



Boundary-aware and cross-modal fusion network for enhanced multi-modal brain tumor segmentation

Tongxue Zhou

School of Information Science and Technology, Hangzhou Normal University, Hangzhou 311121, China

ARTICLE INFO

Keywords:

Brain tumor segmentation
Boundary detection
Multi-modal fusion
Deep learning
MR modalities

ABSTRACT

In recent years, brain tumor segmentation has emerged as a critical area of focus in medical image analysis. Accurate tumor delineation is essential for effective treatment planning and patient monitoring. Many existing algorithms struggle with accurately delineating complex tumor boundaries, particularly in cases where tumors exhibit heterogeneous features or blend with surrounding healthy tissues. In this paper, I propose a novel boundary-aware multi-modal brain tumor segmentation network, which integrates four key contributions to improve segmentation accuracy. First, I introduce a Boundary Extraction Module (BEM) to capture essential boundary information for segmentation. Second, I present a Boundary Guidance Module (BGM) to guide the segmentation process by incorporating boundary-specific information. Third, I design a Boundary Supervision Module (BSM) to enhance segmentation accuracy by providing multi-level boundary supervision. Lastly, I propose a Cross-feature Fusion (CFF) that integrates complementary information from different MRI modalities to enhance overall segmentation performance. Experimental results demonstrate that the proposed model outperforms state-of-the-art methods, achieving superior tumor segmentation accuracy across brain tumor segmentation datasets, thereby indicating its potential for clinical applications in neuroimaging.

1. Introduction

Brain tumors are among the most challenging and life-threatening conditions in neuro-oncology, with gliomas being the most common and aggressive type [1]. Globally, brain tumors affect approximately 308,000 individuals annually, with gliomas accounting for 45.2% of malignant brain tumors. The prognosis for patients with high-grade gliomas remains poor, with a median survival rate of around 15 months despite advances in treatment [2,3]. Accurate and timely detection of brain tumors is critical for patient outcomes, as early diagnosis and precise delineation of tumor boundaries are essential for effective treatment planning, surgical interventions, and radiation therapy.

Brain tumor segmentation plays a critical role in clinical decision-making, treatment planning and patient prognosis [4,5]. Accurately segmenting brain tumors from multi-modal MRI images remains a challenging task due to the high variability in tumor shape, size, and appearance across different modalities [6]. Multi-modal magnetic resonance imaging (MRI) provides complementary information from different modalities, such as T1, T2, FLAIR, and post-contrast T1, presented in Fig. 1, which can capture diverse tissue characteristics essential for comprehensive tumor assessment [7,8]. However, manual segmentation of these tumors is time-consuming, prone to inter-observer variability, and highly dependent on radiologist expertise.

Automated brain tumor segmentation methods, powered by deep learning, aim to address these challenges by providing fast, reliable, and accurate tumor delineation [9]. However, traditional deep learning approaches for brain tumor segmentation often focus on voxel-wise classification without explicitly considering the importance of boundary information [10–13]. Accurate boundary detection in clinical practice is particularly important for differentiating glioma sub-regions, which exhibit heterogeneous and infiltrative growth patterns. It helps remove tumors precisely, minimize damage to healthy tissue, and tailor treatment, improving patient outcomes and survival. Existing methods often struggle with boundary ambiguity, neglecting structural information at tumor boundary, and fail to fully leverage complementary data from multiple MRI modalities, resulting in suboptimal accuracy. The proposed boundary-aware and cross-modal fusion network addresses these challenges by integrating boundary-specific modules to capture, refine, and reinforce intricate tumor boundaries while employing a cross-feature fusion mechanism to leverage multimodal information effectively. By delivering robust and accurate segmentation, the proposed approach not only overcomes the limitations of current techniques but also enhances clinical decision-making, ultimately contributing to better treatment personalization and improved prognostic assessments. The main contributions of this work are summarized as follows:

E-mail address: txzhou@hznu.edu.cn.

<https://doi.org/10.1016/j.patcog.2025.111637>

Received 3 November 2024; Received in revised form 24 February 2025; Accepted 23 March 2025

Available online 4 April 2025

0031-3203/© 2025 Elsevier Ltd. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

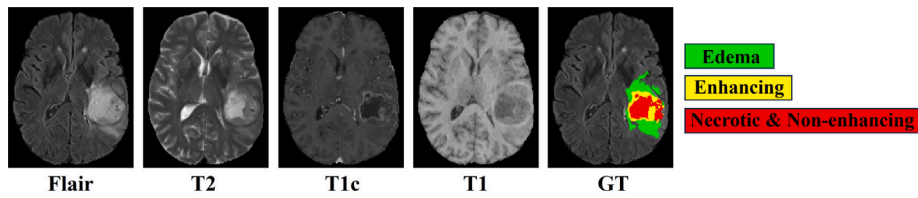


Fig. 1. Illustrations of multi-modal MRI images, comprising Flair, T2, T1c, and T1c sequences. The ground truth (GT) annotations highlight different regions: edema is marked in green, enhancing tumors in yellow, and necrotic tumors alongside non-enhancing tumor cores in red.

(1) A novel boundary-aware and cross-modal fusion network is introduced for multi-modal brain tumor segmentation, addressing the challenges posed by the heterogeneity of glioma sub-regions and their intricate boundaries.

(2) A Boundary Extraction Module (BEM) is proposed to extract multi-scale boundary features, enhancing the model's ability to delineate critical tumor boundary.

(3) A Boundary Guidance Module (BGM) is proposed to leverage boundary-specific features to achieve accurate tumor segmentation.

(4) A Boundary Supervision Module (BSM) is proposed to provide direct boundary supervision at multiple levels, reinforcing the model's capability to capture fine-grained boundary details.

(5) A novel Cross-feature Fusion (CFF) mechanism is designed to integrate complementary information from multiple MRI modalities, enabling optimal utilization of modality-specific features for a more comprehensive tumor representation and improved segmentation performance.

The remainder of this paper is organized as follows: Section 2 provides a detailed overview of related work in brain tumor segmentation. Section 3 introduces the motivation behind the proposed method and presents the proposed boundary-aware multi-modal brain tumor segmentation network. Section 4 describes the datasets, experimental setup and evaluation metrics used in the work. Section 5 discusses the experimental results. Section 6 further analyze the proposed components, addresses limitations and suggests future research directions. Finally, Section 7 concludes the paper.

2. Related works

2.1. Brain tumor segmentation

Brain tumor segmentation is a crucial task in medical imaging, traditionally performed using methods that rely on handcrafted features and conventional machine learning techniques. Classical approaches, such as thresholding, region growing [14], and clustering [15,16], are widely used to segment tumors based on intensity differences in MRI images. Probabilistic models like Gaussian Mixture Models (GMMs) [17] and atlas-based [18] segmentation methods also played a role by leveraging prior knowledge from labeled data to guide the segmentation process. While these methods were effective for segmenting simple tumors, they faced significant challenges with heterogeneous and irregularly shaped tumors, often producing inaccurate results. Moreover, traditional techniques are sensitive to noise and variations in MRI quality, limiting their reliability and generalizability across different clinical settings.

Recent advancements in deep learning have greatly improved brain tumor segmentation, with notable architectures like U-Net [19] becoming widely used due to its effective encoder-decoder structure that captures multi-scale features. For example, Isensee et al. [20] presents nnU-Net as a groundbreaking framework for brain tumor segmentation, achieving state-of-the-art performance and enhancing the efficacy of medical imaging. Luu et al. [21] enhanced the winning nnU-Net framework by implementing several modifications, including a larger network, group normalization, and axial attention in the decoder. The proposed approach achieved the first place on the test

dataset. Moreover, Graph-based methods have emerged as an effective approach for brain tumor segmentation. In these methods, the image is represented as a graph, where pixels serve as nodes and edges denote the relationships between them. This representation facilitates the capture of spatial dependencies and contextual information crucial for accurate segmentation. For instance, Zhou et al. [22] introduced a multi-modal graph convolution network called M2GCNet, which explores relationships across different MRI modalities to tackle the challenges associated with brain tumor segmentation. Similarly, Ma et al. [23] proposed a dual graph reasoning module that models long-range spatial dependencies and captures contextual inter-dependencies between different channels for brain tumor segmentation. Additionally, transformer-based models such as Vision Transformers [24] have gained attention for their ability to capture long-range dependencies and improve feature extraction through self-attention mechanisms. For example, Chen et al. [25,26] proposed a hybrid TransUNet that combines transformers with U-Net to enhance segmentation accuracy by leveraging both local features from CNNs and global contextual information from transformers. Lin et al. [27] introduced a dual swin transformer U-Net that leverages the hierarchical advantages of the Swin Transformer in both the encoder and decoder of the standard U-shaped architecture. Xu et al. [28] proposed a brain tumor segmentation framework with deep nuanced reasoning and Swin Transformer.

2.2. Boundary-aware image segmentation

In recent years, boundary-aware segmentation techniques have attracted significant attention in medical image analysis for their ability to capture complex and irregular tumor boundaries. These techniques address the challenges posed by intricate anatomical structures and variations in tumor morphology, which can complicate accurate delineation. By focusing on boundary information, these methods can effectively reduce segmentation errors in regions where tumors intersect with healthy tissues, leading to improved diagnostic and therapeutic outcomes.

Numerous approaches integrate boundary supervision into neural networks using explicit edge detection algorithms. For example, Wang et al. [29] proposed a boundary-aware context neural network designed specifically for 2D medical image segmentation, which can effectively capture richer contextual information while preserving fine spatial details. Zhu et al. [30] proposed a method for brain tumor segmentation that enhances multimodal spatial information and corrects boundary shapes. The method consists of three components: the modality information extraction (MIE) module, the spatial information enhancement (SIE) module, and the boundary shape correction (BSC) module. Similarly, Tang et al. [31] developed an effective boundary refinement framework that enhances boundary quality through a crop-then-refine strategy for various instance segmentation model, demonstrating the adaptability and effectiveness of boundary-aware techniques. Zhu et al. [32] proposed a sparse dynamic volume TransUNet with multi-level edge fusion for brain tumor segmentation. It effectively integrates both global and local spatial information, enhancing tumor boundary detection and clinical decision-making. Furthermore, other methods focus on designing specific loss functions to emphasize boundary regions. For instance, Huang et al. [33] introduced

a novel loss function that incorporates a contour weighting strategy, effectively combining dice loss and cross-entropy loss to mitigate class imbalance in image segmentation tasks. Borse et al. [34] proposed to integrate a distance-based metric into the boundary detection loss, and utilized an inverse-transformation network to model the spatial distance between boundary maps. In addition to these strategies, some methods utilize edge detection as an auxiliary task to guide the segmentation network in learning more discriminative features. A notable example is the work by Li et al. [35], which proposed a semantic boundary awareness network that effectively generates and applies refined land-cover boundaries to enhance segmentation outcomes. In addition, Wang et al. [36] introduced a multi-task dual boundary-aware network to improve retinal layer segmentation performance in optical coherence tomography (OCT) images. Zhu et al. [37] proposed an innovative deep learning-based method for brain tumor segmentation, which effectively integrates deep semantic features with edge information from multimodal MRI to enhance segmentation accuracy. Such innovations illustrate the ongoing evolution of boundary-aware techniques and their significant contributions to advancing medical image segmentation. These innovations reflect the ongoing evolution of boundary-aware techniques and their significant contributions to advancing medical image segmentation. However, many of the above methods rely on single-modality processing or are tailored primarily for 2D segmentation. Additionally, many existing approaches did not employ specific multi-modal fusion modules, limiting their ability to effectively handle the complexities of 3D brain tumor segmentation, especially when dealing with heterogeneous sub-regions and intricate tumor boundaries. In contrast, the proposed method leverages the power of multi-modal MRI data and multi-scale boundary features to address the challenges of 3D brain tumor segmentation. By integrating these features with advanced boundary supervision at multiple levels, the proposed method ensures a more precise tumor segmentation, particularly in cases involving heterogeneous sub-regions and complex tumor boundaries.

3. Methodology

3.1. Motivation behind the proposed method

Accurate segmentation of gliomas, especially their multiple sub-tumor regions, remains challenging due to the variability in characteristics such as cellular density and complex morphological interfaces. These sub-regions often merge with surrounding healthy tissues, making it difficult for traditional segmentation methods to capture precise boundaries. Existing deep learning approaches primarily focus on voxel-wise classification, which tends to overlook the critical boundary information that is essential for distinguishing between sub-tumor regions. Accurate delineation of these boundaries is vital for precise segmentation, as it directly impacts clinical decisions, such as defining surgical margins and tailoring treatment plans. To address these challenges, the proposed method incorporates a Boundary Extraction Module (BEM) to effectively capture boundary features. This is complemented by the Boundary Guidance Module (BGM) and Boundary Supervision Module (BSM), which ensure that boundary information guides the segmentation process. Additionally, the Cross-feature Fusion (CFF) mechanism leverages complementary information across MRI modalities, providing a holistic representation of tumor structures. By integrating these innovations, the proposed method enhances the model's ability to accurately delineate tumor sub-regions and boundaries, overcoming the limitations of existing techniques. This improvement holds significant potential for improving clinical outcomes in brain tumor diagnosis and treatment.

3.2. Proposed network architecture

The proposed network utilizes a six-level encoder-decoder architecture, as illustrated in Fig. 2. For simplicity, only one encoder is described. Each encoder processes individual MRI modalities, extracting modality-specific features through 3D convolutions, instance normalization (IN), and LeakyReLU activation. After feature extraction at each level, a down-sampling layer reduces the spatial dimensions. In the first three levels of the network, boundary-aware modules are implemented. The Boundary Extraction Module (BEM) extracts boundary features from the input feature maps, which are then combined with primary modality features and refined by the Boundary Guidance Module (BGM) to focus the network on tumor boundaries. In the decoder, up-sampled features are concatenated with corresponding encoder features, and the Cross-Feature Fusion (CFF) integrates modality features for improved segmentation. A convolutional layer adjusts the number of feature channels, while segmentation supervision generates predictions at each stage, combining them for the final segmentation result. In addition, the Boundary Supervision Module (BSM) further ensures accurate boundary predictions, which uses boundary ground truth maps to guide the learning process, emphasizing the network's focus on boundary-specific features.

3.3. The boundary extraction module

The Boundary Extraction Module (BEM) plays a critical role in accurately delineating tumor boundaries. The proposed BEM consists of two distinct components: a coarse boundary extraction module and a refined boundary extraction module, as depicted in Fig. 3.

Coarse boundary extraction module. This module begins by applying three separate convolutional layers to the input feature $F_{in} \in \mathbb{R}^{C \times H \times W \times D}$, each with different dilation rates (1, 2, and 3):

$$C_1 = \text{Conv}(F_{in}, r=1), C_2 = \text{Conv}(F_{in}, r=2), C_3 = \text{Conv}(F_{in}, r=3). \quad (1)$$

These dilated convolutional layers can capture coarse-level features across various scales, ensuring that the boundaries of the tumor are broadly detected.

After convolution, the global average pooling (GAP) is performed to compress each feature map into a single value by averaging the spatial dimensions, which effectively captures the global context. The pooled features can distill high-level spatial information and provide a more abstract understanding of the tumor boundaries:

$$G_1 = \text{GAP}(C_1), G_2 = \text{GAP}(C_2), G_3 = \text{GAP}(C_3). \quad (2)$$

Subsequently, the GAP outputs are subtracted from the respective convolution outputs:

$$S_1 = C_1 - \text{Reshape}(G_1), S_2 = C_2 - \text{Reshape}(G_2), S_3 = C_3 - \text{Reshape}(G_3). \quad (3)$$

This subtraction operation emphasizes critical boundary information by removing the coarse background details, making the edges of the tumor more prominent.

After that, the resulting features S_1, S_2, S_3 are then concatenated and passed through a convolutional layer for channel adjustment:

$$B_{\text{coarse}} = \text{Conv}(\text{Concat}(S_1, S_2, S_3)) \quad (4)$$

This stage focuses on coarse boundary detection, capturing broad features of the tumor.

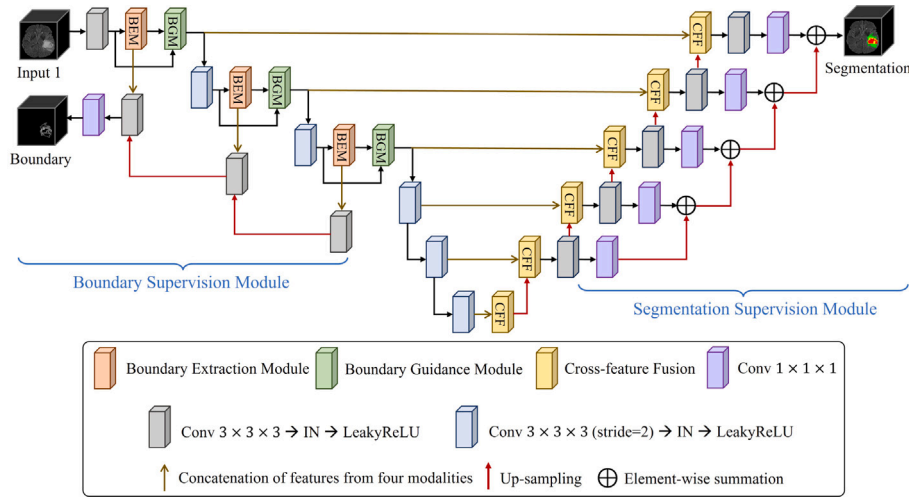


Fig. 2. Overview of the proposed boundary-aware multi-modal brain tumor segmentation network. It consists of a boundary extraction module, a boundary guidance module, a cross-feature fusion module, a boundary supervision module and a segmentation supervision module.

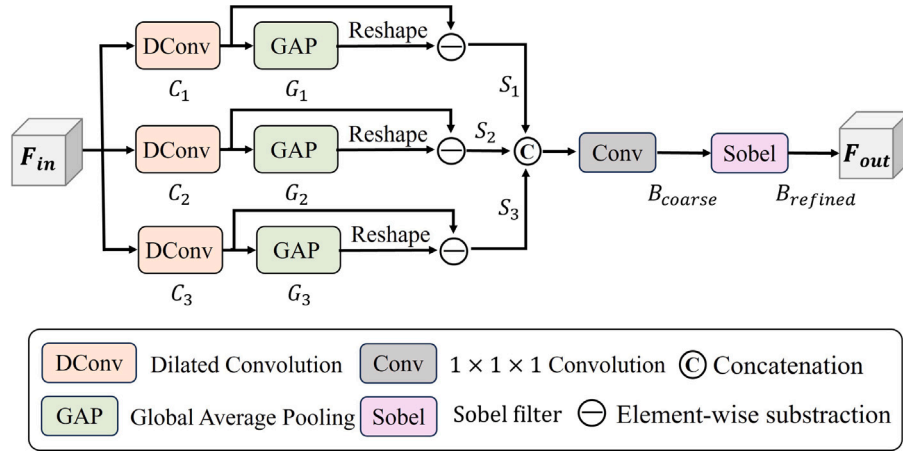


Fig. 3. Overview of the proposed boundary extraction module, consisting of a coarse boundary extraction module and a refined boundary extraction module.

Refined boundary extraction module. To further obtain detailed boundaries, the Sobel filter is applied by computing gradients in the spatial dimensions, highlighting areas of high intensity change, which are indicative of boundaries:

$$B_{refined} = \text{Sobel}(B_{coarse}) \quad (5)$$

This step is critical for isolating the tumor's edges from surrounding tissue, enabling more precise segmentation.

By combining the multi-scale dilated convolutions for coarse boundary detection with the Sobel filter for refined detection, the BEM can provide accurate boundary representations for improving the accuracy of tumor segmentation. It is noted that the ablation experiments are conducted to assess the effectiveness of the proposed BEM, as detailed in Section 5.1. Additionally, a comparative analysis is conducted in Section 6.2 to evaluate the segmentation performance when using either the coarse or refined boundary extraction modules.

3.4. The boundary guidance module

The Boundary Guidance Module (BGM) is designed to enhance the segmentation accuracy by effectively integrating the extracted boundary features with modality-specific features, as depicted in Fig. 4. The input of this module consists of the boundary feature $F_b \in \mathbb{R}^{C \times H \times W \times D}$ obtained from the BEM, and the modality-specific feature $F_m \in \mathbb{R}^{C \times H \times W \times D}$ extracted from each encoder.

First, the modality-specific feature and the boundary feature are combined via element-wise multiplication, and followed by a softmax activation to generate an attention map, which can be represented as:

$$F_{soft} = \text{Softmax}(F_m \otimes F_b), \quad (6)$$

where $\otimes$ denotes the element-wise multiplication operation. This operation aims to emphasize regions of interest in the modality-specific feature based on the boundary information.

Furthermore, an additive combination of the two features is performed, followed by a Sigmoid activation to further refine the attention mechanism:

$$F_{sig} = \sigma(F_m \oplus F_b), \quad (7)$$

where $\oplus$ denotes the element-wise summation operation, σ denotes the Sigmoid activation function.

Next, the attention maps from both the Softmax and Sigmoid operations are combined to yield a final attention map, and then multiplied by the modality-specific feature to produce the attention-enhanced feature:

$$F_{att} = F_{modality} \otimes (F_{soft} \oplus F_{sig}). \quad (8)$$

Finally, the output of the BGM is computed by combining the attention-enhanced feature with the original modality-specific and boundary features:

$$F_{out} = F_{att} \oplus F_m \oplus F_b. \quad (9)$$

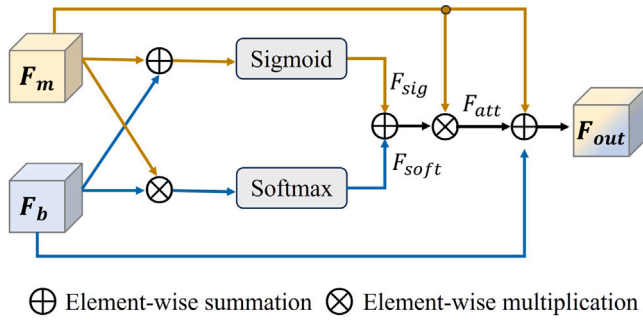


Fig. 4. Overview of the proposed boundary guidance module, consisting of a Softmax attention mechanism and a Sigmoid attention mechanism.

The resulting feature representation F_{out} incorporates both the modality-specific information and the boundary context, thereby enhancing the segmentation performance in subsequent layers of the network. It is noted that an ablation experiment is conducted to assess the effectiveness of the proposed BGM, as detailed in Section 5.1.

3.5. The boundary supervision module

To further guide the network to focus on boundary information, a Boundary Supervision Module (BSM) is designed within the decoder architecture, which uses boundary ground truth maps to guide the learning process, emphasizing the network's focus on boundary-specific features. In this module, the features extracted by the BEM at various network levels are initially concatenated. Then, a up-sampling is applied to match the dimensions of the features from the previous level. Next, a convolution operation is performed to adjust the number of feature channels, ensuring consistency in dimensionality for subsequent processing. Finally, a $1 \times 1 \times 1$ convolution is employed to generate the final boundary output. It is noted that an ablation experiment is conducted to assess the effectiveness of the proposed BSM, as detailed in Section 5.1.

3.6. The cross-feature fusion module

The proposed Cross-Feature Fusion module (CFF) aims to enhance the feature representation by leveraging cross-feature attention mechanism, which allows the network to focus on the most informative features, as presented in Fig. 5. First, the input is fed to GAP and GMP, respectively. The GAP and GMP operations are crucial in capturing global contextual information across all channels. GAP helps in summarizing the global spatial characteristics by averaging the features across the entire spatial domain, while GMP focuses on the most significant feature by selecting the maximum value across the spatial dimensions. These two pooling operations provide complementary perspectives of the input feature map. Given an input feature $F_{in} \in \mathbb{R}^{C \times H \times W \times D}$, the GAP and GMP are first performed to capture global contextual information across channels. The outputs from GAP and GMP are denoted as:

$$G_{avg} = GAP(F_{in}), G_{max} = GMP(F_{in}) \quad (10)$$

Next, these pooled features are passed through Multi-Layer Perceptron (MLP) layers to learn non-linear transformations, which help in generating more expressive attention weights. The MLP layers are then followed by a Sigmoid activation function, denoted as σ , to produce the attention weights for both average and max-pooled features:

$$A_{avg} = \sigma(W_2 \cdot \sigma(W_1 \cdot G_{avg})), A_{max} = \sigma(W_2 \cdot \sigma(W_1 \cdot G_{max})) \quad (11)$$

where W_1 and W_2 represent the weights of the MLP layers.

These attention weights are then used to recalibrate the input features by performing element-wise multiplication ($\odot$) between the initial

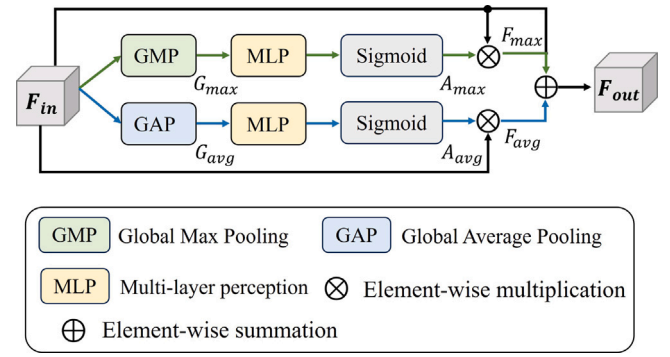


Fig. 5. Overview of the proposed cross-feature fusion module, which incorporates a global max pooling attention mechanism and a global average pooling attention mechanism.

input features and the attention weights. This operation allows the network to emphasize the most informative channels, thus enhancing the overall feature representation:

$$F_{avg} = F_{in} \odot A_{avg}, F_{max} = F_{in} \odot A_{max} \quad (12)$$

Finally, the original input feature F_{in} is fused with the recalibrated features F_{avg} and F_{max} using an element-wise summation operation ($\oplus$). This fusion integrates both average and max-pooled information with the original feature, creating a more informative output for subsequent tasks:

$$F_{out} = F_{in} \oplus F_{avg} \oplus F_{max}. \quad (13)$$

By incorporating cross-feature attention, the CFF module ensures that the most relevant features are highlighted, leading to improved feature discrimination. We also conduct ablation experiments to assess the effectiveness of the proposed CFF, as detailed in Section 5.1.

3.7. Training loss function

In this model, two distinct loss functions are employed tailored for specific tasks: the segmentation task and boundary supervision task. The segmentation task utilizes the Dice loss, which is defined as follows:

$$L_{seg} = \frac{2 \sum_i p_i g_i}{\sum_i p_i + \sum_i g_i} \quad (14)$$

where p_i is the predicted probability of voxel i , and g_i is the ground truth result.

For the Boundary Supervision Module (BSM), the L_2 loss is applied, which is defined as:

$$L_{boundary} = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2 \quad (15)$$

where y_i represents the ground truth result, $\hat{y}_i$ represents the predicted result, and N is the total number of voxels within the image.

$$Loss = L_{seg} + \lambda \cdot L_{boundary} \quad (16)$$

where λ is set to 0.1 in this work, based on the results of the experiment detailed in Section 6.5.

The combination of these loss functions ensures that the model not only excels in accurate segmentation but also emphasizes the integrity of boundary delineation.

4. Experimental setup

4.1. Dataset and implementation settings

The training and evaluation of the proposed model are based on the BraTS 2018, 2019 and 2020 datasets. These datasets consist of

multi-modal MRI images, including T1, T1c, T2, and FLAIR sequences, along with expert-annotated ground truth labels for the segmentation of various tumor regions: enhancing tumor (ET), non-enhancing tumor (NET), and edema (ED). The BraTS 2018 dataset contains 285 training cases, the BraTS 2019 dataset includes 335 training cases, and the BraTS 2020 dataset includes 369 training cases.

For implementation, Keras is chosen for model development, and the model is trained on an NVIDIA GeForce RTX 4090 GPU with 24 GB memory. We preprocess the MRI data using standard normalization techniques and resize them to $128 \times 128 \times 128$. The Adam optimizer is employed, and the initial learning rate is set to $5e-4$, and the training process includes a patience of 10 epochs, which allows the learning rate to be reduced by a factor of 0.5 if there is no improvement in validation loss over 10 consecutive epochs. And an early stopping criterion set to 50 epochs to prevent over-fitting by halting training if the validation loss does not improve over these periods. The training process uses a batch size of 1, and a test split of 0.2 is utilized.

4.2. Evaluation metrics

To assess the performance of the proposed segmentation model, two primary evaluation metrics are utilized: the Dice Similarity Coefficient (DSC) and the Hausdorff Distance (HD).

The Dice Similarity Coefficient quantifies the overlap between the predicted segmentation and the ground truth. It is defined as:

$$DSC = \frac{2|P \cap G|}{|P| + |G|}, \quad (17)$$

where P represents the set of predicted voxels, and G denotes the set of ground truth voxels. A higher Dice coefficient value indicates better agreement between the predicted and actual segmentation.

The Hausdorff Distance measures the maximum distance between the predicted and actual segmentation boundaries. It is defined as:

$$HD(A, B) = \max \left(\sup_{a \in A} \inf_{b \in B} d(a, b), \sup_{b \in B} \inf_{a \in A} d(b, a) \right), \quad (18)$$

where $d(a, b)$ is the Euclidean distance between points a and b . The lower Hausdorff Distance value indicates better performance.

5. Experimental results and analysis

5.1. Ablation experiments

The ablation experiments are first conducted on the BraTS 2018, 2019 and 2020 datasets to validate the effectiveness of the proposed modules, as detailed in Tables 1–3. It is noted that the proposed modules, including BEM, BGM, BSM and CFF have been carefully crafted to utilize operations that are computationally efficient, including addition, multiplication, sigmoid activation, GAP, GMP, concatenation, MLP, and convolutional operations. These lightweight operations ensure the modules achieve a balance between performance and computational feasibility. The baseline method includes the segmentation supervision. From Table 1, it can be observed that the baseline method achieves a DSC of 82.1% and a HD of 7.1 mm. When the Boundary Extraction Module (BEM) is integrated, the DSC improves by 1.1% to 83.0%, and the HD sees a significant improvement of 39.6%, decreasing to 4.3 mm. This improvement can be attributed to BEM's ability to capture crucial boundary features, improving the precision of the segmentation process. Similarly, when the Boundary Guidance Module (BGM) is introduced, the results show the improvements of 1.6% in DSC and 37.3% in HD compared to the baseline method. The BGM helps guide the network by emphasizing boundary information during the segmentation. Incorporating the Boundary Supervision Module (BSM) further enhances the DSC by 2.0% to 83.8% and improves the HD by 47.6% to 3.7 mm compared to the baseline method. The BSM introduces explicit supervision on boundary features, which enhance

the model's ability to distinguish between tumor boundaries and non-tumor regions, leading to more accurate segmentation result. Finally, the proposed full method, which integrates all modules including the Cross-Feature Fusion Module (CFF), achieves the best performance, improving the DSC by 2.4% to 84.1% and reducing the HD by 46.2%, reaching a value of 3.8 mm. This enhancement on DSC is due to the CFF module's ability to combine and recalibrate multi-modal features effectively, which allows the network to focus on the most informative features. Similar findings are observed from the BraTS 2019 and 2020 datasets results in Tables 2 and 3, respectively. For the results of the BraTS 2019 dataset in Table 2, the BEM enhances the baseline by 1.4% in DSC and 14.4% in HD. When compared to the baseline method, the BGM further improves results by 1.6% in DSC and 12.3% in HD. The BSM boosts DSC by 1.8% and improves HD by 26.7%. Finally, incorporating CFF leads to a 2.2% and 21.9% improvement in DSC and HD, respectively. For the results of the BraTS 2020 dataset in Table 3, the BEM improves the baseline by 0.5% in DSC and 25.5% in HD. Compared to the baseline, the BGM further enhances performance by 0.6% in DSC and 27.3% in HD. The BSM increases DSC by 0.8% and reduces HD by 27.3%. Finally, incorporating CFF results in a 1.3% improvement in DSC and a 34.5% reduction in HD. These results demonstrate that each module contributes to the overall segmentation performance, with the proposed method achieving the most substantial improvements.

5.2. Comparison with the state-of-the-art approaches

Tables 4 to 6 compare the proposed method with several state-of-the-art approaches on the BraTS 2018, 2019, 2020 datasets using Dice Similarity Coefficient (DSC) and Hausdorff Distance (HD) across whole tumor, tumor core, and enhancing tumor regions. For the results of the BraTS 2018 dataset in Table 4, Zhou et al. [38] achieved the highest DSC for the whole tumor at 88.7%, the proposed method attains competitive results, achieving the highest DSC across most of the regions: 86.0% for the tumor core, 79.2% for the enhancing tumor, and an average DSC of 84.1%, outperforming other methods. These results indicate the proposed method's capability to accurately segment the most critical tumor areas. In terms of HD, Aboelenein et al. [39] reported a significantly low HD for the tumor core of 3.3 mm, but the proposed method further improves the result with the lowest HD values of 5.3 mm for the whole tumor and 2.4 mm for the enhancing tumor. Additionally, the mean HD of 3.8 mm highlights the model's overall superior segmentation performance across regions. Overall, the comparison demonstrate that the proposed method achieves competitive performance, particularly in the regions of tumor core and enhancing tumor. A similar comparison is conducted on the BraTS 2019 and 2020 datasets, as shown in Tables 5 and 6. For the results of the BraTS 2019 dataset in Table 5, Zhou et al. [40] achieved the highest DSC for the whole tumor at 89.7%, and the method proposed by Wang et al. [41] achieved the highest DSC for the enhancing tumor at 81.8%. However, the proposed method yielded the best DSC of 88.2% for the tumor core, and 85.3% for the average DSC result. Furthermore, our method achieves the lowest HD values, with 5.4 mm for the whole tumor, 3.7 mm for the tumor core, 2.3 mm for the enhancing tumor and 3.8 mm for the average HD result. For the results of the BraTS 2020 dataset in Table 6, Hatamizadeh et al. [42] achieved the highest DSC for the whole tumor at 87.8% and for the enhancing tumor at 81.7%. However, the proposed method attained the best DSC of 88.3% for the tumor core and the highest average DSC of 85.5%. Additionally, it achieved the best HD across all tumor regions. Therefore, the proposed method outperforms the state-of-the-art methods, demonstrating its effectiveness in brain tumor segmentation.

Table 1

Ablation experiments conducted on the BraTS 2018 dataset using Dice Similarity Coefficient (%) and Hausdorff Distance (mm) as evaluation metrics. BEM, BGM, BSM and CFF represent the boundary extraction module, boundary guidance module, boundary supervision module and cross-feature fusion module, respectively. Bold values indicate the highest scores achieved. A checkmark (✓) signifies that the component is included in the evaluation, while a cross (X) indicates its exclusion.

Methods					Dice similarity coefficient (%) ↑				Hausdorff distance (mm) ↓			
Baseline	BEM	BGM	BSM	CFF	Whole	Core	Enh.	Mean	Whole	Core	Enh.	Mean
✓	X	X	X	X	85.4	83.5	77.5	82.1	7.9	9.6	3.7	7.1
✓	✓	X	X	X	86.3	85.3	77.4	83.0	4.7	5.1	3.0	4.3
✓	✓	✓	X	X	86.0	85.9	78.4	83.4	6.3	4.5	2.5	4.4
✓	✓	✓	✓	X	86.3	86.6	78.5	83.8	4.7	3.7	2.7	3.7
✓	✓	✓	✓	✓	87.0	86.0	79.2	84.1	5.3	3.7	2.4	3.8

Table 2

Ablation experiments conducted on the BraTS 2019 dataset using Dice Similarity Coefficient (%) and Hausdorff Distance (mm) as evaluation metrics. BEM, BGM, BSM and CFF represent the boundary extraction module, boundary guidance module, boundary supervision module and cross-feature fusion module, respectively. Bold values indicate the highest scores achieved. A checkmark (✓) signifies that the component is included in the evaluation, while a dash (X) indicates its exclusion.

Methods					Dice similarity coefficient (%) ↑				Hausdorff distance (mm) ↓			
Baseline	BEM	BGM	BSM	CFF	Whole	Core	Enh.	Mean	Whole	Core	Enh.	Mean
✓	X	X	X	X	86.0	85.6	78.8	83.5	6.3	5.1	3.2	4.9
✓	✓	X	X	X	86.8	87.1	80.0	84.6	5.4	4.6	2.5	4.2
✓	✓	✓	X	X	87.0	87.2	80.3	84.8	5.9	4.4	2.5	4.3
✓	✓	✓	✓	X	87.4	87.0	80.5	85.0	4.4	3.7	2.6	3.6
✓	✓	✓	✓	✓	87.2	88.2	80.6	85.3	5.4	3.7	2.3	3.8

Table 3

Ablation experiments conducted on the BraTS 2020 dataset using Dice Similarity Coefficient (%) and Hausdorff Distance (mm) as evaluation metrics. BEM, BGM, BSM and CFF represent the boundary extraction module, boundary guidance module, boundary supervision module and cross-feature fusion module, respectively. Bold values indicate the highest scores achieved. A checkmark (✓) signifies that the component is included in the evaluation, while a dash (X) indicates its exclusion.

Methods					Dice similarity coefficient (%) ↑				Hausdorff distance (mm) ↓			
Baseline	BEM	BGM	BSM	CFF	Whole	Core	Enh.	Mean	Whole	Core	Enh.	Mean
✓	X	X	X	X	87.0	86.6	79.7	84.4	8.6	5.1	2.7	5.5
✓	✓	X	X	X	86.7	87.5	80.3	84.8	5.5	4.0	2.8	4.1
✓	✓	✓	X	X	87.1	87.4	80.3	84.9	5.6	4.3	2.2	4.0
✓	✓	✓	✓	X	87.2	88.0	80.2	85.1	6.5	3.3	2.3	4.0
✓	✓	✓	✓	✓	87.3	88.3	80.9	85.5	5.4	3.5	1.9	3.6

Table 4

Comparison with the state-of-the-art methods based on Dice similarity coefficient (%) and Hausdorff distance (mm) on the BraTS 2018 dataset. Bold values indicate the highest scores achieved.

Methods	Dice similarity coefficient (%) ↑				Hausdorff distance (mm) ↓			
	Whole	Core	Enh.	Mean	Whole	Core	Enh.	Mean
Nestedformer [43]	70.7	73.1	70.5	71.4	27.3	42.2	12.5	27.3
Jungo et al. [44]	88.0	77.8	70.1	78.6	–	–	–	–
Zhou et al. [40]	88.2	78.6	69.4	78.7	6.7	7.6	7.1	7.1
Zhang et al. [45]	87.6	77.2	72.0	78.9	–	–	–	–
Aboelenen et al. [39]	85.2	81.2	74.1	80.2	8.3	3.3	3.3	5.0
Zhou et al. [38]	88.7	79.6	75.0	81.1	8.3	8.4	7.7	8.1
Rehman et al. [46]	86.9	81.4	76.7	81.7	–	–	–	–
M2GCNet [22]	85.1	84.3	76.1	81.8	5.7	6.0	4.3	5.3
TransBTS [41]	85.2	84.4	78.0	82.5	7.3	5.7	4.2	5.7
Ours	87.0	86.0	79.2	84.1	5.3	3.7	2.4	3.8

Table 5

Comparison with the state-of-the-art methods based on Dice similarity coefficient (%) and Hausdorff distance (mm) on the BraTS 2019 dataset. Bold values indicate the highest scores achieved.

Methods	Dice similarity coefficient (%) ↑				Hausdorff distance (mm) ↓			
	Whole	Core	Enh.	Mean	Whole	Core	Enh.	Mean
Zhou et al. [40]	89.7	77.5	70.6	79.3	6.0	7.5	7.6	7.0
Sheng et al. [47]	88.1	79.6	70.7	79.5	8.4	7.1	3.5	6.3
Zhou et al. [48]	88.7	76.5	75.9	80.4	7.9	8.8	8.0	8.2
Nestedformer [43]	77.4	84.9	79.6	80.6	19.4	13.3	4.6	12.4
Pan et al. [49]	89.0	78.0	76.0	81.0	13.3	15.3	15.7	14.8
Jiang et al. [50]	89.8	79.3	74.4	81.2	–	–	–	–
Rehman et al. [46]	88.4	81.4	76.3	82.0	–	–	–	–
M2GCNet [22]	85.6	84.3	78.3	82.7	6.4	5.0	3.8	5.1
TransBTS [41]	86.7	86.2	81.8	84.9	6.5	4.7	2.6	4.6
Ours	87.2	88.2	80.6	85.3	5.4	3.7	2.3	3.8

Table 6

Comparison with the state-of-the-art methods based on Dice similarity coefficient (%) and Hausdorff distance (mm) on the BraTS 2020 dataset. Bold values indicate the highest scores achieved.

Methods	Dice similarity coefficient (%) $\uparrow$				Hausdorff distance (mm) $\downarrow$			
	Whole	Core	Enh.	Mean	Whole	Core	Enh.	Mean
Attention U-Net [51]	79.6	80.8	73.8	78.1	7.2	4.9	3.9	5.3
Nestedformer [43]	83.2	81.4	78.9	81.2	13.5	15.0	3.7	10.7
M2GCNet [22]	85.9	85.3	77.1	82.8	5.7	5.1	3.8	4.9
Multi-BTS [11]	87.3	86.3	77.6	83.7	5.6	4.4	3.1	4.4
VT-UNet [52]	87.6	85.9	80.4	84.6	5.8	5.5	2.6	4.6
TransBTS [41]	86.4	85.5	80.0	84.1	6.6	4.6	3.2	4.8
Swin UNETR [42]	87.8	85.4	81.7	85.0	7.0	5.3	2.9	5.1
Ours	87.3	88.3	80.9	85.5	5.4	3.5	1.9	3.6

5.3. Visualization of the segmentation results

In this section, I present a visual comparison of the segmentation results generated by the proposed method, utilizing ablation experiments, presented in Fig. 6. In the first case, the baseline method inaccurately detects some necrotic and non-enhancing tumor regions. However, the incorporation of the BEM and BGM help to refine these results by enhancing the model's ability to identify tumor boundaries. As integrating the BSM and CFF, we can observe a gradual improvement in segmentation accuracy. These modules contribute to the refinement by providing guidance on boundary information. From the second case, it can be seen that the baseline method detects several false positive regions, particularly in the enhancing, necrotic and non-enhancing tumor regions. The integration of BEM, BGM, and BSM enhances the segmentation performance by incorporating valuable boundary information guidance. Lastly, the CFF effectively integrates multi-modal data, enhancing the model's ability to leverage complementary information from different imaging modalities. In the third case, the baseline method falsely identifies several isolated regions of enhancing, necrotic and non-enhancing tumors. However, with the application of the BEM and BGM can achieve a more satisfactory segmentation. Furthermore, the integration of the BSM and CFF can further refine the results. Similarly, in the last case, the baseline method struggles to complete detect the edema region. However, the inclusion of BEM and BGM can improve the segmentation accuracy. In addition, incorporating BSM and CFF can further enhance the results. In conclusion, the visualization results demonstrate the effectiveness of the proposed modules.

5.4. Visualization of failure segmentation cases

In this section, I present several failure segmentation cases to analyze the limitation of the proposed method, including instances of over-segmentation and under-segmentation, presented in Fig. 7. In the first case (over-segmentation), the model detects an excessive amount of edema in the right part of the brain, where healthy tissue is mistakenly classified as edema. It is primarily due to the similarity in intensity between healthy tissue and the edema region. To mitigate this, I propose employing intensity distribution analysis, which could help differentiate regions with similar intensity by analyzing pixel-level intensity distributions. Additionally, incorporating a post-processing step like conditional random fields could refine boundaries to reduce false positive segmentations. Furthermore, the under-segmentation in the second case, where necrotic and non-enhancing tumor regions are missed, which stems from poor contrast between these regions and surrounding tissue. To address this, attention mechanisms could be incorporated to focus the model's attention on low-contrast areas, highlighting regions that may otherwise be overlooked. Furthermore, region-specific loss functions can be used to focus on challenging-to-segment areas, such as the tumor boundaries, necrotic and non-enhancing tissue. These strategies have the potential to reduce both over-segmentation and under-segmentation issues. Future research directions to further enhance segmentation accuracy are discussed in Section 6.6.

6. Discussion

6.1. Analysis on the effectiveness of the segmentation supervision

To investigate the effectiveness of the proposed segmentation supervision, an ablation experiment is conducted on the BraTS 2018 dataset, with the results presented in Table 7 and Fig. 8. The y-axis of the HD boxplot is set to a range of 0 to 30 mm for clearer visualization. It is evident that incorporating segmentation supervision improves the DSC by 1.3% for the whole tumor, 1.0% for the enhancing tumor, and 0.7% on average. Furthermore, in terms of HD, the improvements are significant, with a 24.0% improvement for the whole tumor and a 9.8% improvement for the enhancing tumor. These results confirm the value of segmentation supervision in enhancing brain tumor segmentation accuracy. However, there is a slight decrease (from 83.6% to 83.5%) observed for the tumor core region in terms of DSC, and there is a major increase for the tumor core (from 6.3 mm to 9.6 mm) and a minor increase (from 6.9 mm to 7.1 mm) for the average result, suggesting that while the segmentation supervision contributes positively to the most part of brain tumor, it may require further refinement to optimize performance specifically for the tumor core. To address this problem, the future research directions are discussed in Section 6.6.

6.2. Analysis on the effectiveness of different boundary extraction methods

To evaluate the impact of the different boundary extraction methods on the segmentation results, an ablation experiment is conducted on the BraTS 2018 dataset. The results are summarized in Table 8 and Fig. 9. The y-axis of the HD boxplot is set to a range of 0 to 30 mm for clearer visualization. The results demonstrate that utilizing only the coarse boundary extraction method yields an average DSC of 82.6% and an average HD of 5.6 mm. When applying the Sobel filter alone for boundary extraction, the model achieves DSC and HD values of 82.8% and 4.6 mm, respectively. Notably, combining both methods leads to significant performance improvements across all tumor regions. For the whole tumor region, the DSC increases from 85.7% to 86.3%, and the HD significantly decreases from 6.9 mm to 4.7 mm. For the tumor core region, the DSC slightly improves from 84.9% to 85.3%, while the HD sees a more substantial reduction from 6.7 mm to 5.1 mm. For the enhancing tumor region, the DSC remains relatively stable (77.2% to 77.4%), but the HD remains at 3.0 mm, indicating no further improvement in boundary accuracy for this region. Across all tumor regions, the average performance improves, with the average DSC increasing from 82.6% to 83.0%, and the average HD significantly decreasing from 5.6 mm to 4.3 mm. Overall, the refined boundary extraction method proves to enhance segmentation performance, especially in the whole tumor and tumor core regions, leading to more accurate segmentation result.

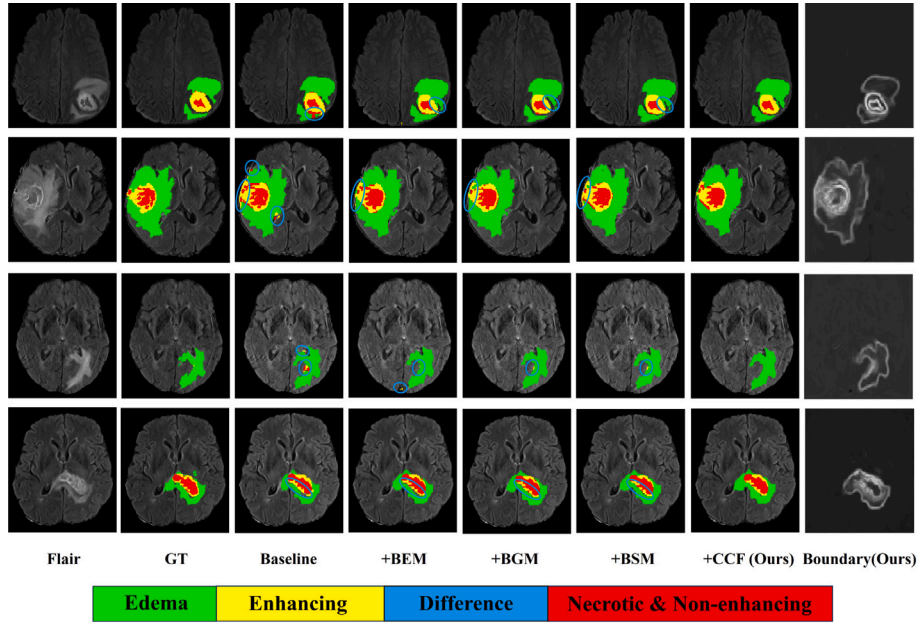


Fig. 6. Visualization of segmentation results. The first column presents the Flair image, while the second column illustrates the ground truth segmentation. The subsequent columns depict segmentation results from the baseline method, followed by the incremental improvements achieved by integrating each module step-by-step: the second column shows the results with the addition of BEM to the baseline, the third column with BEM and BGM, the fourth column with BEM, BGM, and BSM, and the last second column shows the results with all modules combined, including CCF. The last column shows the boundary map achieved using the proposed method. Each row corresponds to different cases, with segmentation regions color-coded as follows: edema (green), enhancing tumor (yellow), and necrotic&non-enhancing core (red).

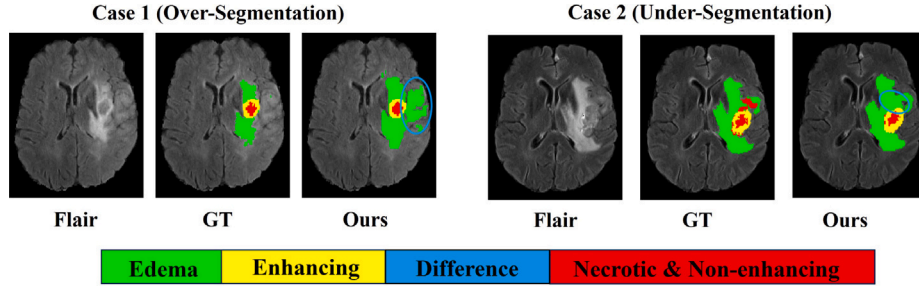


Fig. 7. Examples of failure segmentation cases, illustrating over-segmentation and under-segmentation errors.

Table 7

Comparison results for the effectiveness of the segmentation supervision on the BraTS 2018 dataset. Bold values indicate the highest scores achieved.

Methods	Dice similarity coefficient (%) $\uparrow$				Hausdorff distance (mm) $\downarrow$			
	Whole	Core	Enh.	Mean	Whole	Core	Enh.	Mean
w/o segmentation supervision	84.3	83.6	76.7	81.5	10.4	6.3	4.1	6.9
w segmentation supervision	85.4	83.5	77.5	82.1	7.9	9.6	3.7	7.1

Table 8

Comparison results for the effectiveness of different boundary extraction methods on the BraTS 2018 dataset. Bold values indicate the highest scores achieved.

Methods	Dice similarity coefficient (%) $\uparrow$				Hausdorff distance (mm) $\downarrow$			
	Whole	Core	Enh.	Mean	Whole	Core	Enh.	Mean
Coarse boundary extraction	85.7	84.9	77.2	82.6	6.9	6.7	3.0	5.6
Boundary extraction by Sobel filter	85.7	85.7	77.0	82.8	6.3	4.5	3.1	4.6
Refined boundary extraction	86.3	85.3	77.4	83.0	4.7	5.1	3.0	4.3

6.3. Analysis on the effectiveness of different boundary loss functions

In this section, I analyze the effectiveness of different boundary loss functions employed in the proposed method, focusing on L_1 and L_2 loss functions. The results are shown in Table 9 and Fig. 10. The y-axis of the HD boxplot is set to a range of 0 to 30 mm for clearer visualization. We can observe varying improvements across the whole tumor, tumor core, and enhancing tumor regions when comparing L_1

and L_2 losses. For the whole tumor, the DSC improves from 86.7% to 87.0%, with the HD reducing from 5.8 mm to 5.3 mm, indicating better overall segmentation and boundary accuracy. The core region shows a decrease in DSC (86.6% to 86.0%), but a notable improvement in HD, which drops from 4.1 mm to 3.7 mm. The most significant improvement is seen in the enhancing tumor region, where the DSC rises from 78.1% to 79.2%, and the HD decreases from 2.6 mm to 2.4 mm. And the average DSC increases from 83.8% to 84.1%, and the

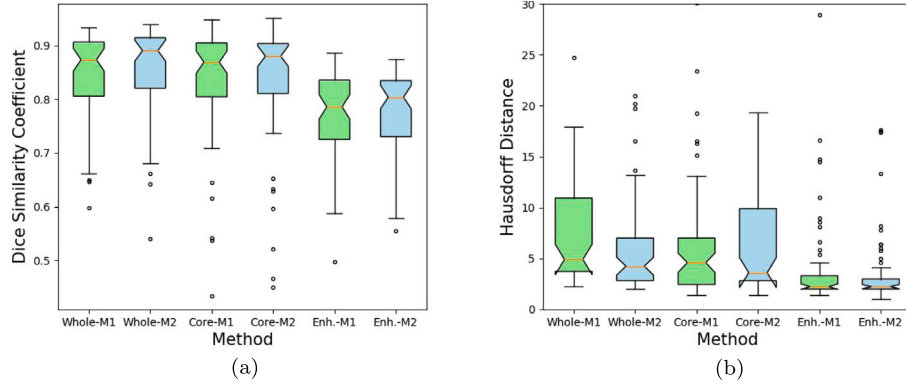


Fig. 8. Boxplots illustrating the Dice Similarity Coefficients and Hausdorff Distance across three tumor regions for various methods. Subfigure (a) displays the Dice Similarity Coefficients, while subfigure (b) presents the Hausdorff Distance. Method 1 which does not utilize segmentation supervision is indicated in light green, whereas Method 2 which incorporates segmentation supervision is shown in light blue.

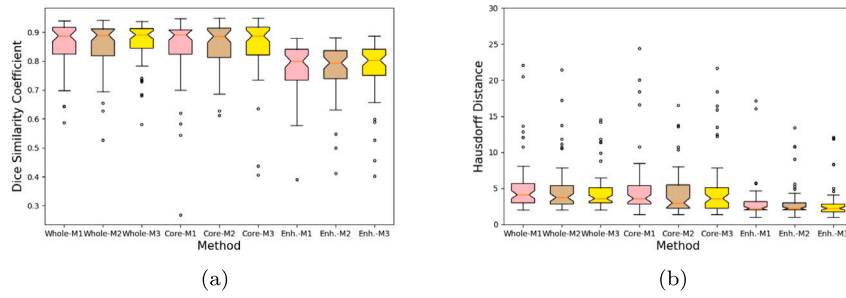


Fig. 9. Boxplots illustrating the Dice Similarity Coefficients and Hausdorff Distance across three tumor regions for various methods. Subfigure (a) displays the Dice Similarity Coefficients, while subfigure (b) presents the Hausdorff Distance. Method 1 which utilizes coarse boundary extraction is indicated in light pink, whereas Method 2 which utilizes only Sobel filter is shown in tan, Method 3 which utilizes refined boundary extraction (coarse&Sobel) is indicated in yellow.

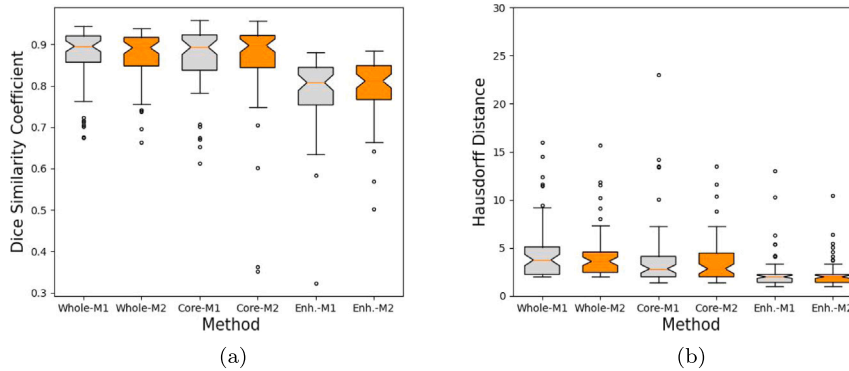


Fig. 10. Boxplots illustrating the Dice Similarity Coefficients and Hausdorff Distance across three tumor regions for various methods. Subfigure (a) displays the Dice Similarity Coefficients, while subfigure (b) presents the Hausdorff Distance. Method 1 which utilizes L_1 is indicated in light gray, whereas Method 2 which utilizes L_2 is shown in orange.

Table 9

Comparison results for different boundary loss functions on the BraTS 2018 dataset. Bold values indicate the highest scores achieved.

Methods	Dice similarity coefficient (%) $\uparrow$				Hausdorff distance (mm) $\downarrow$			
	Whole	Core	Enh.	Mean	Whole	Core	Enh.	Mean
L_1	86.7	86.6	78.1	83.8	5.8	4.1	2.6	4.2
L_2	87.0	86.0	79.2	84.1	5.3	3.7	2.4	3.8

average HD significantly decreases from 4.2 mm to 3.8 mm. Overall, L_2 loss consistently outperforms L_1 loss. The superior performance of L_2 loss can be attributed to its sensitivity to larger discrepancies between predicted and actual boundaries. This sensitivity allows the model to adjust more effectively during training, resulting in a more precise brain tumor segmentation.

6.4. Analysis on the effectiveness of integrating boundary information in different levels

In this section, we analyze the effect of integrating boundary information across various levels of the segmentation network, with results presented in Table 10 and Fig. 11. The y-axis of the HD boxplot is set to a range of 0 to 30 mm for clearer visualization. From Table 10, we can observe that integrating boundary information up to the 3rd level yields a mean DSC of 84.1% and a mean HD of 3.8 mm. Extending the integration to the 4th level results in a large decrease in mean DSC to 83.4% and in mean HD, which increases to 4.7 mm. Specifically, for the whole tumor, the DSC drops from 87.0% to 86.0%, with the HD increasing from 5.3 mm to 6.4 mm. For the tumor core, the DSC slightly increases from 86.0% to 86.1%, while the HD increases from 3.7 mm to 4.6 mm. For the enhancing tumor, the DSC decreases from

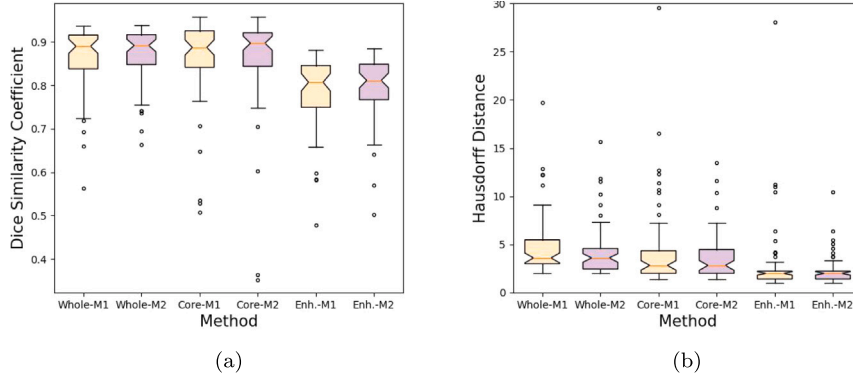


Fig. 11. Boxplots illustrating the Dice Similarity Coefficients and Hausdorff Distance across three tumor regions for various methods. Subfigure (a) displays the Dice Similarity Coefficients, while subfigure (b) presents the Hausdorff Distance. Method 1 which integrates boundary information up to the 3rd level is indicated in bisque, whereas Method 2 which integrates boundary information up to the 4th level is shown in thistle.

Table 10

Analysis on the effectiveness of integrating boundary information in different levels. Bold values indicate the highest scores achieved.

Methods	Dice similarity coefficient (%) $\uparrow$				Hausdorff distance (mm) $\downarrow$			
	Whole	Core	Enh.	Mean	Whole	Core	Enh.	Mean
1st ~ 3rd	87.0	86.0	79.2	84.1	5.3	3.7	2.4	3.8
1st ~ 4th	86.0	86.1	78.2	83.4	6.4	4.6	3.1	4.7

Table 11

Analysis on the hyperparameter λ in the loss function. Bold values indicate the highest scores achieved.

λ	Dice similarity coefficient (%) $\uparrow$				Hausdorff distance (mm) $\downarrow$			
	Whole	Core	Enh.	Mean	Whole	Core	Enh.	Mean
0.1	87.0	86.0	79.2	84.1	5.3	3.7	2.4	3.8
0.3	86.9	86.0	79.1	84.0	4.4	4.1	2.5	3.7
0.5	86.4	85.7	77.6	83.2	4.7	3.9	2.6	3.7
0.7	86.8	86.1	78.9	83.9	4.6	3.9	2.6	3.7
0.9	86.9	85.0	78.4	83.4	4.3	4.6	3.0	4.0
1.0	86.6	86.4	78.0	83.7	4.3	3.9	2.7	3.6

79.2% to 78.2%, and the HD rises from 2.4 mm to 3.1 mm. These results suggest that boundary integration up to the 3rd level achieves a better segmentation performance. We explain that at the 3rd level, the feature map resolution ($32 \times 32 \times 32$) retains sufficient spatial detail to capture fine tumor boundaries. Extending integration to the 4th level reduces the resolution to $16 \times 16 \times 16$, which limits the network's ability to accurately delineate small and intricate structures. Therefore, the proposed network applies the boundary information up to the 3rd level.

6.5. Hyperparameter tuning for λ in the loss function

To optimize the performance of the proposed model, I conducted a grid search to tune the hyperparameter λ used in the loss function. The values for λ are selected based on a range of commonly used coefficients to balance the contribution of different loss terms. The values explored during the grid search are 0.1, 0.3, 0.5, 0.7, 0.9, 1.0. Through the experiment, as detailed in Table 11, it is evident that $\lambda = 0.1$ achieves the highest average DSC of 84.1% and excels in the whole tumor region (87.0%) and enhancing tumor region (79.2%). For HD, it achieves 3.7 mm in tumor core region and 2.4 mm in enhancing tumor region, while maintaining a competitive average HD score of 3.8 mm. Meanwhile, $\lambda = 1.0$ yields the best DSC of 86.4% in tumor core region and best average HD of 3.6 mm, particularly excelling in the whole tumor region (4.3 mm). However, increasing λ leads to a decline in DSC scores, especially in the enhancing tumor region. This analysis suggests that $\lambda = 0.1$ provides the most balanced performance

across all metrics and regions, making it the optimal choice for this work.

6.6. Limitations and future work

While the proposed method demonstrates overall effectiveness in brain tumor segmentation, certain limitations must be addressed to enhance performance further. A notable issue is the observed decrease in segmentation accuracy for the tumor core region when incorporating segmentation supervision in Section 6.1. To mitigate this, future research will implement a region-specific analysis that leverage the unique characteristics of each tumor sub-region. This approach involves designing a more tailored segmentation loss function that assigns specific emphasis to each region, potentially leading to improved segmentation results and more balanced accuracy across all tumor areas. Additionally, I plan to explore advanced techniques such as attention mechanisms, which can enhance the model's focus on critical tumor features. Furthermore, implementing explainable networks may provide insights into the model's decision-making process, allowing us to better understand the factors affecting segmentation accuracy, such as the observed slight decrease in DSC, this understanding could help address limitations like over-segmentation and under-segmentation in the proposed method. Besides, incorporating semi-supervised or self-supervised learning [53,54] as a complementary strategy to further enhance segmentation performance is a promising direction for future work. Lastly, while the proposed method demonstrates good performance on the selected datasets, addressing potential challenges in its application to diverse clinical settings is crucial. Variations in image quality, MRI sequences, patient characteristics, and tumor manifestations may affect its generalizability. Strategies such as integrating domain adaptation techniques, expanding training datasets with diverse clinical cases, and employing transfer learning can significantly enhance its robustness and adaptability to varying clinical environments. Overall, these directions aim to improve both the accuracy and explainability of the segmentation approach, ultimately leading to more reliable segmentation results in brain tumor analysis.

7. Conclusion

In this paper, a novel boundary-aware multi-modal brain tumor segmentation network is designed to address the challenges posed by complex tumor boundaries and heterogeneous features. The method introduces four key components: the Boundary Extraction Module (BEM), Boundary Guidance Module (BGM), Boundary Supervision Module (BSM), and Cross-Feature Fusion (CFF), all of which collectively enhance segmentation performance in multi-modal MRI images. Additionally, the experimental results demonstrate that segmentation supervision substantially improves accuracy, while the use of Sobel

filters further refines the model's performance. Experimental results also reveal that the L_2 loss function outperforms the L_1 loss function in this context. Overall, these findings confirm the effectiveness of the proposed model, showing its superiority over existing algorithms in brain tumor segmentation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China (No. 62206084).

Data availability

Data will be made available on request.

References

- [1] D.N. Louis, A. Perry, P. Wesseling, D.J. Brat, I.A. Cree, D. Figarella-Branger, C. Hawkins, H. Ng, S.M. Pfister, G. Reifenberger, et al., The 2021 WHO classification of tumors of the central nervous system: a summary, *Neuro-Oncol.* 23 (8) (2021) 1231–1251.
- [2] H.A. Wanis, H. Möller, K. Ashkan, E.A. Davies, The incidence of major subtypes of primary brain tumors in adults in England 1995–2017, *Neuro-Oncol.* 23 (8) (2021) 1371–1382.
- [3] M.J. van den Bent, M. Geurts, P.J. French, M. Smits, D. Capper, J.E. Bromberg, S.M. Chang, Primary brain tumours in adults, *Lancet* 402 (10412) (2023) 1564–1579.
- [4] Z. Liu, L. Tong, L. Chen, Z. Jiang, F. Zhou, Q. Zhang, X. Zhang, Y. Jin, H. Zhou, Deep learning based brain tumor segmentation: a survey, *Complex Intell. Syst.* 9 (1) (2023) 1001–1026.
- [5] Z. Zhu, K. Yu, G. Qi, B. Cong, Y. Li, Z. Li, X. Gao, Lightweight medical image segmentation network with multi-scale feature-guided fusion, *Comput. Biol. Med.* 182 (2024) 109204.
- [6] Y. Ding, X. Yu, Y. Yang, RFNet: Region-aware fusion network for incomplete multi-modal brain tumor segmentation, in: *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 3975–3984.
- [7] T. Zhou, S. Ruan, S. Canu, A review: Deep learning for medical image segmentation using multi-modality fusion, *Array* (2019) 100004.
- [8] H. Yang, T. Zhou, Y. Zhou, Y. Zhang, H. Fu, Flexible fusion network for multi-modal brain tumor segmentation, *IEEE J. Biomed. Heal. Inform.* 27 (7) (2023) 3349–3359.
- [9] P. Jyothi, A.R. Singh, Deep learning models and traditional automated techniques for brain tumor segmentation in MRI: a review, *Artif. Intell. Rev.* 56 (4) (2023) 2923–2969.
- [10] Y. Feng, Y. Cao, D. An, P. Liu, X. Liao, B. Yu, DAUnet: A U-shaped network combining deep supervision and attention for brain tumor segmentation, *Knowl.-Based Syst.* 285 (2024) 111348.
- [11] T. Zhou, Multi-modal brain tumor segmentation via disentangled representation learning and region-aware contrastive learning, *Pattern Recognit.* (2024) 110282.
- [12] W. Li, W. Huang, Y. Zheng, CorrDiff: Corrective diffusion model for accurate MRI brain tumor segmentation, *IEEE J. Biomed. Heal. Inform.* (2024).
- [13] T. Zhou, Feature fusion and latent feature learning guided brain tumor segmentation and missing modality recovery network, *Pattern Recognit.* 141 (2023) 109665.
- [14] N.S.M. Raja, S.L. Fernandes, N. Dey, S.C. Satapathy, V. Rajinikanth, Contrast enhanced medical MRI evaluation using Tsallis entropy and region growing segmentation, *J. Ambient. Intell. Humaniz. Comput.* (2024) 1–12.
- [15] A.R. Khan, S. Khan, M. Harouni, R. Abbasi, S. Iqbal, Z. Mehmood, Brain tumor segmentation using K-means clustering and deep learning with synthetic data augmentation for classification, *Microsc. Res. Tech.* 84 (7) (2021) 1389–1399.
- [16] R. Vankdothu, M.A. Hameed, Brain tumor segmentation of MR images using SVM and fuzzy classifier in machine learning, *Meas.: Sens.* 24 (2022) 100440.
- [17] Y. Chen, N. Cheng, M. Cai, C. Cao, J. Yang, Z. Zhang, A spatially constrained asymmetric Gaussian mixture model for image segmentation, *Inform. Sci.* 575 (2021) 41–65.
- [18] J. Wu, X. Tang, Brain segmentation based on multi-atlas and diffeomorphism guided 3D fully convolutional network ensembles, *Pattern Recognit.* 115 (2021) 107904.
- [19] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2015, pp. 234–241.
- [20] F. Isensee, P.F. Jäger, P.M. Full, P. Vollmuth, K.H. Maier-Hein, nnU-Net for brain tumor segmentation, in: *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 6th International Workshop, BrainLes 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 4, 2020, Revised Selected Papers, Part II* 6, Springer, 2021, pp. 118–132.
- [21] H.M. Luu, S.-H. Park, Extending nn-U-Net for brain tumor segmentation, in: *International MICCAI Brainlesion Workshop*, Springer, 2021, pp. 173–186.
- [22] T. Zhou, M2GCNet: Multi-modal graph convolution network for precise brain tumor segmentation across multiple MRI sequences, *IEEE Trans. Image Process.* (2024).
- [23] Q. Ma, S. Zhou, C. Li, F. Liu, Y. Liu, M. Hou, Y. Zhang, DGRUnit: Dual graph reasoning unit for brain tumor segmentation, *Comput. Biol. Med.* 149 (2022) 106079.
- [24] A. Vaswani, Attention is all you need, *Adv. Neural Inf. Process. Syst.* (2017).
- [25] J. Chen, Y. Lu, Q. Yu, X. Luo, E. Adeli, Y. Wang, L. Lu, A.L. Yuille, Y. Zhou, Transunet: Transformers make strong encoders for medical image segmentation, 2021, arXiv preprint [arXiv:2102.04306](https://arxiv.org/abs/2102.04306).
- [26] J. Chen, J. Mei, X. Li, Y. Lu, Q. Yu, Q. Wei, X. Luo, Y. Xie, E. Adeli, Y. Wang, et al., TransUNet: Rethinking the U-Net architecture design for medical image segmentation through the lens of transformers, *Med. Image Anal.* 97 (2024) 103280.
- [27] A. Lin, B. Chen, J. Xu, Z. Zhang, G. Lu, D. Zhang, Ds-transunet: Dual swin transformer u-net for medical image segmentation, *IEEE Trans. Instrum. Meas.* 71 (2022) 1–15.
- [28] Y. Xu, K. Yu, G. Qi, Y. Gong, X. Qu, L. Yin, P. Yang, Brain tumour segmentation framework with deep nuanced reasoning and Swin-T, *IET Image Process.* 18 (6) (2024) 1550–1564.
- [29] R. Wang, S. Chen, C. Ji, J. Fan, Y. Li, Boundary-aware context neural network for medical image segmentation, *Med. Image Anal.* 78 (2022) 102395.
- [30] Z. Zhu, Z. Wang, G. Qi, N. Mazur, P. Yang, Y. Liu, Brain tumor segmentation in MRI with multi-modality spatial information enhancement and boundary shape correction, *Pattern Recognit.* 153 (2024) 110553.
- [31] C. Tang, H. Chen, X. Li, J. Li, Z. Zhang, X. Hu, Look closer to segment better: Boundary patch refinement for instance segmentation, in: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 13926–13935.
- [32] Z. Zhu, M. Sun, G. Qi, Y. Li, X. Gao, Y. Liu, Sparse dynamic volume TransUNet with multi-level edge fusion for brain tumor segmentation, *Comput. Biol. Med.* (2024) 108284.
- [33] Z. Huang, Y. Sui, Contour-weighted loss for class-imbalanced image segmentation, in: *2024 IEEE International Conference on Image Processing, ICIP, IEEE*, 2024, pp. 3084–3090.
- [34] S. Borse, Y. Wang, Y. Zhang, F. Porikli, Inverseform: A loss function for structured boundary-aware segmentation, in: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 5901–5911.
- [35] A. Li, L. Jiao, H. Zhu, L. Li, F. Liu, Multitask semantic boundary awareness network for remote sensing image segmentation, *IEEE Trans. Geosci. Remote Sens.* 60 (2021) 1–14.
- [36] Y. Wang, W. Ding, R. Zhang, H. Li, Boundary-aware multitask learning for remote sensing imagery, *IEEE J. Sel. Top. Appl. Earth Obs. Remote. Sens.* 14 (2020) 951–963.
- [37] Z. Zhu, X. He, G. Qi, Y. Li, B. Cong, Y. Liu, Brain tumor segmentation based on the fusion of deep semantics and edge information in multimodal MRI, *Inf. Fusion* 91 (2023) 376–387.
- [38] T. Zhou, S. Ruan, P. Vera, S. Canu, A Tri-Attention fusion guided multi-modal segmentation network, *Pattern Recognit.* 124 (2022) 108417.
- [39] N.M. Abolenein, P. Songhao, A. Koubaa, A. Noor, A. Afifi, HTTU-Net: Hybrid two track U-net for automatic brain tumor segmentation, *IEEE Access* 8 (2020) 101406–101415.
- [40] T. Zhou, S. Canu, P. Vera, S. Ruan, Latent correlation representation learning for brain tumor segmentation with missing MRI modalities, *IEEE Trans. Image Process.* 30 (2021) 4263–4274.
- [41] W. Wang, C. Chen, M. Ding, H. Yu, S. Zha, J. Li, Transbts: Multimodal brain tumor segmentation using transformer, in: *Medical Image Computing and Computer Assisted Intervention-MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part I* 24, Springer, 2021, pp. 109–119.
- [42] A. Hatamizadeh, V. Nath, Y. Tang, D. Yang, H.R. Roth, D. Xu, Swin unetr: Swin transformers for semantic segmentation of brain tumors in mri images, in: *International MICCAI Brainlesion Workshop*, Springer, 2021, pp. 272–284.
- [43] Z. Xing, L. Yu, L. Wan, T. Han, L. Zhu, NestedFormer: Nested modality-aware transformer for brain tumor segmentation, in: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2022, pp. 140–150.
- [44] A. Jungo, F. Balsiger, M. Reyes, Analyzing the quality and challenges of uncertainty estimations for brain tumor segmentation, *Front. Neurosci.* 14 (2020) 501743.

- [45] J. Zhang, Z. Jiang, J. Dong, Y. Hou, B. Liu, Attention gate resU-Net for automatic MRI brain tumor segmentation, *IEEE Access* 8 (2020) 58533–58545.
- [46] M.U. Rehman, J. Ryu, I.F. Nizami, K.T. Chong, RAAGR2-Net: A brain tumor segmentation network using parallel processing of multiple spatial frames, *Comput. Biol. Med.* 152 (2023) 106426.
- [47] N. Sheng, D. Liu, J. Zhang, C. Che, J. Zhang, Second-order ResU-Net for automatic MRI brain tumor segmentation, *Math. Biosci. Eng.* 18 (5) (2021) 4943–4960.
- [48] T. Zhou, S. Canu, P. Vera, S. Ruan, A dual supervision guided attentional network for multimodal MR brain tumor segmentation, in: *International Conference on Medical Imaging and Computer-Aided Diagnosis*, Springer, 2021, pp. 3–11.
- [49] M. Pan, Y. Shi, Z. Song, Segmentation of gliomas based on a double-pathway residual convolution neural network using multi-modality information, *J. Med. Imaging Heal. Inform.* 10 (11) (2020) 2784–2794.
- [50] Y. Jiang, Y. Zhang, X. Lin, J. Dong, T. Cheng, J. Liang, SwinBTS: A method for 3D multimodal brain tumor segmentation using swin transformer, *Brain Sci.* 12 (6) (2022) 797.
- [51] O. Oktay, J. Schlemper, L.L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N.Y. Hammerla, B. Kainz, et al., Attention u-net: Learning where to look for the pancreas, 2018, arXiv preprint [arXiv:1804.03999](https://arxiv.org/abs/1804.03999).
- [52] H. Peiris, M. Hayat, Z. Chen, G. Egan, M. Harandi, A robust volumetric transformer for accurate 3D tumor segmentation, in: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2022, pp. 162–172.
- [53] Y. Ye, J. Zhang, Z. Chen, Y. Xia, CADs: A self-supervised learner via cross-modal alignment and deep self-distillation for CT volume segmentation, *IEEE Trans. Med. Imaging* (2024).
- [54] Q. Zeng, Y. Xie, Z. Lu, M. Lu, J. Zhang, Y. Zhou, Y. Xia, Consistency-guided differential decoding for enhancing semi-supervised medical image segmentation, *IEEE Trans. Med. Imaging* (2024).

Tongxue Zhou received the Ph.D. degree in Computer Science from INSA Rouen, France, in 2022. She is currently an Associate Professor in the School of Information Science and Technology at Hangzhou Normal University, China. Her current research interests include medical image analysis, data fusion and deep learning.