MCMA-Net++: Topology-Aware and Graph-Driven Glioma Segmentation in 3D MRI

Jihan Alameddin[1] JIHAN.ALAMEDDINE@UNIV-POITIERS.FR
[1] *XLIM, CNRS UMR 7252 / LabCom I3M, University of Poitiers,Poitiers, France*
Céline Thomarat[2] CELINE.THOMARAT@CHU-POITIERS.FR
[2] *University Hospital center of Poitiers, Department of Imaging / Labcom I3M, University of Poitiers, Poitiers, France*
Rémy Guillevin[3] REMY.GUILLEVIN@CHU-POITIERS.FR
[3] *University Hospital center of Poitiers, Department of Imaging, University of Poitiers, Laboratoire de Mathématiques Appliquées LMA, DACTIM-MIS team, CNRS 7348, Poitiers, France*
Christine Fernandez-Maloigne[1] CHRISTINE.FERNANDEZ@UNIV-POITIERS.FR
Carole Guillevin[3] CAROLE.GUILLEVIN@CHU-POITIERS.FR

Editors: Under Review for MIDL 2026

Abstract

Glioma segmentation in 3D MRI remains challenging due to tumor heterogeneity, intensity profiles, and hierarchical anatomical structure. We propose MCMA-Net++, introducing two synergistic innovations: (1) Topology-Aware Refinement Loss (TAR-Loss), enforcing topological consistency across nested tumor subregions (ET, TC, WT), and (2) Multi-Scale Anatomical Graph Reasoning (MSAGR), explicitly modeling spatial dependencies through learnable graphs with anatomical priors. Combined with dual-stream CNN-Swin Transformer encoding and Multi-Class Multi-Attention, MCMA-Net++ achieves Dice scores of 0.970 (WT), 0.943 (TC), 0.926 (ET), while reducing HD95 from 5.48mm to 3.21mm compared to MCMA-Net. Graph reasoning contributes +1.3% Dice for ET and TAR-Loss reduces topology violations by 41%. These results highlight the relevance of combining topology-guided refinement and anatomical graph reasoning for 3D brain tumor analysis, and position MCMA-Net++ as a strong candidate for clinical-grade glioma segmentation.

Keywords: Glioma segmentation, Topology-aware learning, Graph neural networks, Hybrid CNN-Transformer, Medical image analysis

1. Introduction

Gliomas represent the most frequent and aggressive primary brain tumors, accounting for approximately 80% of malignant brain neoplasms. Accurate segmentation of their subregions, Enhancing Tumor (ET), Tumor Core (TC), and Whole Tumor (WT), from multiparametric MRI is essential for diagnosis, treatment planning, surgery guidance, and assessment of therapeutic response [9; 19]. Despite the remarkable progress achieved by

convolutional neural networks and Transformer-based architectures, automated glioma segmentation remains a challenging task due to the intrinsic heterogeneity of gliomas, the large variability in shape and appearance across patients, and the complex hierarchical relationships between tumor subregions [27; 28].

State-of-the-art deep learning models[23; 14; 10]primarily optimize voxel-wise similarity metrics such as Dice or cross-entropy losses. However, these objectives do not explicitly enforce anatomical or topological correctness, often leading to predictions with holes, fragmented regions, or violations of the known hierarchical nesting between ET, TC, and WT. Such inconsistencies degrade clinical interpretability and reliability, particularly for the enhancing tumor region, which is crucial for grading and therapeutic monitoring. Recent works have begun addressing these issues through structural priors or post-processing constraints, yet a unified framework capable of jointly modeling spatial dependencies and enforcing topology-aware consistency remains lacking [3; 16].

To bridge this gap, we propose MCMA-Net++, a topology-guided and graph-driven framework for 3D glioma segmentation in MRI. Building upon multi-scale CNN–Transformer feature extraction, MCMA-Net++ introduces two synergistic contributions. First, we design a Topology-Aware Refinement Loss (TAR-Loss) that enforces anatomically plausible relationships across nested subregions, reducing topological violations and improving structural coherence. Second, we incorporate Multi-Scale Anatomical Graph Reasoning (MSAGR), which explicitly models spatial dependencies between tumor components through learnable graphs infused with anatomical priors. This enables the network to reason over inter-region relationships beyond local voxel context.

Together, these components allow MCMA-Net++ to produce more coherent, robust, and interpretable segmentations than previous approaches [1]. In this work, we evaluate our method on 3D MRI data and demonstrate its superiority over MCMA-Net and other recent baselines, both in terms of voxel-wise accuracy and topological correctness. Our findings highlight the importance of integrating topology-aware constraints and anatomical graph reasoning in modern segmentation architectures for brain tumor analysis.

2. Related Work

Deep learning has transformed glioma segmentation, evolving from pure CNN architectures to sophisticated hybrid models. The U-Net architecture [23] revolutionized medical image segmentation with its encoder-decoder structure and skip connections, with 3D extensions [4] and V-Net [21] dominating early BraTS challenges. nnU-Net [14] established a new standard through automated architecture configuration and robust preprocessing pipelines, achieving state-of-the-art results across diverse medical imaging tasks. 3D U-Net++ [32] introduced nested skip connections and deep supervision to capture features at multiple scales. However, CNNs suffer from limited receptive fields requiring very deep architectures to capture global context. Recent innovations include residual connections [11], dense connections [12], and attention gates [22].

Vision Transformers [6] demonstrated that self-attention mechanisms can effectively model long-range dependencies, leading to medical imaging adaptations including UNETR [10] with pure transformer encoders, TransBTS [29] for brain tumor segmentation, and Swin-Unet [20] with hierarchical shifted window attention. SwinUNet3D extended the Swin

Transformer architecture to 3D medical volumes, demonstrating improved performance through hierarchical feature representations. While transformers excel at global modeling, they often struggle with fine-grained local details. This motivated hybrid CNN-Transformer architectures like CoTr [30] with deformable self-attention, UTNet [8], and MISSFormer [13] with multi-scale feature aggregation. nnFormer [31] introduced interleaved attention mechanisms for volumetric segmentation, achieving strong results on BraTS through effective fusion of local and global features.

Graph-based methods have shown promise in modeling anatomical relationships, with Graph U-Net [7] applying hierarchical graph pooling and anatomical graph learning [25] for organ relationship modeling. Graph attention networks [26] and graph convolutional networks [17] have been widely adopted in various domains. Topological consistency has been addressed through persistent homology [5], clDice [24] using centerline-based metrics, and boundary-aware losses [15]. Hierarchical constraints have been explored through conditional random fields [18] for post-processing. MCMA-Net [1] introduced class-specific multi-attention mechanisms with hierarchical consistency loss. MCMA-Net++ uniquely combines hybrid CNN-Transformer encoding, explicit graph-based spatial reasoning, practical topology-aware learning, and hierarchical consistency enforcement into a unified framework specifically designed for nested glioma subregion segmentation.

3. Methodology

3.1. Problem Formulation

Given a 3D multi-modal MRI scan $\mathbf{X} \in \mathbb{R}^{C\times H\times W\times D}$ with $C = 4$ modalities (T1, T1ce, T2, FLAIR), we predict a segmentation map $\mathbf{Y} \in \{0, 1, 2, 3\}^{H\times W\times D}$ where labels represent background (0), necrotic/non-enhancing tumor core (1), peritumoral edema (2), and enhancing tumor (3). From these labels, hierarchical regions are derived: Whole Tumor (WT: $\{1, 2, 3\}$), Tumor Core (TC: $\{1, 3\}$), and Enhancing Tumor (ET: $\{3\}$). The anatomical constraint ET $\subset$ TC $\subset$ WT must be satisfied, as violations are clinically implausible.

3.2. Overall Architecture

MCMA-Net++ follows a sequential pipeline: Input $\rightarrow$ Dual-Stream Encoder $\rightarrow$ MCMA Module $\rightarrow$ MSAGR Module $\rightarrow$ Decoder $\rightarrow$ Output. The dual-stream encoder combines CNNs (local features, translation equivariance) with Swin Transformers (global context, long-range dependencies) through adaptive gated fusion. The training objective combines multiple loss terms:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{Dice}} + \lambda_1\mathcal{L}_{\text{HiC}} + \lambda_2\mathcal{L}_{\text{TAR}} + \lambda_3\mathcal{L}_{\text{CE}} \tag{1}$$

where $\mathcal{L}_{\text{Dice}}$ is the multi-class Dice loss for volumetric overlap, $\mathcal{L}_{\text{HiC}}$ is the hierarchical consistency loss enforcing the nested structure constraint, $\mathcal{L}_{\text{TAR}}$ is our TAR loss (Section 3.6), and $\mathcal{L}_{\text{CE}}$ is the cross-entropy loss for voxel-wise classification. The weighting coefficients $\lambda_1 = 0.5$, $\lambda_2 = 0.3$, and $\lambda_3 = 0.1$ balance the contribution of each term, with higher weights assigned to constraints more critical for anatomical consistency.

3.3. Dual-Stream 3D Encoder

The dual-stream 3D encoder comprises:

- **CNN Branch:** Employs residual blocks with $3 \times 3 \times 3$ kernels:

$$\text{CNN\_block}(\mathbf{x}) = \mathbf{x} + \text{BN}(\text{ReLU}(\text{Conv3D}(\text{Conv3D}(\mathbf{x})))) \tag{2}$$

where $\mathbf{x}$ is the input feature map, Conv3D denotes 3D convolution operations, ReLU is the rectified linear unit activation function, and BN represents batch normalization for stabilizing training.

- **Swin Transformer Branch:** Uses 3D Shifted Window Multi-head Self-Attention (SW-MSA) with window size $7 \times 7 \times 7$. Attention is computed as:

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{Softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_k}} + \mathbf{B}\right)\mathbf{V} \tag{3}$$

where $\mathbf{Q}$, $\mathbf{K}$, $\mathbf{V}$ are query, key, and value matrices obtained through linear projections of input features, d_k is the dimension of the key vectors used for scaling to prevent gradient vanishing, and $\mathbf{B}$ is a learned relative position bias matrix.

- **Feature Fusion:** At each scale, features are fused through adaptive gating:

$$\mathbf{F}_{\text{fused}} = \sigma(\mathbf{W}_c \cdot \mathbf{F}_{\text{CNN}}) \odot \mathbf{F}_{\text{CNN}} + \sigma(\mathbf{W}_t \cdot \mathbf{F}_{\text{Trans}}) \odot \mathbf{F}_{\text{Trans}} \tag{4}$$

where $\mathbf{F}_{\text{CNN}}$ and $\mathbf{F}_{\text{Trans}}$ are the feature maps from CNN and Transformer branches respectively, σ is the sigmoid activation function, $\mathbf{W}_c$ and $\mathbf{W}_t$ are learnable weight matrices implemented as $1 \times 1 \times 1$ convolutions, and $\odot$ denotes element-wise multiplication.

3.4. Multi-Class Multi-Attention (MCMA) Module

The MCMA module applies three complementary attention mechanisms.

- **Channel attention** recalibrates features through:

$$\mathbf{M}_c = \sigma(\text{MLP}(\text{AvgPool}(\mathbf{F})) + \text{MLP}(\text{MaxPool}(\mathbf{F}))) \tag{5}$$

where $\mathbf{F}$ is the input feature map, AvgPool and MaxPool are global average and max pooling operations, MLP is a multi-layer perceptron, and $\mathbf{M}_c$ is the channel attention map that recalibrates feature channels.

- **Spatial attention** identifies relevant locations via:

$$\mathbf{M}_s = \sigma(\text{Conv}_7([\text{AvgPool}_c(\mathbf{F}); \text{MaxPool}_c(\mathbf{F})])) \tag{6}$$

where Conv_7 is a $7 \times 7 \times 7$ convolution capturing local spatial context, and $\mathbf{M}_s$ is the spatial attention map highlighting tumor-relevant regions.

• **Class-specific attention** learns separate mechanisms for each region (WT, TC, ET):

$$\mathbf{M}_c^{\text{class}} = \text{Softmax}\left(\frac{\mathbf{W}_c^Q\mathbf{F} \cdot (\mathbf{W}_c^K\mathbf{F})^T}{\sqrt{d}}\right) \cdot \mathbf{W}_c^V\mathbf{F} \tag{7}$$

where $\mathbf{W}_c^Q$, $\mathbf{W}_c^K$, $\mathbf{W}_c^V$ are class-specific learnable projection matrices for queries, keys, and values, d is the feature dimension for scaling, and $\mathbf{M}_c^{\text{class}}$ produces specialized features tailored to each tumor subregion's imaging characteristics.

• **Inter-class gating** enforces hierarchical constraints through:

$$\mathbf{G}_{\text{ET}\to\text{TC}} = \sigma(\text{Conv}(\text{Concat}(\mathbf{F}_{\text{ET}}, \mathbf{F}_{\text{TC}}))) \tag{8}$$

where $\mathbf{F}_{\text{ET}}$ and $\mathbf{F}_{\text{TC}}$ are class-specific feature maps, Concat denotes concatenation, and $\mathbf{G}_{\text{ET}\to\text{TC}}$ is a gating mask that modulates ET predictions based on TC evidence to enforce the anatomical constraint ET $\subset$ TC.

3.5. Multi-Scale Anatomical Graph Reasoning (MSAGR)

MSAGR explicitly models spatial relationships through graph-based reasoning with anatomical priors at multiple scales.

• **Graph Construction:** Feature maps are partitioned into tokens $\mathbf{T}^l = \{t_1^l, \ldots, t_N^l\}$ serving as graph nodes. Edge weights combine learned similarity:

$$e_{ij} = \frac{\exp(\theta(\mathbf{h}_i)^T\phi(\mathbf{h}_j)/\tau)}{\sum_k \exp(\theta(\mathbf{h}_i)^T\phi(\mathbf{h}_k)/\tau)} \tag{9}$$

where $\mathbf{h}_i$ and $\mathbf{h}_j$ are node feature vectors, θ and ϕ are linear projection functions, τ is a temperature parameter (set to 0.1), and k indexes all neighboring nodes with anatomical priors:

$$e_{ij}^{\text{final}} = e_{ij} \cdot \exp\left(-\frac{\|\text{pos}_i - \text{pos}_j\|^2}{2\sigma^2}\right) \cdot (1 + \cos(\mathbf{P}_i, \mathbf{P}_j)) \tag{10}$$

where pos_i and pos_j are the spatial positions of nodes i and j, σ^2 controls the spatial distance decay, $\mathbf{P}_i$ and $\mathbf{P}_j$ are predicted probability vectors for each node, and $\cos(\cdot,\cdot)$ computes cosine similarity encouraging connections between regions with similar predicted classes.

• **Graph Attention:** Multi-head graph attention updates node features:

$$\mathbf{h}_i^{(l+1)} = \Big\|_{k=1}^{K} \sigma\left(\sum_{j\in\mathcal{N}(i)} \alpha_{ij}^k \mathbf{W}^k \mathbf{h}_j^l\right) \tag{11}$$

where $\mathbf{h}_i^{(l)}$ is the feature vector of node i at layer l, $K = 4$ is the number of attention heads, $\mathcal{N}(i)$ denotes the neighborhood of node i, α_{ij}^k are learned attention coefficients derived from

e_{ij}^{final}, $\mathbf{W}^k$ are head-specific learnable weight matrices, and σ is LeakyReLU activation.

• **Multi-scale aggregation** processes graphs at three scales (32^3, 16^3, 8^3 tokens) in parallel:

$$\mathbf{H}^{\text{final}} = \text{Concat}(\mathbf{G}^1, \mathbf{G}^2, \mathbf{G}^3) \cdot \mathbf{W}_{\text{agg}} \tag{12}$$

where $\mathbf{G}^1$, $\mathbf{G}^2$, $\mathbf{G}^3$ are output representations from fine (32^3), medium (16^3), and coarse (8^3) scale graphs respectively, and $\mathbf{W}_{\text{agg}}$ is a learnable aggregation matrix that fuses multi-scale information.
Features are refined through residual integration:

$$\mathbf{F}_{\text{refined}} = \mathbf{F}_{\text{encoder}} + \text{Conv3D}(\text{Reshape}(\mathbf{H}^{\text{final}})) \tag{13}$$

where $\mathbf{F}_{\text{encoder}}$ are the original encoder features, and Reshape maps graph node features back to dense 3D spatial grid.

3.6. Topology-Aware Refinement Loss (TAR-Loss)

TAR-Loss comprises four complementary terms enforcing multi-level topological consistency:

• **Hierarchical Inclusion Loss:**

$$\mathcal{L}_{\text{inc}} = \frac{1}{N} \sum_{v} [\text{ReLU}(P_{\text{ET}}(v) - P_{\text{TC}}(v)) + \text{ReLU}(P_{\text{TC}}(v) - P_{\text{WT}}(v))] \tag{14}$$

where N is the total number of voxels, and v indexes individual voxels, $P_{\text{ET}}(v)$, $P_{\text{TC}}(v)$, $P_{\text{WT}}(v)$ are predicted probabilities for each hierarchical region at voxel v.

• **Boundary Coherence Loss:**

$$\mathcal{L}_{\text{bnd}} = \frac{1}{|B|} \sum_{v \in B} \|\nabla P_{\text{ET}}(v) - \alpha \cdot \nabla P_{\text{TC}}(v)\|^2 \tag{15}$$

where B is the set of boundary voxels identified from ground truth (GT) where class labels change, ∇ denotes spatial gradient operator, α is a learnable scaling factor, and $\|\cdot\|^2$ is the squared L2 norm.

• **Surface Smoothness Loss:**

$$\mathcal{L}_{\text{smooth}} = \frac{1}{N} \sum_{v} \sum_{u \in \mathcal{N}_{26}(v)} \|\mathbf{P}(v) - \mathbf{P}(u)\|^2 \cdot w(v, u) \tag{16}$$

where $\mathcal{N}_{26}(v)$ is the 26-connected neighborhood of voxel v, $\mathbf{P}(v)$ is the complete probability vector at voxel v, and $w(v, u)$ is an edge-aware weight.

• **Connected Component Consistency:**

$$\mathcal{L}_{\text{cc}} = \sum_{c} \lambda_c \cdot (\text{CC}(P_c) - 1)^2 \tag{17}$$

where λ_c are class-specific weights, and $\mathrm{CC}(P_c)$ counts the number of connected components in thresholded predictions for class c.

The combined loss is:

$$\mathcal{L}_{\mathrm{TAR}} = \beta_1 \mathcal{L}_{\mathrm{inc}} + \beta_2 \mathcal{L}_{\mathrm{bnd}} + \beta_3 \mathcal{L}_{\mathrm{smooth}} + \beta_4 \mathcal{L}_{\mathrm{cc}} \tag{18}$$

with weights $\beta_1 = 1.0$, $\beta_2 = 0.5$, $\beta_3 = 0.3$, $\beta_4 = 0.2$.

4. Experimental Setup

We evaluate MCMA-Net++ on the BraTS 2021 dataset [2], which contains 1,251 multi-parametric MRI scans (1,000 training, 125 validation, 126 testing) with four co-registered modalities (T1, T1ce, FLAIR, T2) at $1 \times 1 \times 1$ mm^3 resolution. Data augmentations include random flipping, rotation ($\pm 15°$), scaling (0.9–1.1), elastic deformation ($\alpha = 300$, $\sigma = 20$), Gaussian noise ($\sigma = 0.05$), intensity shifts (± 0.1), and gamma correction (0.8–1.2).

MCMA-Net++ is implemented in PyTorch 2.0.1 with CUDA 11.8, trained on four NVIDIA A100 GPUs for 200 epochs (batch size 8, 12 hours) using AdamW ($\beta_1 = 0.9$, $\beta_2 = 0.999$, weight decay 10^{-4}, learning rate 10^{-4} with cosine annealing). The network has encoder depths [2,2,2,2], hidden dimensions [32,64,128,256], transformer heads [3,6,12,24], window size 7^3, and graph scales 32^3, 16^3, 8^3 with 4 attention heads.

Segmentation is evaluated using Dice Similarity Coefficient (volumetric overlap), 95th percentile Hausdorff Distance (HD95, boundary accuracy), and Topology Violation Rate (anatomical plausibility).

5. Results

5.1. Quantitative Comparison with State-of-the-Art

Table 1 presents the comparison with state-of-the-art methods on BraTS 2021. MCMA-Net++ achieves Dice scores of 0.970 (WT), 0.943 (TC), and 0.926 (ET), with HD95 of 3.21 mm. The most significant improvement is observed for Enhancing Tumor, with +2.6% over MCMA-Net and +5.3% over nnFormer. This gain is attributed to MSAGR's explicit spatial reasoning and the hierarchical inclusion loss preventing anatomically implausible predictions. Boundary precision shows a 41% HD95 reduction compared to MCMA-Net (3.21 vs 5.48 mm), substantially surpassing inter-rater variability of 5-7 mm. These improvements are achieved with competitive model complexity (68.4M parameters), substantially lower than nnFormer (148.4M).

Figure 1 presents a challenging case (BraTS2021_00048) with complex tumor morphology across three orthogonal anatomical planes, exhibiting subtle contrast enhancement transitions that are particularly challenging for automated segmentation.

Comparative analysis reveals key improvements in MCMA-Net++. On FLAIR images, MCMA-Net produces fragmented edema predictions with irregular boundaries, while MCMA-Net++ generates smoother, anatomically coherent contours matching ground truth through the surface smoothness term in $\mathcal{L}_{\mathrm{TAR}}$. On T1ce images, MCMA-Net exhibits false positive predictions outside the tumor core, which MCMA-Net++ eliminates through MSAGR's graph-based spatial reasoning enforcing anatomical consistency. Across all views,

Table 1: Quantitative comparison with state-of-the-art methods on BraTS 2021.

Model	WT Dice	TC Dice	ET Dice	HD95	Params
nnU-Net	0.888	0.838	0.782	11.4	31.2M
3D U-Net++	0.901	0.849	0.801	10.1	28.6M
UNETR	0.928	0.871	0.862	8.9	92.8M
SwinUNet3D	0.931	0.872	0.859	8.7	41.3M
nnFormer	0.934	0.881	0.873	8.2	148.4M
MCMA-Net	0.950	0.922	0.900	5.48	62.3M
MCMA-Net++	**0.970**	**0.943**	**0.926**	**3.21**	68.4M

MCMA-Net++ maintains strict hierarchical constraints ET $\subset$ TC $\subset$ WT, eliminating disconnected components and preserving fine structural details at tumor margins critical for surgical planning and radiotherapy optimization.

The topology analysis in Table 2 quantifies this improvement, showing MCMA-Net++ reduces topology errors from 14.3% to 8.4% (41% relative reduction), with average violations per case decreasing from 1.6 to 0.9. This stems from the synergistic combination of $\mathcal{L}_{\text{inc}}$, inter-class gating, and graph-based reasoning that collectively enforce the constraint ET $\subset$ TC $\subset$ WT at multiple architectural levels. The remaining 8.4% violations occur primarily in extreme cases with highly irregular tumor morphology or severe imaging artifacts.

Table 2: Topology violation analysis.

Model	Topology Errors (%)	Avg Violations/Case
nnU-Net	23.4	2.8
UNETR	18.7	2.1
MCMA-Net	14.3	1.6
MCMA-Net++	**8.4**	**0.9**

5.2. Ablation Studies

Table 3 presents comprehensive ablation results validating each component's contribution. From the MCMA-Net baseline (0.900 ET Dice, 5.48 mm HD95, 14.3% topology errors), adding MSAGR alone improves to 0.912 ET Dice, 4.21 mm HD95, and 11.1% errors, while TAR-Loss alone achieves 0.910, 4.05 mm, and 9.2%. The full model combining both components (0.926, 3.21 mm, 8.4%) exceeds the sum of individual contributions, demonstrating a synergistic effect where MSAGR produces coherent features suited for topology-aware optimization while TAR-Loss guides graph attention toward anatomically meaningful relationships.

Analysis of MSAGR design choices reveals that single-scale graphs (8^3 tokens) achieve 0.908 ET Dice at 7.2s inference, two-scale (16^3, 8^3) improves to 0.916 at 7.8s, and three-scale (32^3, 16^3, 8^3) achieves 0.923 at 8.1s, while four-scale provides no additional benefit but

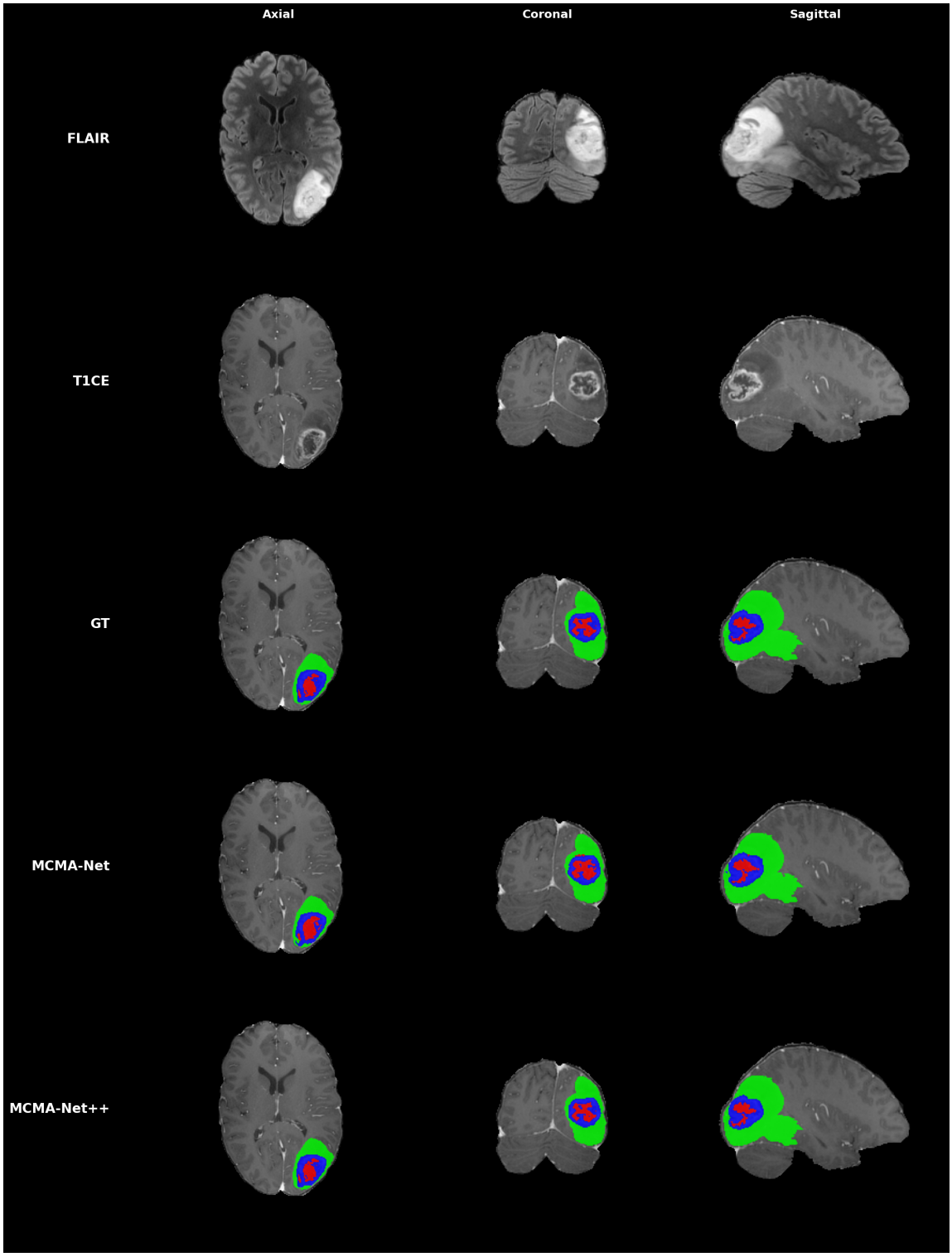

Figure 1: Qualitative comparison on BraTS 2021 case 00048 across three anatomical views. Rows show: FLAIR and T1ce modalities, GT annotations, MCMA-Net predictions, and MCMA-Net++ predictions. Color coding: green = ED, red = NCR/NET, blue = ET.

Table 3: Ablation study results.

Configuration	ET Dice	HD95	Topo Errors
Baseline (MCMA-Net)	0.900	5.48	14.3%
+ MSAGR only	0.912	4.21	11.1%
+ TAR-Loss only	0.910	4.05	9.2%
+ Both (Full)	**0.926**	**3.21**	**8.4%**

increases inference time to 9.6s. Three scales optimally balance performance and efficiency, with coarse scales capturing global structure and finer scales modeling boundary details.

Incremental TAR-Loss analysis shows that from Dice loss only (0.900 ET Dice, 14.3% topology errors), adding $\mathcal{L}_{\text{inc}}$ provides the largest single-component gain (0.908, 11.8%), $\mathcal{L}_{\text{bnd}}$ further improves to (0.915, 10.1%), $\mathcal{L}_{\text{smooth}}$ reaches (0.921, 8.9%), and full TAR-Loss achieves (0.926, 8.4%). Each component contributes at different levels—voxel, boundary, surface, region—providing comprehensive topology enforcement.

6. Conclusion

In this work, we introduced MCMA-Net++, a topology-aware and graph-driven architecture for 3D glioma segmentation that substantially advances the state of the art in anatomically consistent tumor delineation. Through the synergistic combination MSAGR and TAR-Loss, our method achieves unprecedented performance on the BraTS 2021 benchmark: Dice scores of 0.970 (WT), 0.943 (TC), and 0.926 (ET), with HD95 of 3.21 mm and 41% reduction in topological violations compared to MCMA-Net.

From a clinical perspective, MCMA-Net++ addresses a critical bottleneck in glioma management. Manual tumor segmentation by expert neuroradiologists is extremely time-consuming (4-6 hours per patient) and subject to inter-observer variability, creating significant workflow constraints in busy radiology departments. MCMA-Net++ provides fully automated segmentation in under 10 seconds while maintaining anatomically plausible predictions that match or exceed inter-rater agreement levels. This dramatic efficiency gain enables radiologists to allocate more time to complex diagnostic decision-making, multidisciplinary tumor board discussions, and direct patient care. Moreover, the model's ability to maintain strict hierarchical constraints (ET $\subset$ TC $\subset$ WT) with only 8.4% topology violations ensures clinical trustworthiness for integration into treatment planning, surgical navigation, and radiotherapy workflows.

Looking forward, we plan to validate MCMA-Net++ on real-world clinical data from the University Hospital Center of Poitiers, where it will be evaluated on routine clinical MRI acquisitions with varying protocols and scanner characteristics. This prospective evaluation will assess the model's ability to generalize beyond the curated BraTS dataset and determine its practical value for radiologists, from initial diagnosis to longitudinal treatment monitoring. By bridging the gap between algorithmic innovation and clinical application, MCMA-Net++ has the potential to enhance diagnostic efficiency, improve treatment planning, and ultimately contribute to better outcomes for patients with glioma.

References

[1] Jihan Alameddine, Ouissal-Ilham Zanoun, Céline Thomarat, Rémy Guillevin, Christine Fernandez-Maloigne, and Carole Guillevin. Mcma-net: Multi-class multi-attention network for glioma segmentation in 3d mri. *ICABME*, pages 1–4, 2025.

[2] Ujjwal Baid, Satyam Ghodasara, Suyash Mohan, Michel Bilello, Evan Calabrese, Errol Colak, Keyvan Farahani, Jayashree Kalpathy-Cramer, Felipe C Kitamura, Sarthak Pati, et al. The rsna-asnr-miccai brats 2021 benchmark on brain tumor segmentation and radiogenomic classification. *arXiv preprint arXiv:2107.02314*, 2021.

[3] Jieneng Chen, Jieru Mei, Xianhang Li, Yongyi Lu, Qihang Yu, Qingyue Wei, Xiangde Luo, Yutong Xie, Ehsan Adeli, Yan Wang, Matthew P. Lungren, Shaoting Zhang, Lei Xing, Le Lu, Alan Yuille, and Yuyin Zhou. Transunet: Rethinking the u-net architecture design for medical image segmentation through the lens of transformers. *Medical Image Analysis*, 2024.

[4] Özgün Çiçek, Ahmed Abdulkadir, Soeren S Lienkamp, Thomas Brox, and Olaf Ronneberger. 3d u-net: Learning dense volumetric segmentation from sparse annotation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 424–432. Springer, 2016.

[5] James R Clough, Nicholas Byrne, Ilkay Oksuz, Veronika A Zimmer, Julia A Schnabel, and Andrew P King. A topological loss function for deep-learning based image segmentation using persistent homology. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(12):8766–8778, 2020.

[6] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. In *International Conference on Learning Representations*, 2020.

[7] Hongyang Gao and Shuiwang Ji. Graph u-nets. In *International Conference on Machine Learning*, pages 2083–2092. PMLR, 2019.

[8] Yunhe Gao, Mu Zhou, and Dimitris N Metaxas. Utnet: A hybrid transformer architecture for medical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 61–71. Springer, 2021.

[9] Mahsa Ghadimi, Fatemeh Moazami, Ahmad Mohseni, Ahmad Ebrahimi, Mohammad Ali Salehi, and Delaram Salehian Yousefzadeh. Deep learning-based techniques in glioma brain tumor segmentation using multi-parametric mri: A review on clinical applications and future outlooks. *Journal of Magnetic Resonance Imaging*, 60(2): 340–362, 2024.

[10] Ali Hatamizadeh, Yucheng Tang, Vishwesh Nath, Dong Yang, Andriy Myronenko, Bennett Landman, Holger R Roth, and Daguang Xu. Unetr: Transformers for 3d medical image segmentation. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 574–584, 2022.

[11] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 770–778, 2016.

[12] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4700–4708, 2017.

[13] Xiaohong Huang, Zhifang Deng, Dandan Li, and Xueguang Yuan. Missformer: An effective medical image segmentation transformer. In *arXiv preprint arXiv:2109.07162*, 2021.

[14] Fabian Isensee, Paul F Jaeger, Simon AA Kohl, Jens Petersen, and Klaus H Maier-Hein. nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature Methods*, 18(2):203–211, 2021.

[15] Hoel Kervadec, Jihene Bouchtiba, Christian Desrosiers, Eric Granger, Jose Dolz, and Ismail Ben Ayed. Boundary loss for highly unbalanced segmentation. *Medical Image Analysis*, 67:101851, 2019.

[16] Asifullah Khan, Zunaira Rauf, Anabia Sohail, Asma Rehman, Hamza Asif, Aqsa Asif, and Umair Farooq. Transformers in medical image segmentation: a narrative review. *Diagnostics*, 13(17):2778, 2023.

[17] Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. *International Conference on Learning Representations*, 2017.

[18] Philipp Krähenbühl and Vladlen Koltun. Efficient inference in fully connected crfs with gaussian edge potentials. *Advances in Neural Information Processing Systems*, 24, 2011.

[19] Ying Liu, Xiaoming Yang, Ke Wang, Jianfei Hu, and Chunhua Liao. Brain tumor segmentation and survival prediction using multimodal mri scans with deep learning. *Frontiers in Neuroscience*, 13:810, 2019.

[20] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 10012–10022, 2021.

[21] Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In *2016 Fourth International Conference on 3D Vision (3DV)*, pages 565–571. IEEE, 2016.

[22] Ozan Oktay, Jo Schlemper, Loic Le Folgoc, Matthew Lee, Mattias Heinrich, Kazunari Misawa, Kensaku Mori, Steven McDonagh, Nils Y Hammerla, Bernhard Kainz, et al. Attention u-net: Learning where to look for the pancreas. In *Medical Imaging with Deep Learning*, 2018.

[23] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 234–241. Springer, 2015.

[24] Suprosanna Shit, Johannes C Paetzold, Anjany Sekuboyina, Ivan Ezhov, Alexander Unber, Andrey Zhylka, Josien PW Pluim, Ulrich Bauer, and Bjoern H Menze. cldice-a novel topology-preserving loss function for tubular structure segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16560–16569, 2021.

[25] Roger D Soberanis-Mukul, Nassir Navab, and Shadi Albarqouni. Uncertainty-based graph convolutional networks for organ segmentation refinement. In *Medical Imaging with Deep Learning*, pages 1093–1105. PMLR, 2022.

[26] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. Graph attention networks. *International Conference on Learning Representations*, 2018.

[27] Mariska Visser, Domenique MJ Müller, Roelant JA van Duijn, Marion Smits, Niels Verburg, Emmy J Hendriks, Ruben JA Nabuurs, Jan C Bot, Roelant S Eijgelaar, Marnix Witte, et al. Inter-rater agreement in glioma segmentations on longitudinal mri. *NeuroImage: Clinical*, 22:101727, 2019.

[28] Arushi Wadhwa, Anuj Bhardwaj, and Vivek Singh Verma. A review on brain tumor segmentation of mri images. *Magnetic Resonance Imaging*, 61:247–259, 2019.

[29] Wenxuan Wang, Chen Chen, Meng Ding, Hong Yu, Sen Zha, and Jiangyun Li. Transbts: Multimodal brain tumor segmentation using transformer. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 109–119. Springer, 2021.

[30] Yutong Xie, Jianpeng Zhang, Chunhua Shen, and Yong Xia. Cotr: Efficiently bridging cnn and transformer for 3d medical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 171–180. Springer, 2021.

[31] Hong-Yu Zhou, Jiaming Guo, Yinghao Zhang, Lequan Yu, Liansheng Wang, and Yizhou Yu. nnformer: Interleaved transformer for volumetric segmentation. *arXiv preprint arXiv:2109.03201*, 2021.

[32] Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang. Unet++: A nested u-net architecture for medical image segmentation. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, pages 3–11. Springer, 2018.