Journal of Computer Science and Artificial Intelligence
ISSN: 3078-8242 (Print), ISSN: 3078-2309 (Online) | Vol. 6, No. 3, 2026

TriFNet: A Tri-Domain Frequency-Aware Network for 3D Brain Tumor Segmentation

Hengbo Hao *

School of Computer Science and Technology, Henan Polytechnic University, Jiaozuo, Henan, 454000, PR China

Abstract: Brain tumor segmentation from multimodal MRI is a challenging task due to large inter-patient heterogeneity, irregular lesion geometry, and ambiguous boundaries between pathological and normal tissues. Although existing 3D segmentation networks have achieved promising performance, most of them primarily operate in the spatial domain and often rely on coarse feature fusion, which limits their ability to explicitly disentangle global structural information from fine-grained boundary details. To address this issue, we propose TriFNet, a Tri-domain Frequency-aware network for 3D brain tumor segmentation. Specifically, a Radial Frequency Decomposition (RFD) module is introduced in the encoder to decompose intermediate features into low-frequency and high-frequency components for separate modeling of global structural cues and local detail information. At the bottleneck, a Cross-Stream Synergistic Fusion (CSF) module is designed to adaptively integrate the complementary responses of the two frequency streams. In the decoder, a Multi-scale Tri-domain Gating (MTG) module jointly exploits spatial-domain, Fourier-domain, and wavelet-domain representations to enhance multi-scale feature reconstruction and boundary refinement. Extensive experiments on the BraTS 2019, BraTS 2020, and BraTS 2021 datasets demonstrate that TriFNet achieves average Dice scores of 86.62%, 88.90%, and 91.88%, with corresponding average HD95 values of 3.32 mm, 2.98 mm, and 3.51 mm, respectively. Ablation studies further verify the effectiveness and complementarity of the proposed modules. These results indicate that TriFNet provides an accurate and robust solution for 3D brain tumor segmentation, especially for challenging tumor core and enhancing tumor regions.

Keywords: Brain tumor segmentation, Multimodal MRI, 3D medical image segmentation, Frequency-aware learning, Fourier transform, Wavelet transform.

1. Introduction

Brain tumor segmentation is a fundamental task in medical image analysis, as it provides essential support for clinical diagnosis, treatment planning, surgical navigation, and prognosis assessment. In multimodal magnetic resonance imaging (MRI), different sequences, such as T1, T1ce, T2, and FLAIR, capture complementary aspects of tumor appearance. Effective segmentation therefore requires the model to jointly exploit cross-modal information, preserve global structural consistency, and delineate subtle tumor boundaries [1]. However, this remains highly challenging due to the heterogeneous appearance of brain tumors, the irregular geometry of lesion regions, and the ambiguous boundaries between pathological and normal tissues.

Recent years have witnessed substantial progress in brain tumor segmentation driven by deep learning. In particular, encoder-decoder architectures based on 3D convolution have become the dominant paradigm because of their ability to learn hierarchical volumetric representations. Despite their success, most existing methods primarily operate in the spatial domain and rely on progressively stacked convolutions to extract contextual information [2]. Such designs are effective for local representation learning, yet they often struggle to explicitly disentangle global structural patterns from fine-grained boundary details. As a result, segmentation performance may degrade in challenging subregions such as the tumor core (TC) and enhancing tumor (ET), where structural irregularity and class imbalance are particularly severe.

Another limitation of current approaches lies in feature interaction. Multi-scale and multi-level representations are typically fused through simple concatenation, summation, or attention-based weighting. Although these strategies improve information aggregation, they do not always model the complementary relationship between different feature streams sufficiently well. In practice, this may lead to feature redundancy, suppression of informative responses, and inadequate refinement of fine structures during decoding. Moreover, high-resolution feature recovery is commonly performed by standard upsampling operations and skip connections, which may be insufficient to fully exploit complementary cues from different representation domains.

To address these issues, this paper proposes a novel Tri-domain Frequency Collaborative Network (TriFNet) for 3D brain tumor segmentation. The proposed framework introduces dedicated modules at three key stages of the network. First, a Radial Frequency Decomposition (RFD) module is incorporated in the encoder to explicitly separate high-frequency and low-frequency components in the Fourier domain. Second, a Cross-Stream Synergistic Fusion (CSF) module is designed at the bottleneck to adaptively integrate the complementary information carried by the two frequency streams. Third, a Multi-scale Tri-domain Gating (MTG) module is introduced in the decoder to jointly exploit spatial-domain, Fourier-domain, and wavelet-domain representations for refined feature reconstruction. Through this unified design, TriFNet simultaneously strengthens structural modeling, feature interaction, and boundary recovery.

The main contributions of this work can be summarized as follows:

- We propose a Frequency-aware 3D segmentation framework that explicitly disentangles structural and detail information by decomposing intermediate features into low- and high-frequency components.
- We design a cross-stream fusion mechanism that

dynamically balances complementary frequency responses according to semantic requirements at different spatial locations.

• We introduce a Tri-domain gated reconstruction strategy that integrates spatial, Fourier, and wavelet representations to improve multi-scale decoding and boundary refinement.

• Extensive experiments on BraTS 2019, BraTS 2020, and BraTS 2021 demonstrate that the proposed method achieves competitive and stable performance, particularly on challenging tumor subregions.

2. Related Work

(1) CNN-Based Brain Tumor Segmentation

Convolutional neural networks have long served as the foundation of brain tumor segmentation. Representative encoder-decoder architectures, such as 3D U-Net [3] and its variants, have shown strong capability in volumetric representation learning by combining hierarchical feature extraction with skip connections. These models are effective at capturing local appearance patterns and spatial continuity, which makes them suitable for dense medical image prediction tasks. Subsequent improvements, including residual connections, dense feature reuse, and automated network configuration, have further advanced the performance of CNN-based segmentation frameworks [4].

Nevertheless, convolution-based models are inherently limited by the locality of convolutional operators. Although stacking multiple layers enlarges the receptive field, long-range dependencies are still modeled only implicitly. This makes it difficult to preserve global structural coherence while simultaneously emphasizing boundary-sensitive details. In brain tumor segmentation, where lesion regions vary substantially in size, shape, and internal composition, such limitations often result in incomplete segmentation of complex tumor cores or blurred delineation of enhancing regions.

(2) Global Context Modeling and Feature Fusion

To overcome the locality of convolutions, many studies have introduced global context modeling mechanisms into medical image segmentation. Transformer-based models [5], for example, leverage self-attention to capture long-range dependencies and have shown strong performance in volumetric segmentation tasks. These methods improve the modeling of global interactions, but their computational complexity and data dependency may hinder efficiency and robustness in practical 3D medical settings [6].

Beyond global context modeling, feature fusion has also become a central topic in recent segmentation research. Multi-scale and multi-branch designs aim to aggregate complementary information from different levels of abstraction [7, 8]. However, many fusion strategies remain relatively coarse, treating all feature streams in a similar manner or relying on fixed combinations. For heterogeneous targets such as brain tumors, the utility of different features can vary significantly across anatomical locations [9, 10]. Therefore, effective segmentation requires a more adaptive mechanism that can selectively emphasize structural cues, textural responses, and semantic context.

(3) Frequency-Aware and Multi-Domain Representation Learning

Unlike conventional segmentation networks that mainly operate in the spatial domain, Frequency-aware methods attempt to explicitly model spectral information. Frequency decomposition provides a natural way to separate slowly varying structural components from rapidly changing boundary and texture patterns. This property is particularly relevant for brain tumor segmentation, where global tumor extent and fine lesion margins contribute differently to the final prediction [11, 12]. Despite its potential, frequency-domain modeling remains underexplored in 3D medical segmentation.

In parallel, multi-domain representation learning has attracted growing attention. Wavelet transforms are effective for capturing localized edge and multi-scale textural information, while Fourier transforms are better suited to describing global spectral structures. Spatial-domain representations, on the other hand, preserve direct semantic correspondence to anatomical regions. Integrating these complementary domains offers an appealing direction for improving segmentation accuracy [13]. However, existing studies seldom establish a unified framework that explicitly coordinates frequency decomposition, cross-stream interaction, and Tri-domain reconstruction in a single end-to-end network.

(4) CNN-Based Brain Tumor Segmentation

The above discussion highlights three major gaps in current brain tumor segmentation methods. First, most approaches do not explicitly disentangle structural information from fine details during feature extraction. Second, the interaction between complementary feature streams is often insufficiently modeled. Third, decoder-side reconstruction typically lacks an effective mechanism to jointly exploit heterogeneous representation domains. Motivated by these observations, we propose TriFNet, which combines radial frequency decomposition, synergistic dual-stream fusion, and Tri-domain gated decoding into a unified architecture for more accurate and robust 3D brain tumor segmentation.

3. Method

(1) Overview of TriFNet

TriFNet follows a standard encoder-decoder architecture for volumetric segmentation, while introducing Frequency-aware and multi-domain collaborative modeling at different stages of the network. Given a multimodal 3D MRI volume as input, the encoder progressively extracts hierarchical features and reduces spatial resolution to enlarge the receptive field. At the early encoding stage, the proposed RFD module decomposes intermediate representations into low-frequency and high-frequency streams, allowing the network to separately learn global structural information and fine boundary details. At the bottleneck, the CSF module adaptively integrates the complementary responses from the two streams. During decoding, the MTG module refines feature reconstruction by jointly modeling spatial-domain, Fourier-domain, and wavelet-domain information. In this way, the entire network forms a coherent pipeline of frequency decomposition, cross-stream interaction, and Tri-domain reconstruction, as shown in Fig. 1.

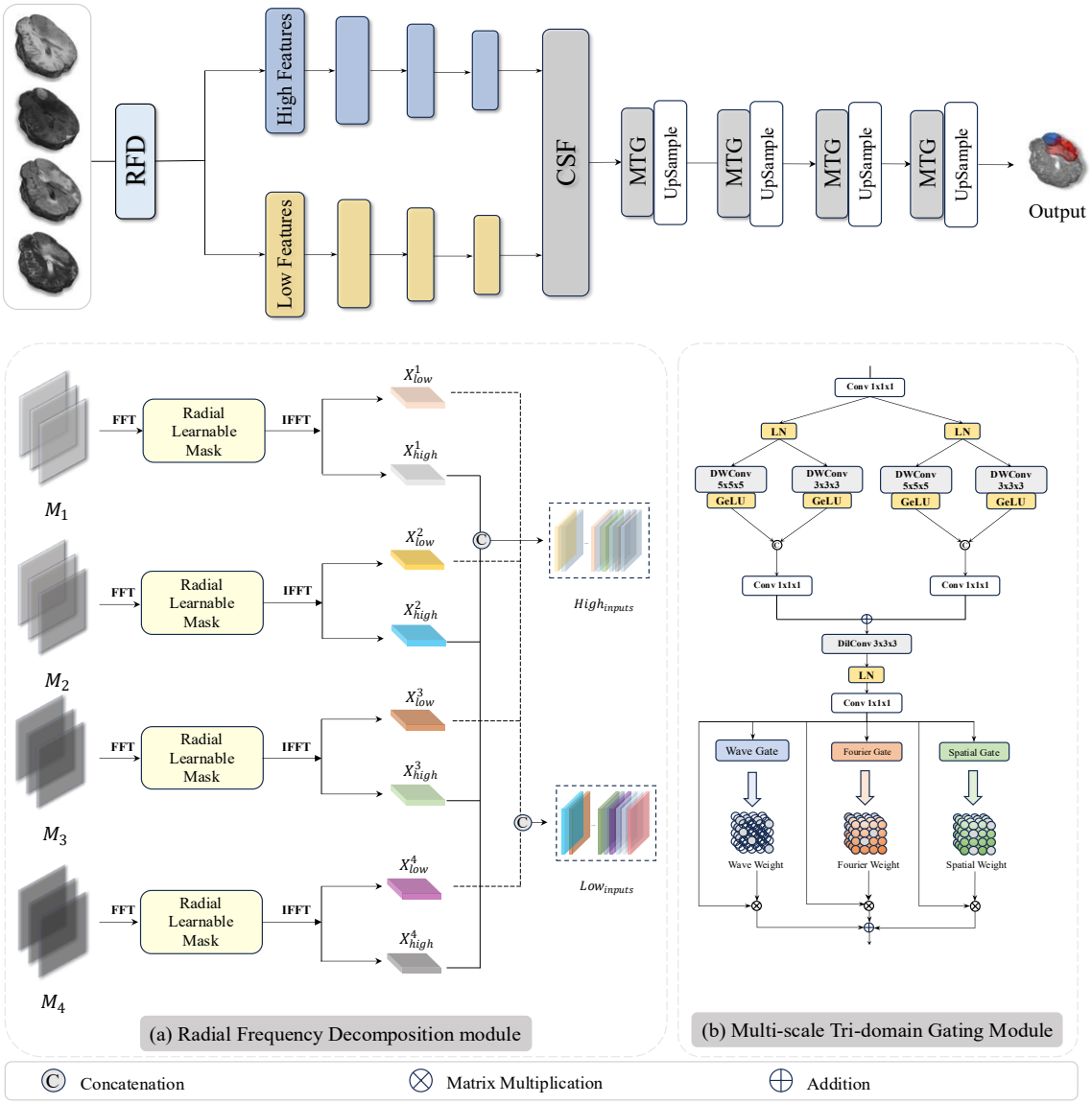

Figure 1. Overview of our proposed model framework

(2) Radial Frequency Decomposition Module

To explicitly model frequency information in medical images, this paper introduces a Radial Frequency Decomposition (RFD) module during the encoding stage, as illustrated in Fig. 1(a). By incorporating a learnable radial frequency mask in the Fourier domain, the module adaptively partitions feature representations in the frequency space, thereby enabling an effective separation of high-frequency and low-frequency components. As shown in Fig. 2, given an intermediate feature representation, a three-dimensional fast Fourier transform is first applied to project it into the frequency domain. A learnable radial frequency mask then softly decomposes the spectrum into low-frequency and high-frequency components according to the distance from the spectrum center, where the cutoff frequency r_0 is optimized in a data-driven manner.

Let the input feature at a certain encoding layer be denoted as $X \in \mathbb{R}^{C\times D\times H\times W}$, where C denotes the number of channels, and D, H, and W represent the depth, height, and width of the feature map, respectively. The feature is first transformed into the frequency domain via a three-dimensional fast Fourier transform (FFT): $F = \mathcal{F}(X)$. In the frequency domain, each frequency component corresponds to a coordinate (u, v, w). To measure its distance from the center of the spectrum, the radial frequency distance is defined as:

$$r = u2 + v2 + w2 \tag{1}$$

Based on this radial formulation, a learnable frequency partition function is introduced to construct the low-frequency mask:

$$M_{low}(r) = \sigma\big(\alpha(r_0 - r)\big) \tag{2}$$

Where r_0 is the learnable cutoff radius, α controls the smoothness of the transition region, and $\sigma(\cdot)$ denotes the Sigmoid function. The corresponding high-frequency mask is defined as:

$$M_{high}(r) = 1 - M_{low}(r) \tag{3}$$

The spectrum is then decomposed into low-frequency and high-frequency components through element-wise multiplication:

$$\begin{cases} F_{low} = M_{low} \odot F \\ F_{high} = M_{high} \odot F \end{cases} \tag{4}$$

Finally, the frequency-domain features are mapped back to the spatial domain through the inverse Fourier transform:

$$\begin{cases} X_{low} = \mathcal{F}^{-1}(F_{low}) \\ X_{high} = \mathcal{F}^{-1}(F_{high}) \end{cases} \tag{5}$$

Where X_{low} denotes the low-frequency structural feature, which mainly characterizes the overall anatomical structure and semantic content of the image, while X_{high} emphasizes edge details and local texture patterns. In this way, the RFD module explicitly disentangles structural information from detail information during the encoding stage, thereby providing a clearer representational basis for subsequent feature fusion.

(3) Cross-Stream Synergistic Fusion Module

After passing through the RFD module, the encoder features are decomposed into a high-frequency feature stream and a low-frequency feature stream. Since these two types of features are highly complementary in terms of information representation, effectively integrating them becomes crucial for improving the overall model performance. To this end, we propose a Cross-Stream Synergistic Fusion (CSF) module.

Let the high-frequency and low-frequency features be denoted by F_h and F_l, respectively. The two feature streams are first concatenated, and global contextual information is then extracted global average pooling:

$$S = AvgPool(F_h \parallel F_l) \tag{6}$$

Next, a multilayer perceptron is used to generate two normalized gating weights:

$$\alpha, \beta = Softmax(MLP(S)) \tag{7}$$

Where $\alpha + \beta = 1$, and $\alpha$and βrepresent the contributions of the high-frequency and low-frequency features, respectively. The final fused feature is expressed as:

$$F_{fused} = Conv_{1\times1\times1}((\alpha F_h) \oplus (\beta F_l)) \tag{8}$$

Through this adaptive gating mechanism, the network can dynamically adjust the weights of high-frequency and low-frequency information according to the semantic requirements of different regions. For instance, in tumor boundary regions, the contribution of high-frequency features is automatically enhanced, whereas in the interior of the tumor, the network relies more heavily on low-frequency structural information. This mechanism not only reduces the redundancy introduced by naive concatenation, but also significantly improves the representational capacity of feature fusion.

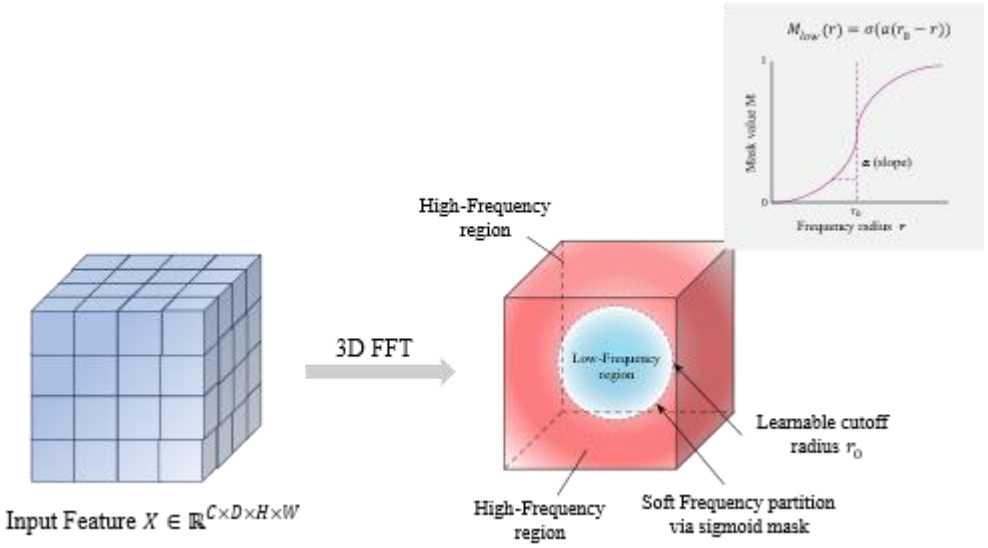

Figure 2. Radially Learnable Mask

(4) Multi-scale Tri-domain Gating Module

During the decoding stage, to further enhance feature reconstruction, we propose a Multi-scale Tri-domain Gating (MTG) module, as illustrated in Fig. 1(b). The module jointly models features in three complementary representation domains, namely the spatial domain, the Fourier domain, and the wavelet domain, so as to fully exploit the mutual complementarity across different domains. By enabling adaptive fusion of multi-domain information, the proposed module strengthens the network's ability to represent complex tumor structures and fine boundary details.

Let the input feature be denoted as $F_{in} \in \mathbb{R}^{C\times D\times H\times W}$. The MTG module first extracts contextual information through multi-scale convolutions. The resulting feature is then processed by three parallel branches, namely the Wavelet, Fourier, and Spatial branches, and dynamically fused through a gating mechanism.

First, the input feature is projected in the channel dimension, and local structural information at different scales is extracted by parallel depthwise convolutions:

$$F_{ms} = \text{Proj}(P_1(F_{in}) \parallel P_2(F_{in})) \tag{9}$$

$$P_i(F) = Conv_{1\times1\times1}[GeLU(DWConv_{5\times5\times5}(LN(F)) \parallel GeLU(DWConv_{3\times3\times3}(LN(F))))] \tag{10}$$

This design captures spatial structural information at multiple scales while maintaining computational efficiency. To further enlarge the receptive field, a dilated convolution is introduced for contextual modeling:

$$F_{mid} = Conv_{1\times1\times1}(LN(DilConv_{3\times3\times3}(F_{ms}))) \tag{11}$$

Where F_{mid} serves as the input feature for the subsequent Tri-domain gating process.

The wavelet branch is designed to capture local multi-scale edge information. Different from conventional wavelet transforms, this work adopts learnable wavelet filters to approximate wavelet-based feature extraction. The wavelet-domain representation is formulated as:

$$F_{wave} = \psi_{wave}(F_{mid}) \tag{12}$$

Where $\psi_{wave}(\cdot)$ denotes the learnable wavelet filtering operation, which can be written as:

$$\psi_{wave}(F) = Conv_{DoG}(F) + Conv_{MH}(F) \quad (13)$$

Where Conv_{DoG} denotes a Difference-of-Gaussian convolution kernel, and Conv_{MH} denotes a Mexican Hat convolution kernel. A wavelet gating map is then generated through a Sigmoid function:

$$W_w = \sigma\big(Conv_{1\times1\times1}(F_{wave})\big) \quad (14)$$

This weight map highlights regions with significant intensity variation, thereby enhancing edge structures and fine-grained texture information.

The Fourier branch is used to capture the global frequency structure of the image. Specifically, the feature is first transformed into the frequency domain by a three-dimensional Fourier transform:

Where $\mathcal{F}(\cdot)$ denotes the 3D FFT. The spectrum is then modulated by a learnable frequency weighting matrix:

$$\hat{F}_{mod} = W_f \odot \hat{F} \quad (15)$$

Where W_f is the learnable frequency weighting matrix and $\odot$ denotes element-wise multiplication. The spatial-domain feature is then recovered through the inverse Fourier transform:

$$F_{fourier} = \mathcal{F}^{-1}\big(\hat{F}_{mod}\big) \quad (16)$$

And the corresponding gating weight is computed as:

$$W_f = \sigma\Big(Conv_{1\times1\times1}\big(F_{fourier}\big)\Big) \quad (17)$$

This branch captures global structural patterns and long-range dependencies in the image.

The spatial branch is designed to preserve the original spatial feature information, and is formulated as:

$$F_{spatial} = Conv_{1\times1\times1}(F_{mid}) \quad (18)$$

With its corresponding gating weight defined as:

$$W_s = \sigma\Big(Conv_{1\times1\times1}\big(F_{spatial}\big)\Big) \quad (19)$$

This branch emphasizes spatial regions containing important semantic information.

After obtaining the feature weights from the three domains, gated fusion is performed as:

$$F_{out} = (F_{mid} \otimes W_w) + \big(F_{mid} \otimes W_f\big) + (F_{mid} \otimes W_s) \quad (20)$$

And the final output feature is given by:

$$F_{out} = Conv_{1\times1\times1}(F_{out}) \quad (21)$$

Where $\otimes$ enotes element-wise multiplication and $\oplus$ denotes element-wise addition. Through this Tri-domain gating mechanism, the network can adaptively regulate the contributions of spatial information, frequency information, and wavelet-based edge information according to the characteristics of different regions. As a result, it not only improves high-resolution feature reconstruction, but also enhances the representation of complex tumor structures and subtle boundary details.

4. Experiment

(1) Dataset and implementation details

Dataset. This study uses the publicly available datasets released by the Brain Tumor Segmentation (BraTS) Challenge, which is organized in conjunction with MICCAI. Specifically, BraTS 2019, BraTS 2020, and BraTS 2021 are employed for training and evaluation. These datasets consist of multi-institutional, multimodal brain MRI scans collected from real clinical settings and annotated at the voxel level by expert neuroradiologists, providing reliable reference labels for quantitative assessment. The scale and diversity of the datasets increase progressively across years, comprising 335, 369, and 1,251 cases for BraTS 2019, 2020, and 2021, respectively, as summarized in Table 1. Each case includes four MRI modalities, namely FLAIR, T1, T1ce, and T2, which offer complementary anatomical and pathological information. Since the official validation and test sets of BraTS do not provide publicly accessible annotations, the official training set is further divided into training, validation, and test subsets with a ratio of 8:1:1. This split is used only for local experiments, model selection, and early stopping, and does not compromise the reproducibility of the study.

Evaluation Metrics. Segmentation performance is quantitatively evaluated using two standard metrics widely adopted in medical image analysis: the Dice Similarity Coefficient (Dice) and the 95th percentile Hausdorff Distance (HD95). Dice measures the degree of overlap between the predicted mask and the reference annotation, with higher values indicating better segmentation accuracy. HD95 is used to assess boundary agreement by calculating the 95th percentile of the bidirectional surface distances between the predicted and ground-truth segmentations, thereby alleviating the influence of extreme outliers. In general, a higher Dice score and a lower HD95 value correspond to superior segmentation quality. The combination of these two metrics enables a comprehensive evaluation from both region-level and boundary-level perspectives.

Implementation Details. To improve data management efficiency and accelerate loading during training, the original NIfTI files are converted into the HDF5 format after preprocessing. Intensity normalization is performed using Z-score normalization on non-background voxels only. The segmentation labels are remapped into consecutive class indices $\{0,1,2,3\}$, corresponding to background, necrotic/non-enhancing tumor, edema, and enhancing tumor, respectively. Each preprocessed sample contains an $4 \times H \times W \times D$ image tensor stored in float32 format and a corresponding label map stored in uint8 format. All experiments are conducted on a Linux workstation using a single NVIDIA RTX 4090 GPU with 24 GB of memory. The model is implemented in PyTorch 2.5.0 and trained for 300 epochs with a batch size of 2.

Training Objective. Since the main contribution of this work lies in the proposed network architecture rather than the design of a novel optimization objective, we train TriFNet using a composite segmentation loss that combines weighted cross-entropy loss, region-specific Dice loss, and boundary-aware Focal loss. The weighted cross-entropy term alleviates class imbalance at the voxel level, the Dice term improves optimization consistency with the BraTS evaluation regions (WT, TC, and ET), and the boundary-aware Focal term enhances supervision on difficult boundary voxels. The overall loss is defined in Eqs. (22)–(25). Optimization is performed using the Adam optimizer with an initial learning rate of 0.001, a weight decay of 5×10^{-4}, and a cosine annealing learning rate schedule.

$$\mathcal{L}_{CE} = -\sum_{c=1}^{C} \omega_c \cdot y_c \cdot \log(p_c) \quad (22)$$

$$\mathcal{L}_{Dice} = \sum_{r\in WT,TC,ET} \lambda_r \cdot (1 - \mathcal{D}_r) \quad (23)$$

$$\mathcal{L}_{BFocal} = \frac{1}{|\sigma\Omega|}\sum_{i\in\sigma\Omega}(1-p_i)^{\gamma}\cdot\log p_i \quad (24)$$

$$\mathcal{L}_{seg} = (1-\alpha)\cdot\mathcal{L}_{CE} + \alpha\cdot\mathcal{L}_{Dice} + \beta\cdot\mathcal{L}_{BFocal} \quad (25)$$

Table 1. The details of the three BraTS datasets.

	BraTS 2019	BraTS 2020	BraTS 2021
Training set	268	295	1000
Validation set	33	37	125
Testing set	34	37	126
Total	335	369	1251

(2) Comparison results

To comprehensively evaluate our approach, state-of-the-art (SOTA) methods on the BraTS 2019, 2020, and 2021 datasets are included for comparison. Considering that numerous existing methods are not fully open-source and retraining could introduce reproducibility issues, we instead relied on published quantitative results for comparison. Although slight variations may arise due to implementation differences, this ensures an equitable and reproducible evaluation across all methods under a consistent benchmark. The evaluation methods are listed in Table 2. The quantitative evaluation results for TriFNet on the three BraTS benchmark datasets are presented in Table 3, Table 4, and Table 5, respectively. The best-performing values are highlighted in bold, while the second-best results are underlined.

Table 2. Brief description of the methods used for performance comparison in the experiment.

Method	Brief description
nnU-Net [5]	An automatic self-configuring U-Net framework that adapts preprocessing, architecture, and training to the dataset.
TransBTS [9]	Combines a CNN-based encoder with a Transformer bottleneck and CNN decoder for hybrid feature modeling.
Swin UNETR [10]	Employs Swin Transformer as encoder and CNN-based decoder in a U-shaped architecture for 3D segmentation tasks.
U-Net++ [14]	Enhances U-Net with nested and dense skip pathways to improve gradient flow and multi-scale feature fusion.
3D ResU-Net [15]	A 3D U-Net variant that incorporates residual connections to improve training stability and gradient propagation.
Swin-BTS [16]	Adopts 3D Swin Transformer blocks in the encoder to extract hierarchical features for brain tumor segmentation.
VT-UNet [17]	Introduces a Unet-shaped Volume Transformer to model volumetric relationships across slices using 3D attention mechanisms.
Segtransvae [18]	Combines Transformer-based encoder-decoder architecture with a variational autoencoder (VAE) for uncertainty-aware segmentation.
CKD-TransBTS [19]	Introduces clinical knowledge guidance into the TransBTS framework to better capture domain-specific tumor patterns.
ACMINet [20]	Proposes an aligned cross-modal interaction network for enhanced feature fusion from multiple MRI modalities.
SDV-TUNet [21]	Designs a sparse dynamic voxel-aware TransUNet to enhance segmentation of irregular tumor shapes with fewer parameters.
MSDMAT-BTS [22]	Proposes a multi-scale deformable attention-based BTS variant that captures rich context while maintaining model compactness.

Results on BraTS 2019. As shown in Table 3, TriFNet achieves the best overall performance on the BraTS 2019 dataset, with an average Dice of 86.62% and the lowest average HD95 of 3.32 mm among all competing methods. Compared with the 3D ResU-Net baseline, TriFNet improves the average Dice by 4.27 percentage points and reduces the average HD95 by 3.75 mm, demonstrating clear gains in both region overlap and boundary accuracy. Specifically, TriFNet attains the best WT Dice of 92.51%, outperforming the second-best MSDMAT-BTS (91.95%). In the TC region, although its Dice score (86.89%) is marginally lower than that of Segtransvae (86.91%), it achieves the lowest HD95 of 3.067 mm, indicating more accurate boundary localization in the tumor core. In the ET region, TriFNet obtains a Dice of 80.46%, slightly lower than SDV-TUNet (80.96%), yet its superior average Dice and HD95 indicate a more balanced overall segmentation performance. As illustrated in Fig. 3, TriFNet produces segmentation maps that are closer to the ground truth, with clearer and more continuous delineation of the tumor core and enhancing tumor regions. These results suggest that TriFNet not only preserves strong regional overlap, but also yields more reliable and consistent boundary reconstruction.

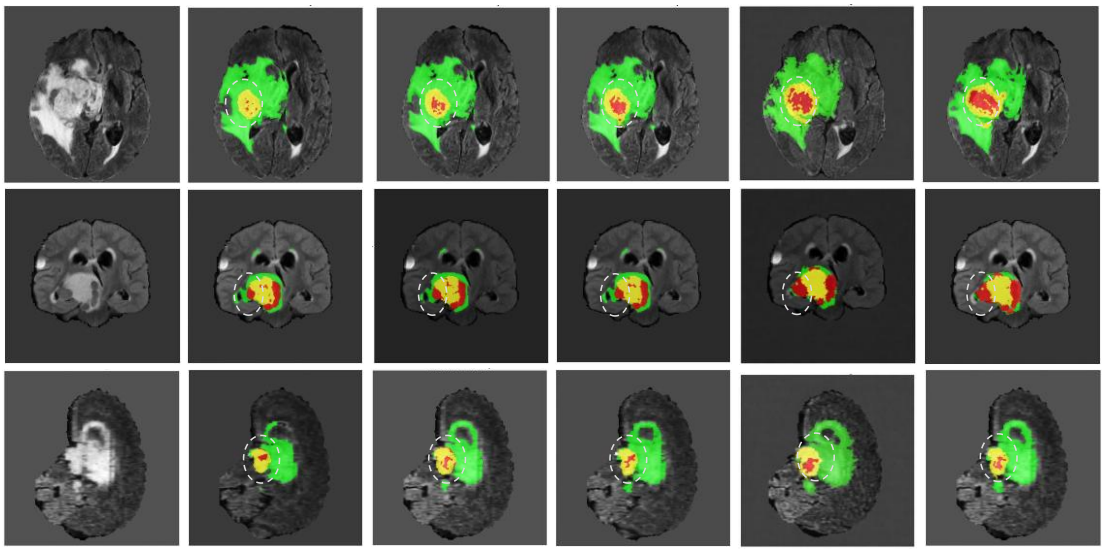

Figure 3. Visual comparison of brain tumor segmentation results by TriFNet and baseline models on the BraTS 2019 dataset.

Table 3. Performance Comparison of Different Methods on Dice and 95HD in BraTS 2019.

Method	WT		TC		ET		Average	
	Dice (%)	HD95 (mm)	Dice	HD95	Dice	HD95	Dice	HD95
3D ResU-Net	88.61	7.756	81.32	8.258	77.13	5.202	82.35	7.07
nnU-Net	90.12	6.345	84.21	5.521	78.32	3.313	84.21	5.06
TransBTS	90.01	5.644	81.94	7.599	78.93	3.736	83.62	5.14
Swin UNETR	90.85	5.013	85.63	4.862	79.65	3.421	85.38	4.43
ACMINet	90.21	5.322	84.71	5.103	79.22	3.881	84.71	4.77
Segtransvae	90.56	4.892	86.91	4.353	80.33	3.241	85.93	4.16
CKD-TransBTS	90.71	4.713	85.80	4.521	80.01	3.313	85.51	4.18
MSDMAT-BTS	91.95	3.922	86.52	4.201	79.87	3.112	86.11	3.74
SDV-TUNet	91.11	4.136	86.40	5.520	80.96	3.690	86.16	4.45
Ours (TriFNet)	92.51	3.462	86.89	3.067	80.46	3.441	86.62	3.32

Results on BraTS 2020. On the BraTS 2020 dataset, as reported in Table 4, TriFNet shows even more pronounced advantages, achieving the highest average Dice of 88.90% and the lowest average HD95 of 2.98 mm. Compared with the 3D ResU-Net baseline, the proposed model improves Dice by 6.66 percentage points and reduces HD95 by 14.08 mm, demonstrating excellent segmentation accuracy and strong generalization on a more challenging benchmark. Although MSDMAT-BTS attains the highest WT Dice (92.92%), TriFNet achieves the lowest WT HD95 of 2.431 mm, indicating more precise boundary localization for the whole tumor. More importantly, TriFNet achieves the best Dice scores in the two more challenging subregions, reaching 91.02% in TC and 83.89% in ET. In particular, the TC Dice exceeds that of the second-best SDV-TUNet by 1.82 percentage points, while the ET Dice is also the highest among all compared methods. The visual results in Fig. 4 further show that TriFNet produces more complete and smoother tumor contours with fewer false positives and false negatives in edema, tumor core, and enhancing regions. These findings confirm that TriFNet is especially effective in handling complex lesion structures and improving segmentation reliability in difficult subregions.

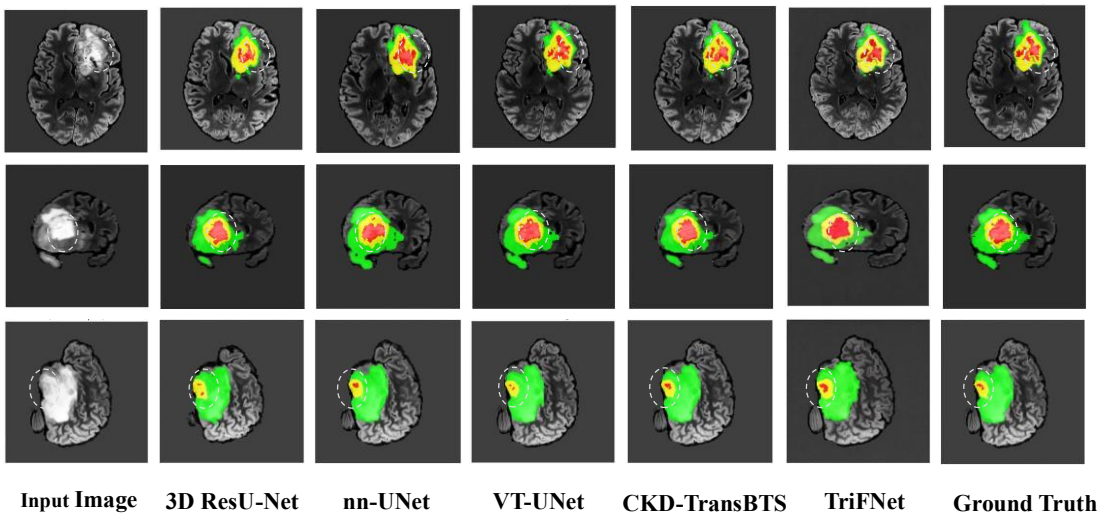

Figure 4. Qualitative visualization of brain tumor segmentation on the BraTS 2020 dataset.

Table 4. Performance Comparison of Different Methods on Dice and 95HD in BraTS 2020.

Method	WT		TC		ET		Average	
	Dice (%)	HD95 (mm)	Dice	HD95	Dice	HD95	Dice	HD95
3D ResU-Net	89.06	8.56	80.3	15.78	77.36	26.84	82.24	17.06
TransBTS	89.77	6.299	85.57	5.483	79.83	4.328	85.06	5.37
Swin UNETR	88.72	9.312	84.3	8.504	78.96	6.233	83.99	8.02
nnU-Net	88.95	8.501	85.06	17.34	82.03	17.81	85.35	14.55
ACMINet	90.61	4.45	84.7	8.63	81.13	17.5	85.48	10.19
Segtransvae	90.34	4.922	86.25	5.663	81.42	4.911	86.0	5.16
CKD-TransBTS	90.11	4.881	86.03	5.112	81.96	4.723	86.03	4.9
SDV-TUNet	90.22	4.032	89.2	3.302	82.48	2.297	87.3	3.21
MSDMAT-BTS	92.92	2.542	85.71	4.245	77.14	3.925	85.05	3.57
Ours (TriFNet)	91.79	2.431	91.02	3.084	83.89	3.412	88.90	2.98

Results on BraTS 2021. As presented in Table 5, TriFNet again delivers the best overall performance on the BraTS 2021 benchmark, achieving an average Dice of 91.88% and the lowest average HD95 of 3.51 mm among all competing methods. Compared with the 3D ResU-Net baseline, TriFNet improves the average Dice by 6.25 percentage points and reduces the average HD95 by 7.18 mm, which further verifies the effectiveness of the proposed multi-domain design. TriFNet obtains Dice scores of 93.47%, 93.16%, and 89.02% for WT, TC, and ET, respectively, all of which are the best results among the compared methods. Notably, the Dice in the TC region is 0.55 percentage points higher than that of the second-best SDV-TUNet, while the Dice in the ET region is improved by 1.38 percentage points, indicating a stronger capability in segmenting small, irregular, and boundary-ambiguous tumor regions. Although TriFNet does not achieve the minimum HD95 in every individual subregion, it maintains the most balanced performance across all categories and yields the best overall HD95. As shown in Fig. 5, TriFNet generates finer and more accurate boundaries, particularly in the TC and ET regions, producing segmentation results

that are visually closer to the ground truth. This consistent improvement across BraTS 2019, 2020, and 2021 demonstrates the robustness and generalizability of TriFNet for 3D brain tumor segmentation.

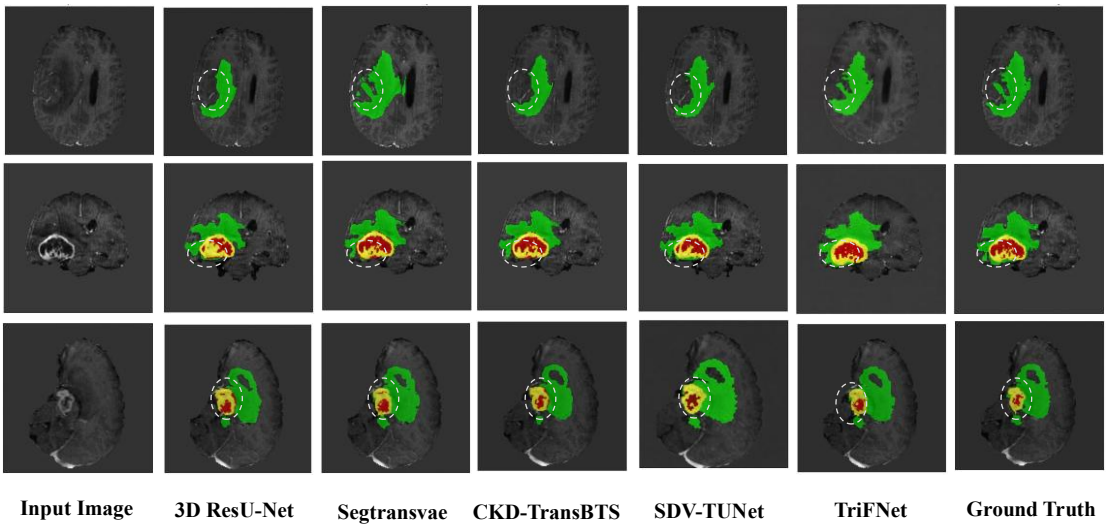

Figure 5. Qualitative visualization of brain tumor segmentation on the BraTS 2021 dataset.

Table 5. Performance Comparison of Different Methods on Dice and 95HD in BraTS 2021.

Method	WT		TC		ET		Average	
	Dice (%)	HD95 (mm)	Dice	HD95	Dice	HD95	Dice	HD95
3D ResU-Net	89.90	4.300	85.03	9.890	81.96	17.890	85.63	10.69
nn-UNet	92.75	3.470	87.81	7.623	84.51	20.730	88.36	10.61
TransBTS	91.93	4.121	87.23	5.912	84.90	6.831	88.02	5.62
Swin UNETR	92.60	5.831	88.50	3.770	85.80	6.016	88.97	5.20
ACMINet	91.20	10.03	87.41	6.590	85.59	6.230	88.07	7.52
Segtransvae	90.52	3.570	92.60	5.840	85.48	2.890	89.53	4.10
CKD-TransBTS	92.33	4.230	88.07	4.390	84.76	3.160	88.39	3.92
MSDMAT-BTS	93.21	3.881	89.52	4.352	85.21	3.411	89.31	3.88
SDV-TUNet	93.10	4.069	92.61	3.274	87.64	4.069	90.58	3.80
Ours (TriFNet)	93.47	3.914	93.16	3.201	89.02	3.428	91.88	3.51

(3) Ablation study

To validate the effectiveness of the proposed components, we conducted an ablation study on the BraTS 2021 dataset by progressively introducing RFD, CSF, and MTG into the baseline network. As reported in Table 6, all three modules consistently improve segmentation performance, while their full combination yields the best results.

When the RFD module is introduced alone, the average Dice increases from 85.63% to 88.20%, while the average HD95 decreases from 10.69 mm to 5.305 mm, indicating that explicit frequency decomposition substantially improves both feature discriminability and boundary localization. Incorporating CSF alone also leads to a clear gain, achieving an average Dice of 88.07%. However, its average HD95 remains 7.520 mm, suggesting that cross-stream fusion mainly enhances regional overlap but is less effective in boundary refinement when used independently. Similarly, adding MTG alone improves the average Dice to 87.96% and reduces the average HD95 to 5.672 mm, demonstrating the benefit of multi-domain gated decoding for contextual aggregation and structural recovery.

A further performance gain is observed when RFD and CSF are jointly employed. This configuration achieves an average Dice of 88.97% and an average HD95 of 5.206 mm, outperforming either single-module variant. This result confirms that frequency decomposition and cross-stream fusion are complementary, as they enhance feature representation and information interaction from different perspectives.

The full model, integrating RFD, CSF, and MTG, achieves the best overall performance, with Dice scores of 93.47%, 93.16%, and 89.02% for WT, TC, and ET, respectively, corresponding to an average Dice of 91.88%. Meanwhile, the HD95 values are reduced to 3.914 mm, 3.201 mm, and 3.428 mm, with an average of only 3.51 mm. Compared with the baseline, the complete TriFNet improves the average Dice by 6.25 percentage points and reduces the average HD95 by 7.18 mm, demonstrating the effectiveness of jointly optimizing Frequency-aware encoding, cross-stream feature fusion, and Tri-domain gated reconstruction.

Notably, the largest improvement is achieved in the ET region, where the Dice score increases from 81.96% to 89.02%, and the HD95 decreases sharply from 17.890 mm to 3.428 mm. Since ET is typically characterized by small volume, irregular shape, and ambiguous boundaries, this result further highlights the superiority of TriFNet in handling challenging tumor subregions. Overall, the ablation results confirm that the three proposed modules are individually beneficial and jointly contribute to more accurate and robust brain tumor segmentation.

Table 6. Ablation study results of TriFNet on the BraTS 2021 dataset under different module configurations.

Modules				WT		TC		ET		Average	
Baseline	RFD	CSF	MTG	Dice (%)	HD95 (mm)	Dice	HD95	Dice	HD95	Dice	HD95
√				89.90	4.300	85.03	9.890	81.96	17.890	85.63	10.69
√	√			91.14	6.003	88.43	4.512	85.02	5.401	88.20	5.305
√		√		91.20	10.03	87.41	6.590	85.59	6.230	88.07	7.520
√			√	90.95	6.245	88.02	4.780	84.91	5.990	87.96	5.672
√	√	√		92.60	5.831	88.50	3.770	85.80	6.016	88.97	5.206
√	√	√	√	93.47	3.914	93.16	3.201	89.02	3.428	91.88	3.51

5. Conclusion

In this paper, we proposed TriFNet, a Tri-domain Frequency-aware network for 3D brain tumor segmentation. The proposed framework integrates three key components: the RFD module for explicit decomposition of low-frequency structural information and high-frequency detail information in the encoder, the CSF module for adaptive cross-stream feature interaction at the bottleneck, and the MTG module for spatial-Fourier-wavelet collaborative reconstruction in the decoder. Through this unified design, TriFNet strengthens structural modeling, complementary feature fusion, and boundary-aware decoding within a single end-to-end framework.

Extensive experiments on the BraTS 2019, BraTS 2020, and BraTS 2021 datasets demonstrate that the proposed method achieves consistently competitive and robust performance, yielding superior average Dice and HD95 results compared with representative state-of-the-art methods. In particular, TriFNet shows clear advantages in challenging tumor subregions such as tumor core and enhancing tumor, indicating its effectiveness in handling small, irregular, and boundary-ambiguous lesions. In addition, the ablation study confirms that the three proposed modules are individually beneficial and jointly complementary.

Overall, the results validate the effectiveness of combining frequency decomposition, cross-stream synergistic fusion, and Tri-domain gated reconstruction for 3D brain tumor segmentation. In future work, we will further investigate lightweight model design, cross-dataset generalization, and broader validation in real clinical scenarios to improve the practical applicability of the proposed framework.

References

[1] J. Yu, Q. Liu, C. Xu, Q. Zhou, J. Xu, L. Zhu, C. Chen, Y. Wang, H. Zhang, Deep learning-driven modality imputation and subregion segmentation to enhance high-grade glioma grading, BMC Med. Inform. Decis. Mak.25 (2025) 1–16.

[2] A. Wadhwa, A. Bhardwaj, V.S. Verma, A review on brain tumor segmentation of MRI images, Magn. Reson. Imaging. 61 (2019) 247–259.

[3] Ö. Çiçek, A. Abdulkadir, S.S. Lienkamp, T. Brox, O. Ronneberger, 3D U-Net: Learning dense volumetric segmentation from sparse annotation, in: Proc. Int. Conf. Med. Image Comput. Comput. Assist. Interv., Athens, Greece, 2016, pp. 424–432.

[4] F. Milletari, N. Navab, S.-A. Ahmadi, V-Net: Fully convolutional neural networks for volumetric medical image segmentation, in: 2016 Fourth Int. Conf. 3D Vis., Stanford, CA, USA, 2016, pp. 565-571.

[5] F. Isensee, P. Kickingereder, W. Wick, M. Bendszus, K.H. Maier-Hein, nnU-Net for brain tumor segmentation, in: Brainlesion: Glioma, Multiple Sclerosis, Stroke Traumatic Brain Injuries (LNCS, vol. 12658), Cham: Springer, 2021, pp. 118–132.

[6] Z. Liu, L. Tong, L. Chen, W. Wang, F. Li, Deep learning based brain tumor segmentation: a survey, Complex Intell. Syst.9 (2023) 1001–1026.

[7] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, Q. Hu, ECA-Net: Efficient channel attention for deep convolutional neural networks, in: Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., 2020, pp. 11534–11542.

[8] S. Woo, J. Park, J.Y. Lee, I.S. Kweon, CBAM: Convolutional block attention module, in: Proc. Eur. Conf. Comput. Vis., Munich, Germany, 2018, pp. 3–19.

[9] W. Wang, C. Chen, M. Ding, H. Yu, S. Zha, J. Li, TransBTS: Multimodal brain tumor segmentation using transformer, in: Proc. Int. Conf. Med. Image Comput. Comput. Assist. Interv., Strasbourg, France, 2021, pp. 109–119.

[10] A. Hatamizadeh, V. Nath, Y. Tang, D. Yang, D. Xu, Swin UNETR: Swin transformers for semantic segmentation of brain tumors in MRI images, in: Proc. Int. Conf. Med. Image Comput. Comput. Assist. Interv., Singapore, 2022, pp. 272–284.

[11] Z. Liu, H. Hu, Y. Lin, Z. Yao, Z. Xie, Y. Wei, J. Cao, Z. Zhang, L. Zhang, H. Dong, J. Guo, B. Guo, Swin transformer V2: Scaling up capacity and resolution, in: Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., New Orleans, LA, USA, 2022, pp. 12009–12019.

[12] Y. Jiang, Y. Huang, X. Li, Z. Wang, M. Wang, GLIM: Global and local attention for medical image segmentation, IEEE Trans. Med. Imaging42 (2023) 1298–1310.

[13] Z. Lu, C. She, W. Wang, Y. Li, X. Zhang, LM-Net: A light-weight and multi-scale network for medical image segmentation, Comput. Biol. Med.168 (2024) 107717.

[14] Z. Zhou, M.M.R. Siddiquee, N. Tajbakhsh, J. Liang, UNet++: a nested U-Net architecture for medical image segmentation, in: Proc. Deep Learn. Med. Image Anal. Multimodal Learn. Clin. Decis. Support, 2018, pp. 3–11.

[15] X. Li, J. Zhang, Y. Chen, K. Wang, L. Liu, 3D ResUNet: a residual encoder-decoder network for brain tumor segmentation, in: Proc. IEEE Int. Symp. Biomed. Imaging, 2022, pp. 1–5.

[16] H. Wang, Y. Lu, Y. Wang, Z. Han, Q. Yang, Swin-BTS: a Swin Transformer-based brain tumor segmentation network, in: Proc. Int. Conf. Med. Image Comput. Comput. Assist. Interv., 2022, pp. 467–477.

[17] J. Gao, Y. Tang, H. Lin, X. Yang, Z. Li, Y. Zhang, VT-UNet: a Vision Transformer-based UNet for 3D brain tumor segmentation, Med. Image Anal.84 (2023) 102683.

[18] W. Huang, X. Zhao, Y. Ma, L. Chen, H. Li, S. Wang, SegTransVAE: a Transformer and VAE based hybrid model for brain tumor segmentation, Comput. Biol. Med.152 (2023) 106355.

[19] Y. Liu, M. Zhang, H. Liu, Z. Wu, K. Chen, L. Zhao, CKD-TransBTS: clinical knowledge-driven TransBTS for brain

tumor segmentation, in: Proc. Int. Conf. Med. Image Comput. Comput. Assist. Interv., 2023, pp. 234–245.

[20] Y. Zhang, L. Wang, X. Xu, H. Liu, B. Yu, ACMINet: aligned cross-modal interaction network for brain tumor segmentation, IEEE Trans. Med. Imaging42 (2023) 2105–2116.

[21] Z. Zhu, Sparse dynamic volume TransUNet with multi-level edge fusion for brain tumor segmentation, Comput. Biol. Med. (2024).

[22] X. Wu, J. Wang, H. Chen, Y. Li, K. Zhang, M. Liu, MSDMAT-BTS: Multi-scale deformable attention transformer for brain tumor segmentation, in: Proc. Int. Conf. Med. Image Comput. Comput. Assist. Interv., 2024, pp. 100–111.