi-modal segmentation is how to integrate information from multi-modal medical images effectively.

In most existing brain tumor segmentation methods, multi-modal MRI scans are simply stacked as inputs to the model for semantic segmentation. This may lead to insufficient utilization of multi-modal information. Recently, Some multi-modal fusion methods have been proposed. Zhang et al.[7] directly integrate modalities during the input stage. Wang et al.[5] utilize densely connected multiple modal branches to learn complementary information, and Zhou et al.[13] propose independent encoders for four modalities. While these methods can learn the complementary information between modalities, they ignore exploring the semantic relationship between multi-categories in brain tumor segmentation. Moreover, multi-modal fusion may introduce challenges such as blurred boundaries and inaccurate localization of tumor areas, as the same area of multi-category tumor subregions exhibits diversity across different modalities. In fact, contours are strongly structurally correlated with multi-category subregions: The precise localization and differentiation between multi-category subregions will enhance the clarity of contours. On the other hand, contour can make the boundary information of multiple category areas more accurate. Therefore, integrating multi-category and contour information can assist the localization of tumor regions and alleviate the problem of edge blurring.

This paper divides brain tumor segmentation into two branches: regions-based and contours-based. It fully utilizes multi-modal and multi-category information through feature extraction and graph reasoning stages. In the feature extraction stage, a cross-attention encoder effectively fuses multi-modal features, and a Contour Attention Gate (CAG) is proposed to highlight contour information in brain tumor images. In the graph reasoning stage, multi-category graphs are built based on semantic relationships among subregions, and a Multi-category Region-guided Graph Reasoning Network is developed to explore relationships between regions and contours. Transformer-based Multi-category Interaction Module (TMIM) is designed to learn semantic relationships among subregions. Then, we utilize Graph Neural Networks for remote information propagation and cross-domain feature updating to optimize contour features. The model is learned by a combined loss function that integrates constraints on both region and contour, promoting comprehensive learning across both aspects.

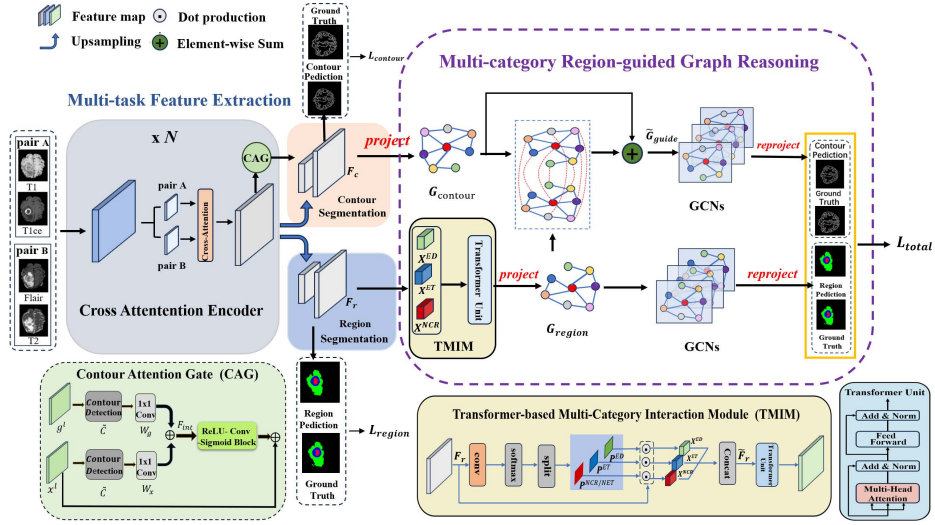The main contributions in this paper are summarized as:

Fig. 1: An Overview of the proposed Multi-category Graph Reasoning Network.

1. We propose a Multi-category Region-guided Graph Reasoning Network that leverages multi-modal and multi-category information in brain tumor images and introduces contour information to assist segmentation.
2. We develop a Transformer-based Multi-Category Interaction Module (TMIM) to capture the multi-category feature relationships among brain tumor subregions of NCT, ET, and ED.
3. We conduct a series of experiments on the BraTS2019 and BraTS2020 datasets, whose results show that the proposed method achieves promising performance on multi-categories brain tumor segmentation.

2   Method

2.1   Overview

An overview of the proposed method is illustrated in Fig. 1. Given an input MRI scan $X \in \mathbb{R}^{C \times H \times W}$, where C represents the modality number of brain images. Firstly, it passes through a Multi-task Feature Extraction Network to obtain feature representations for contours and regions. Subsequently, the features are input into a Region-guided Graph Reasoning Network to examine the relationship between regions and contours. The region's features are processed through a transformer-based Multi-category Interaction Module (TMIM) to learn the multi-category relationships among brain tumor subregions and generate enhanced representations highlighting tumor subregion features. Enhanced region and contour features are projected into graph nodes. The Multi-category graphs capture relationships between tumor subregions, while region-guided graph learning facilitates feature transfer from regions to contours, re-

sulting in a contour graph that incorporates region information. Finally, graph nodes are projected back into the original coordinate space.

2.2 Multi-task Feature Extraction Network

According to expertise [6], there exists a strong structural correlation between paired brain tumor modalities. The information between different modalities is diverse and complementary based on their imaging principles. We divide the input modalities $X \in \mathbb{R}^{C \times H \times W}$ into two pairs along the channel dimension: $X^{T1,T1ce}$ and $X^{T2,FLAIR} \in \mathbb{R}^{C/2 \times H \times W}$. The Multi-task Feature Extraction Network adopts an encoder-decoder architecture with a shared encoder. Subsequently, two independent segmentation branches generate region predictions F_r and contours prediction F_c.

Cross-Attention Encoder. In order to adaptively capture interactions between modalities, In multi-modality brain tumor images, the information between different modalities is complementary and correlated. To adaptively capture interactions between modalities, cross-attention encoders are embedded into the feature extraction process. By processing modalities in pairs, closely related modalities are processed with the same branches, while the complementary modalities are associated using cross-attention. This design not only reduces the computational cost associated with multiple branches but also facilitates the efficient learning of multi-modal information.

Contour Attention Gate. During convolution, image or signal contour features may be lost or blurred as kernels typically detect local patterns rather than contours directly. Inspired by the information reweighing ability of attention gates [8], this paper incorporates Contour Attention Gate (CAG) modules into skip connections to extract multi-scale contour features during contour segmentation. Details of CAG is illustrated in Figure 1. CAG selects the feature map of the current layer x^l and the gating signal g^l of the next layer decoder as inputs. After passing through the contour detector, features focused on contour information are obtained. Then, they are element-wise added to obtain the updated features:

$$F_{int} = W_x^T(x^l * \tilde{C}_{contour}) + W_g^T(g^l * \tilde{C}_{contour}) \tag{1}$$

where $*$ is the convolution operation and $\tilde{C}_{contour}$ is the contour detection operator, $\tilde{C}_{contour} = \frac{1}{2\pi\sigma^2}e^{-\frac{x^2+y^2}{2\sigma^2}}$ i.e. σ denotes the standard deviation and (x,y) is spatial coordinates. The features W_x and W_g are linear transformations

We adopt multi-dimensional attention coefficients [11] to selectively focus on the contours of brain tumor sub-regions at multiple scales.The multi-dimensional attention coefficient formula is as follows:

$$\alpha_i^l = \sigma_2(\psi^T(\sigma_1(F_{int} + b_u)) + b_\psi) \tag{2}$$

where $\sigma_1(\cdot)$ is the Sigmoid activation function and $\sigma_2(\cdot)$ is the ReLU activation function.. ψ^T denotes linear transformations and b_ψ, b_u are bias terms.

2.3   Multi-category Region-guided Graph Reasoning Network

A multi-category Region-guided Graph Reasoning Network is designed to investigate the semantic associations between diverse categories, facilitating the identification of tumor locations and guiding contour learning. By utilizing a graph-based strategy to explicitly model the relationship between contour features and region features, we achieve propagation of knowledge from region to contour over long distances to improve the definition of contours in the contour-based branch. Then, leveraging Graph Convolutional Networks for relational reasoning on graphs enhances the expression capability of both regional features and contours, as well as multi-category discrimination.

Transformer-based Multi-category Interaction Module. After obtaining the multi-modal feature map F_r, we pass F_r through convolutional layers, softmax, and segmentation operations to obtain separate tumor sub-regions representations X^{NCR}, X^{ED}, X^{ET} and concatenate them:

$$\tilde{F}_r = concat[F_r \odot X^{NCR}, F_r \odot X^{ED}, F_r \odot X^{ET}] \tag{3}$$

where $\odot$ is dot production,*concat* denotes concatenation operation. Then, standard transformer encoder structure[12] is applied as a transformer unit to build long-term relationships between tumor subregions. The structure of the transformer unit is shown in Figure 1.

Region-guided Graph Convolution Network. Considering that learning from regions can provide additional supervision for contour segmentation tasks, we construct graphs separately for regions and contours, then guide contour learning with the knowledge obtained by region. Following [15], we employ graph projection operations to project features of regions and contours onto graph nodes. Subsequently, we compute inter-graph dependencies using attention mechanisms to propagate node information from regions to contours.The updated contour graph $\tilde{G}_{guide}$ can be formulated as:

$$\tilde{G}_{guide} = \mathcal{W}(LN(X_q^T \otimes X_k) \otimes X_v^T) + G_{contour} \tag{4}$$

where X_q, X_k, X_v are query graph, key graph value graph obtained by multi-layer perceptrons (MLPs)[1]. $LN(\cdot)$ denotes the layer normalization and $\mathcal{W}$ is the weighting parameter. $\otimes$ acts as a matrix product. Subsequently, we perform graph convolutions (GCN) [4] on both the region graph and the contour graph

$$\tilde{G}_{guide} = \chi(A^T \tilde{G}_{guide}\mathcal{W}) \tag{5}$$

χ is a non-linear activation function,and A^T denotes the graph adjacent matrices for $\tilde{G}_{guide}$. Finally, the graph is reprojected to the original coordinate space by utilizing assignment matrix M:

$$\tilde{F}_{contour} = M\tilde{G}_{guide} + F_c \tag{6}$$

The reprojection of the region graph G_{region} is similar as above.

2.4   Loss Function

We optimize the proposed network with three losses: L_{region}, L_{contour}, and L_{final}. During the feature extraction stage, we introduced L_{region} and L_{contour} to respectively learn specific features for contour branches and region branches, guiding the training of the Multi-task Segmentation Network. L_{region} is employed by a combination of soft dice loss and cross-entropy loss (BCE) [24]. L_{contour} enhances segmentation accuracy by leveraging both boundary detail from edge loss [16] and contour overlap from dice loss, mathematically,

$$L_{\text{region}} = \sum \left(0.5 \cdot L_{\text{BCE}} + L_{\text{Dice}}\right) \tag{7}$$

$$L_{\text{contour}} = 0.5 \cdot L_{\text{CRF}} + L_{\text{Dice}} \tag{8}$$

whe   In the stage of graph reasoning, we combine L_{contour} and L_{region} to encourage mutual learning between regions and contours and to maintain common features between positional features, formulated as,

$$L_{\text{final}} = L'_{region} + \lambda \cdot L'_{contour} \tag{9}$$

where L'_{region} and L'_{contour} are the losses of predictions regarding region and contour with the same definitions as L_{region} and L_{contour} above. λ denotes a hyperparameter controlling the balance between L'_{region} and L'_{contour}.

3   Experiment

3.1   Data and Implementation Details

We select the BraTS 2019 and BraTS 2020 datasets [21] as the brain glioma segmentation benchmarks. The two datasets comprise 335 and 369 annotated brain tumor samples, respectively. Each subject in the datasets contains four modalities: T1, T1ce, T2, and Flair, which have undergone skull-stripping, resampling, and co-registration processes. The dimensions of each MRI volume are $155 \times 240 \times 240$, and we perform data preprocessing involving common z-score normalization techniques, with each modality scan being sliced and cropped into sizes of 160×160. Each dataset is divided into 80% for model training and 20% for testing. Our model is trained end-to-end. We employ the Adam optimizer to update model parameters with a learning rate of 10^{-4}. We assess the model's performance using two widely adopted metrics in medical image segmentation: the Dice score and the 95% Hausdorff distance (HD).

3.2   Comparison with SOTA Methods

We conduct a comparative analysis of our network with five state-of-the-art (SOTA) segmentation methods. These include two approaches that emphasize multi-modal feature fusion (SF-Net[17] and ACM-Net[18]), as well as three recent Transformer-based methods (Nestedformer[19], TransBTS[14], EoFormer[20]),

Table 1: Result comparison of brain tumor segmentation methods on BraTS2019

Model	Dice(%) ↑				HD95(mm) ↓			
	ET	TC	WT	Ave	ET	TC	WT	Ave
TransBTS[14]	80.86	81.19	89.35	83.80	5.642	6.048	4.332	5.460
Nestedformer[19]	82.11	86.42	**91.18**	86.57	5.534	5.906	5.317	5.585
SF-Net[17]	80.08	82.33	88.61	83.67	4.787	7.440	7.288	6.505
ACM-Net[18]	80.63	87.15	88.08	85.28	4.564	7.774	3.862	5.400
Eoformer[20]	82.94	86.83	90.39	86.72	**4.053**	**5.843**	5.822	**5.239**
Ours	**83.23**	**89.10**	90.44	**87.59**	5.110	7.523	**3.775**	5.469

Table 2: Result comparison of brain tumor segmentation methods on BraTS2020

Model	Dice(%) ↑				HD95(mm) ↓			
	ET	TC	WT	Ave	ET	TC	WT	Ave
TransBTS[14]	80.89	83.25	90.10	84.08	5.873	6.875	4.876	5.824
Nestedformer[19]	82.85	86.48	**91.20**	86.84	5.721	6.115	4.598	5.528
SF-Net[17]	81.10	83.84	89.01	84.65	**4.305**	7.661	7.720	6.562
ACM-Net[18]	82.42	87.75	90.08	86.75	4.492	7.624	3.956	5.375
Eoformer[20]	83.54	87.12	90.87	87.17	5.911	**6.041**	3.852	**5.268**
Ours	**84.38**	**89.21**	90.77	**88.12**	5.413	7.759	**3.844**	5.672

with one of them focusing on edge-oriented segmentation. As depicted in Table 1, our method demonstrates Dice scores of 83.23%, 89.10%, and 90.44% for the ET, TC, and WT regions, respectively, in the BraTS2019 dataset. Through the incorporation of multi-category interaction and Region-guided Graph Reasoning Modules, our method achieves the highest Dice scores for the Enhance Tumor and Tumor Core regions, indicating its efficacy in capturing the interrelationships among brain tumor subregions. Similarly, as illustrated in Table 2, our method continues to outperform others by achieving the highest Dice scores in the ET and TC regions in the BraTS 2020 dataset. Although the Dice scores of our method in the WT region slightly trail behind Nestedformer and Eoformer, the difference from the highest score is acceptable, and our results notably surpass the others.

It is worth noting that the average Dice score of our proposed method is the highest among all methods. Figure 2 visually shows the visual segmentation results of our method on BraTS2019. It can be observed that we can accurately segment the contours of brain tumors and the positions of sub-regions. Based on our results, the model exhibits outstanding performance in handling multi-category dependencies and achieving precise delineation of brain tumor contours. This success is due to its adept integration of constraints on both region and contour, showcasing its effectiveness in accurate tumor segmentation.
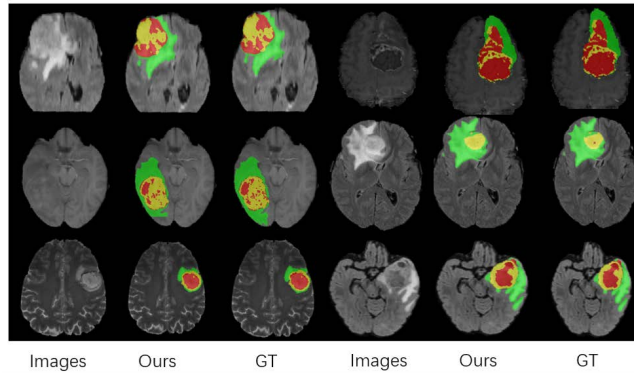
Fig. 2: The visual results on BraTS2019. Green, yellow and red colors indicate regions of ED, ET and NCR, respectively.

3.3   Ablation Study

This section uses the standard Unet [9] as the baseline model and then adds different components one by one to validate their effect. Specifically, we mainly compare the performance of the four models: (1) Unet, (2) Unet+CAG, (3) Unet+GCN, and (4) Unet+CAG+GCN. We can see that each of the above components leads to some improvements in the segmentation results. Comparing (1) with (2), the Unet+CAG model improves segmentation accuracy for all classes, with significant enhancements observed in ET and TC segmentation, along with a minor improvement in WT segmentation, which validates that the contour information extracted by CAG can serve as complementary information for brain tumor segmentation, thereby enhancing performance. Comparing (1) with (3), The Unet+GCN model slightly outperforms the Unet model overall. Significant improvements are seen in ET and TC segmentation, with marginal improvements in WT, confirming that performing graph inference can improve segmentation effectiveness. By comparing (4) with (2) and (3), it is evident that combining region and contour branches for region-guided graph reasoning enhances the segmentation performance for all three categories of brain tumors. Comparing the results of the whole model with those of (4), the proposed model surpasses the other four models in all classes, achieving a notably higher average Dice score of 87.59% compared to the alternative models. It concludes that the TMIM module can effectively learn multi-class relationships and transmit knowledge from the region to the contour graph.

4   Conclusion

This paper introduces a novel approach leveraging graph learning to capture semantic relationships between multi-categories from region and contour and aspects for brain tumor segmentation. A TMIM is proposed to accurately identify

Table 3: Ablation study of critical components

Model	Dice(%)			
	ET	TC	WT	Ave
(1) Unet	79.10	81.00	87.93	82.67
(2) Unet+CAG	80.73	83.23	89.62	84.52
(3) Unet+GCN	80.93	85.09	88.12	84.71
(4) Unet+CAG+GCN	81.69	88.12	90.11	86.64
Proposed	**83.23**	**89.10**	**90.44**	**87.59**

and locate brain tumor subregions, addressing the challenge of contour overlap. In addition, cross-attention encoders together with CAG, are introduced to enhance the utilization of contour information, which has shown effectiveness in reducing boundary errors and providing complementary information for subregion localization. The quantitative and qualitative results on the BraTS 2019 and BraTS 2020 datasets demonstrate the effectiveness of our method

Acknowledgement

Disclosure of Interests

The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Charles R Qi, Hao Su, Kaichun Mo, and Leonidas J Guibas.: Pointnet: Deep learning on point sets for 3d classification and segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp.652-660. (2017)
2. Deng-Ping Fan, Tao Zhou, Ge-Peng Ji, Yi Zhou, Geng Chen, Fu. Huazhu,Jianbing Shen, Ling Shao.:Inf-net: Automatic covid-19 lung infection segmentation from ct images. In: Trans.Med.Imag.pp.2626-2637. IEEE (2020)
3. F. Milletari, N. Navab, S.-A. Ahmadi.:V-net: Fully convolutional neural networks for volumetric medical image segmentation. In: 4th international conference on 3D vision (3DV). pp. 565-571. IEEE (2016)
4. Yazhou Yao, Tao Chen, Guo-Sen Xie, Chuanyi Zhang, Fumin Shen, Qi Wu, Zhenmin Tang, Jian Zhang.:Non-Salient Region Object Mining for Weakly Supervised Semantic Segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2623-2632 (2021)

5. Y. Wang et al., "Modality-pairing learning for brain tumor segmentation," in Proc. Int. MICCAI Brainlesion Workshop, pp. 230–240. (2021)
6. J. Lin, J. Lin, C. Lu, H. Chen, H. Lin, B. Zhao, Z. Shi, B. Qiu, X. Pan, Z. Xu, B. Huang, C. Liang, G. Han, Z. Liu, C. Han.: CKD-TransBTS: clinical knowledge-driven hybrid transformer with modality-correlated cross-attention for brain tumor segmentation. In: Trans. Med. Imaging. pp.2451-2461. IEEE (2023)
7. Zhang, Y., Pei, Y., Zha, H.: Learning dual transformer network for diffeomorphic registration. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 129–138. Springer (2021)
8. Oktay, O., Schlemper, J., Folgoc, L.L., Lee, M.: Attention U-Net: learning where to look for the pancreas. In: MIDL. pp. 2626-2637. IEEE (2018)
9. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. pp. 234-241. Springer (2015)
10. S. Mo, M. Cai, L. Lin, R. Tong, Q. Chen, F. Wang, H. Hu, Y. Iwamoto, X.-H. Han, Y.-W. Chen.: multi-modal priors guided segmentation of liver lesions in MRI using mutual information based graph co-attention networks. In: J International Conference on Medical Image Computing and Computer-Assisted Intervention. pp.429-438. (2020)
11. Niu Z, Zhong G, Hui Yu. A review on the attention mechanism of deep learning. Neurocomputing. pp.48–62. (2021)
12. TAlexey D, Lucas B, Alexander K,Dirk W, Xiaohua Z, Thomas U, et al.:Disan: An image is worth 16x16 words: Transformers for image recognition at scale. In: ICLR. (2020)
13. T. Zhou, S. Canu, P. Vera, and S. Ruan.:Latent correlation representation learning for brain tumor segmentation with missing MRI modalities. In: Trans.Image Process. pp. 4263-4274. IEEE (2021)
14. Wang, W., Chen, C., Ding, M., Yu, H., Zha, S., Li, J.: Transbts: multi-modal brain tumor segmentation using transformer. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp.109-119. (2021)
15. Huang, Z., Li, Y.: Interpretable and accurate fine-grained recognition via region grouping. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2020)
16. Y. Liu, M.-M. Cheng, X. Hu, K. Wang, X. Bai.: Richer convolutional features for edge detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp.3000-3009. (2017)
17. Y. Liu, F. Mu, Y. Shi, X. Chen., Li, J.: Sf-net: A multi-task model for brain tumor segmentation in multi-modal mri via image fusion. In: Signal Processing Letters. pp.1799-1803. (2022)
18. Yuzhou Z, Hong L, Enmin S, and Chih-Cheng H.: Sf-net: A 3d cross-modality feature interaction network with volumetric feature alignment for brain tumor and tissue segmentation. In: Journal of Biomedical and Health Informatics. pp.1799-1803. (2022)
19. Z. Xing, L. Yu, L. Wan, T. Han, and L. Zhu.: NestedFormer: Nested modality-aware transformer for brain tumor segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp.140-150. (2022)
20. Dong She, Yueyi Zhang, Zheyu Zhang, Hebei Li, ZihanYan, and Xiaoyan Sun .:Eo-Former: Edge-Oriented Transformer for Brain Tumor Segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 333-343. Springer (2023)

21. B. H. Menze, A. Jakab, S. Bauer, and et al., "The multi-modal brain tumor image segmentation benchmark (brats)," IEEE Transactions on Medical Imaging, vol. 34, no. 10, pp. 1993-2024.(2015)
22. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation.In: Proceedings of the IEEE conference on computer vision and pattern recognition pp. 4313440 (2015)
23. B. Yu, L. Zhou, L. Wang, W. Yang, M. Yang, P. Bourgeat, and J. Fripp, "Learning sample-adaptive intensity lookup table for brain tumor segmentation," in International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 216–226. (2020)
24. Xiaoya Li, Xiaofei Sun, Yuxian Meng, Junjun Liang, Fei Wu, and Jiwei Li, "Dice Loss for Data-imbalanced NLP Tasks," in arXiv:1911.02855 (2019)