Information Fusion 129 (2026) 104085

Contents lists available at ScienceDirect

Information Fusion

journal homepage: www.elsevier.com/locate/inffus

Full Length Article

A reliable framework for brain tumor segmentation via multi-modal fusion and uncertainty modeling

Tongxue Zhou [a,*], Mingyang Li [a], Su Ruan [b], Tingjin Luo [c,*], Bingbing Jiang [a], Junan Zhu [d], Ping Ma [d], Defu Yang [a], Guang Yang [e,f,g,h,*]

[a] *School of Information Science and Technology, Hangzhou Normal University, Hangzhou, 311121, Zhejiang, China*
[b] *Laboratoire AIMS, Université de Rouen Normandie, Rouen, 76183, France*
[c] *College of Science, National University of Defense Technology, Changsha, 410073, Hunan, China*
[d] *School of Internet, Anhui University, Hefei, 230039, Anhui, China*
[e] *Bioengineering Department and Imperial-X, Imperial College London, London, W12 7SL, UK*
[f] *National Heart and Lung Institute, Imperial College London, London, SW7 2AZ, UK*
[g] *Cardiovascular Research Centre, Royal Brompton Hospital, London, SW3 6NP, UK*
[h] *School of Biomedical Engineering and Imaging Sciences, King's College London, London, WC2R 2LS, UK*

ARTICLE INFO

Keywords:
Brain tumor segmentation
Uncertainty quantification
Multi-modality fusion
Deep learning
MRI

ABSTRACT

Accurate brain tumor segmentation from MRI scans is critical for effective diagnosis and treatment planning. Recent advances in deep learning have significantly improved brain tumor segmentation performance. However, these models still face challenges in clinical adoption due to their inherent uncertainties and potential for errors. In this paper, we propose a novel MR brain tumor segmentation approach that integrates multi-modal data fusion and uncertainty quantification to improve the accuracy and reliability of brain tumor segmentation. Recognizing that each MR modality contributes unique insights into the tumor's characteristics, we propose a novel modality-aware guidance by explicitly categorizing the modalities into "teacher" (FLAIR and T1c) and "student" (T2 and T1) groups. Since the teacher modalities are the most informative modalities for identifying brain tumors, we propose a multi-modal teacher-student fusion strategy. This strategy leverages the teacher modalities to guide the student modalities in both spatial and channel feature representation aspects. To address prediction reliability, we employ Monte Carlo dropout during training to generate multiple uncertainty estimates. Additionally, we develop a novel uncertainty-aware loss function that optimizes segmentation accuracy while quantifying the uncertainty in predictions. Experimental results conducted on three BraTS datasets demonstrate the effectiveness of the proposed components and the superior performance compared to the state-of-the-art methods, highlighting their potential for clinical application.

1. Introduction

A brain tumor refers to an abnormal mass or collection of cells that develop within the brain, these tumors can either be malignant or benign. According to the World Health Organization (WHO), gliomas are classified into four grades based on microscopic images and tumor behaviours. Grades I and II are categorized as Low-Grade Gliomas (LGGs), which exhibit characteristics similar to benign tumors with a slow growth rate. In contrast, Grades III and IV are designated as High-Grade Gliomas (HGGs), characterized by aggressive and cancerous behaviour [1,2]. Magnetic Resonance Imaging (MRI) has emerged as a gold standard in the diagnosis and evaluation of brain tumors. With its superior soft tissue contrast and multi-planar imaging capabilities, MRI provides detailed anatomical information essential for accurate tumor localization, characterization, and assessment of surrounding structures [3]. Commonly used MR modalities include fluid attenuation inversion recovery (FLAIR), T1-weighted (T1), contrast-enhanced T1-weighted (T1c) and T2-weighted (T2), as depicted in Fig. 1. It can be observed that each modality highlights different tissue properties and can reveal distinct aspects of brain anatomy and pathology. For instance, FLAIR and T2 are particularly useful for detecting lesions in regions with high water content, such as edema. T1c and T1 are typically used

* Corresponding authors.
E-mail addresses: zhoutongxue1992@163.com (T. Zhou), limingyang@stu.hznu.edu.cn (M. Li), su.ruan@univ-rouen.fr (S. Ruan), tingjinluo@hotmail.com (T. Luo), jiangbb@hznu.edu.cn (B. Jiang), zhujunan@ahu.edu.cn (J. Zhu), maping@ahu.edu.cn (P. Ma), dfyang@hznu.edu.cn (D. Yang), g.yang@imperial.ac.uk (G. Yang).

https://doi.org/10.1016/j.inffus.2025.104085
Received 6 September 2025; Received in revised form 13 November 2025; Accepted 18 December 2025
Available online 20 December 2025

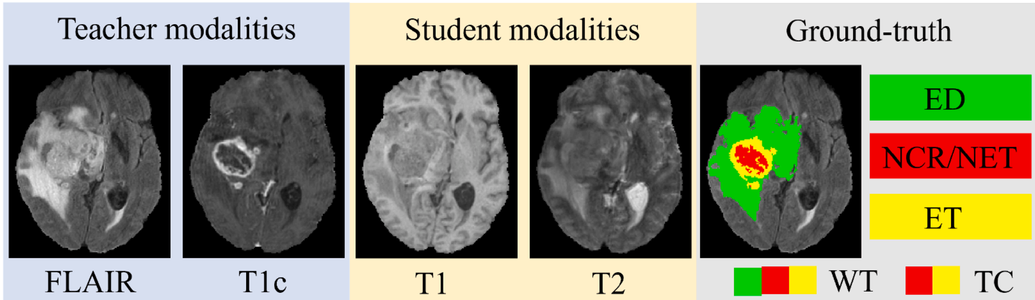

Fig. 1. The multi-modal MR modalities include FLAIR, T1c, T2, and T1, categorized into teacher (FLAIR, T1c) and student (T2, T1) modalities. Ground truth (GT) annotations are: edema (ED, green), necrotic and non-enhancing core (NCR&NET, red), and enhancing tumor (ET, yellow).

for identifying tumor core, such as enhancing tumor, necrotic and non-enhancing tumor. In addition, FLAIR and T1c are the most informative modalities for identifying brain tumor, providing a comprehensive view that combines the advantages of both high water content sensitivity and enhanced anatomical detail. Most existing teacher-student frameworks in medical imaging primarily address domain adaptation (e.g., transferring knowledge from labeled to unlabeled data) or model compression (e.g., distilling large models into smaller ones). However, these frameworks do not fully exploit the hierarchical diagnostic relationships among MRI modalities. Therefore, our method introduces a novel modality-aware guidance by explicitly categorizing MRI modalities into teacher-student groups based on their diagnostic value, where teacher modalities (FLAIR and T1c) guide student modalities (T2 and T1) at both spatial and channel levels. By combining information from these modalities, clinicians can obtain a comprehensive assessment of brain structure and pathology, aiding in diagnosis, treatment planning, and monitoring of disease progression [4,5].

Accurate segmentation of brain tumors is a critical task in medical image analysis, essential for diagnosis, treatment planning, and monitoring of patient progress. Traditional methods for brain tumor segmentation often rely on manual delineation by radiologists, which is not only time-consuming but also prone to inter-observer variability. Recent advances in deep learning have significantly improved the automation and accuracy of segmentation tasks [6–8]. However, these models still face challenges in clinical adoption due to their inherent uncertainties and potential for errors. In this context, leveraging uncertainty quantification becomes crucial, it can provide a quantifiable measure of confidence in the model's predictions, allowing clinicians to better assess the reliability of automated segmentation results [9]. For example, in critical treatment planning steps such as radiotherapy dose mapping or surgical margin definition, understanding the uncertainty associated with the tumor boundary can help clinicians avoid under-treatment or over-treatment. Moreover, in follow-up scenarios, tracking changes in uncertain regions over time can support early identification of recurrence or treatment response.

To this end, this paper explores the integration of uncertainty quantification into a brain tumor segmentation network. Specifically, we employ Monte Carlo (MC) dropout during training to generate uncertainty estimates. Moreover, we introduce a novel uncertainty-aware loss function tailored to this specific task. By minimizing prediction errors while simultaneously quantifying uncertainty, the proposed method aims to produce both segmentation results and uncertainty quantification. These outputs can be visualized and reviewed by radiologists to support decision-making, especially in borderline or ambiguous cases. Additionally, we propose a multi-modal teacher-student fusion strategy to leverage complementary information across multiple MR modalities. This strategy enables a more informed feature alignment across modalities, unlike conventional fusion strategies (e.g., concatenation, attention) that lack explicit modality hierarchies. By leveraging this modality-aware supervision, our approach better mimics how radiologists correlate sequences to reach a comprehensive diagnostic understanding. By modeling inter-modality dependencies more effectively, the network can produce clinically consistent and structurally coherent tumor segmentation. The contributions of this work are summarized as follows:

(1) A novel deep learning-based neural network is proposed for segmenting brain tumors using multiple MR modalities. By leveraging the complementary information from sequences such as T1, T1c, T2, and FLAIR, the model achieves more accurate and anatomically consistent segmentation, supporting critical clinical tasks like diagnosis, surgical planning, and radiotherapy targeting.

(2) A multi-modal teacher-student fusion strategy is proposed to exploit the complementary information across different MR modalities. This design enables more effective cross-modality feature alignment, enhancing the model's ability to mimic the way radiologists integrate multi-sequence MRI data for comprehensive tumor assessment.

(3) A novel uncertainty-aware loss function is proposed to not only optimize segmentation accuracy but also quantify voxel-wise predictive uncertainty, providing valuable insights into the reliability of segmentation results. These uncertainty maps offer clinicians interpretable confidence estimates, which can guide further manual inspection, assist in risk-aware decision-making, and improve trust in automated tools.

This paper is structured as follows: Section 2 comprehensively reviews current methodologies in brain tumor segmentation and uncertainty quantification. Section 3 details our proposed segmentation network. Section 4 describes the experimental setup. Section 5 provides a thorough analysis of the experimental results. Section 6 discusses the proposed method with the limitations and future work. Section 7 concludes this paper.

2. Related works

2.1. Brain tumor segmentation

Brain tumor segmentation plays a critical role in medical image analysis, aiding diagnosis and treatment planning. Deep learning methods have gained prominence due to their ability to learn complex patterns from multi-modal imaging data, enhancing segmentation accuracy and efficiency. The U-Net [10] architecture pioneered this field, providing an effective framework with encoder-decoder architecture and skip connections for precise localization and contextual preservation. Various adaptations have emerged to address specific challenges in brain tumor segmentation. For instance, Allah et al. [11] proposed an edge U-Net to enhance tumor localization by integrating boundary-related MRI data with main MRI data. Zhang et al. [12] proposed to combine transformer and U-Net for spatial dependency capture and computational efficiency. Ma et al. [13] proposed a dual graph reasoning unit, consisting of a spatial reasoning module and a channel reasoning module, to refine brain tumor segmentation. Zhou et al. [7] proposed a multi-modal brain tumor segmentation network via disentangled representation learning and region-aware contrastive learning for multi-modal brain tumor segmentation. Zhou et al. [14] proposed a boundary-aware multi-modal brain tumor segmentation network that integrates boundary-specific modules to capture, refine, and enhance complex tumor boundaries. Additionally, a cross-feature fusion mechanism is employed to effectively exploit complementary information from multiple MRI modalities. Similarly, Zhu et al. [15] proposed a novel framework that integrates multi-module information enhancement and boundary shape correction within a deep convolutional network, demonstrating its effectiveness in diverse medical image segmentation tasks. Li et al. [16] proposed a lightweight Vision Mamba U-Net that incorporates a rotation-enhanced Mamba module and a correlated feature fusion mechanism. This design effectively addresses the Mamba module's limitations in capturing fine-grained local details while simultaneously reducing computational costs. Zhu et al. [17] proposed a probability map-guided network for 3D volumetric medical image segmentation, which leverages probability-based supervision, adaptive reconstruction, and dynamic feature fusion to achieve superior segmentation accuracy and robustness.

In addition, attention mechanisms have emerged as powerful tools for improving brain tumor segmentation, such as spatial attention [18], channel attention [19] and hybrid attention [20]. These mechanisms enable the network to dynamically adjust the importance of different parts of the input image, thereby improving segmentation accuracy and robustness. For example, Zhou et al. [21] proposed a cross-task guided attention module to learn category-specific channel attention for brain tumor segmentation. Ghazouani et al. [22] proposed to include channel squeeze and spatial excitation blocks for capturing more informative features from both spatially and channel-wise for brain tumor segmentation. Liu et al. [23] designed a spatial-channel fusion block to effectively aggregate multi-modal features for brain tumor segmentation. Zhao et al. [24] introduced a prior attention module that guides multi-lesion segmentation in a coarse-to-fine manner. The proposed lesion-aware spatial attention enhances can focus on relevant regions. In addition, an intermediate supervision strategy is proposed to further refine the attention and improve segmentation accuracy. Zhu et al. [25] introduced a dual-branch U-Net for ultrasound image segmentation, combining multi-component modules to facilitate cross-branch feature fusion and enhance noise resilience.

In contrast to prior approaches, this study introduces a novel multi-modal teacher-student fusion strategy based on prior knowledge of MR modalities. Unlike conventional fusion methods that treat all modalities equally or employ generic attention-based integration, our approach proposes a modality-aware guidance by explicitly categorizing MRI modalities into teacher (FLAIR and T1c) and student (T2 and T1) groups, leveraging their distinct diagnostic roles. FLAIR and T1c provide stronger tumor contrast and structural details, making them well-suited as teacher modalities, while T2 and T1, though informative, benefit from additional guidance due to their relatively lower contrast specificity for tumor boundaries. This structured grouping facilitates more effective feature learning, where the student modalities receive spatial and channel-wise supervision from the teacher modalities. By enforcing hierarchical guidance, the proposed method enhances the model's ability to learn discriminative and robust tumor features, leading to improved brain tumor segmentation performance.

2.2. Uncertainty quantification

Uncertainty quantification in deep learning has gained significant attention in recent years due to its importance in improving model reliability and decision-making. Various techniques have been proposed to quantify uncertainty, including Bayesian methods [26], MC dropout [27], and ensemble methods [28]. Bayesian methods treat model parameters as random variables and estimate their posterior distribution, allowing for the quantification of uncertainty in predictions. MC dropout [29] extends this approach by sampling multiple predictions with dropout applied during inference, providing a computationally efficient way to quantify uncertainty. Ensemble methods aggregate predictions from multiple independently trained models to capture different sources of variability and uncertainty. The quantification of uncertainty involves two fundamental types: aleatoric and epistemic uncertainty. Aleatoric uncertainty arises from inherent variability in the observed data itself, such as noise, measurement errors, or natural variability in the phenomena being studied. In contrast, epistemic uncertainty originates from limitations in the model's capacity to accurately represent the underlying data distribution. It reflects uncertainty about the model's parameters or structure and can be reduced with additional data or model improvements.

In the domain of medical image segmentation, uncertainty quantification plays a crucial role in assessing the confidence of model predictions and identifying areas of ambiguity. By incorporating uncertainty information into the training process or post-processing steps, researchers have demonstrated enhanced segmentation accuracy and the ability to identify challenging cases where the model's confidence is low. For example, Nair et al. [30] developed a 3D CNN for MS lesion segmentation with voxel-based uncertainties. Furthermore, uncertainty-based filtering is utilized as the post-processing. Kown et al. [31] proposed a Bayesian neural network to quantify uncertainties in classification problems by decomposing the moment-based predictive uncertainty into aleatoric and epistemic uncertainty. Arega et al. [32] proposed a cardiac MR segmentation model which utilizes the MC dropout method to generate uncertainty estimates during training. These uncertainty estimates are then incorporated into the loss function to enhance the segmentation results. Lee et al. [33] proposed leveraging four uncertainty estimates for brain tumor segmentation. They introduced a designed uncertainty-attention module to correct uncertain predictions. Zhou et al. [9] introduced a multi-stage brain tumor segmentation network, integrating uncertainty maps to refine results as additional constraint information.

In this work, we propose utilizing the MC dropout method to generate uncertainty estimates during training. Additionally, we integrate an uncertainty-aware loss into the Dice-based segmentation framework, allowing the model to dynamically adjust its learning process based on uncertainty levels. By explicitly incorporating uncertainty into the training loss, the proposed method not only enhances segmentation performance but also improves the reliability of predictions. This uncertainty-guided learning ensures that the model is more aware of its confidence in different regions, leading to more trustworthy outputs, particularly in challenging cases with ambiguous tumor boundaries or missing modalities.

3. Methodology

3.1. Motivation behind the proposed approach

Brain tumor segmentation typically relies on multiple MR modalities, each providing unique and complementary insights into tumor characteristics. However, traditional segmentation methods often treat all modalities equally and fail to leverage the distinct strengths of each, leading to suboptimal performance. To address this, we propose a multi-modal teacher-student fusion strategy, where the more informative modalities (FLAIR and T1c) serve as teacher modalities, guiding the student modalities (T2 and T1). This approach ensures that the most critical modality information is maximally exploited, enhancing feature representation and ultimately improving segmentation accuracy. In addition, traditional segmentation methods often struggle to quantify uncertainty in their predictions. However, uncertainty quantification is crucial in medical imaging tasks, where incorrect segmentation results can have significant consequences. To address this challenge, we integrate uncertainty quantification into the segmentation network, providing four different uncertainty maps that offer invaluable insights into the reliability of the segmentation predictions. Furthermore, the complexities of brain tumor segmentation require to take into account not only prediction errors but also uncertainty. Existing loss functions often do not explicitly account for both prediction errors and uncertainty, limiting their effectiveness in complex tumor segmentation tasks. To overcome this limitation, we introduce a novel uncertainty-aware loss function, enabling the model to jointly optimize segmentation accuracy and uncertainty reduction.

3.2. Proposed network architecture

The proposed network architecture is a multi-encoder-decoder network, as illustrated in Fig. 2. It employs four separate encoders to extract modality-specific feature representations. Each encoder consists of a convolutional layer with Instance normalization and LeakyReLU activation, followed by down-sampling achieved through a convolutional layer with a stride of 2. The decoder includes an up-sampling layer, a multi-modal teacher-student fusion module, and a convolutional layer with Instance normalization and LeakyReLU activation. To generate uncertainty estimates, we apply dropout with a rate of 0.3 in the bottom layers of the network. Additionally, the segmentation result from each layer is processed through a $1 \times 1 \times 1$ convolutional layer followed by an up-sampling operation, and then combined via element-wise summation. This strategy, referred to as deep supervision, allows for the generation of multi-scale segmentation results, enhancing the overall performance and accuracy of the network. During training, we utilize MC dropout to generate N segmentation predictions. The average result is taken as the final segmentation output, while the variance is incorporated into the loss function. The entire network is trained using a hybrid loss function that combines Dice loss and variance-based uncertainty loss. These N segmentation predictions are also used to obtain four different uncertainty maps: aleatoric uncertainty, epistemic uncertainty, entropy, and mutual information. These maps offer insights into the trustworthiness of the segmentation results, guiding further refinement of the network. Consequently, this approach enables the network to produce both segmentation results and uncertainty estimates, enhancing its effectiveness in applications that require reliable prediction confidence.

3.3. Multi-modal teacher-student fusion strategy

To learn complementary information across modalities, we propose a multi-modal teacher-student fusion (MTSF) strategy, consisting of a spatial attention fusion module (SAFM) and a channel attention fusion module (CAFM), as depicted in Fig. 3. The former is utilized to learn fused feature representation from a spatial perspective, the latter is to extract feature representation from a channel perspective. The teacher-student grouping is based on the discriminative power of each modality: FLAIR and T1c are known to provide the most informative signals for identifying tumor regions, particularly edema and enhancing tumor areas, and are therefore designated as teacher modalities. T2 and T1, which provide additional structural context but are less discriminative individually, are treated as student modalities. This grouping allows the network to exploit the strong guidance of teacher modalities to enhance the representation of student modalities, improving overall multi-modal fusion. Initially, FLAIR and T1c are concatenated to form the teacher modality $f_t \in \mathbb{R}^{2C\times H\times W\times D}$. Similarly, T2 and T1 are concatenated to obtain the student modality $f_s \in \mathbb{R}^{2C\times H\times W\times D}$, where C, H, W, and D represent the channel, height, width, and depth of the 3D volume, respectively.

In the spatial attention fusion module, separate convolution blocks with kernel sizes of $3\times3\times3$ and $5\times5\times5$, each followed by a Sigmoid operation, are first applied to f_s and f_t, respectively. This process can yield student attention weights $w_{s1} = Sig(Conv_3(f_s))$, $w_{s2} = Sig(Conv_5(f_s))$, and teacher attention weights $w_{t1} = Sig(Conv_3(f_t))$, $w_{t2} = Sig(Conv_5(f_t))$, here Sig denotes Sigmoid operation. These weights highlight the regions that contribute more significantly to the final fused feature representation. Then, the attention-weighted feature representations can be obtained via element-wise multiplication of the original teacher modality f_t and student modality f_s, followed by addition:

$$f_{sps} = (f_s \otimes w_{s1}) \oplus (f_s \otimes w_{s2}), \quad (1)$$

$$f_{spt} = (f_t \otimes w_{t1}) \oplus (f_t \otimes w_{t2}), \quad (2)$$

where $\otimes$ denotes element-wise multiplication, $\oplus$ denotes element-wise summation,

Subsequently, the student spatial feature representation and teacher spatial feature representation are concatenated to yield the final spatial feature representation $f_{sp} \in \mathbb{R}^{2C\times H\times W\times D}$. In this way, the spatial attention fusion module ensures that the fused feature representation captures the most informative spatial details from each modality. Additionally, we evaluate the impact of different kernel sizes in the spatial

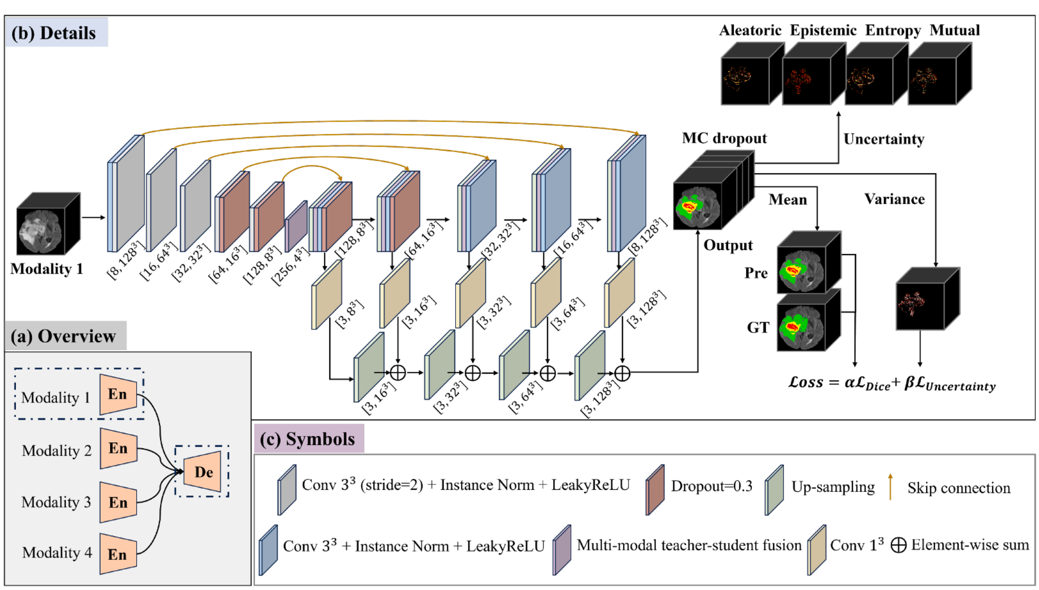

Fig. 2. Overview of the proposed multi-encoder-decoder network architecture for brain tumor segmentation. Each of the four MR modalities (FLAIR, T1c, T2, and T1) is processed by a dedicated encoder, enabling the extraction of modality-specific feature representations. In this simplified illustration, only one encoder is shown. The decoder path incorporates deep supervision to facilitate multi-scale segmentation learning. The network outputs both segmentation results and uncertainty quantifications for each prediction. The entire network is trained using a hybrid loss function that combines traditional Dice-based segmentation loss with the proposed uncertainty-aware penalties.

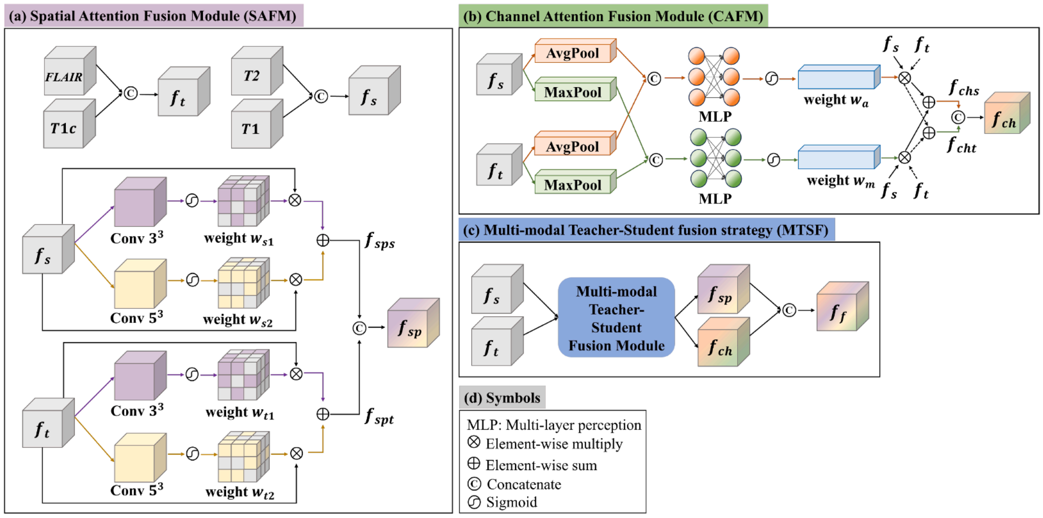

Fig. 3. The proposed multi-modal teacher-student fusion strategy, consisting of a spatial attention fusion module and a channel attention fusion module. FLAIR and T1c are concatenated to form the teacher modality, while T2 and T1 are concatenated to obtain the student modality.

attention fusion module, as discussed in Section 6.1.

$$f_{sp} = [f_{sps}, f_{spt}], \tag{3}$$

where $[\cdot]$ denotes concatenation.

In the channel attention module, global average pooling (GAP) and max pooling operations (GMP) are applied to student modality f_s and teacher modality f_t, respectively. GAP computes the average value of the feature maps across spatial dimensions, producing a channel descriptor for each modality. GMP, on the other hand, captures the most prominent features by taking the maximum value within each feature map. The results of each kind of pooling operation are then concatenated, $[GAP(f_s), GAP(f_t)]$, $[GMP(f_s), GMP(f_t)]$. The reason is that the teacher and student modalities may capture different aspects of the same data. This integration of their average/max pooling results can help the model leverage the strengths of both modalities, leading to better performance. Following this, a multi-layer perception with a Sigmoid function is employed to obtain the attention weights:

$$w_a = Sig\big(MLP\big([GAP(f_s), GAP(f_t)]\big)\big), \tag{4}$$

$$w_m = Sig\big(MLP\big([GMP(f_s), GMP(f_t)]\big)\big). \tag{5}$$

These weights emphasize the importance of different channels, allowing the network to focus on the most relevant features across channels. Then, multiplying these weights with the original teacher modality f_t and student modality f_s, followed by addition, yields the student and teacher attention-weighted feature representations:

$$f_{chs} = (f_s \otimes w_a) \oplus (f_s \otimes w_m), \tag{6}$$

$$f_{cht} = ((f_t \otimes w_a) \oplus ((f_t \otimes w_m). \tag{7}$$

Subsequently, these feature representations are concatenated to achieve the final channel attention feature $f_{ch} \in \mathbb{R}^{2C\times H\times W\times D}$. Thus, the channel attention fusion module facilitates the extraction of complementary information by highlighting important channels while suppressing less relevant ones. Moreover, we investigate the impact of different configurations in the channel attention fusion module, as detailed in Section 6.2.

$$f_{ch} = [f_{chs}, f_{cht}]. \tag{8}$$

Finally, the spatial feature representation f_{sp} and channel attention feature f_{ch} are concatenated to achieve the fused feature representation $f_f \in \mathbb{R}^{4C\times H\times W\times D}$. By combining these two attention modules, the proposed multi-modal teacher-student fusion strategy can effectively learn a comprehensive and complementary feature representation. This dual attention approach leverages the strengths of both spatial and channel information, improving the overall performance of the multi-modal fusion process.

$$f_f = [f_s, f_t]. \tag{9}$$

3.4. Proposed uncertain-aware loss function

In brain tumor segmentation, we propose to integrate Dice loss with variance-based uncertainty loss to enhance segmentation accuracy while simultaneously quantifying uncertainty. The Dice loss measures the overlap between predicted and ground truth segmentation (refer to (10)). The variance-based uncertainty loss is derived from N times MC dropout samples, where each voxel i has N segmentation predictions: $P^n = \{p_i^1, p_i^2, \dots, p_i^N\}$. Hence, the sampling mean μ_i and variance σ_i^2 can be computed using (11) and (12), respectively. The variance-based uncertainty loss (UL) is then calculated via (13), which can capture the uncertainty inherent in the segmentation predictions. Additionally, we compare different uncertainty-aware loss functions, as discussed in Section 6.3. The final proposed hybrid loss function is illustrated in (14), aiming to not only improve segmentation results but also provide clinicians with valuable insights into the reliability of automated segmentation.

$$L_{Dice} = 1 - 2\frac{\sum_{i=1}^{M} p_i g_i + \epsilon}{\sum_{i=1}^{M} p_i + \sum_{i=1}^{M} g_i + \epsilon}, \tag{10}$$

where M refers to the number of voxel in the volume, $p_i \in [0,1]$ and $g_i \in [0,1]$ represent the prediction probability and ground truth value of voxel i, respectively. Additionally, ϵ is a small constant included to prevent division by zero.

$$\mu_i = \frac{1}{N}\sum_{n=1}^{N} p_i^n, \tag{11}$$

where N refers to the number of MC dropout sampling, $N = 10$, p_i^n and μ_i refer to the predicted probability of voxel i for nth MC dropout sampling, and the mean predicted probability of voxel i for N MC dropout sampling.

$$\sigma_i^2 = \frac{1}{N}\sum_{n=1}^{N} (p_i^n - \mu_i)^2, \tag{12}$$

where σ_i^2 refers to the variance of predicted probabilities over N MC dropout sampling.

$$\mathcal{L}_{Uncertainty} = \frac{1}{M}\sum_{i=1}^{M} \sigma_i^2, \tag{13}$$

where M refers to the number of voxel in the volume.

$$\mathcal{L}_{Hybrid} = \alpha\mathcal{L}_{Dice} + \beta\mathcal{L}_{Uncertainty}, \tag{14}$$

where the hyper-parameters are set as $\alpha = 1$ and $\beta = 1$, the section of hyper-parameters is further discussed in Section 6.4.

3.5. Uncertainty quantification

In the context of deep learning and probabilistic modeling, understanding and quantifying uncertainty are crucial for making reliable predictions. In this work, we utilized MC dropout during training by randomly disconnecting 30 % of neuronal connections after the convolution layer in the bottom layers of the network. This approach enabled each input to generate N segmentation predictions, which are utilized to obtain four different uncertainty maps: aleatoric uncertainty, epistemic uncertainty, entropy, and mutual information. These maps provide valuable insights into the model's confidence and reliability in its predictions, enhancing the interpretability and robustness of the segmentation results.

Aleatoric uncertainty: It arises from the inherent randomness in the data, which cannot be reduced by gathering more data. It can be expressed as the expected variance of the predictions given the model parameters θ:

$$\text{Aleatoric Uncertainty} = \mathbb{E}_{p(\theta|D)}[\mathrm{Var}(\hat{y}|\theta)], \tag{15}$$

where $\mathbb{E}_{p(\theta|D)}$ denotes the expectation with respect to the posterior distribution of the model parameters θ, given the data D. $\mathrm{Var}(\hat{y}|\theta)$ represents the variance of the predictions $\hat{y}$ given the model parameters θ.

Epistemic uncertainty: It stems from a lack of knowledge about the model parameters, which can be reduced by collecting more data or improving the model. It can be quantified as the variance of the expected predictions over the posterior distribution of the model parameters:

$$\text{Epistemic Uncertainty} = \mathrm{Var}_{p(\theta|D)}[\mathbb{E}(\hat{y}|\theta)], \tag{16}$$

where $\mathrm{Var}_{p(\theta|D)}$ denotes the variance with respect to the posterior distribution of the model parameters θ, given the data D. $\mathbb{E}(\hat{y}|\theta)$ represents the expected predictions $\hat{y}$ given the model parameters θ.

Entropy: It is a measure of the uncertainty or randomness in a random variable. For a discrete random variable X with probability mass function $P(X)$, the entropy $H(X)$ is defined as:

$$H(X) = -\sum_{x\in X} P(x)\log P(x), \tag{17}$$

where $P(x)$ is the probability mass function of X evaluated at x.

Mutual information: It quantifies the amount of information obtained about one random variable through another random variable. For two random variables X and Y, the mutual information $I(X;Y)$ is given by:

$$I(X;Y) = \sum_{x\in X}\sum_{y\in Y} P(x,y)\log\frac{P(x,y)}{P(x)P(y)}, \tag{18}$$

where $P(x)$ and $P(y)$ are the marginal probability mass functions of X and Y, respectively. $P(x,y)$ is the joint probability mass function of X and Y evaluated at x and y.

4. Experimental setup

4.1. Dataset and training details

We evaluate the proposed method using the BraTS 2018, 2019 and 2020 datasets [34], which are publicly available multi-modal MR brain tumor segmentation datasets. Specifically, the BraTS 2018, 2019 and 2020 datasets include 285, 335 and 369 training cases, respectively. Each case contains annotated images with four MR modalities: T1, T1c, T2, and FLAIR. The brain tumor segmentation task involves classifying tumors into three regions: whole tumor, tumor core, and enhancing tumor. Before training, each MRI scan undergoes additional preprocessing steps to ensure uniformity in input data, including bias field correction, intensity normalization, skull stripping, registration across modalities, and resizing to a uniform volume size of $128 \times 128 \times 128$ pixels. For intensity normalization, each modality is normalized independently to zero mean and unit variance. The proposed model is optimized using the Adam optimizer with an initial learning rate of 0.0005, which is halved every 10 epochs if no improvement in validation performance is observed. To prevent over-fitting, early stopping is implemented; the training terminates if there is no validation improvement after 50 epochs. The model training is conducted on an NVIDIA GeForce RTX 4090 GPU (24GB) using the Keras platform, employing an 8:2 train-test split.

4.2. Evaluation metrics

In this study, we employ the Dice Similarity Coefficient (DSC) and the Hausdorff Distance (HD) to assess the performance of the proposed method. The DSC measures the spatial overlap between the predicted segmentation and the ground truth, with values ranging from 0 to 1. A higher DSC indicates a more accurate segmentation. The HD quantifies the maximum distance between points on the predicted tumor boundary and the real tumor boundary, with lower HD values indicating better segmentation accuracy.

$$DSC = \frac{2|P\cap G|}{|P|+|G|}, \tag{19}$$

where P and G represent the sets of voxel in the prediction result and the ground truth, respectively.

$$HD = \max\left\{\max_{x\in X}\min_{y\in Y} d(x,y), \max_{y\in Y}\min_{x\in X} d(x,y)\right\}, \tag{20}$$

where X and Y denote the predicted tumor boundary and the true tumor boundary, respectively, and $d(\cdot,\cdot)$ represents the minimum Euclidean distance.

5. Experimental results and analysis

5.1. Ablation experiments

In this section, we conduct ablation experiments to evaluate the effectiveness of the proposed components to the baseline method, which includes the channel attention fusion module (CAFM), spatial attention fusion module (SAFM), and uncertainty loss function (UL). The results are detailed in Tables 1–3. Initially, the baseline method achieves an average DSC of 82.3 % and an average HD of 5.4 mm, as shown in Table 1. The inclusion of each proposed component individually improves the results. Specifically, integrating the CAFM increases the average DSC by 0.5 % and the average HD by 11.1 % compared to the baseline method. This highlights the module's capability to enhance feature extraction by emphasizing informative channels while suppressing irrelevant ones. Furthermore, the incorporation of the SAFM results in a 1.1 % improvement in terms of average DSC and a 9.3 % improvement in terms of average HD compared to the baseline method. In addition, It can be observed that the SAFM improves segmentation performance more than the CAFM. This is because the SAFM focuses on specific spatial locations, capturing fine details and separating objects from complex backgrounds more effectively. When both attention modules are integrated, the segmentation performance improves by 2.2 % in terms of average DSC and 16.7 % in terms of average HD compared to the baseline method. Finally, the combination of all three components yields the best results, which achieves 84.2 % in terms of average DSC and 4.2 mm in terms of average HD, with 2.3 % improvement in terms of average DSC and 22.2 % in terms of average HD, compared to the baseline method.

Similar ablation experiments are conducted on the BraTS 2019 dataset, as shown in Table 2. The baseline method can obtain 83.8 % in terms of average DSC and 3.9 mm in terms of average HD. Compared to the baseline method, the proposed CAFM and SAFM can achieve 0.4 %, and 0.7 % improvements in terms of average DSC, respectively. When

Table 1
Comparison results of ablation experiments on BraTS 2018 dataset. * indicates statistically significant improvements of the proposed methods over the baseline (Wilcoxon signed-rank test, p-value < 0.05).

Method				Dice Similarity Coefficient (% ↑)				Hausdorff Distance (mm ↓)			
Baseline	CAFM	SAFM	UL	WT	TC	ET	Mean	WT	TC	ET	Mean
✓	✗	✗	✗	85.7	84.5	76.7	82.3	7.9	5.3	3.1	5.4
✓	✓	✗	✗	86.0	84.4	77.8	82.7	5.7	5.4	3.3	4.8
✓	✗	✓	✗	86.4*	85.7	77.4	83.2*	7.0	4.5*	3.3	4.9*
✓	✓	✓	✗	**87.0***	86.8*	78.4*	84.1*	6.5*	**4.0***	**3.0***	4.5*
✓	✓	✓	✓	86.6*	**87.1***	**79.0***	**84.2***	**5.3***	4.3*	**3.0***	**4.2***

Table 2
Comparison results of ablation experiments on BraTS 2019 dataset. * indicates statistically significant improvements of the proposed methods over the baseline (Wilcoxon signed-rank test, p-value < 0.05).

Method				Dice Similarity Coefficient (% ↑)				Hausdorff Distance (mm ↓)			
Baseline	CAFM	SAFM	UL	WT	TC	ET	Mean	WT	TC	ET	Mean
✓	✗	✗	✗	86.9	85.7	78.8	83.8	**4.9**	4.2	2.7	**3.9**
✓	✓	✗	✗	86.9	86.6	78.8	84.1	6.1	4.0	2.9	4.3
✓	✗	✓	✗	87.1*	86.4*	79.8*	84.4*	5.9	4.1	2.7	4.2
✓	✓	✓	✗	**87.3***	87.3*	79.9*	84.8*	6.3	4.2	2.8	4.4
✓	✓	✓	✓	**87.3***	**87.8***	**80.3***	**85.1***	5.7	**3.6***	**2.5***	**3.9**

Table 3
Comparison results of ablation experiments on BraTS 2020 dataset. * indicates statistically significant improvements of the proposed methods over the baseline (Wilcoxon signed-rank test, p-value < 0.05).

Method				Dice Similarity Coefficient (% ↑)				Hausdorff Distance (mm ↓)			
Baseline	CAFM	SAFM	UL	WT	TC	ET	Mean	WT	TC	ET	Mean
✓	✗	✗	✗	86.6	86.7	79.0	84.1	8.3	4.0	2.7	5.0
✓	✓	✗	✗	86.6	87.4	79.5	84.5	7.7	4.5	2.6*	4.9
✓	✗	✓	✗	87.2*	86.7	80.2*	84.7*	6.3*	4.3	2.6*	4.4*
✓	✓	✓	✗	87.4*	87.2*	80.3*	85.0*	6.0*	4.3	**2.4***	4.2*
✓	✓	✓	✓	**87.6***	**87.8***	**80.4***	**85.3***	**5.8***	**3.9***	2.7	**4.1***

both attention modules are introduced to the baseline, the segmentation performance improves by 1.2 % in terms of average DSC. However, the HD score does not improve compared to the baseline method. This observation can be attributed to the attention modules primarily enhancing the overlap between predicted and ground truth segmentation (as measured by DSC) without a significant impact on boundary accuracy (measured by HD). With the integration of the hybrid loss function, 1.6 % improvement is observed in terms of average DSC. These experimental results demonstrate the significant contributions of each component to the overall model performance, confirming their importance in enhancing segmentation accuracy and robustness.

Ablation experiments on the BraTS 2020 dataset, as shown in Table 3, demonstrate the effectiveness of our proposed components. Integrating CAFM improves the baseline by 0.5 % in terms of average DSC and 2.0 % in terms of average HD. SAFM further enhances performance, achieving gains of 0.7 % in terms of average DSC and 12.0 % in terms of average HD. The combination of CAFM and SAFM yields an improvement of 1.1 % in terms of average DSC and 16.0 % in terms of average HD. Finally, incorporating UL leads to the highest performance boost, outperforming the baseline by 1.4 % in terms of average DSC and 18.0 % in terms of average HD.

5.2. Quantitative analysis compared to the state-of-the-art methods

To evaluate the performance of the proposed method, we conduct a comprehensive quantitative analysis and compared the results with the state-of-the-art methods, the results are shown in the Tables 4–6. The method described in UQ-BTS [9] integrates uncertainty maps to refine segmentation results in a three-stage brain tumor segmentation network. In addition, the methods introduced in Attention U-Net, [35], M2GC-Net [36] and SuperLightNet [37] utilize attention-based fusion methods for brain tumor segmentation. On the other hand, the methods introduced in SegFormer3D [38], Slim UNETR [39], Swin UNETR [40], VT-UNet [41] and TransBTS [42] employ transformer-based methods. From Table 4, the results from BraTS 2018 dataset, it can be observed that the proposed method outperforms the compared methods in terms of average DSC and average HD. Specifically, for WT, the proposed method achieves the highest DSC at 86.6 % as in method VT-UNet [41]. For TC, the proposed method leads with a DSC of 87.1 %, as in method UQ-BTS [9]. In terms of HD, the proposed method achieves the lowest values across all tumor regions: 5.3 mm for WT, 4.3 for TC and 3.0 mm for ET, resulting in the average HD of 4.2 mm. From Table 5, the results from BraTS 2019 dataset, it can be observed that the highest DSC on WT is achieved by our method and Swin UNETR [40] at 87.3 %, and the method from TransBTS [42] achieves the highest DSC of 81.8 % on ET. However, the proposed method achieves the highest DSC of 87.8 % on TC and the best average DSC of 85.1 %. In addition, the proposed method achieves the lowest values on TC, ET and average result in terms of HD. From Table 6, the results from BraTS 2020 dataset, it can be observed that the method from Swin UNETR [40] achieves the highest DSC of 87.8 % on WT and 81.7 % on ET. However, the proposed method achieves the highest DSC of 87.8 % on TC and the best average DSC of 85.3 %. Moreover, in terms of HD, UQ-BTS [9] achieves the lowest value of 5.5 mm on WT , VT-UNet [41] achieves the lowest of 2.6 mm on ET , while our proposed method achieves the lowest of 3.9 mm on TC and the best average HD of 4.1 mm. These results demonstrate the effectiveness of the proposed multi-modal teacher-student fusion strategy and the incorporation of uncertainty estimates in brain tumor segmentation.

5.3. Qualitative analysis of segmentation performance

In addition to quantitative metrics, we present several representative cases to demonstrate the performance of the proposed method, pre-

Table 4
Comparison results with other state-of-the-art methods on BraTS 2018 dataset. * indicates statistically significant improvements of the propose method over others (Wilcoxon signed-rank test, p-value < 0.05).

	Dice Similarity Coefficient (% ↑)				Hausdorff Distance (mm ↓)			
Methods	WT	TC	ET	Mean	WT	TC	ET	Mean
UQ-BTS [9]	86.1	**87.1**	78.9	84.0	6.2	5.2	3.1	4.8
Attention U-Net [35]	85.9	77.1*	74.5*	79.2*	**5.3**	4.9	4.7*	5.0*
M2GCNet [36]	85.1*	84.3*	76.1*	81.8*	5.7	6.0	4.3*	5.3
SuperLightNet [37]	85.7	85.7*	**79.5**	83.6*	6.1	4.8	3.4	4.8
SegFormer3D [38]	79.8*	80.1*	70.8*	76.9*	14.5*	11.5*	6.9*	11.0*
Slim UNETR [39]	83.0*	83.0*	76.0	80.7*	20.6*	7.9	5.7*	11.4*
Swin UNETR [40]	86.1	81.2*	78.9	82.1*	8.7*	7.1*	3.9*	6.6*
VT-UNet [41]	**86.6**	84.1*	78.0	82.9*	6.7	6.1*	3.3	5.4*
TransBTS [42]	85.2	84.4	78.0*	82.5	7.3	5.7	4.2	5.7
Ours	**86.6**	**87.1**	79.0	**84.2**	**5.3**	**4.3**	**3.0**	**4.2**

Table 5
Comparison results with other state-of-the-art methods on BraTS 2019 dataset. * indicates statistically significant improvements of the proposed method over others (Wilcoxon signed-rank test, p-value < 0.05).

	Dice Similarity Coefficient (% ↑)				Hausdorff Distance (mm ↓)			
Methods	WT	TC	ET	Mean	WT	TC	ET	Mean
UQ-BTS [9]	86.4*	86.9	79.8	84.4*	6.6	4.6	3.1*	4.8
Attention U-Net [35]	81.3	77.8*	76.7*	78.6*	7.5	7.2*	3.2	6.0*
M2GCNet [36]	85.6*	84.3*	78.3*	82.7*	6.4	5.0*	3.8*	5.1*
SuperLightNet [37]	85.9*	86.4	81.6	84.6	6.6*	4.5	3.0	4.7*
SegFormer3D [38]	80.8*	82.0*	71.2*	78.0*	10.4*	9.2*	6.5*	8.7*
Slim UNETR [39]	85.3*	85.7*	76.6*	82.5*	**5.5**	4.7	3.8*	4.7*
Swin UNETR [40]	**87.3**	84.1*	81.3	84.2	8.2*	5.5*	3.4*	5.7*
VT-UNet [41]	87.2	84.8*	79.7	83.9*	5.8	5.7*	3.1	4.9*
TransBTS [42]	86.7	86.2	**81.8**	84.9	6.5	4.7*	2.6	4.6
Ours	**87.3**	**87.8**	80.3	**85.1**	5.7	**3.6**	**2.5**	**3.9**

Table 6
Comparison results with other state-of-the-art methods on BraTS 2020 dataset. * indicates statistically significant improvements of the proposed method over others (Wilcoxon signed-rank test, p-value < 0.05).

	Dice Similarity Coefficient (% ↑)				Hausdorff Distance (mm ↓)			
Methods	WT	TC	ET	Mean	WT	TC	ET	Mean
UQ-BTS [9]	86.9*	86.5*	79.0*	84.1*	**5.5**	4.8	2.9	4.4
Attention U-Net [35]	79.6	80.8	73.8	78.1	7.2	4.9*	3.9	5.3*
M2GCNet [36]	85.9*	85.3*	77.1*	82.8*	5.7	5.1*	3.8*	4.9*
SuperLightNet [37]	86.9	86.8	81.5	85.1*	5.7	4.3	2.9	4.3
SegFormer3D [38]	79.9*	82.1*	70.8*	77.6*	14.7*	7.7*	5.4*	9.3*
Slim UNETR [39]	86.2*	85.6*	77.2*	83.0*	8.2*	4.7*	3.3*	5.4*
Swin UNETR [40]	**87.8**	85.4*	**81.7**	85.0	7.0	5.3*	2.9	5.1*
VT-UNet [41]	87.6	85.9*	80.4	84.6	5.8	5.5*	**2.6**	4.6*
TransBTS [42]	86.4	85.5	80.0*	84.1	6.6	4.6	3.2	4.8
Ours	87.6	**87.8**	80.4	**85.3**	5.8	**3.9**	2.7	**4.1**

sented in Fig. 4. In the first case, the baseline method detects some isolated false edema regions. However, the proposed fusion module with the uncertainty loss function helps focus on the relevant regions, ensuring precise delineation of the tumor regions. Additionally, in the second case, the baseline method fails to detect a small isolated enhancing tumor region, whereas the proposed method achieves a better segmentation result. In the third case, the baseline method segments some false NCR&NET regions, but the proposed method effectively mitigates this issue. The proposed fusion strategy and uncertainty integration contribute to this accurate segmentation. The qualitative analysis highlights the superior performance of the proposed method compared to baseline approaches. By integrating a fusion module and an uncertainty loss function, the proposed method focuses on relevant regions and achieves more precise tumor delineation.

Furthermore, we compare the segmentation performance of different methods in Fig. 5. As shown in Fig. 5, in the first case, UQ-BTS, SegFormer3D and VT-UNet yield incomplete segmentation of the edema region (green). In the second case, the irregular shape of the edema region presents additional challenges, both Attention U-Net and Swin UNETR misclassify parts of it as the NCR&NET region (red). In the third case, several methods falsely segment the NCR&NET and enhancing tumor regions, resulting in over-segmentation. Notably, VT-Net and TransBTS under-segment the NCR&NET region (red). Overall, most compared methods are prone to segmentation errors, whereas our method achieves results more consistent with the ground truth, demonstrating superior performance.

5.4. Uncertainty visualization of segmentation results

We conduct an uncertainty analysis to compare the baseline method with the proposed method, presented in Fig. 6. For this purpose, we select two cases and analyzed their segmentation results alongside four corresponding uncertainty maps: aleatoric uncertainty, epistemic uncertainty, entropy uncertainty, and mutual information uncertainty. In the first case, the baseline method shows false segmentation in the NCR&NET region, leading to high aleatoric uncertainty, particularly along the tumor boundaries. Moderate epistemic uncertainty and significant entropy and mutual information uncertainty along the tumor boundaries can be observed. In contrast, the proposed method shows reduced uncertainty in all four maps, indicating more reliable segmentation. In the second case, the baseline method incorrectly detects the NCR&NET region, again resulting in high aleatoric uncertainty along the boundaries, noticeable epistemic uncertainty, high entropy at the boundaries, and uncertainty in mutual information around the tumor boundaries. Our proposed method significantly lowers uncertainty across all maps, demonstrating improved segmentation accuracy and reduced prediction variability. Overall, the main source of uncertainty is the aleatoric component, prevalent at tumor boundaries, while the proposed method consistently reduces uncertainty, demonstrating its effectiveness in handling complex brain tumor segmentation tasks.

5.5. Visualization of feature maps

To analyze the model's decision-making process, we compare the attention heatmaps of our method with those of the baseline across six representative cases, as illustrated in Fig. 7. The visualization includes input FLAIR images (FLAIR), ground truth annotations (GT), baseline predictions, and our method's outputs (Ours) for key tumor sub-regions: enhancing tumor, edema, necrotic and non-enhancing tumor (NCR&NET), with red regions indicating higher model focus on tumor areas. Across all cases, the baseline exhibits relatively scattered and diffused attention, often failing to fully capture the tumor extent or erroneously activating non-tumor areas. In contrast, our method consistently produces stronger and more focused attention aligned closely with the true tumor regions. Notably, in the first four cases, our method captures the tumor core more precisely, while the baseline spreads its attention too broadly. In the last two cases, our method shows better boundary localization and structure awareness, even in challenging shapes, as seen in the last case. These results highlight the effectiveness of our attention mechanism in enhancing tumor localization and providing more clinically relevant feature focus than the baseline.

5.6. Visualization of failure cases

Although the proposed method achieves competitive overall performance, certain challenging cases still result in suboptimal segmentation. As illustrated in Fig. 8, these failure cases mainly correspond to tumors with ambiguous boundaries, irregular shapes, or small volumes. In such situations, the model occasionally under-segments the edema and NCR&NET regions or mis-classifies diffuse tumor boundaries as surrounding edema. To further interpret these results, we visualize the corresponding uncertainty maps, which highlight regions of high predictive uncertainty, particularly along low-contrast boundaries. This observation suggests that the model is aware of its potential errors in these regions. In future work, we will focus on enhancing boundary sensitivity modeling, such as through dedicated boundary refinement modules to improve segmentation robustness in these extreme cases.

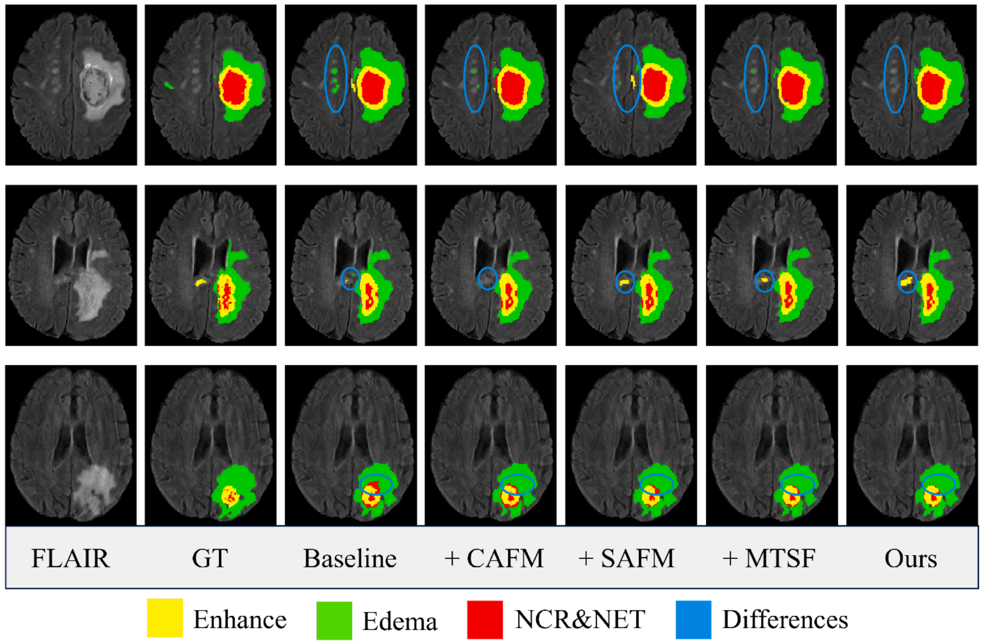

Fig. 4. The qualitative analysis compares segmentation performance across various methods, including baseline, baseline + CAFM, baseline + SAFM, baseline + MTSF (a combination of CASF and SAFM), and the proposed method.

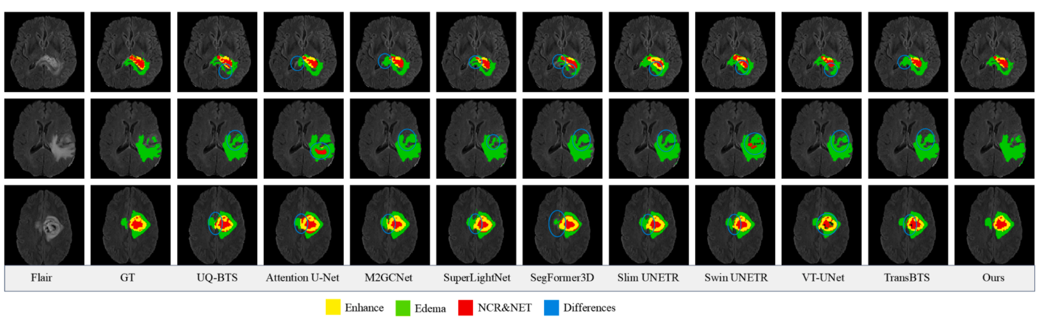

Fig. 5. The qualitative analysis compares segmentation performance across various methods, including UQ-BTS, Attention U-Net, M2GCNet, SuperLightNet, SegFormer3D, Slim UNETR, Swin UNETR, VT-UNet, TransBTS, and the proposed method.

6. Discussion

In this section, we evaluate the effectiveness of the proposed strategies through experiments on the BraTS 2018 dataset. We first analyze the contributions of the spatial and channel attention fusion modules to brain tumor segmentation. We then investigate the influence of the uncertainty-aware loss function and the effect of the hyperparameter β in the hybrid loss formulation. Furthermore, we assess the role of modality-aware supervision under different teacher-student modality combinations, demonstrating that structured modality guidance can effectively enhance segmentation performance. In addition, we compare various dropout configurations, particularly the application of modality-level dropout in each encoder layer to simulate attenuation or noise in modality signals. We also conduct experiments with missing modalities to evaluate the model's robustness when certain MRI sequences are unavailable. These analyses offer valuable insights into the contributions of each component and their role in improving overall segmentation accuracy.

6.1. Analysis of the spatial attention fusion module

In this section, we analyze the effectiveness of the spatial attention fusion module with different kernel sizes: $1 \times 1 \times 1, 3 \times 3 \times 3$ and $5 \times 5 \times 5$. The goal is to determine how varying kernel sizes impact the model's ability to focus on relevant spatial regions and enhance segmentation performance. The selection of these specific sizes is motivated by their capacity to capture spatial information and tumors of varying sizes and shapes within MRI volumes. The results are detailed in Table 7. From Table 7, it can be observed that combining the $3 \times 3 \times 3$ and $5 \times 5 \times 5$ kernel sizes yields the optimal segmentation results, achieving superior performance in terms of both DSC and HD. By integrating both sizes, the spatial attention module can effectively capture both detailed local features and broader spatial contexts. This combination leads to an average DSC of 84.2 % and an average HD of 4.2 mm, surpassing the other two combinations.

Table 7
Comparison results of different kernel sizes in the spatial attention fusion module. * indicates statistically significant improvements of the proposed method over others (Wilcoxon signed-rank test, p-value < 0.05).

	Dice Similarity Coefficient (% ↑)				Hausdorff Distance (mm ↓)			
Kernel size	WT	TC	ET	Mean	WT	TC	ET	Mean
$Conv1^3$ and $Conv3^3$	86.3	84.3*	77.8*	83.0*	6.4	5.2*	3.4	5.0*
$Conv1^3$ and $Conv5^3$	86.4*	86.1	78.7	83.7*	5.7*	6.2	3.2	5.0
$Conv3^3$ and $Conv5^3$	**86.6**	**87.1**	**79.0**	**84.2**	**5.3**	**4.3**	**3.0**	**4.2**

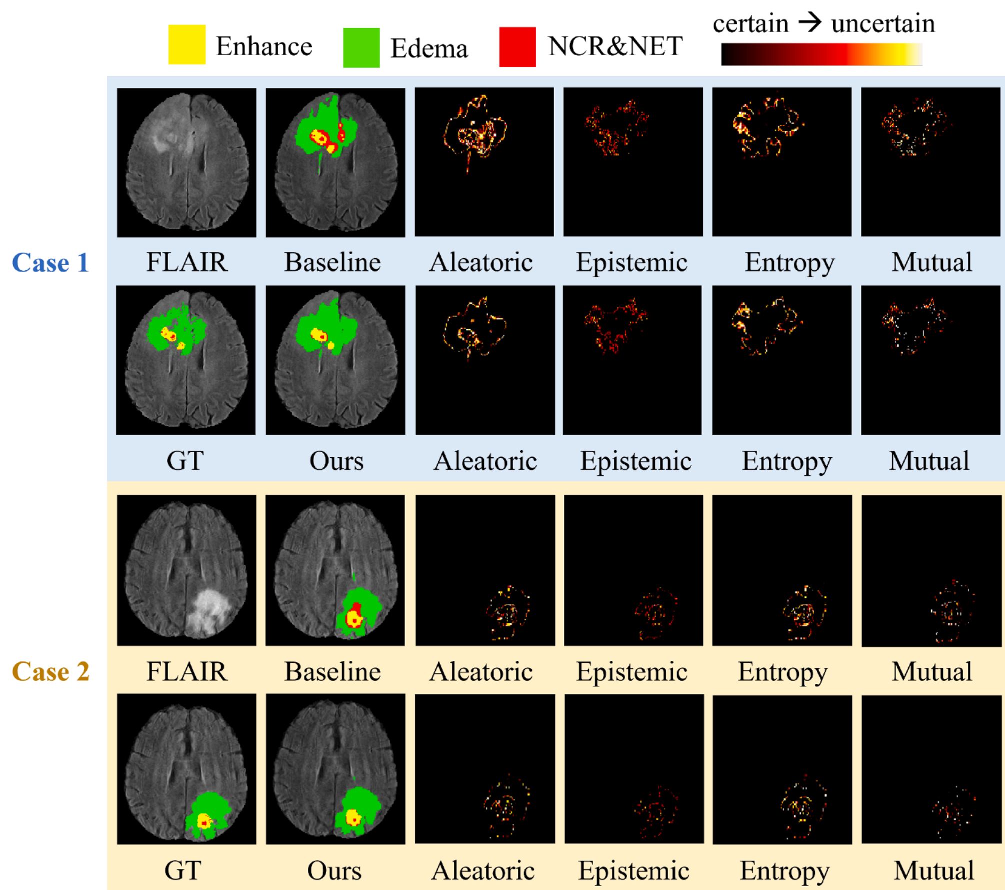

Fig. 6. The uncertainty analysis between the baseline and proposed method.

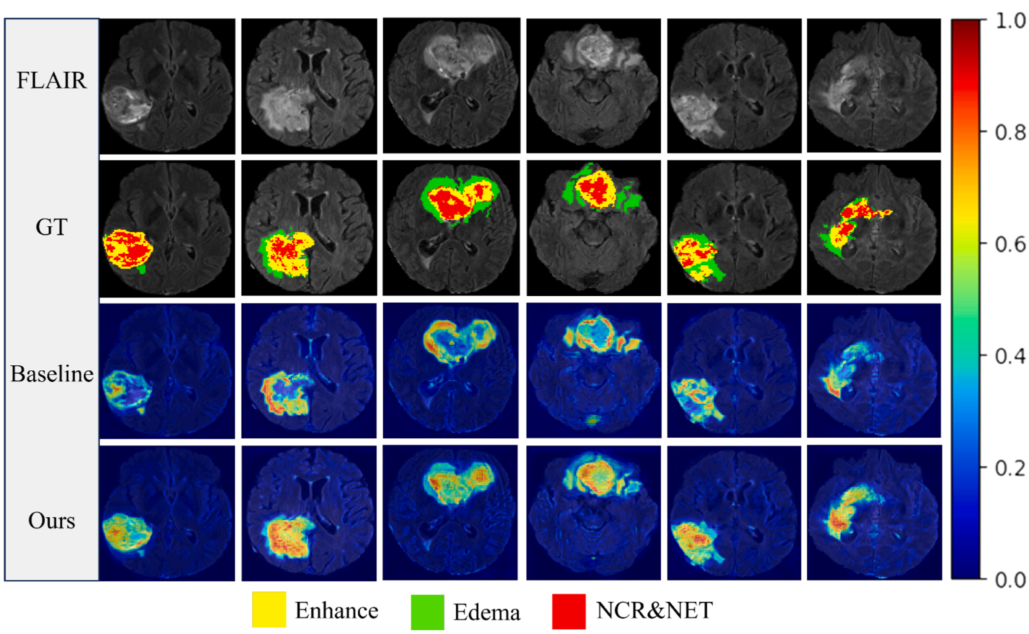

Fig. 7. The feature visualization between the baseline and the proposed method.

6.2. Analysis of the channel attention fusion module

We also analyze the different settings in the channel attention fusion module, the results are illustrated in Table 8. In the first setting, we apply global average pooling to compute weights for the student modalities, and global max pooling to compute weights for the teacher modalities: $f_{chs} = w_a \times f_s$, $f_{cht} = w_m \times f_t$. This approach leverages the average pooling's ability to highlight general feature importance for the student, while the max pooling emphasizes the most prominent features for the teacher. In the second setting, we reverse this process by ap-

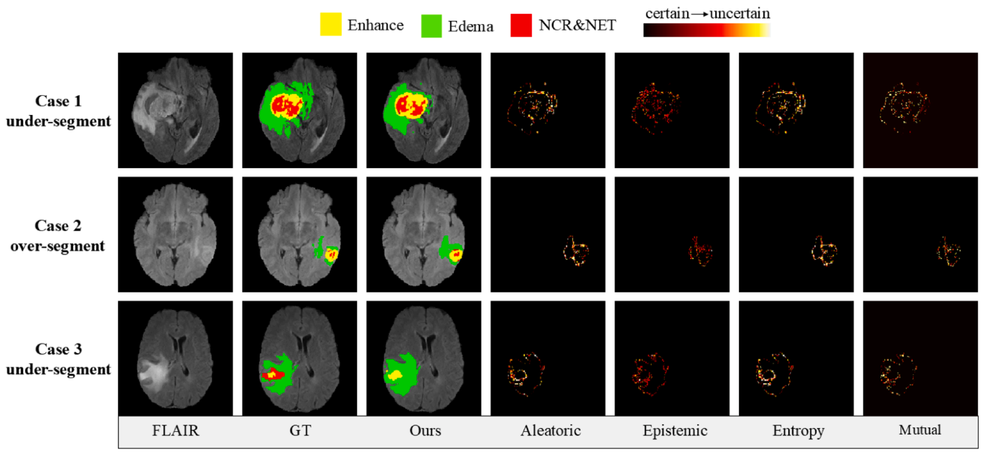

Fig. 8. Visualization of some failure cases.

Table 8
Comparison results of different settings in the channel attention fusion module. * indicates statistically significant improvements of the proposed method over others (Wilcoxon signed-rank test, p-value < 0.05).

	Dice Similarity Coefficient (% ↑)				Hausdorff Distance (mm ↓)			
Connection method	WT	TC	ET	Mean	WT	TC	ET	Mean
$f_{chs} = w_a \times f_s$, $f_{cht} = w_m \times f_t$	**87.0**	85.9	77.0	83.3	**5.1**	**4.1**	3.0	**4.1**
$f_{cht} = w_a \times f_t$, $f_{chs} = w_m \times f_s$	86.9	85.1*	78.6	83.5	6.1	5.7	**2.8**	4.9
$f_{chs} = w_a \times f_s + w_m \times f_s$, $f_{cht} = w_m \times f_t + w_a \times f_t$	86.6	**87.1**	**79.0**	**84.2**	5.3	4.3	3.0	4.2

plying global average pooling weights to the teacher modalities and global max pooling weights to the student modalities: $f_{cht} = w_a \times f_t$, $f_{chs} = w_m \times f_s$. This configuration allows the teacher to benefit from the general feature importance highlighted by average pooling, while the student gains from the prominent feature emphasis provided by max pooling. In the third setting, we apply both global average pooling and global max pooling weights to both the student and teacher modalities: $f_{chs} = w_a \times f_s + w_m \times f_s$, $f_{cht} = w_m \times f_t + w_a \times f_t$. This combined approach aims to capture a comprehensive feature representation by leveraging the strengths of both pooling methods across both modalities, ensuring that both the student and teacher networks benefit from detailed and prominent feature information. From Table 8, it can be observed that the first setting obtains the best DSC on WT, the best HD on WT, TC and average result. The best HD on ET comes from the second setting. However, the proposed method achieves the best DSC on TC, ET and average result. The proposed method achieves an average HD of 4.2 mm, which is comparable to the best-performing methods (4.1 mm). Since the DSC evaluation metric is more critical for accurate segmentation, we choose the third setting, which combines global average pooling and global max pooling weights for both student and teacher modalities. This choice optimally balances general feature importance and prominent feature emphasis, resulting in superior segmentation performance.

6.3. Analysis of the uncertainty-aware loss function

The proposed network utilizes a hybrid loss function that incorporates both Dice loss and variance-based uncertainty loss to quantify prediction uncertainty. This section analyzes the differences between variance-based uncertainty loss (refer to (21)), entropy-based uncertainty loss (refer to (22)), and their combination (refer to (23)), and the comparison results are illustrated in Table 9. Variance-based uncertainty loss measures the variability of predicted probabilities, and aims to reduce the variability to produce more reliable predictions. Entropy-based uncertainty loss measures the randomness in the predicted probability distribution, and high entropy indicates greater uncertainty. The hybrid loss function combines variance and entropy measures to capture a more comprehensive view of uncertainty. The experimental results in Table 9 indicate that the entropy-based uncertainty loss function achieves the best HD on TC and ET. However, the variance-based approach surpasses both the entropy-based and hybrid methods in terms of DSC, and it also achieves the best HD on WT and average results. Therefore, the proposed method adopts a variance-based uncertainty loss function to improve prediction reliability.

Table 9
Comparison results of different uncertainty-aware loss functions. * indicates statistically significant improvements of the proposed method over others (Wilcoxon signed-rank test, p-value < 0.05).

	Dice Similarity Coefficient (% ↑)				Hausdorff Distance (mm ↓)			
Loss function	WT	TC	ET	Mean	WT	TC	ET	Mean
$\mathcal{L}_{entropy}$	86.0*	86.0	77.9*	83.3*	6.7	**4.1**	**2.7**	4.5
$\mathcal{L}_{variance} + \mathcal{L}_{entropy}$	**86.6**	85.5	77.9*	83.3	6.4	4.7	2.8	4.6
$\mathcal{L}_{variance}$	**86.6**	**87.1**	**79.0**	**84.2**	**5.3**	4.3	3.0	**4.2**

$$\mathcal{L}_{variance} = \frac{1}{M}\sum_{i=1}^{M}\sigma_i^2, \tag{21}$$

$$\mathcal{L}_{entropy} = -\frac{1}{M}\sum_{i=1}^{M} p_i \log p_i, \tag{22}$$

$$\mathcal{L}_{uncertainty} = L_{variance} + L_{entropy}, \tag{23}$$

where M refers to the number of voxel in the volume, σ_i^2 refers to the variance of predicted probabilities over MC dropout sampling. $p_i \in [0, 1]$ represents the prediction probability for the ith voxel in the volume.

6.4. Analysis of the hyper-parameter β in the hybrid loss function

In this section, we conduct experiments to evaluate how different values of the hyper-parameter β in the hybrid loss function, $\mathcal{L}_{Hybrid} =$

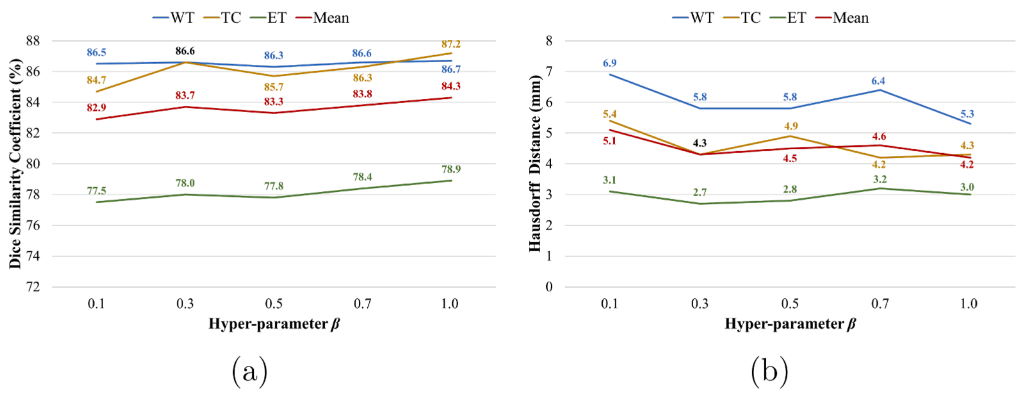

Fig. 9. Impact of hyper-parameter β in the hybrid loss function. Subfigure (a) presents the DSC results, while subfigure (b) shows the HD results.

Table 10
Comparison results of different hyper-parameter β in the hybrid loss function $\mathcal{L}_{Hybrid} = \alpha\mathcal{L}_{Dice} + \beta\mathcal{L}_{Uncertainty}$, where $\alpha = 1$. * indicates statistically significant improvements of the proposed method over others (Wilcoxon signed-rank test, p-value < 0.05).

	Dice Similarity Coefficient (% ↑)				Hausdorff Distance (mm ↓)			
Hyper-parameter β	WT	TC	ET	Mean	WT	TC	ET	Mean
0.1	86.5	84.7*	77.5*	82.9*	6.9	5.4	3.1	5.1
0.3	**86.6**	86.6	78.0	83.7	5.8	4.3	**2.7**	4.3
0.5	86.3	85.7	77.8*	83.3*	5.8	4.9	2.8	4.5
0.7	**86.6**	86.3	78.4	83.8	6.4	**4.2**	3.2	4.6
1.0	**86.6**	**87.1**	**79.0**	**84.2**	**5.3**	4.3	3.0	**4.2**

$\alpha\mathcal{L}_{Dice} + \beta\mathcal{L}_{Uncertainty}$, affect the model's performance. In this study, we focus on varying β while keeping α constant as 1 to explore how the relative emphasis on uncertainty influences the model's behavior, and β is varied over a range of values: $\beta \in \{0.1, 0.3, 0.5, 0.7, 1.0\}$. The results presented in Table 10 and Fig. 9 show that the value $\beta = 1.0$ provides the best trade-off between segmentation performance and uncertainty control, while smaller values result in a decrease in segmentation accuracy in terms of both DSC and HD. Therefore, the proposed method adopts $\mathcal{L}_{Hybrid} = \mathcal{L}_{Dice} + \mathcal{L}_{Uncertainty}$ as the training loss function.

6.5. Analysis of the teacher-student fusion strategy

The selection of teacher-student modality pairs in the proposed framework is critical for achieving effective brain tumor segmentation, as each MRI modality captures distinct yet complementary tumor characteristics. To validate the effectiveness of the modality-aware guidance in multi-modal teacher-student fusion strategy (MTSF), we first conduct an ablation study by replacing the teacher-student grouping with direct concatenation of the four modalities in the "Baseline + CAFM + SAFM" model, with results summarized in Table 11. It can be observed that the proposed MTSF achieves an impressive 87.0 % DSC for WT, 86.8 % for TC, and 78.4 % for ET, resulting in a mean DSC of 84.1 %, indicating improved segmentation accuracy. In terms of HD, MTSF reduces HD to 4.0 mm on TC, but it slightly increases HD to 6.5 mm from 6.4 mm for WT and from 2.5 mm to 3.0 mm for ET, yielding a mean HD of 4.5 mm. This result is comparable to the method without MTSF, which has a mean HD of 4.3 mm. These findings demonstrate that MTSF effectively improves segmentation accuracy while maintaining competitive boundary precision.

Additionally, we systematically assess the impact of different teacher-student combinations on segmentation accuracy, with results summarized in Table 12. The findings indicate that using FLAIR and T1c as teacher modalities, and T1 and T2 as student modalities yields superior segmentation performance. This improvement can be attributed to the unique strengths of these modalities: T1c provides high-contrast enhancement that precisely delineates the tumor core, while FLAIR excels at highlighting peritumoral edema, which is crucial for assessing tumor infiltration. By structuring the fusion process based on these inherent modality characteristics, the proposed framework effectively exploits their complementary information, leading to more accurate segmentation results.

6.6. Analysis of the dropout settings

To evaluate the impact of dropout on model generalization and robustness, we conducted experiments with different dropout configurations. In particular, we applied modality-level dropout in each encoder layer to simulate attenuation or noise in modality signals, thereby mimicking real-world acquisition variability. As shown in Table 13, applying dropout throughout the entire encoder does not yield better performance than applying it only in the bottleneck layers. This is mainly because excessive dropout in shallow layers disrupts the learning of low-level structural features that are crucial for accurate tumor boundary localization. In contrast, applying dropout only in the deeper (bottleneck) layers, as adopted in our method, achieves a better balance between feature regularization and representation stability, leading to improved segmentation accuracy and more reliable uncertainty estimation.

6.7. Analysis of the missing modalities

We further evaluate the model's robustness under missing modality conditions, a common issue in clinical practice where certain MRI sequences may be unavailable or corrupted. Given the large number of possible missing-modality combinations (15 in total for four modalities), we focus on representative cases in which a single modality was absent. To simulate these scenarios, we randomly omit the selected modality during inference. As summarized in Table 14, the proposed framework maintains satisfactory performance even when one modality (e.g., T1 or T2) is missing, demonstrating its ability to effectively exploit the remaining modalities through robust feature fusion. However, the performance drops notably when either FLAIR or T1c is absent. Specifically, the absence of FLAIR leads to a broad performance decrease, particularly for whole tumor, whereas the absence of T1c specifically degrades the segmentation of the tumor core and enhancing tumor. This demonstrates that FLAIR and T1c contribute most significantly to identifying edema and tumor core, respectively. This finding consistently supports our design choice of treating FLAIR and T1c as teacher modalities and T1-T2 as student modalities. Although the results under missing FLAIR or T1c conditions are less favorable, this is expected since our proposed method is not specifically designed to handle missing-modality

Table 11
Analysis of the effectiveness of the modality-aware guidance in multi-modal teacher-student fusion strategy (MTSF). * indicates statistically significant improvements of the proposed method over the method w/o MTSF (Wilcoxon signed-rank test, p-value < 0.05).

	Dice similarity coefficient (% ↑)				Hausdorff Distance (mm ↓)			
Multi-modal teacher-student fusion	WT	TC	ET	Mean	WT	TC	ET	Mean
w/o MTSF	86.3*	85.7*	78.4	83.5	**6.4**	4.1	**2.5**	**4.3**
w MTSF	**87.0**	**86.8**	78.4	**84.1**	6.5	**4.0**	3.0	4.5

Table 12
Comparison results of different teacher-student combinations. * indicates statistically significant improvements of the proposed method over others (Wilcoxon signed-rank test, p-value < 0.05).

	Dice Similarity Coefficient (% ↑)				Hausdorff Distance (mm ↓)			
Teacher-student combinations	WT	TC	ET	Mean	WT	TC	ET	Mean
FLAIR and T1 as teachers, T1c and T2 as students	**87.0**	85.8*	78.4	83.7	5.6	4.5	3.0	4.4
FLAIR and T2 as teachers, T1c and T1 as students	86.9	86.1	78.7	83.9	**5.0**	6.1	**2.6**	4.6
FLAIR and T1c as teachers, T1 and T2 as students	86.6	**87.1**	**79.0**	**84.2**	5.3	**4.3**	3.0	**4.2**

Table 13
Analysis of the dropout settings. * indicates statistically significant improvements of the proposed method over other method (Wilcoxon signed-rank test, p-value < 0.05).

	Dice Similarity Coefficient (% ↑)				Hausdorff Distance (mm ↓)			
Dropout settings	WT	TC	ET	Mean	WT	TC	ET	Mean
All layers	**86.6**	84.9*	77.8*	83.1*	6.7	4.9	**2.7**	4.8
Bottleneck layers (Ours)	**86.6**	**87.1**	**79.0**	**84.2**	**5.3**	**4.3**	3.0	**4.2**

Table 14
Analysis of the missing modalities. ∘ denotes a missing modality, and • denotes an available modality. Best results are highlighted in bold. * indicates significant improvement over other methods (p < .05, Wilcoxon test).

Method				Dice Similarity Coefficient (% ↑)				Hausdorff Distance (mm ↓)			
T1	FLAIR	T1c	T2	WT	TC	ET	Mean	WT	TC	ET	Mean
∘	•	•	•	84.6*	84.0*	75.6*	81.4*	8.0*	6.2*	3.4*	5.9*
•	∘	•	•	53.5*	60.7*	62.6*	58.9*	37.5*	42.2*	26.4*	35.4*
•	•	∘	•	85.9*	49.9*	2.3*	46.0*	5.6*	15.7*	25.3*	15.5*
•	•	•	∘	79.6*	68.9*	75.3*	74.6*	13.3*	12.5*	4.7*	10.2*
•	•	•	•	**86.6**	**87.1**	**79.0**	**84.2**	**5.3**	**4.3**	**3.0**	**4.2**

scenarios. In future work, we will focus on developing segmentation methods that can effectively handle missing-modality scenarios.

6.8. *Limitations and future work*

The proposed approach shows promising improvements in the accuracy and reliability of brain tumor segmentation. However, several limitations and future directions remain. First, although multiple uncertainty maps are generated to assess prediction confidence, effectively integrating these maps into the segmentation network to further boost performance is still an open challenge. Future work may explore new fusion strategies or refine existing frameworks to better utilize these estimates. Second, investigating different uncertainty quantification techniques, such as Bayesian models, ensemble methods, and evidential approaches, could offer deeper insights into improving model robustness against noise, missing data, and clinical variability. Third, future work will explore incorporating boundary information and disentangled representation learning to further reduce uncertainty and improve segmentation accuracy. Finally, handling missing-modality scenarios in a systematic way will be an important focus, enabling the model to maintain reliable performance even when certain MRI sequences are unavailable or corrupted.

7. Conclusion

In this paper, we propose a novel MR brain tumor segmentation approach that integrates multi-modal data fusion with uncertainty quantification to improve both accuracy and reliability. The proposed multi-modal teacher-student fusion strategy can effectively exploit spatial and channel-wise information across modalities. Additionally, the uncertainty-aware loss function not only improves segmentation accuracy but also quantifies uncertainty. The proposed network can generate both segmentation maps and uncertainty estimates, offering critical insights for evaluating the confidence of predictions, which is essential for informed clinical decision-making. Experimental results on three public brain tumor segmentation datasets validate the effectiveness of the proposed approach, highlighting its potential to advance more reliable and interpretable segmentation methods.

CRediT authorship contribution statement

Tongxue Zhou: Writing – review & editing, Writing – original draft, Visualization, Supervision, Methodology, Investigation, Funding acquisition, Conceptualization; **Mingyang Li:** Visualization, Software, Methodology; **Su Ruan:** Writing – review & editing, Methodology; **Tingjin Luo:** Writing – review & editing; **Bingbing Jiang:** Writing – review & editing; **Junan Zhu:** Visualization, Validation; **Ping Ma:** Visualization; **Defu Yang:** Writing – review & editing; **Guang Yang:** Writing – review & editing, Visualization, Methodology.

Data availability

Data will be made available on request.

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.

Acknowledgment

Tongxue Zhou is supported by the National Natural Science Foundation of China (62206084). Guang Yang is supported in part by the ERC IMI (101005122), the H2020 (952172), the MRC (MC/PC/21013), the Royal Society (IEC\NSFC\211235), the NVIDIA Academic Hardware Grant Program, the SABER project supported by Boehringer Ingelheim Ltd, NIHR Imperial Biomedical Research Centre (RDA01), Wellcome Leap Dynamic Resilience, UKRI guarantee funding for Horizon Europe MSCA Postdoctoral Fellowships (EP/Z002206/1), UKRI MRC Research Grant, TFS Research Grants (MR/U506710/1), and the UKRI Future Leaders Fellowship (MR/V023799/1).

References

[1] D.N. Louis, A. Perry, P. Wesseling, D.J. Brat, I.A. Cree, D. Figarella-Branger, C. Hawkins, H.K. Ng, S.M. Pfister, G. Reifenberger, et al., The 2021 WHO classification of tumors of the central nervous system: a summary, Neuro-oncology 23 (8) (2021) 1231–1251.
[2] R.L. Siegel, K.D. Miller, N.S. Wagle, A. Jemal, et al., Cancer statistics, 2023, CA Cancer J. Clin. 73 (1) (2023) 17–48.
[3] Z. Liu, L. Tong, L. Chen, F. Zhou, Z. Jiang, Q. Zhang, Y. Wang, C. Shan, L. Li, H. Zhou, Canet: context aware network for brain glioma segmentation, IEEE Trans. Med. Imaging 40 (7) (2021) 1763–1777.
[4] T. Zhou, S. Ruan, S. Canu, A review: deep learning for medical image segmentation using multi-modality fusion, Array (2019) 100004.
[5] Z. Zhu, Z. Wang, G. Qi, Y. Zhao, Y. Liu, Visually stabilized mamba U-shaped network with strong inductive bias for 3D brain tumor segmentation, IEEE Trans. Instrum. Meas. 74 (2025) 1–11
[6] P. Wang, A.C.S. Chung, Relax and focus on brain tumor segmentation, Med. Image Anal. 75 (2022) 102259.
[7] T. Zhou, Multi-modal brain tumor segmentation via disentangled representation learning and region-aware contrastive learning, Pattern Recogn. 149 (2024) 110282.
[8] D. Maji, P. Sigedar, M. Singh, Attention Res-UNet with guided decoder for semantic segmentation of brain tumors, Biomed. Signal Process. Control 71 (2022) 103077.
[9] T. Zhou, S. Zhu, Uncertainty quantification and attention-aware fusion guided multi-modal MR brain tumor segmentation, Comput. Biol. Med. (2023) 107142.
[10] 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.
[11] A.M.G. Allah, A.M. Sarhan, N.M. Elshennawy, Edge U-Net: brain tumor segmentation using MRI based on deep U-Net model with boundary information, Expert Syst. Appl. 213 (2023) 118833.
[12] W. Zhang, S. Chen, Y. Ma, Y. Liu, X. Cao, ETUNet: Exploring efficient transformer enhanced UNet for 3D brain tumor segmentation, Comput. Biol. Med. 171 (2024) 108005.
[13] 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.
[14] T. Zhou, Boundary-aware and cross-modal fusion network for enhanced multi-modal brain tumor segmentation, Pattern Recogn. (2025) 111637.
[15] 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 Recogn. 153 (2024) 110553.
[16] Y. Li, Y. Duan, G. Qi, B. Cong, L. Zhang, Z. Zhu, A lightweight vision mamba coding UNet for medical image segmentation, Eng. Appl. Artif. Intell. 162 (2025) 112676.
[17] Z. Zhu, Z. Zhang, G. Qi, Y. Li, P. Yang, Y. Liu, Probability map-guided network for 3D volumetric medical image segmentation, IEEE Transactions on Image Processing: A Publication of the IEEE Signal Processing Society (2025).
[18] S. Woo, J. Park, J.-Y. Lee, I.S. Kweon, Cbam: convolutional block attention module, in: Proceedings of the European Conference on Computer Vision (ECCV), 2018, pp. 3–19.
[19] J. Hu, L. Shen, G. Sun, Squeeze-and-excitation networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 7132–7141.
[20] T. Zhou, Modality-level cross-connection and attentional feature fusion based deep neural network for multi-modal brain tumor segmentation, Biomed. Signal Process. Control 81 (2023) 104524.
[21] C. Zhou, C. Ding, X. Wang, Z. Lu, D. Tao, One-pass multi-task networks with cross-task guided attention for brain tumor segmentation, IEEE Trans. Image Process. 29 (2020) 4516–4529.
[22] F. Ghazouani, P. Vera, S. Ruan, Efficient brain tumor segmentation using swin transformer and enhanced local self-attention, Int. J. Comput. Assist. Radiol. Surg. 19 (2) (2024) 273–281.
[23] C. Liu, W. Ding, L. Li, Z. Zhang, C. Pei, L. Huang, X. Zhuang, Brain tumor segmentation network using attention-based fusion and spatial relationship constraint, in: International MICCAI Brainlesion Workshop, Springer, 2020, pp. 219–229.
[24] X. Zhao, P. Zhang, F. Song, C. Ma, G. Fan, Y. Sun, Y. Feng, G. Zhang, Prior attention network for multi-lesion segmentation in medical images, IEEE Trans. Med. Imaging 41 (12) (2022) 3812–3823.
[25] Z. Zhu, Z. Zhang, G. Qi, Y. Li, Y. Li, L. Mu, A dual-branch network for ultrasound image segmentation, Biomed. Signal Process. Control 103 (2025) 107368.
[26] T. Pearce, F. Leibfried, A. Brintrup, Uncertainty in neural networks: approximately bayesian ensembling, in: International Conference on Artificial Intelligence and Statistics, PMLR, 2020, pp. 234–244.
[27] D. Milanés-Hermosilla, R. Trujillo Codorniú, R. López-Baracaldo, R. Sagaró-Zamora, D. Delisle-Rodriguez, J.J. Villarejo-Mayor, J.R. Núñez-Álvarez, Monte Carlo dropout for uncertainty estimation and motor imagery classification, Sensors 21 (21) (2021) 7241.
[28] R. Rahaman, et al., Uncertainty quantification and deep ensembles, Adv. Neural Inf. Process. Syst. 34 (2021) 20063–20075.
[29] Y. Gal, Z. Ghahramani, Dropout as a bayesian approximation: representing model uncertainty in deep learning, in: International Conference on Machine Learning, PMLR, 2016, pp. 1050–1059.
[30] T. Nair, D. Precup, D.L. Arnold, T. Arbel, Exploring uncertainty measures in deep networks for multiple sclerosis lesion detection and segmentation, Med. Image Anal. 59 (2020) 101557.
[31] Y. Kwon, J.-H. Won, B.J. Kim, M.C. Paik, Uncertainty quantification using Bayesian neural networks in classification: application to biomedical image segmentation, Computational Statistics & Data Analysis 142 (2020) 106816.
[32] T.W. Arega, S. Bricq, F. Meriaudeau, Leveraging uncertainty estimates to improve segmentation performance in cardiac MR, in: Uncertainty for Safe Utilization of Machine Learning in Medical Imaging, and Perinatal Imaging, Placental and Preterm Image Analysis: 3rd International Workshop, UNSURE 2021, and 6th International Workshop, PIPPI 2021, Held in Conjunction with MICCAI 2021, Strasbourg, France, October 1, 2021, Proceedings 3, Springer, 2021, pp. 24–33.
[33] J. Lee, D. Shin, S.-H. Oh, H. Kim, Method to minimize the errors of AI: quantifying and exploiting uncertainty of deep learning in brain tumor segmentation, Sensors 22 (6) (2022) 2406.
[34] B.H. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, Y. Burren, N. Porz, J. Slotboom, R. Wiest, et al., The multimodal brain tumor image segmentation benchmark (BRATS), IEEE Trans. Med. Imaging 34 (10) (2014) 1993–2024.
[35] 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, arXiv preprint arXiv:1804.03999 (2018).
[36] T. Zhou, M2GCNet: multi-modal graph convolution network for precise brain tumor segmentation across multiple MRI sequences, IEEE Trans. Image Process. 33 (2024) 4896–4910
[37] F. Yu, J. Cao, L. Liu, M. Jiang, Superlightnet: lightweight parameter aggregation network for multimodal brain tumor segmentation, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2025, pp. 5197–5206.
[38] S. Perera, P. Navard, A. Yilmaz, SegFormer3D: an efficient transformer for 3D medical image segmentation, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024, pp. 4981–4988.
[39] Y. Pang, J. Liang, T. Huang, H. Chen, Y. Li, D. Li, L. Huang, Q. Wang, Slim UNETR: scale hybrid transformers to efficient 3D medical image segmentation under limited computational resources, IEEE Trans. Med. Imaging 43 (3) (2023) 994–1005.
[40] 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.
[41] 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.
[42] 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.