Journal of Artificial Intelligence and Technology, 2024, 4, 160-168
https://doi.org/10.37965/jait.2023.0442

ISTP
RESEARCH ARTICLE

Brain Tumor Segmentation Based on the Learning Statistical Texture

Yufeng Guo,[1,3] Feiba Chang,[1] Xiaoyu Chen,[1] Fengjun Sun,[1,2] and Zihong Wang[1]
[1]Department of Medical Engineering, The First Hospital Affiliated to Army Medical University, Chongqing, China
[2]Department of Pharmacy, The First Hospital Affiliated to Army Medical University, Chongqing, China
[3]Department of Purchase, The First Hospital Affiliated to Army Medical University, Chongqing, China

(Received 11 August 2023; Revised 27 August 2023; Accepted 27 August 2023; Published online 27 November 2023)

Abstract: Achieving accurate segmentation of brain tumors in Magnetic Resonance Imaging (MRI) is important for clinical diagnosis and accurate treatment, and the efficient extraction and analysis of MRI multimodal feature information is the key to achieving accurate segmentation. In this paper, we propose a multimodal information fusion method for brain tumor segmentation, aimed at achieving full utilization of multimodal information for accurate segmentation in MRI. In our method, the semantic information processing module (SIPM) and Multimodal Feature Reasoning Module (MFRM) are included: (1) SIPM is introduced to achieve free multiscale feature enhancement and extraction; (2) MFRM is constructed to process both the backbone network feature information layer and semantic feature information layer. Using extensive experiments, the proposed method is validated. The experimental results based on BraTS2018 and BraTS2019 datasets show that the method has unique advantages over existing brain tumor segmentation methods.

Keywords: brain tumor segmentation; convolutional neural networks; edge feature; multimodal information fusion; transformer

I. INTRODUCTION

Semantic segmentation is a technique to segment different objects in an image from a pixel perspective and to annotate each pixel in the original image. It is one of the most basic techniques of computer vision processing and has a wide range of applications such as automatic driving, unmanned aerial vehicle (UAV) navigation [1], remote sensing images [2], medical diagnosis [3], etc. Among them, medical image segmentation is an important topic in the medical image processing community [4].

Magnetic Resonance Imaging (MRI) is based on the fading time of the magnetic field of hydrogen atoms recorded by the magnetic field induction coil after the external magnetic force is withdrawn. MRI is mostly utilized in diagnosing and treating brain tumor disease because of its excellent ability to provide high-resolution views of soft tissue anatomy. To obtain comprehensive information for accurate segmentation, multimodal MRI scans with different imaging parameters are usually required in brain tumor segmentation. Commonly used modalities include fluid-attenuated inversion recovery (FLAIR), T1-weighted (T1), contrast-enhanced T1-weighted (T1ce), and T2-weighted (T2). Different imaging techniques are suitable for different situations. Images of different modalities capture different pathological information and they can effectively complement each other [5,6], which plays a key role in segmenting multiple types of brain tumor regions such as edema (ED), necrotic non-enhancing tumors (NCR/NET), and enhancing tumors (ET). Therefore, the detection of brain tumors in MRI images using semantic segmentation methods has become the focus of many studies [7].

Deep learning-based semantic segmentation methods have become the mainstream of medical image processing, and many semantic segmentation methods have been studied and proposed. Fully Convolutional Networks (FCN) [8] is the pioneer of deep learning in semantic segmentation, proving that neural networks can be trained for end-to-end semantic segmentation of images, and now it becomes the most popular method. U-Net method [8] proposes an "encoder-decoder" architecture based on FCN, which uses deconvolution in up-sampling to increase the number of feature maps while reducing the number of feature maps. By adding a jump connection between the corresponding layers of encoding and decoding to preserve the low-level features of the image. The SegNet method [9] also uses an "encoder-decoder" architecture with maximum pooling of nonlinear upsampling and then convolution of sparse upsampling maps to improve the resolution of segmentation. However, the CNN-based method mentioned above all have problems such as low accuracy and insensitivity to details limited by the receptive field of convolutions, leading to difficulty in characterizing the global dependencies of features.

To alleviate this problem, researchers have employed Recurrent Neural Networks (RNN) to model inter-pixel dependencies, enabling sequential processing of pixels to establish global contextual relationships and improve segmentation. Transformer-based models have been introduced to the study of brain tumor segmentation and receive great attention in the field of computer vision. Dosovitskiy [10]. Proposed the Vision Transformer (ViT) model, applying the Transformer model from the text domain to the image domain. The Transformer-based models exhibit a better ability to capture global contextual information by employing a self-focused mechanism, but there is still the problem of the high computational cost of the change method. And then Swin Transformer [11] introduces a network architecture with a sliding window and hierarchical design, which reduces the computation

Corresponding author: Zihong Wang (e-mails: wangzihong@tmmu.edu.cn; 35163916@qq.com).

while considering the acquisition of local features and achieves advanced performance in semantic segmentation. However, there also exists problems of low locality inductive bias in the Swin transformer model, which means that large amounts of training data are required to achieve satisfactory visual performance.

In this paper, we propose a method of multimodal information fusion segmentation for brain tumor segmentation, aimed at achieving fuller utilization of multimodal information for accurate segmentation in MRI. Specifically, the segmentation framework proposed in this paper can be divided into a feature extraction backbone network, a semantic information processing sub-network, a feature reasoning module, and an upsampling network. The feature extraction backbone network adopts the ResNet50 architecture for preliminary feature extraction to extract deep features and multiple shallow features. Moreover, a Semantic Information Processing Module (SIPM) is introduced to achieve free Multiscale feature enhancement and extraction, the core of which is to build a new SI-EE (Semantic Information Enhancement and Extraction) module. Then, a novel Multimodal Feature Reasoning Module (MFRM) is constructed to process both the backbone network feature information layer and semantic feature information layer, and the features of different modalities can be further interactively fused to obtain finer lesion segmentation results.

In summary, the main contribution of this paper can be concluded as follows:

1. First and foremost, we proposed a method of multimodal information fusion segmentation to better implement brain tumor segmentation, and a feature extraction backbone network, a semantic information processing sub-network, a feature reasoning module, and an upsampling network are included. The different modules built above result in the following two contributions.
2. To effectively extract the deep features and multiple shallow features, the SIPM is designed to achieve free multiscale feature enhancement and extraction by fusing the information of the two feature layers with a new Semantic Information Enhancement and Extraction (SI-EE) module.
3. Both the semantic feature information layer and backbone network feature information layer can be obtained with the help of SIPM, and a novelMFRM is introduced to further interactively fuse the obtained features, which can lead to a refined lesion segmentation result.

II. RELATED WORKS

Image Semantic Segmentation: In recent years, with the rapid development of deep learning technology and its wide application in various fields, image semantic segmentation based on deep learning (ISSDL) has also received a lot of attention from researchers across the world [12–16]. And ISSDL can be divided into two major categories [17–19]: ISS based on the regional (ISSR) and ISS based on the pixel (ISSP).

For the ISSR techniques, a regional convolutional neural network (regions with CNN features, RCNN) was proposed [20,21], which fuses candidate regions generated by the selective search (SS) algorithm on the basis of visual features extracted by CNN to achieve a multi-task objective for target detection and semantic segmentation. SPPNet [22,23] network was introduced, which inserts a spatial pyramid pooling player [24] (SPP player) after the RCNN convolutional layer to reduce the repetitive computation process in feature extraction. But these methods still have the limitations of generating too many candidate regions, irregular shape of regions, and a large amount of network operations.

For the ISSP technique, Shelhamer [8] designed the classical fully convolutional network (FCN), which performs pixel-level classification of images in fully supervised learning, and lots of methods based on FCN are proposed [25–27]. Chen [28–30] proposed a DeepLab series improvement scheme, DeepLab network introduced fully connected conditional random field (FCCRF), structured prediction of the coarse segmentation map after FCN, and did image smoothing operation to achieve edge optimization, used convolution with holes to expand the perceptual field of the feature map, and finally output a complete semantic segmentation result of the image. and further optimized to DeepLab-V2 and DeepLab-V3.

Feature Encoding: The encoder-decoder architecture approach has also been proposed to address the balance between performance and efficiency of semantic segmentation, as well as the high computational complexity and memory consumption problems posed on high-resolution feature maps. Ronneberger *et al.* [13] proposed the classical U-Net network for implementing semantic segmentation of biomedical images, which uses a downsampling operation to gradually reduce the resolution of the feature map in the encoding phase, and a downsampling operation to gradually restore image detail and resolution in the decoding phase. The U-net++ was further developed as a more powerful architecture for medical image segmentation. Badrinarayanan *et al.* [9,31] proposed a SegNet-Basic network for solving image semantic segmentation tasks in the fields of autonomous driving and intelligent robotics to achieve end-to-end pixel-level image segmentation. And then SegNet was applied to automated brain tumor segmentation for four MRI modalities by Alqazzaz *et al.* [32]. Noh *et al.* [33] propose a fully symmetric DeconvNet network based on VGG16, which alleviates the limitations of existing methods based on fully convolutional networks. And DeconvNet was also utilized for Simultaneous Localization and Mapping (SLAM) to improve the drift in long-run odometry [31].

Feature Fusion: Feature fusion is studied to solve the problems of excessive computation, long training time, and severe memory consumption. Liu *et al.* proposed a technique to add global context to a fully convolutional network for semantic segmentation, and performed many critical tasks [34,35]. Their method uses the average features of the layers to add features at each location.

Ghiasi [36] designed the Laplacian pyramid reconstruction and refinement model (LRR) using the Laplacian pyramid algorithm for the reconstruction of shallow features. Attention mechanism [37–39] with a soft weighting of multiscale features at each pixel location and proposed an advanced semantic image segmentation model that is jointly trained with multiscale input images and an attention model. RefineNet [40] was first proposed by Lin [41], a generalized multi-path refinement network for high-resolution semantic segmentation, which utilizes all available information from the downsampling process to perform high-resolution prediction using remote residual connections. And lightweight RefineNet was further developed for effective and efficient semantic image segmentation [42]. Zhu [43] proposed a brain tumor segmentation method based on the fusion of deep semantics and edge information in multimodal MRI, achieving a more sufficient utilization of multimodal information for accurate segmentation by introducing a semantic segmentation module, an edge detection module, and a feature fusion module.

III. METHODS

A. OVERALL STRUCTURE

The overall framework of multimodal information fusion segmentation proposed in this paper for brain tumor segmentation is shown in Fig. 1. It can be divided into two parts: encoder and decoder. With more detail, the process can be divided into a feature extraction backbone network, a semantic information processing sub-network, a feature reasoning module, and an upsampling network. The feature extraction backbone network adopts the ResNet50 architecture for preliminary feature extraction to extract deep features and multiple shallow features, and then send these features to SIPM to achieve free Multi-scale feature enhancement and extraction. SIPM first fuses the information of the two feature layers, the core of which is to build a new SI-EE (Semantic Information Enhancement and Extraction) module, because the information obtained after fusion contains a lot of semantics information. Thus this module mainly enhances and then extracts semantic information to better realize the segmentation task. Through the first two parts, a backbone network feature information layer and a semantic feature information layer can be obtained, and then through the constructed MFRM (Multimodal Feature Reasoning Module) to process it, its main task is to further interactively fuse the obtained features of different modalities to obtain a more refined lesion segmentation result, and finally obtain the segmentation result through the upsampling network.

B. SEMANTIC INFORMATION PROCESSING MODULE

SIPM is to process semantic information, the core of which is the SI-EE (Semantic Information Enhancement and Extraction) module proposed in this paper. Its structure is shown in Fig. 2. The task of this module is mainly to target a large amout of the semantic information of the method is enhanced first, and then extracted, so that the obtained feature information is more detailed and more suitable for subsequent segmentation tasks. Semantic feature information reflects a global feature of homogeneous phenomena in the image and describes the slowly changing or periodically changing surface organization structure and arrangement rules in the image. However, the quality and contrast of the low-level information extracted by the ResNet backbone network (such as pixel values or local area attributes) are usually low, and images captured in dark environments usually have poor visualization quality, resulting in unclear texture details, which cannot be well acquired. And use this low-level information. This paper exploits the statistical properties of textures, which focus on the distribution analysis of images, such as intensity histograms. The first part enhances the texture details of low-level features by histogram equalization after the enhancement step, making it more detailed and suitable for segmentation.

First, in the generated histogram, its horizontal axis and vertical axis represent gray level and count value respectively, which are expressed as feature vectors H and F respectively. For the histogram quantization, the statistical information of the count value F is reconstructed, and the reconstructed is H'. Each corresponding level H_n is also converted by the formula H'_n:

$$H'_n = \frac{(N-1)\sum_{i=1}^{n} F_n}{\sum_{i=1}^{n} F_n} \tag{1}$$

where N represents the total number of gray levels.

Then we introduce the concept of quantization and counting operators to describe the intensity of semantic information. Input the feature map, quantized coding map, $E \in R^{N\times HW}$ and statistical features obtained $D \in R^{C_1\times N}$ by QCO, where D plays the role of the histogram. And get the new quantization level L' from the original quantization level L via D:

$$Y = \text{Soft max}(\phi_1(D)^T \cdot \phi_2(D)) \tag{2}$$

$$L' = \phi_3(D) \cdot Y \tag{3}$$

where ϕ_1, ϕ_2 and ϕ_3 denote three different 1×1 convolutions and a Softmax performed on the first dimension acts as a non-linear normalization function.

We then update each node by fusing features from all other nodes, resulting in the reconstructed quantization level $L' \in R^{C_2\times N}$

Afterward, the reconstructed level is L' assigned to each pixel to $E \in R^{N\times HW}$ obtain the final output R using the quantization encoding map, since E can reflect the original quantization level of each pixel. R is obtained from the following formula:

$$R = L' \cdot E \tag{4}$$

Finally, R was reshaped as $R^{C_2\times H\times W}$.

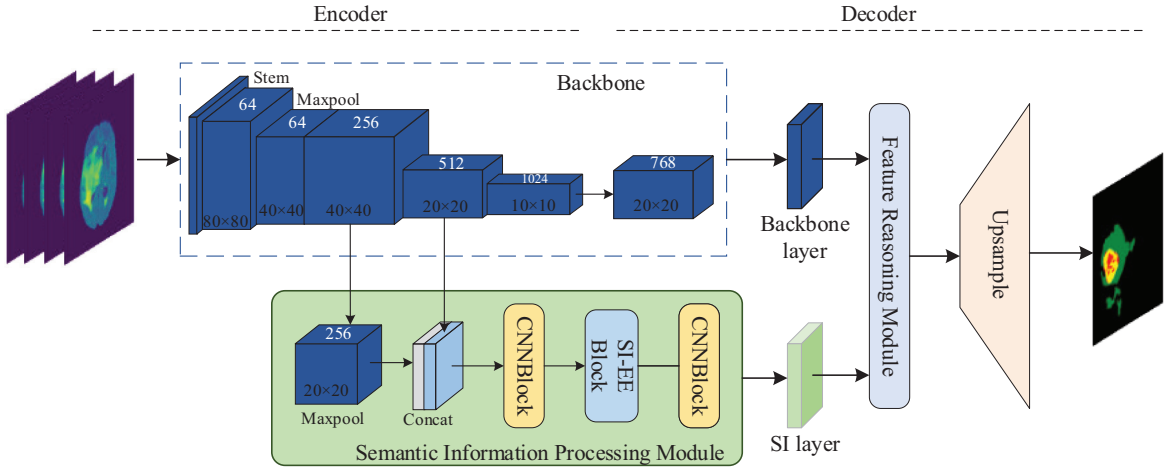

Fig. 1. An illustration of the method proposed in this paper and two parts are included: the encoding part and the decoding part.

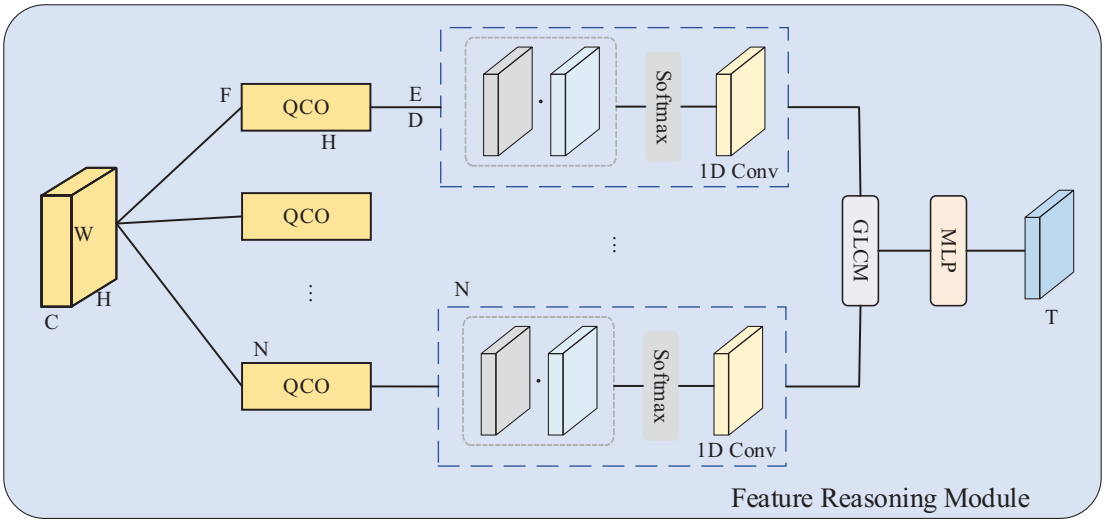

Fig. 2. Semantic information enhancement and extraction module.

The second part performs semantic information feature extraction after underlying feature enhancement, exploiting texture-related information from multiple scales in a feature map containing rich texture details. Different from color features, texture features are not pixel-based features. It needs to be statistically calculated in an area containing multiple pixels, which is related to the statistical information of the spatial relationship. A gray-level co-occurrence matrix (GLCM) is used to capture the texture. The co-occurrence matrix is first generated by GLCM $M \in R^{H\times W}$, and the texture information is represented by the contrast uniformity of the statistical description. In the processing of the feature map, it is first input into the two-dimensional quantization and counting operator to obtain the co-occurrence statistical features $F \in R^{C\times N\times N}$, and then use the multi-layer perceptron (MLP) and step-by-step averaging to generate the texture feature T of the processing area:

$$F' = MLP(F), \qquad F' \in R^{C'\times N\times N} \tag{5}$$

$$T = \frac{\sum_{m=1}^{N}\sum_{n=1}^{N} F'_{:,m,n}}{N \cdot N} \tag{6}$$

This paper also adopts the image pyramid structure, which is a structure that expresses images at multiple scales and interprets images at multiple resolutions. Enhanced texture features can be described from multiple scales. The designed pyramid structure adopts four parallel but different scale branch input feature maps. The feature map of each different branch is divided into a different number of sub-regions, and each sub-region passes through the texture feature extraction unit to utilize the corresponding texture representation of the region. Then, the obtained texture feature map of each branch is upsampled to the original size by nearest interpolation as the input map, and finally the outputs of the four branches are concatenated.

C. MULTIMODAL FEATURE REASONING MODULE

After obtaining the feature information of different modes, this feature information needs to be further processed to capture the complementary information between them. Therefore, we propose a feature fusion module called Multimodal Feature Reasoning Module (MFRM) to perform feature reasoning and information dissemination, as shown in Fig. 3. Finally, the information obtained by reasoning is upsampled to obtain segmentation results.

The features obtained by the given feature extraction backbone network and SIPM network are, respectively, represented as: $X_{\text{bacbone}} \in R^{H\times W\times C}$, $X_{SI} \in R^{H\times W\times C}$. In this paper, the input feature map is X mapped from the spatial domain to the graph domain $G \in R^{N\times F}$, where N represents the number of nodes in the graph and F represents the features contained in a node. In this way, pixels with similar features can be aggregated into a node as an anchor to generate a semantically aware graph feature. Specifically, the input features are mapped to X_{bacbone},X_{SI} image-domain features G_{bacbone},G_{SI} through two convolutional layers:

$$G = v(X; W_v) \times w(X; W_w) \tag{7}$$

These represent $v(\cdot)$,$w(\cdot)$ the convolution operations used for graph projection and feature dimensionality reduction, and W_v,W_w represent the learnable convolution kernels of each layer.

After projection, in order to learn the relationship between the semantic graph and the associated node features of the edge graph, we adopt graph convolution [44] to learn the edge weights corresponding to the features of each node to reason on the fully connected graph. The input is G and the output is:

$$G = ((I - A_g)G)W_g \tag{8}$$

Among them, $I \in R^{N\times F}$ represents the identity matrix; $A_g \in R^{N\times N}$ represents the adjacency matrix; W_g represents the update parameters.

The fused map to the original spatial domain via the projection matrix obtained in the mapping step $v(\cdot)$ to obtain new features X_{bacbone},$\hat{X}_{SI}$. The obtained new features are passed through the upsampling network to obtain segmentation results.

IV. RESULTS

In this section, we will introduce the multimodal information fusion segmentation network structure proposed in this paper in detail. First, we will describe the overall structure of the proposed network in Section IV.A; then, in Sections IV.B and IV.C, we will discuss the innovations of this paper, SIPM and MFRM.

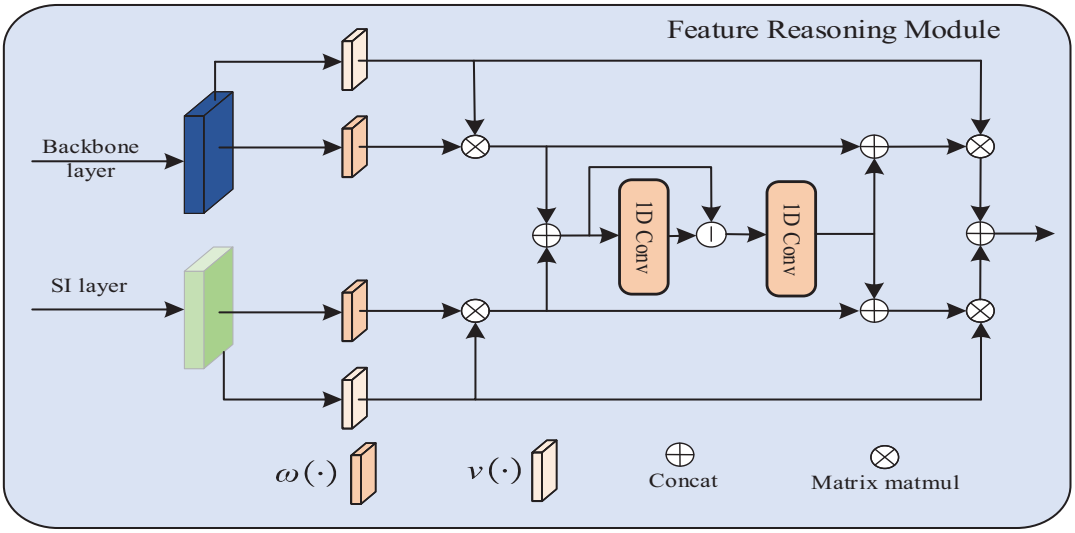

Fig. 3. Multimodal feature reasoning module.

A. DATASET

In the experiments, the training and testing data sets are all from BraTS2018 and BraTS2019. As important public datasets for multimodal brain tumor segmentation, BraTS are used in MICCAI's medical image segmentation competition. The data set will be added, deleted, or replaced in the competition every year to enrich the scale of the data sets. BraTS2018 and 2019 have 285 and 335 annotated brain tumor samples, respectively, they and are divided into highgrade glioma and low-grade glioma cases. Each case has 3D data of 4 different modes (Flair, t1, t1ce, t2). The datasets can differentiate three different brain tumor regions, namely Whole Tumor (WT), Tumor Core (TC), and Enhance Tumor (ET). The preprocessed datasets are composed of slices of different cases. During training, slices with tumors are used as data input, and all slices are cropped to reduce the amount of data and speed up training. The tested datasets are the brain tumor cases newly added in BraTS 2019 compared to BraTS 2018, a total of 50 labeled brain tumor samples.

B. EXPERIMENTAL DETAILS

In the preprocessing stage, the size of the source dataset is 240*240*155. In this paper, the data sets of all modalities are sliced. The size of each slice is the original 240*240, and the size of each slice after cropping is 160*160. And the original data will be standardized and normalized to realize the feature extraction and segmentation tasks of images in different modalities.

All the programs in this article are implemented under the PyTorch framework. The training process uses a GeForce RTX 3090 GPU. The optimizer during the experiment is Adam, the optimizer momentum is 0.9, and other parameters are default parameters. The initial learning rate, weight decay, and batch size are 0.001, $1e^{-4,}$ and 32, respectively.

C. COMPARATIVE EXPERIMENT

First, model training and prediction are carried out on the method proposed in this paper. During the training process, the training loss (loss), IoU index of the training set and the test set, and the loss (val_loss) and val_IoU index of the test set are obtained, as follows. As shown in Fig. 4, for the previous training, the loss drops very quickly while the curve of the IoU index rises rapidly, and the subsequent training loss gradually decreases and becomes stable. Meanwhile, the rising curve of the IoU index gradually slows down and stabilizes, indicating that the training effect of the proposed method is ideal. To verify the superiority of the proposed multimodal information fusion segmentation network overall framework, this paper compares it with some classic brain tumor segmentation methods on the BraTS18,19 dataset. These methods include some classic segmentation methods FCN8 s, U-net, U-net++, Deep ResUnet (DRU) [45–48], and the experimental index comparison results are shown in Table I. The article uses several common indicators in image segmentation, Dice Score, Positive Predictive Value (PPV), and Hausdorff Score (HD). According to the experimental results, the method proposed in this paper has achieved considerable results in brain tumor segmentation.

It can be observed from the experimental results that the method proposed in this paper achieves considerable performance compared with the existing classical and advanced methods. Specifically, in terms of the Dice similarity coefficient score, the method based on the BraTS2018 and 2019 datasets achieved 87.2%, 87.6%, and 78.1% on the three segmentation tasks of whole tumor (WT), tumor core (TC) and enhancing tumor (ET), which are better than other classical segmentation methods, which are better than 2.50%~3.50%, 1.3%~5.5% and −0.1%~3.2%, respectively. Compared with Unet, which also uses Resnet50 as the feature extraction network for semantic segmentation, this paper achieves a more effective improvement, among which the improvement in the tumor core is the most obvious, and its Dice coefficient is 5.5% better than that. At the same time, compared with other brain tumor segmentation networks using different modal fusions, the method in this paper has achieved improvements in the tumor core and enhanced tumors. The extraction of different features using different modalities in this paper has also been proven to be effective. The segmentation indicators of Dice, HD and PP have been improved. In addition, for the Dice index, as shown in the histogram of the segmentation index in Fig. 5, an intuitive comparison result is obtained, and the proposed method is superior to the classic advanced segmentation method.

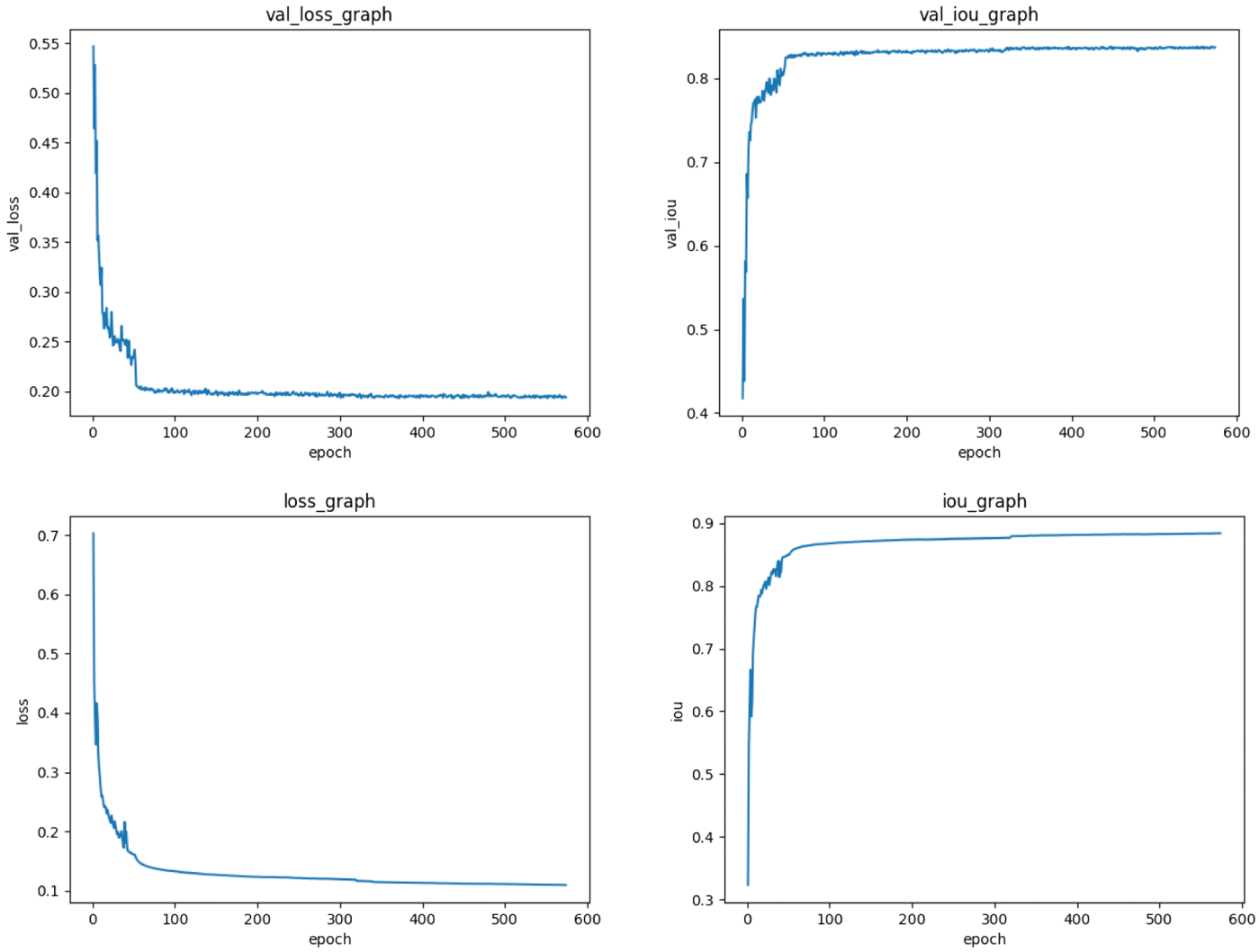

Fig. 4. Convergence of loss, val_loss, iou, val_iou during training.

Table I. Comparison of segmentation indicators on Bra TS2018 and 2019

	WT			TC			ET		
Bra TS	**Dice**	**HD**	**PPV**	**Dice**	**HD**	**PPV**	**Dice**	**HD**	**PPV**
FCN8s	0.837	2.744	0.884	0.834	1.716	0.850	0.749	2.928	0.761
U-net	0.838	2.662	0.856	0.821	1.751	0.853	0.769	2.847	0.774
Unet ++	0.842	2.638	0.866	0.827	1.722	0.850	0.774	2.815	0.792
DRU	0.847	2.589	0.888	0.834	1.693	0.860	**0.782**	2.779	**0.814**
Ours	**0.872**	**2.584**	**0.905**	**0.876**	**1.511**	**0.956**	0.781	2.778	0.812

D. ABLATION EXPERIMENT

To further verify the importance and actual contribution of the backbone network used in this paper and the designed modules, relevant ablation experiments are carried out in this paper. The index comparison of the ablation experiment is shown in Table II, and the experimental results are shown in Fig. 7.

In this paper, Resnet50 is used as the feature extraction base network (Base) for segmentation, and then the modules designed in this paper are added one by one on this basis for experiments, and SI-EE is added to the feature extraction network to improve and realize the enhancement of semantic information. And extraction, add an FRM module for feature fusion and feature reasoning. As shown in the performance index results in Table II, woSI-EE means that the SI-EE module is not added, and woFRM means that the FRM module is not added. According to the experimental results and data, it can be seen that the modules proposed in this paper can improve the Base Segmentation performance, among which the improvement

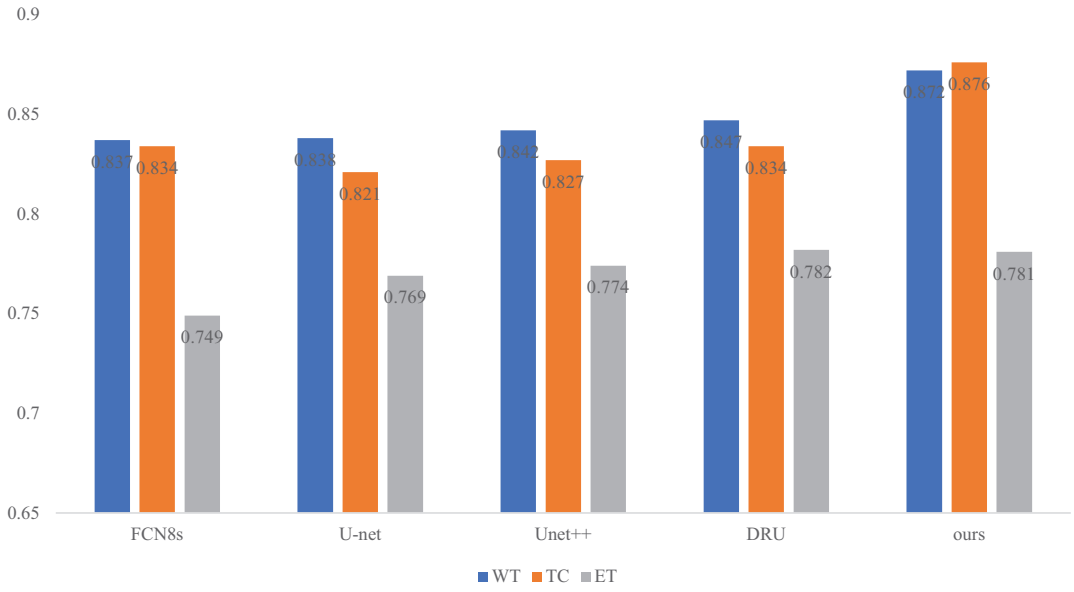

Fig. 5. Histogram of comparison experiment based on Dice indicator.

Table II. Comparison of ablation experiments of segmentation indicators on BraTS2018 and 2019

	WT			TC			ET		
Bra TS	**Dice**	**HD**	**PPV**	**Dice**	**HD**	**PPV**	**Dice**	**HD**	**PPV**
Base	0.825	2.780	0.839	0.820	1.753	0.846	0.748	2.938	0.754
woSI-EE	0.761	3.043	0.831	0.687	2.297	0.778	0.651	3.319	0.706
woFRM	0.807	2.944	0.808	0.782	1.939	0.816	0.712	3.168	0.700
Ours	0.872	2.584	0.905	0.876	1.511	0.956	0.781	2.778	0.812

of the FRM feature reasoning module on the above basis is the most significant. In addition, as shown in Fig. 6, based on the histogram comparison of the segmentation index Dice, the effectiveness of the proposed method for brain tumor segmentation can be obtained.

Figure 7 shows the visual images of the brain tumor segmentation obtained by each method in the ablation experiment. It can be seen from the figure that when the FRM module is not added, the edge segmentation of the brain tumor and the segmentation effect of the red area are not very good. At this time, under the action of the SI-EE module, compared with the Base, the segmentation results are greatly improved; When the SI-EE module is not added, the segmentation effect of the red area of the brain tumor is not very good, but it is easy to find that under the FRM module, the segmentation result of the edge area is closer to Ground Truth; in short, in the method proposed in this paper, the obtained brain tumor segmentation effect is smoother.

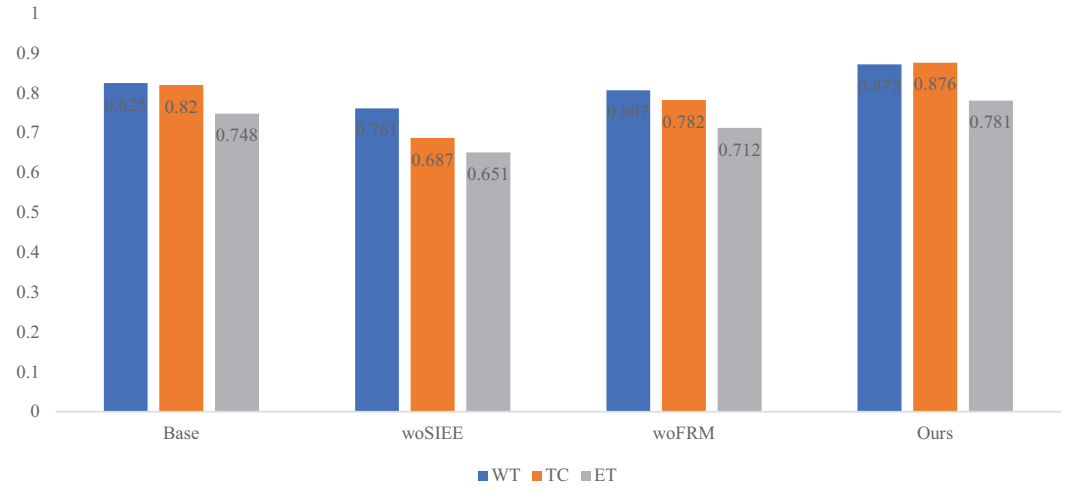

Fig. 6. Histogram of ablation experiment based on Dice index.

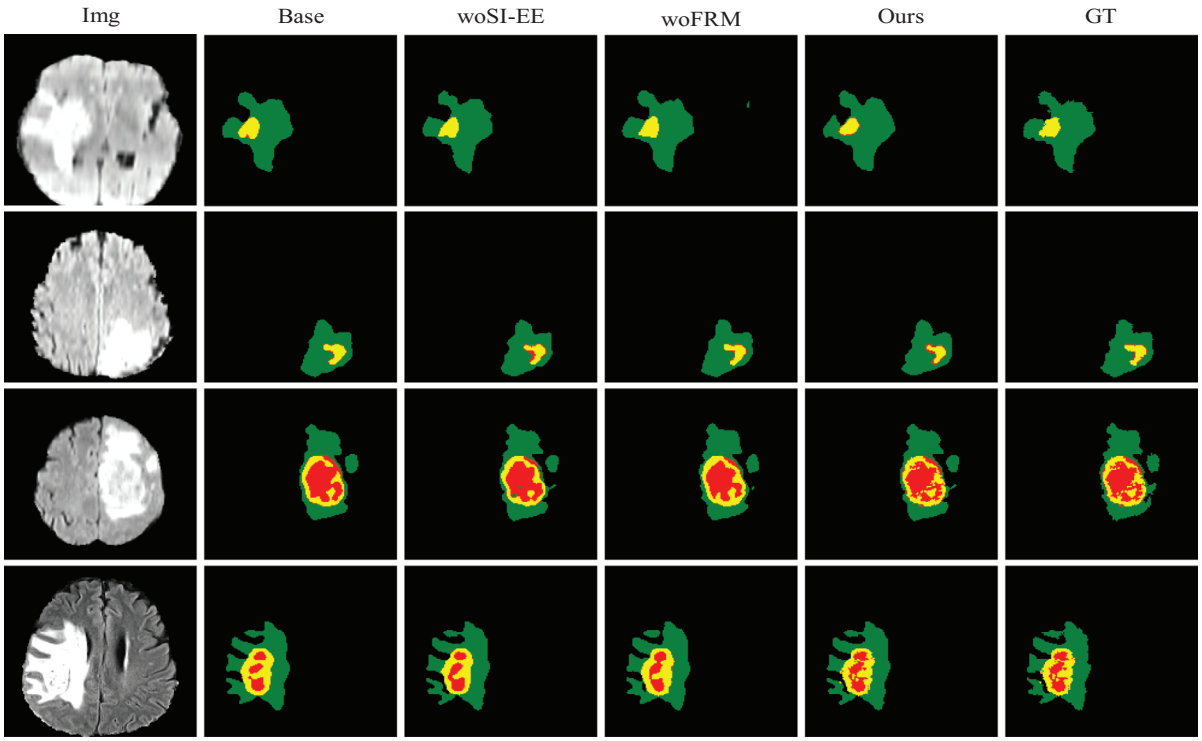

Fig. 7. Comparison of ablation experiments. Here green represents edematous areas, yellow represents enhancing tumors, and red represents necrotic and non-enhancing tumors.

V. CONCLUSIONS

In this paper, we proposed a novel method of multimodal information fusion method for brain tumor segmentation. More comprehensive multimodal information from MRI was utilized to ensure accurate segmentation. Specifically, a Semantic Information Processing Module was introduced to achieve free Multiscale feature enhancement and extraction. The core of module formed a new SI-EE (Semantic Information Enhancement and Extraction) module. A new Multimodal Feature Reasoning Module was constructed to process both the backbone network feature information layer and semantic feature information layer. According to the experiments based on BraTS2018 and 2019 datasets, the method proposed in this paper achieved better results than the classic advanced segmentation methods.

CONFLICT OF INTEREST STATEMENT

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

REFERENCES

[1] Y. Lu, Z. Xue, G. Xia, and L. Zhang, "A survey on vision-based UAV navigation," *Geo-Spatial Inf. Sci.,* vol. 21, pp. 21–32, 2018.

[2] G. Cheng and J. Han, "A survey on object detection in optical remote sensing images," *Isprs J. Photogramm.*, vol. 117, pp. 11–28, 2016.

[3] J. G. Richens, C. M. Lee, and S. Johri, "Improving the accuracy of medical diagnosis with causal machine learning," *Nat. Commun.*, vol. 11, p. 3923, 2020.

[4] Y. Liu, L. Wang, J. Cheng, C. Li, and X. Chen, "Multi-focus image fusion: a survey of the state of the art," *Inform Fusion*, vol. 64, pp. 71–91, 2020.

[5] Y. Liu, Y. Shi, F. Mu, J. Cheng, and X. Chen, "Glioma segmentation-oriented multi-modal MR image fusion with adversarial learning," *IEEE/CAA J. Automat. Sin.*, vol. 9, pp. 1528–1531, 2022.

[6] Y. Liu, F. Mu, Y. Shi, and X. Chen, "Sf-net: a multi-task model for brain tumor segmentation in multimodal mri via image fusion," *IEEE Signal Proc. Lett.*, vol. 29, pp. 1799–1803, 2022.

[7] Y. Xu, X. He, G. Xu, G. Qi, K. Yu, L. Yin, P. Yang, Y. Yin, and H. Chen, "A medical image segmentation method based on multi-dimensional statistical features," *Front. Neurosci.-Switz.*, vol. 16, p. 1009581, 2022.

[8] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recogn.*, 2015, pp. 3431–3440.

[9] V. Badrinarayanan, A. Handa, and R. Cipolla, "Segnet: a deep convolutional encoder-decoder architecture for robust semantic pixel-wise labelling," arXiv preprint arXiv:1505.07293, 2015.

[10] A. Dosovitskiy et al., "An image is worth 16x16 words: Transformers for image recognition at scale," arXiv preprint arXiv:2010.11929, 2020.

[11] Z. Liu et al., "Swin transformer: hierarchical vision transformer using shifted windows," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2021, pp. 10012–10022.

[12] Y. Li, Z. Wang, L. Yin, Z. Zhu, G. Qi, and Y. Liu, "X-Net: a dual encoding–decoding method in medical image segmentation," *Visual Comput.*, vol. 39, pp. 1–11, 2021.

[13] O. Ronneberger, P. Fischer, and T. Brox, "U-net: convolutional networks for biomedical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Munich, Germany: Springer, 2015, pp. 234–241.

[14] A. Garcia-Garcia, S. Orts-Escolano, S. Oprea, V. Villena-Martinez, and J. Garcia-Rodriguez, "A review on deep learning techniques applied to semantic segmentation," arXiv preprint arXiv:1704.06857, 2017.

[15] Q. Zhou, X. Wu, S. Zhang, B. Kang, Z. Ge, and L. J. Latecki, "Contextual ensemble network for semantic segmentation," *Pattern Recogn.*, vol. 122, p. 108290, 2022.

[16] X. He, G. Qi, Z. Zhu, Y. Li, B. Cong, and L. Bai, "Medical image segmentation method based on multi-feature interaction and fusion over cloud computing," *Simul. Model. Pract. Ther.*, vol. 126, p. 102769, 2023.

[17] B. S. Reddy and A. Sathish, "A hybrid method for magnetic resonance brain images classification and segmentation using soft computing techniques," *J. Artif. Intell. Technol.*, vol. 3, pp. 134–141, 2023.

[18] X. Shi, and M. Abisado, "Diagnostic segmentation based on kidney medical image," *J. Artif. Intell. Technol.*, vol. 3, no. 4, pp. 173–180, 2023.

[19] J. Tian, Y. Zhang, J. Lei, C. Sun, and G. Hu, "Lightweight classification network for pulmonary tuberculosis based on CT images," *J. Artif. Intell. Technol.*, vol. 3, pp. 25–31, 2023.

[20] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recogn.*, 2014, pp. 580–587.

[21] N. Kesav and M. G. Jibukumar, "Efficient and low complex architecture for detection and classification of brain tumor using RCNN with two channel CNN," *J. King Saud Univ.-Comput. Inf. Sci.*, vol. 34, pp. 6229–6242, 2022.

[22] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 2961–2969.

[23] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE T. Pattern Anal.*, vol. 37, pp. 1904–1916, 2015.

[24] P. J. Burt and E. H. Adelson, "The Laplacian pyramid as a compact image code," in *Readings in Computer Vision*. Texas: Elsevier, 1987, pp. 671–679.

[25] J. Dai, Y. Li, K. He, and J. Sun, "R-fcn: Object detection via region-based fully convolutional networks," *Adv. Neural Inf. Process. Syst.*, vol. 29, pp. 379–387, 2016.

[26] S. Wu, J. Shi, and Z. Chen, "HG-FCN: hierarchical grid fully convolutional network for fast VVC intra coding," *IEEE T. Circ. Syst. Video Technol.*, vol. 32, pp. 5638–5649, 2022.

[27] K. Chaiyasarn, A. Buatik, H. Mohamad, M. Zhou, S. Kongsilp, and N. Poovarodom, "Integrated pixel-level CNN-FCN crack detection via photogrammetric 3D texture mapping of concrete structures," *Automat. Constr.*, vol. 140, p. 104388, 2022.

[28] L. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," arXiv preprint arXiv:1706.05587, 2017.

[29] L. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Deeplab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFS," *IEEE T. Pattern Anal.*, vol. 40, pp. 834–848, 2017.

[30] L. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 801–818.

[31] V. Badrinarayanan, A. Kendall, and R. Cipolla, "Segnet: a deep convolutional encoder-decoder architecture for image segmentation," *IEEE T. Pattern Anal.*, vol. 39, pp. 2481–2495, 2017.

[32] S. Alqazzaz, X. Sun, X. Yang, and L. Nokes, "Automated brain tumor segmentation on multi-modal MR image using SegNet," *Comput. Vis. Media*, vol. 5, pp. 209–219, 2019.

[33] H. Noh, S. Hong, and B. Han, "Learning deconvolution network for semantic segmentation," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2015, pp. 1520–1528.

[34] Y. Liu, F. Mu, Y. Shi, J. Cheng, C. Li, and X. Chen, "Brain tumor segmentation in multimodal MRI via pixel-level and feature-level image fusion," *Multimodal Brain Image Fusion: Methods, Eval. Appl.*, vol. 16648714, p. 62, 2023.

[35] Y. Liu, Y. Shi, F. Mu, J. Cheng, C. Li, and X. Chen, "Multimodal MRI volumetric data fusion with convolutional neural networks," *IEEE T. Instrum. Meas.*, vol. 71, pp. 1–15, 2022.

[36] G. Ghiasi and C. C. Fowlkes, "Laplacian pyramid reconstruction and refinement for semantic segmentation," in *European Conference on Computer Vision*. Amsterdam, The Netherlands: Springer, 2016, pp. 519–534.

[37] L. Chen, Y. Yang, J. Wang, W. Xu, and A. L. Yuille, "Attention to scale: scale-aware semantic image segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recogn.*, 2016, pp. 3640–3649.

[38] M. Guo et al., "Attention mechanisms in computer vision: a survey," *Comput. Visual Media*, vol. 8, no. 3, pp. 1–38, 2022.

[39] A. M. Obeso, J. Benois-Pineau, M. S. G. Vázquez, and A. Á. R. Acosta, "Visual vs internal attention mechanisms in deep neural networks for image classification and object detection," *Pattern Recogn.*, vol. 123, p. 108411, 2022.

[40] H. Zhou et al., "Refine-net: normal refinement neural network for noisy point clouds," *IEEE T. Pattern Anal.*, vol. 45, no. 1, pp. 946–963, 2022.

[41] G. Lin, A. Milan, C. Shen, and I. Reid, "Refinenet: multi-path refinement networks for high-resolution semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recogn.*, 2017, pp. 1925–1934.

[42] V. Nekrasov, C. Shen, and I. Reid, "Light-weight refinenet for real-time semantic segmentation," arXiv preprint arXiv:1810.03272, 2018.

[43] Z. Zhu, X. He, G. Qi, Y. Li, B. Cong, and Y. Liu, "Brain tumor segmentation based on the fusion of deep semantics and edge information in multimodal MRI," *Inform Fusion*, vol. 91, pp. 376–387, 2023.

[44] X. He, K. Deng, X. Wang, Y. Li, Y. Zhang, and M. Wang, "Lightgcn: simplifying and powering graph convolution network for recommendation," in *Proc. 43rd Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval*, 2020, pp. 639–648.

[45] A. S. Darshik, A. Dev, M. Bharath, B. A. Nair, and G. Gopakumar, "Semantic segmentation of spectral images: a comparative study using FCN8s and U-NET on RIT-18 dataset," in *2020 11th Int. Conf. Comput., Commun. Netwk. Technol. (ICCCNT)*. Kharagpur, India: IEEE, 2020, pp. 1–6.

[46] N. Siddique, S. Paheding, C. P. Elkin, and V. Devabhaktuni, "U-net and its variants for medical image segmentation: a review of theory and applications," *IEEE Access*, vol. 9, pp. 82031–82057, 2021.

[47] Z. Zhang, Q. Liu, and Y. Wang, "Road extraction by deep residual u-net," *IEEE Geosci. Remote Sens.*, vol. 15, pp. 749–753, 2018.

[48] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "Unet++: redesigning skip connections to exploit multiscale features in image segmentation," *IEEE T. Med. Imaging*, vol. 39, pp. 1856–1867, 2019.