M2GCNet: Multi-Modal Graph Convolution Network for Precise Brain Tumor Segmentation Across Multiple MRI Sequences

Tongxue Zhou

***Abstract*— Accurate segmentation of brain tumors across multiple MRI sequences is essential for diagnosis, treatment planning, and clinical decision-making. In this paper, I propose a cutting-edge framework, named multi-modal graph convolution network (M2GCNet), to explore the relationships across different MR modalities, and address the challenge of brain tumor segmentation. The core of M2GCNet is the multi-modal graph convolution module (M2GCM), a pivotal component that represents MR modalities as graphs, with nodes corresponding to image pixels and edges capturing latent relationships between pixels. This graph-based representation enables the effective utilization of both local and global contextual information. Notably, M2GCM comprises two important modules: the spatial-wise graph convolution module (SGCM), adept at capturing extensive spatial dependencies among distinct regions within an image, and the channel-wise graph convolution module (CGCM), dedicated to modelling intricate contextual dependencies among different channels within the image. Additionally, acknowledging the intrinsic correlation present among different MR modalities, a multi-modal correlation loss function is introduced. This novel loss function aims to capture specific nonlinear relationships between correlated modality pairs, enhancing the model's ability to achieve accurate segmentation results. The experimental evaluation on two brain tumor datasets demonstrates the superiority of the proposed M2GCNet over other state-of-the-art segmentation methods. Furthermore, the proposed method paves the way for improved tumor diagnosis, multi-modal information fusion, and a deeper understanding of brain tumor pathology.**

***Index Terms*— Brain tumor segmentation, graph convolution network, feature fusion, MRI, deep learning.**

I. Introduction

BRAIN tumors are the abnormal growth or mass of cells within the brain or its surrounding structures [1], [2], which is considered one of the most aggressive and challenging cancers in the world. World Health Organization (WHO) categorizes brain tumors into high-grade glioma (HGG) and low-grade glioma (LGG). The former refers to a group of malignant brain tumors originating from glial cells, characterized by aggressive growth and a high likelihood of recurrence. The latter refers to a group of less malignant or indolent gliomas that grow more slowly [3]. Early diagnosis and intervention can help control the tumor's growth and prevent further harm to the brain. MRI is a widely used imaging modality for diagnosing brain tumors due to its ability to provide multi-dimensional images of the brain and its excellent soft tissue contrast. Multi-modal MRI typically includes various imaging sequences, such as T1-weighted (T1), T1-weighted contrast-enhanced (T1c) sequences, T2-weighted (T2) and Fluid-attenuated inversion recovery (Flair), shown in Figure 1. Combining information from these modalities can provide a more comprehensive view of the tumor, as each modality highlights different aspects of the brain tumor. For example, T1 and T1c sequences can help to distinguish the enhancing tumor region from surrounding brain tissue; Flair and T2 images are effective in delineating tumor boundaries and locating the entire tumor region, especially the edema region [4], [5]. In this work, I refer to these MRI sequences as MR modalities.

Multi-modal Brain tumor segmentation plays a critical role in the diagnosis, treatment planning, and monitoring of brain tumor patients. It is also challenging due to the variability in tumor appearance, size, and location across different patients and imaging modalities [6]. Deep learning-based algorithms, such as convolution neural networks (CNNs), have shown great potential in segmenting brain tumors. However, they may not capture long-range dependencies between pixels, which is critical for segmentation. On the other hand, graph convolution network (GCN) [7] can be a powerful choice for segmentation, which is designed to capture relationships between nodes (pixels) in a graph and modelling long-range dependencies in non-grid data. To this end, in this paper, I propose a novel multi-modal graph convolution network (M2GCNet) to explore the relationships across different MR modalities and improve the brain tumor segmentation performance. Different from the recent GCN-based methods [8], [9], [10] dedicated to modelling the global context of the input feature, the proposed M2GCNet focuses on capturing the potential information across different modalities by considering two different graphs, spatial-wise graph and channel-wise graph, in a single network. The main contributions of this paper are concluded as follows:

(1) To the best of our knowledge, this is the first work to propose a multi-modal graph convolution network (M2GCNet)

Manuscript received 11 December 2023; revised 30 May 2024; accepted 26 August 2024. Date of publication 5 September 2024; date of current version 11 September 2024. This work was supported by the National Natural Science Foundation of China under Grant 62206084. The associate editor coordinating the review of this article and approving it for publication was Prof. Dong Liu.

The author is with the School of Information Science and Technology, Hangzhou Normal University, Hangzhou 311121, China (e-mail: txzhou@hznu.edu.cn).

Digital Object Identifier 10.1109/TIP.2024.3451936

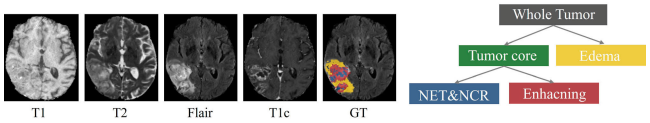

Fig. 1. The left panel presents examples of multiple MR sequences and their corresponding ground-truth annotations. The right panel presents the hierarchical relationships between different brain tumor regions.

to explore the relationships across different MR modalities and achieve brain tumor segmentation.

(2) The proposed method introduces a spatial-wise graph convolution module (SGCM) to capture long-range spatial dependencies among distinct regions and a channel-wise graph convolution module (CGCM) to model intricate contextual dependencies among different channels within the image.

(3) A novel multi-modal correlation loss is proposed to exploit the latent multi-modal correlation and facilitate learning effective feature representations for brain tumor segmentation.

(4) Experimental evaluations conducted on two widely used BraTS benchmark datasets demonstrate the effectiveness of the proposed method but also underline its potential to significantly contribute to the field of brain tumor segmentation and data fusion issues.

The remainder of the paper is organized as follows: Section II introduces the existing literature and related research works. Section III introduces the proposed method; Section IV describes the experimental setup; Section V presents and analyzes the experimental results. Section VI provides a further discussion about the model design. Section VII summarizes the work and provides suggestions for future research directions.

II. Related Works

A. Brain Tumor Segmentation

Accurate segmentation of brain tumors is essential for a range of clinical applications, including diagnosis, treatment planning, and monitoring disease progression. Manual segmentation is labour-intensive and time-consuming, making it impractical for large datasets or time-sensitive situations. Automated segmentation has emerged as a pivotal advancement to alleviate the challenges associated with manual brain tumor segmentation, especially in contexts involving extensive datasets or time-sensitive scenarios. Early automated brain tumor segmentation methods primarily relied on traditional machine learning methods with hand-crafted features [3], [11], including support vector machine (SVM) [12], [13], random forest [14], [15], K-nearest neighbours (KNN) [16] and fuzzy C-means (FCM) [17], [18]. However, the limitations of these early methods are threefold: (1) Traditional machine learning models often require manual feature engineering, which can be time-consuming, may not capture all the intricate details in medical images, and might not generalize well across diverse datasets; (2) Machine learning models, especially those using linear classifiers or shallow architectures, may struggle to capture highly complex and nonlinear patterns present in medical images; (3) Machine learning models are often sensitive to variations in image acquisition settings, resolutions, and other factors. This sensitivity can affect the model's generalisation across different scanners or imaging protocols, making it less robust in real-world clinical scenarios.

By embracing more advanced techniques, such as deep learning, we can effectively overcome the limitations associated with manual feature engineering, and address the complexities inherent in the brain tumor segmentation task. Deep learning-based methods have seen significant advancements with the use of various neural network architectures, such as CNN [19], FCN [20], and U-Net [21]. For example, Yang et al. [22] proposed a contour-aware 3D convolution neural network with adversarial training for segmenting brain tumor. Kamnitsas et al. [23] introduced a parallel convolution pathway to efficiently incorporate both local and contextual information. A 3D fully connected conditional random field (CRF) is introduced as a post-processing step to refine the segmentation results. Myronenko et al. [24] proposed adding a variational autoencoder (VAE) branch to the network to reconstruct the input images jointly with segmentation to regularize the shared encoder. Isensee et al. [25] designed a U-Net-based network named nnU-Net, a general-purpose segmentation method that automatically configures segmentation pipelines for diverse biomedical datasets, and demonstrated the effectiveness for brain tumor segmentation. Zhang et al. [26] proposed to integrate residual modules and attention gates into U-Net architecture for brain tumor segmentation. Zhou and Zhu [27] proposed integrating uncertainty quantification into multi-modal brain tumor segmentation to enhance performance. Allah et al. [28] developed an edge U-Net model for brain tumor segmentation, which incorporates boundary-related MRI data with primary brain MRI data. Rehman et al. [29] proposed a brain tumor segmentation network using parallel processing of multiple spatial frames. Zhang et al. [30] introduced a multi-branch structure with high-resolution feature extraction for precise brain tumor segmentation, particularly beneficial for fuzzy boundaries. Li et al. [31] proposed a brain tumor segmentation network integrating attention mechanism and transformer, leveraging prior knowledge to enhance segmentation accuracy. Zhou et al. [32] introduced a multi-modal brain tumor segmentation network by leveraging disentangled representation learning and region-aware contrastive learning.

Although deep learning-based segmentation models have made significant strides in achieving high performance, a persistent challenge remains in their limited ability to capture intricate long-range dependencies and nuanced relationships within images. This limitation is largely attributed to the constrained receptive field of CNNs, hindering their capacity to comprehensively grasp the broader contextual information crucial for precise segmentation. This paper introduces a novel multi-modal graph convolution network to address this inherent constraint. This advanced framework is specifically designed to harness the graph structure, enabling the capture of long-range dependencies within diverse MR modalities. This approach significantly contributes to improved feature representation learning and enhanced segmentation performance.

B. Graph Convolution Network

Conventional CNNs are tailored for data organized in grid-like data with fixed-size receptive fields, which imposes restrictions on its ability to effectively process graph-structured data, where entities (nodes) are connected by relationships (edges) [33]. Therefore, Kipf et al. [7] proposed GCN to effectively capture intricate relationships for accurately understanding and modelling real-world systems. Furthermore, GCN has gained significant attention in various computer vision tasks, including visual tracking [34], point cloud analysis [35], human pose estimation [36] and image segmentation [37], due to their ability to capture complex relationships and dependencies within graph-structured data. Recently, GCN has also been employed in medical image analysis. For example, Zhao et al. [38] proposed to adopt a graph-based residual convolution network to achieve pancreatic mass segmentation, diagnosis, and disease management. Li et al. [8] introduced a dual encoder-based dynamic-channel graph convolution network for retinal vessel segmentation. Huang and Chung [39] proposed to utilise GCN to address population-based disease prediction issues. Ma et al. [40] presented a dual graph reasoning unit for brain tumor segmentation which adopted GNNs to capture relational and contextual information. Mishra and Verma [41] proposed a graph attention auto-encoder convolution neural network for brain tumor classification. Liu et al. [42] proposed a novel brain tumor segmentation network considering different contextual information with standard and graph convolutions.

Although GCN-based methods have successfully extracted semantic and structural features from images in the domain of brain tumor segmentation. However, a comprehensive review of the existing literature highlights a notable lack of previous research efforts utilizing GCN to reveal the intricate relationships across different MR modalities. To this end, this paper introduces a novel M2GCNet to address this research gap. The primary aim is to explore and leverage the inherent relationships within diverse MR modalities, and to enhance the overall performance of brain tumor segmentation. With the introduction of M2GCNet, this study aims to go beyond current methods, seeking to enhance the effectiveness of brain tumor segmentation through a detailed exploration of multi-modal relationships.

III. Methodology

A. Motivation of the Proposed Method

The driving force behind this work is the fact that different MR modalities can provide diverse views and information about the same region of interest, which is the target brain tumor. Building on the findings of my previous work [43], a strong correlation exists in latent feature representation between each pair of MR modalities. This suggests that there's valuable information hidden within these relationships and can be leveraged for a more comprehensive feature representation. GCN adept at exploiting the graph structure inherent in the data. Applying GCN to the multi-modal MRI data can create a more comprehensive feature representation that considers the complex relationships across the modalities. Furthermore, the richer feature representation can enhance the accuracy of brain tumor segmentation by capturing subtle patterns and information that may not be evident when analyzing each modality independently. Hence, this paper introduces the M2GCNet, consisting of a spatial-wise graph convolution module (SGCM) and a channel-wise graph convolution module (CGCM), to harness the complementary information from various MRI modalities, extracting structured features for precise and robust brain tumor segmentation.

B. Network Architecture Overview

The proposed network is an encoder-decoder-based architecture, and the architecture is illustrated in Figure 2. Four different MR modalities are first concatenated in the input space and then fed to the network. The encoder consists of multiple convolution layers, each followed by instance normalization and a non-linear activation function, LeakyReLU. In addition, the strided convolutions ($stride = 2$) are employed as the downsampling operations to capture contextual information across different scales. In the bottom level of the architecture, the proposed M2GCM is employed, which allows the network to focus on relevant information across different MR modalities while suppressing noise or irrelevant details. Section VI-A discusses the impact of applying the proposed M2GCM at various levels of the network on overall performance. The decoder consists of up-sampling layers to gradually increase the spatial dimensions of the feature maps. Instance normalization and LeakyReLU are applied after each convolution layer to enhance the expressiveness of the decoder. It is noted that dilated convolutions are employed in both encoder and decoder parts to capture contextual information across different scales and receptive fields. Skip connections are incorporated between the encoder and decoder layers to facilitate information flow and gradient propagation. These connections enable the model to access features from earlier stages, aiding in preserving fine-grained details. Furthermore, instead of having a single output result, multiple segmentation results at different depths of the decoder are combined through $1 \times 1 \times 1$ convolution layer and element-wise summation operations to produce the final segmentation result. By having multiple intermediate supervision layers, gradients can be better preserved, allowing for more effective training and improving the final segmentation performance.

C. Multi-Modal Graph Convolution Module (M2GCM)

The GCN primarily comprises two essential components: the nodes within the graph and the edges representing the relationships or connections between the nodes [40]. In the context of brain tumor segmentation, each node is linked to a feature vector encoding information about the corresponding region in the MR modalities, while the edges have the capacity to capture semantic relationships between these regions. The proposed M2GCM consists of a spatial-wise graph convolution module (SGCM) and a channel-wise graph convolution module (CGCM). The motivation for incorporating the SGCM is to enhance the network's ability to capture long-range

spatial dependencies within the image. Traditional convolutional layers are limited to local receptive fields, which can hinder the model's ability to understand broader contextual information crucial for accurate segmentation. By leveraging graph convolution, SGCM can effectively model spatial relationships among distinct regions in an image, improving the overall segmentation performance. Similarly, the CGCM is designed to address the intricate contextual dependencies among different feature channels. In conventional convolutional networks, channels are treated independently during spatial convolutions, potentially missing out on valuable inter-channel correlations. CGCM employs graph convolution to capture these dependencies, allowing the network to learn more robust and comprehensive feature representations, which contributes to better segmentation accuracy.

Furthermore, it's important to note that the proposed M2GCM is designed to capture potential information interactions across different modalities. M2GCM takes two pairs of correlated MR modalities (T1 & T1c, Flair & T2) as its inputs. The selection of these two pairs of MR modalities is deliberate since T1 and T1c modalities are effective at distinguishing the enhancing tumor region, while Flair and T2 images excel at identifying the edema region, as introduced in Section I. Combining the two correlated pairs allows the network to leverage the synergy between modalities, facilitating the distinction between tumor and healthy tissue. Consequently, M2GCM can exploit the cross-modality relationships to enhance segmentation performance. For a more in-depth understanding of the network architecture, please refer to the following section, which explains the detailed network architecture.

1) Spatial-Wise Graph Convolution Module (SGCM): In the SGCM illustrated in Figure 3, the above-mentioned two pairs of MR modality (T1 & T1c, Flair & T2) are used to construct a dual-graph network, respectively. First, let's consider a pair of MR modalities, T1 and T1c, as an example. It is noted that each modality (T1 and T1c) produces its own feature, denoted as X_i for T1 and X_j for T1c. To find a specific sample X_i that belongs only to the T1 modality, if T1 has one channel ($C_i = 1$), then X_i would be the first channel of X, i.e., $X[:1, :, :, :]$. The input features $\{X_i, X_j\} \in \mathbb{R}^{C\times H\times W\times D}$ from these two modalities are first passed through two convolution layers ϕ and ψ separately. This generates two sets of features: $\{Q_i = \phi(X_i), K_i = \psi(X_i)\}$ and $\{Q_j = \phi(X_j), K_j = \psi(X_j)\}$, where $\{Q_i, K_i, Q_j, K_j\} \in \mathbb{R}^{C\times H\times W\times D}$. Next, these features $\{Q_i, K_i, Q_j, K_j\}$ are reshaped to $\{Q_i, K_i, Q_j, K_j\} \in \mathbb{R}^{C\times N}$, where $N = H \times W \times D$, representing the number of pixels. Then, the adjacent matrices A_{ii} and A_{ij} can be obtained using matrix multiplication and a softmax layer for features $\{Q_i, K_i\}$ and $\{Q_i, K_j\}$:

$$A_{ii} = \frac{exp(Q_i \cdot K_i)}{\sum_{i=1}^{N} exp(Q_i \cdot K_i)} \tag{1}$$

$$A_{ij} = \frac{exp(Q_i \cdot K_j)}{\sum_{i=1}^{N} exp(Q_i \cdot K_j)} \tag{2}$$

where $A_{ii} \in \mathbb{R}^{N\times N}$ represents the correlation between the ith feature and itself, and $A_{ij} \in \mathbb{R}^{N\times N}$ represents the correlation between the ith feature and the jth feature.

Since different MR sequences capture complementary information about the brain tumor. Thus, these two adjacent matrices A_{ii} and A_{ij} are combined via element-wise summation to achieve the final adjacent matrix $A_i \in \mathbb{R}^{N\times N}$. By adding their respective adjacent matrices, the network can integrate information from each modality, resulting in a more comprehensive representation of the brain tumor.

$$A_i = A_{ii} + A_{ij} \tag{3}$$

Meanwhile, the input features $\{X_i, X_j\}$ are passed through a third convolution layer ξ to generate two new features $\{V_i = \xi(X_i), V_j = \xi(X_j)\}$, where $\{V_i, V_j\} \in \mathbb{R}^{C\times H\times W\times D}$. At this point, a sub-graph network has been constructed for modality T1, where V_i contains the features of each node and the adjacency matrix A_i represents the connections between nodes. The output feature is computed by multiplying A_i and V_i after reshaping and transposition operations:

$$S_i = V_i \cdot A_i \tag{4}$$

Unlike previous work that employed a single graph network, in this study, following a similar procedure, another sub-graph is established for modality T1c. Here, V_j contains the features of each node, and the adjacency matrix A_j represents the connections between nodes. The corresponding calculations are as follows:

$$A_{jj} = \frac{exp(Q_j \cdot K_j)}{\sum_{i=1}^{N} exp(Q_j \cdot K_j)} \tag{5}$$

$$A_{ji} = \frac{exp(Q_j \cdot K_i)}{\sum_{i=1}^{N} exp(Q_j \cdot K_i)} \tag{6}$$

$$A_j = A_{jj} + A_{ji} \tag{7}$$

$$S_j = V_j \cdot A_j \tag{8}$$

By considering complementary information, the proposed dual-graph networks can help reduce false positives in the segmentation, resulting in a more accurate segmentation result. Finally, S_i and S_j are passed through a $1D$ convolution layer and reshaped, and the final output feature is calculated as the sum of S_i and S_j: $S_{ij} = S_i + S_j$, allowing us to capture spatial-wise graph structure information and potential cross-modal relationships between T1 and T1c modalities.

The same operations are also implemented on T2 and Flair modalities, and the results from these two pairs of MR modalities are concatenated to produce the final result $X_{os} = S_{ij} \parallel S_{mn} \in \mathbb{R}^{C\times H\times W\times D}$, $\parallel$ denotes concatenation. The comparison of applying concatenation or addition operation on features S_{ij} and S_{mn} is discussed in Section VI-C.

2) Channel-Wise Graph Convolution Module (CGCM): In the CGCM illustrated in Figure 4, the asymmetry in the CGCM is designed to handle the inherent differences in feature distributions between two modalities. This design allows the module to adaptively focus on the most relevant features from each modality, leading to more robust feature fusion. The above-mentioned two pairs of MR modality (T1 & T1c, Flair & T2) are used to construct a dual-graph network, respectively. Similar to SGCM, T1 and T1c are chosen as examples to illustrate the architecture of CGCM. First, on one hand, the input feature $X_i \in \mathbb{R}^{C\times H\times W\times D}$ from T1 modality is projected

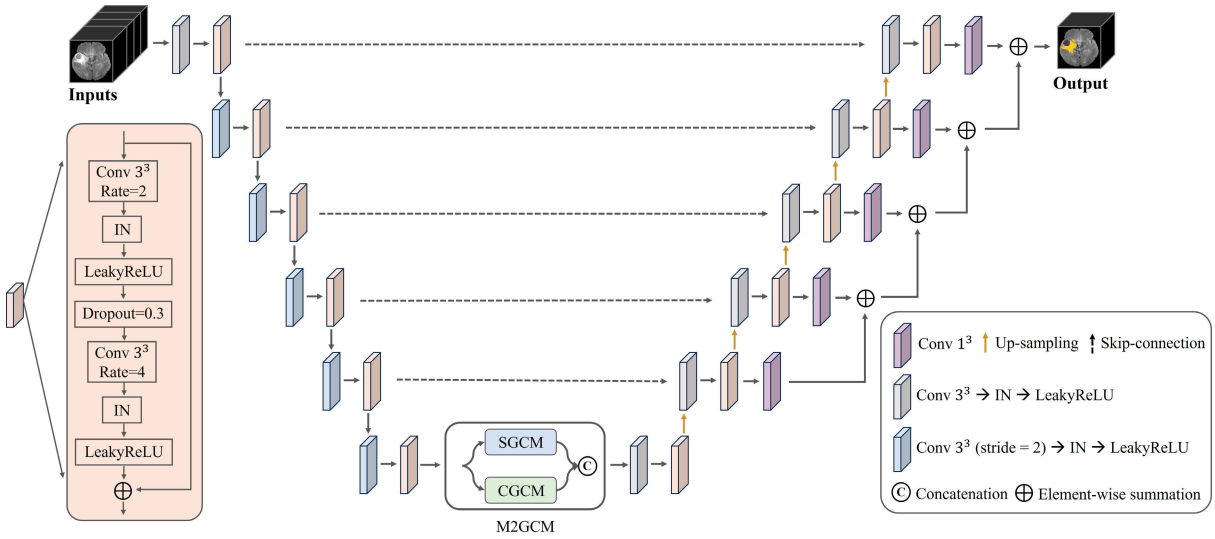

Fig. 2. The architecture of the proposed multi-modal graph convolution network (M2GCNet), which can explore the relationships across different MR modalities, and achieve brain tumor segmentation. IN denotes instance normalization.

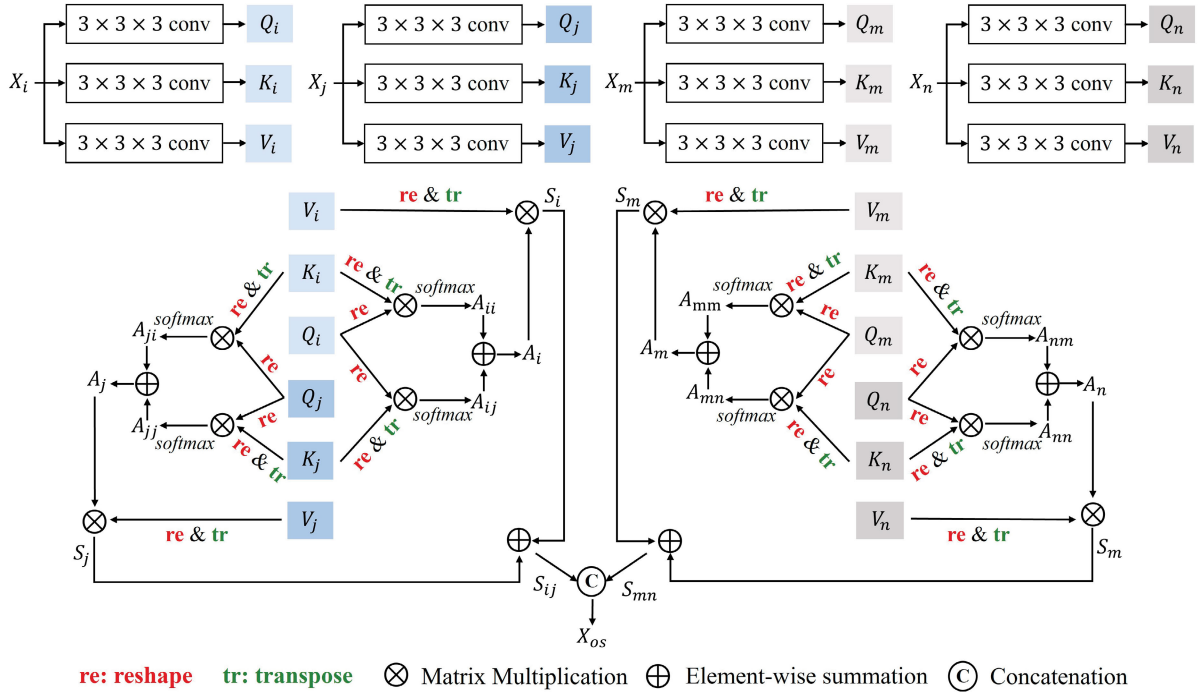

Fig. 3. The architecture of the proposed spatial-wise graph convolution module (SGCM), which can capture extensive spatial dependencies that exist between distinct regions within the image.

to a new feature using the projection function $\eta(X_i) \in \mathbb{R}^{N \times C_1}$. On the other hand, the input feature $X_j \in \mathbb{R}^{C \times H \times W \times D}$ from T1c modality is passed through a function $\theta(X_j) = H_{\mathcal{C}}^T(X_i) \in \mathbb{R}^{N \times C_2}$ to reduce its dimension from C to C_2. Consequently, a new feature $V_{\mathcal{C}}$ is obtained using the following equation:

$$V_{\mathcal{C}} = H_{\mathcal{C}}^T \eta(X_i) = \theta(X_j)\eta(X_i) \tag{9}$$

where $\eta(\cdot)$ and $\theta(\cdot)$ are implemented using $1 \times 1 \times 1$ convolution layers. The resulting $V_{\mathcal{C}} \in \mathbb{R}^{C_2 \times C_1}$ comprises C_2 nodes, with each node containing C_1 features. In this work, C_1 is set to $C/2$, and C_2 is set to $C/4$.

Subsequently, this new feature $V_{\mathcal{C}}$ is passed through a convolution layer and combined with itself to create a channel-wise graph network denoted as $Z_{\mathcal{C}}$ for T1 and T1c modalities. Following this, the feature graph $Z_{\mathcal{C}}$ is processed through a 1D convolution layer and multiplied with the feature $\theta(X_j)$. It is noted that the 1D convolution within the CGCM is employed to adjust the channel dimension, ensuring that the feature maps from different modalities are appropriately aligned and integrated. The final output feature is obtained after passing through a $1 \times 1 \times 1$ convolution layer. To this end, this channel-wise graph network enables us to capture

channel-wise graph structure information hidden between T1 and T1c modalities.

The same operations are also applied to T2 (X_m) and Flair (X_n) modalities, and the results from these two pairs of MR modalities are then concatenated to yield the final result $X_{oc} = C_{ij} \parallel C_{mn} \in \mathbb{R}^{C\times H\times W\times D}$, $\parallel$ denotes concatenation operation. The comparison of applying concatenation or addition operation on features C_{ij} and C_{mn} is discussed in Section VI-C.

Finally, the SGCM and CGCM are concatenated to form the M2GCM, contributing to the final feature representation. Section VI-A discusses the impact of incorporating M2GCM at various levels of the segmentation network.

D. Proposed Multi-Modal Segmentation Loss Function

It has been demonstrated that an inherent correlation exists among different MR modalities within the same tumor region for a given patient [43], [44]. Building on this insight, I introduce a multi-modal nonlinear correlation loss function to the proposed M2GCNet. This addition provides extra guidance during training, encouraging the model to capture specific nonlinear relationships crucial for accurate segmentation. The proposed correlation loss (CLoss) function is defined in Equation 10, where α_{ij}, β_{ij}, and γ_{ij} represent the correlation parameters of the features from MR modalities X_i and X_j. These parameters are obtained by passing the concatenation of the two multi-modal features through a global average pooling layer, followed by two fully connected layers with LeakyReLU activation, as shown in Figure 5. A lower CLoss indicates higher multi-modal similarity. This approach ensures that the model learns to capture the complementary information present between different modalities, leading to improved performance. Conversely, if these relationships are not well captured, the model's performance would degrade, reflecting its inability to effectively integrate multi-modal information.

$$CLoss = \mathcal{F}(X_i) log \frac{\mathcal{F}(X_i)}{\alpha_{ij}\mathcal{G}^2(X_j) + \beta_{ij}\mathcal{G}(X_j) + \gamma_{ij}} \tag{10}$$

where $\mathcal{F}$ and $\mathcal{G}$ are probability distributions of the feature representation of modality i, and correlated feature representation of modality j in the bottom level of the network, $(i \neq j)$, respectively. α_{ij}, β_{ij} and γ_{ij} are the correlation parameters of the correlated pair of MR modalities.

Traditional Dice loss function Equation 11 measures the degree of spatial overlap between the predicted and ground truth segmentation masks, which is primarily concerned with the spatial alignment of segmented regions. However, the proposed CLoss focuses on capturing relationships and dependencies among features from different modalities. It encourages the model to capture intricate, nonlinear dependencies that might not be adequately addressed by traditional overlap-based metrics like Dice. Hence, I propose to combine the two loss functions to achieve a balanced optimization method (defined in Equation 12), where the Dice loss emphasizes spatial accuracy in segmentation, and the CLoss focuses on capturing complex feature relationships. It is noted that the proposed training loss function is adaptable and can be applied to a broad range of multi-source correlation problems. The selection of a specific correlation expression should be tailored to the unique demands of each application. Section VI-B discusses the impact of incorporating CLoss at various levels of the segmentation network.

$$Dice\ Loss = 1 - \frac{2|A \cap B|}{|A| + |B|} \tag{11}$$

where A represents the set of voxels in the predicted segmentation. B represents the set of voxels in the ground truth annotation. $|\cdot|$ denotes the number of elements of a set.

$$Loss = Dice\ Loss + \lambda \sum_{i=1}^{n} CLoss_n \tag{12}$$

where λ is the trade-off parameter weighting the importance of each component, $\lambda = 0.1$, n denotes the number of correlation pair, in this work, $n = 2$, two most correlated pairs are used: T1-T1c and T2-FLAIR.

IV. Experimental Setup

A. Dataset

The BraTS (Brain Tumor Image Segmentation) dataset is a widely recognized resource for brain tumor image segmentation research. I employ the BraTS 2018 and BraTS 2019 datasets in this study to assess the proposed method. The BraTS 2018 dataset comprises 285 training cases, and BraTS 2019 dataset comprises 335 training cases. Each case contains four different imaging modalities (T1, T1ce, T2, and Flair), and the corresponding ground truth labels provided by the dataset organizers. The ground truth labels include three tumor types: necrotic and non-enhancing tumor (NCR&NET), enhancing tumor (ET), and peritumoral edema (ED). The segmentation task involves delineating three specific regions: The whole tumor (WT) region, consisting of NCR&NET, ET, and ED. The tumor core (TC) region, consisting of NCR&NET and ET. The enhancing tumor (ET) region, shown in Figure 1. The co-registration and skull-stripping procedures have been applied on these images by dataset organizers. The original image size is $155 \times 240 \times 240$, with a resolution of $1mm^3$. Additionally, bias field correction using N4ITK [45] and Z-score normalization are carried out. The images are resized to dimensions of $128 \times 128 \times 128$ for the training stage.

B. Evaluation Metrics

To quantitatively assess the performance of the proposed method, two evaluation metrics are employed: the Dice similarity coefficient and the 95% Hausdorff Distance.

The Dice similarity coefficient is a widely used metric for evaluating the degree of overlap between the predicted tumor region and the real tumor region. A higher Dice similarity coefficient score indicates more accurate segmentation performance. Small segmentation errors directly impact the Dice similarity coefficient value as they affect the overlap between the segmented and ground truth regions. It is calculated as:

$$Dice\ Similarity\ Coefficient = \frac{2|A \cap B|}{|A| + |B|} \tag{13}$$

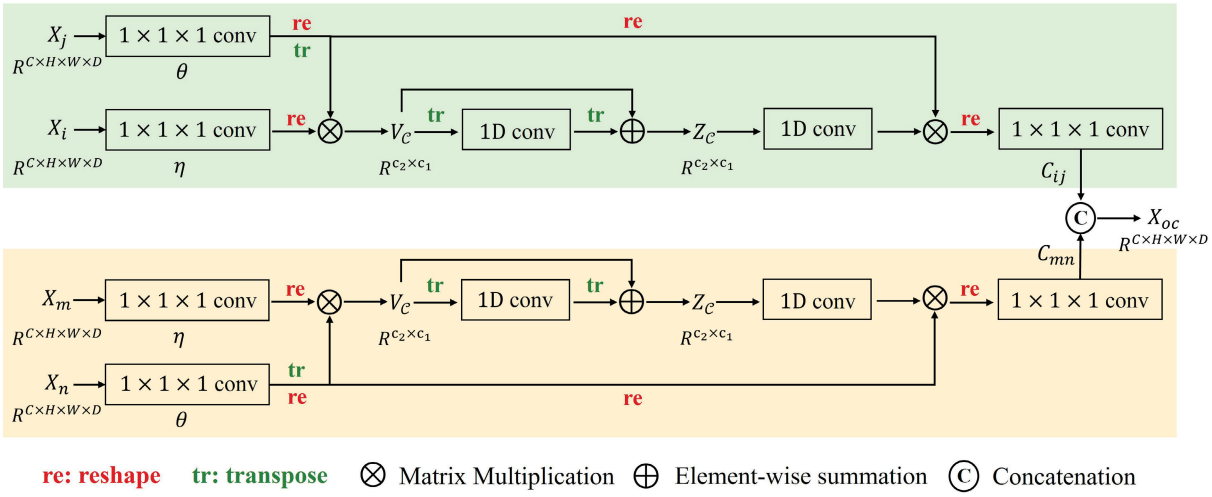

Fig. 4. The architecture of the proposed channel-wise graph convolution network (CGCM), which is dedicated to modelling the intricate contextual dependencies among different channels within the image.

Fig. 5. The multi-modal features X_i and X_j are first concatenated, then passed into a global average pooling followed by two fully connected layers with LeakyReLU to obtain the multi-modal correlation parameters α_{ij}, β_{ij} and γ_{ij}.

where A and B represent the set of voxels in the predicted and ground truth mask, respectively.

The Hausdorff Distance is a metric that quantifies the maximum distance between the predicted tumor boundary and the real tumor boundary. A lower Hausdorff Distance value indicates more accurate segmentation, representing a closer match between the boundaries. Hausdorff Distance is sensitive to outliers since it considers the maximum distance, which means even a single large discrepancy can lead to a high Hausdorff Distance value. It is calculated as:

$$Hausdorff\ \ Distance = \max\{\max_{x\in X}\min_{y\in Y} d(x, y), \max_{y\in Y}\min_{x\in X} d(x, y)\} \quad (14)$$

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

C. Implementation Details

The proposed model is implemented using the Keras deep learning framework and executed on a Nvidia Geforce RTX 4090 graphics processing unit (GPU) with 24 GB of memory. The Adam optimizer in used to train the network. The initial learning rate is set as 5×10^{-4}, and the learning rate is halved when the validation loss did not improve for 10 consecutive epochs. To avoid overfitting, an early stopping mechanism is implemented, training is terminated if no improvement occurs in the validation loss for 50 consecutive epochs. The batch size is also set to 1, and the dataset is randomly split into 80% for training and 20% for testing.

V. Experimental Results and Analysis

A. Ablation Experiments

In this section, a series of ablation experiments are conducted to analyze the impact of the proposed components, and the comparison results are illustrated in Table I. The baseline is established without incorporating any proposed modules, utilizing Dice loss as the default training metric. In contrast, the methods involving the addition of the SGCM, the CGCM, or both modules simultaneously to the baseline maintain consistency by employing Dice loss as their respective training loss. However, for the configuration integrating SGCM, CGCM, and CLoss, the hybrid loss is adopted. From Table I, it can be observed that the baseline method obtains 78.1% and 6.9 in terms of average Dice Similarity Coefficient and average Hausdorff Distance, respectively. Upon incorporating the Spatial-wise Graph Convolution Module (SGCM) to capture long-range spatial dependencies, performance improves by 2.2% and 20.3% in terms of average Dice Similarity Coefficient and average Hausdorff Distance, respectively. Furthermore, compared to the baseline method, the integration of the Channel-wise Graph Convolution Module (CGCM) to capture intricate contextual dependencies among different channels enhances segmentation results by 3.5% and 26.1% in terms of average Dice Similarity Coefficient and average Hausdorff Distance, respectively. Furthermore, the combination of SGCM and CGCM can outperform the baseline method by 3.8% and 27.5% in terms of average Dice Similarity Coefficient and average Hausdorff Distance, respectively. Applying the proposed multi-modal correlation loss function yields the best Dice Similarity Coefficient of 81.8% on average and a Hausdorff Distance of 5.7 on the whole tumor region. Moreover, it can be observed that the Dice Similarity Coefficient improves significantly with the addition of each module (SGCM, CGCM, CLoss), suggesting enhanced overall overlap between the predicted and actual segmentations. Similarly, the Hausdorff Distance improves with each module addition (SGCM, CGCM), indicating more accurate

TABLE I
ABLATION EXPERIMENTS ON THE BRATS 2018 DATASET IN TERMS OF DICE SIMILARITY COEFFICIENT (%) AND 95% HAUSDORFF DISTANCE (MM), WT, TC, ET DENOTE WHOLE TUMOR, TUMOR CORE AND ENHANCING TUMOR, RESPECTIVELY; BOLD RESULTS DENOTE THE BEST SCORES

Method				Dice Similarity Coefficient (%) ↑				Hausdorff Distance (mm) ↓			
Baseline	SGCM	CGCM	CLoss	WT	TC	ET	Avg	WT	TC	ET	Avg
✓	✗	✗	✗	82.4	78.6	73.2	78.1	6.9	8.4	5.4	6.9
✓	✓	✗	✗	84.3	81.0	74.1	79.8	6.4	5.7	4.3	5.5
✓	✗	✓	✗	84.8	82.6	75.1	80.8	5.9	5.5	**3.9**	5.1
✓	✓	✓	✗	84.5	83.1	75.7	81.1	6.0	**5.1**	4.0	**5.0**
✓	✓	✓	✓	**85.1**	**84.3**	**76.1**	**81.8**	**5.7**	6.0	4.3	5.3

boundary delineation. However, the introduction of CLoss leads to a slight increase in Hausdorff Distance. I explain that the Hausdorff Distance measures the maximum distance between corresponding boundary points of the predicted and ground truth segmentations. While the CLoss module may improve the overall Dice Similarity Coefficient by enhancing segmentation overlap, it might introduce discrepancies in boundary delineation, leading to a higher maximum distance between boundary points. These findings are consistently supported by similar conclusions drawn from experiments conducted on the BraTS 2019 dataset, as illustrated in Table II. The proposed method attains a notable performance with an average Dice Similarity Coefficient of 82.7% and an average Hausdorff Distance of 5.1. Compared to the baseline method, it outperforms by 2.4% and 5.6% in terms of average Dice Similarity Coefficient and average Hausdorff Distance, respectively. In addition, the observation that the Dice Similarity Coefficient improves with the addition of each module (SGCM, CGCM, CLoss) suggests an improvement in the overall overlap. The Hausdorff Distance improves with the addition of the SGCM. However, introducing the CGCM results in the same value in the Hausdorff Distance as the baseline, while there is an improvement in the Dice Similarity Coefficient. This discrepancy between the metrics implies that the CGCM may contribute to better segmentation overlap. Interestingly, when comparing SGCM + CGCM with and without CLoss, Hausdorff Distance remains unchanged. This observation implies that the improvements in boundary accuracy achieved by SGCM and CGCM may already capture the majority of the benefits that CLoss would offer in terms of boundary refinement. Another notable observation is the significantly improved performance on the BraTS 2019 dataset, attributed to the inclusion of a larger number of training cases. In summary, the ablation experiments demonstrate the efficacy of the proposed components.

B. Comparison With the State-of-the-Art Methods

To evaluate the effectiveness of the proposed method, a comprehensive comparison is conducted against current leading methods, and the results are presented in Table III and Table IV. As shown in Table III, the method from [22] achieves the best Dice Similarity Coefficient of 78.1% on the enhancing tumor region, the method from [46] obtains the best Dice Similarity Coefficient of 88.8% on the whole tumor region. Nevertheless, the proposed method showcases superior performance on the tumor core region with 84.3% in terms of Dice Similarity Coefficient. In addition, the proposed method achieves the highest average Dice Similarity Coefficient of 81.8% and the best Hausdorff Distance of 6.0, 4.3, and 5.3 on the tumor core, enhancing tumor, and average result, respectively. This superiority is attributed to the innovative multi-modal graph-based framework, which effectively captures spatial and contextual information from different MR modalities. From Table IV, it can be observed that the method from [43] achieves the best Dice Similarity Coefficient of 89.7% and the best Hausdorff Distance of 6.0 on the whole tumor region. The method from [59] achieves the best Hausdorff Distance of 3.5 on the enhancing tumor region. However, the proposed method achieves the highest Dice Similarity Coefficient of 84.3% on tumor core, 78.3% on enhancing tumor, and 82.7% on average, best Hausdorff Distance of 5.0 on tumor core, and 5.1 on average.

To underscore the significance of the proposed method, additional comparisons are conducted with attention-based methods, specifically the squeeze-and-excitation (SE) block [48], convolutional block attention module (CBAM) [49], and the non-local network [50]. The SE block introduces a mechanism for channel-wise information capture and adaptive recalibration of feature maps, enhancing representational ability. CBAM combines channel and spatial attention modules to learn relevant channels and spatial locations, while the non-local network utilizes a non-local operation to capture long-range dependencies in feature maps. Observing the results, it is noted that the method from [48] achieves the best Hausdorff Distance of 5.6 on the whole tumor region. However, the proposed multi-modal graph convolution network demonstrates superior performance, achieving the highest Dice Similarity Coefficient of 84.3% on tumor core and 81.8% on average, best Hausdorff Distance of 6.0 on tumor core, 4.3 on enhancing tumor and 5.3 on average. It is essential to emphasize that the GCN-based method, representing the data as a graph, effectively captures relationships and dependencies between different regions. Furthermore, from Table IV, it can be observed that the method from [50] achieves the best Hausdorff Distance of 6.0 on the whole tumor region. However, the proposed

TABLE II

ABLATION EXPERIMENTS ON THE BRATS 2019 DATASET IN TERMS OF DICE SIMILARITY COEFFICIENT (%)AND 95% HAUSDORFF DISTANCE (MM), WT, TC, ET DENOTE WHOLE TUMOR, TUMOR CORE AND ENHANCING TUMOR, RESPECTIVELY; BOLD RESULTS DENOTE THE BEST SCORES

Method				Dice Similarity Coefficient (%) ↑				Hausdorff Distance (mm) ↓			
Baseline	SGCM	CGCM	CLoss	WT	TC	ET	Avg	WT	TC	ET	Avg
✓	✗	✗	✗	84.5	82.4	75.4	80.8	6.8	5.8	**3.7**	5.4
✓	✓	✗	✗	85.0	82.7	77.2	81.6	**6.2**	6.0	**3.7**	5.3
✓	✗	✓	✗	85.2	83.5	77.2	82.0	6.8	5.5	3.8	5.4
✓	✓	✓	✗	85.1	84.0	77.5	82.2	6.3	5.3	**3.7**	**5.1**
✓	✓	✓	✓	**85.6**	**84.3**	**78.3**	**82.7**	6.4	**5.0**	3.8	**5.1**

TABLE III

COMPARISON RESULTS WITH THE STATE-OF-THE-ART METHODS IN TERMS OF DICE SIMILARITY COEFFICIENT (%) AND 95% HAUSDORFF DISTANCE (MM) ON THE BRATS 2018 DATASET. WT, TC, ET DENOTE WHOLE TUMOR, TUMOR CORE AND ENHANCING TUMOR, RESPECTIVELY; BOLD RESULTS DENOTE THE BEST SCORES

Method	Dice Similarity Coefficient (%) ↑				Hausdorff Distance (mm) ↓			
	WT	TC	ET	Avg	WT	TC	ET	Avg
Hu et al. [51]	83.0	73.0	61.0	72.3	47.2	41.1	41.5	43.3
Zhou et al. [52]	80.4	75.9	67.0	74.4	7.0	8.6	6.7	7.4
Hu et al. [53]	85.7	76.9	72.0	78.2	10.8	10.0	5.5	8.8
Zhou et al. [43]	88.2	78.6	69.4	78.7	6.7	7.6	7.1	7.1
Zhang et al. [26]	87.6	77.2	72.0	78.9	-	-	-	-
Mehta et al. [46]	**88.8**	79.3	69.0	79.0	6.6	7.9	7.3	7.3
Kermi et al. [54]	86.7	79.8	71.7	79.4	8.7	6.4	4.7	6.6
Yang et al. [22]	83.8	78.6	**78.1**	80.2	5.9	7.1	4.7	5.9
Zhou et al. [5]	88.7	79.6	75.0	81.1	8.3	8.4	7.7	8.1
Weninger et al. [55]	86.0	81.7	76.3	81.3	7.0	7.9	5.6	6.8
Yao et al. [56]	88.5	83.4	72.9	81.6	17.0	7.1	5.7	9.9
Rehman et al. [29]	86.9	81.4	76.7	81.7	-	-	-	-
SE [48]	85.2	82.6	74.8	80.9	**5.6**	6.2	4.9	5.6
CBAM [49]	84.1	80.9	71.3	78.7	6.2	6.5	4.7	5.8
Non-local [50]	84.6	81.9	74.5	80.3	6.0	6.3	4.4	5.6
Ours	85.1	**84.3**	76.1	**81.8**	5.7	**6.0**	**4.3**	**5.3**

method achieves the highest Dice Similarity Coefficient of 84.3% on tumor core, 78.3% on enhancing tumor, and 82.7% on average, best Hausdorff Distance of 5.0 on tumor core, and 5.1 on average. This reinforces the effectiveness of the proposed method in leveraging multi-modal information and graph-based representation for improved medical image segmentation. However, the increase in Dice Similarity Coefficient may seem modest, its clinical significance lies in its potential to enhance treatment planning, disease monitoring, and overall patient care. Furthermore, in research settings, small increases in Dice Similarity Coefficient can signal advancements in algorithmic performance and pave the way for further innovations in medical image analysis.

C. Visualization of the Segmentation Results

In this section, I present the visualization of the segmentation results obtained from the proposed method on the BraTS 2018 dataset, illustrated in Figure 6. The visualization results aim to provide a qualitative assessment of the model's performance in delineating brain tumor regions. I selected

TABLE IV
COMPARISON RESULTS WITH THE STATE-OF-THE-ART METHODS IN TERMS OF DICE SIMILARITY COEFFICIENT (%) AND 95% HAUSDORFF DISTANCE (MM) ON THE BRATS 2019 DATASET. WT, TC, ET DENOTE WHOLE TUMOR, TUMOR CORE AND ENHANCING TUMOR, RESPECTIVELY; BOLD RESULTS DENOTE THE BEST SCORES

Method	Dice Similarity Coefficient (%) ↑				Hausdorff Distance (mm) ↓			
	WT	TC	ET	Avg	WT	TC	ET	Avg
Weninger et al. [57]	82.0	76.0	65.0	74.3	11.4	10.5	9.2	10.4
Zhou et al. [52]	82.9	79.5	72.1	78.2	7.4	7.7	6.1	7.1
Di et al. [58]	86.0	77.3	71.5	78.3	-	-	-	-
Zhou et al. [43]	**89.7**	77.5	70.6	79.3	**6.0**	7.5	7.6	7.0
Sheng et al. [47]	88.1	79.6	70.7	79.5	8.4	7.1	**3.5**	6.3
Zhou et al. [59]	88.7	76.5	75.9	80.4	7.9	8.8	8.0	8.2
Rehman et al. [29]	88.4	81.4	76.3	82.0	-	-	-	-
SE [48]	84.6	83.6	76.9	81.7	6.6	6.5	4.3	5.8
CBAM [49]	85.3	83.4	77.3	82.0	6.5	5.9	4.0	5.5
Non-local [50]	85.5	83.4	77.2	82.0	**6.0**	5.6	4.6	5.4
Ours	85.6	**84.3**	**78.3**	**82.7**	6.4	**5.0**	3.8	**5.1**

some examples and compared the ground truth annotations with the segmentation predictions produced by the proposed method. It can be observed that in the first case, the baseline method reveals some false predictions, highlighted in green. The gradual integration of the proposed SGCM and CGCM visibly enhances results. Integrating the proposed multi-modal segmentation loss function can further achieve superior performance with Dice Similarity Coefficients of 92.3%, 94.8%, and 86.0%. Moving to the second and third cases, the baseline method exhibits false predictions in the edema region. However, the proposed components contribute to refining these predictions. We can also observe an increase in Dice Similarity Coefficient. Hence, these visualization results serve as additional evidence, further reinforcing the superior segmentation performance achieved by the proposed method.

Moreover, I have selected specific examples to conduct a comparative analysis of their visualization results with attention-based methods, and the findings are detailed in Figure 7. In the first case, all three attention-based methods (SE, CBAM, and Non-local) exhibit a tendency to generate false edema regions around the tumor boundary, whereas the proposed GCN-based method yields more accurate results, achieving the Dice Similarity Coefficient Dice of 91.0%, 93.3% and 85.4% on the whole tumor, tumor core and enhancing tumor regions, respectively. In the second case, the SE method outperforms CBAM and Non-local in detecting edema regions, but the proposed method excels in capturing superior results within the tumor core region (highlighted in blue and red), achieving the Dice Similarity Coefficient Dice of 87.3%, 91.0% and 83.2% on the whole tumor, tumor core and enhancing tumor regions, respectively. In the last case, where the enhancing and NET&NCR regions are notably small, attention-based methods fail to detect the two regions, however, the proposed method demonstrates improved accuracy, as indicated by the Dice Similarity Coefficient with 92.5%, 89.2% and 81.9% on the whole tumor, tumor core and enhancing tumor regions, respectively. Thus, the comprehensive analysis of these results underscores the efficacy of the proposed method.

VI. DISCUSSION

A. Efficacy of the Proposed M2GCM

To assess the impact of applying the proposed M2GCM at various levels of the network on overall performance, experiments are conducted using the BraTS 2018 dataset, and the results are illustrated in Table V. The first two layers of the network have 16 and 32 filters, respectively. When M2GCM is set in the first level, the parameters in the CGCM are configured as $C_1 = 2$ and $C_2 = 1$. In the second level, within the CGCM, $C_1 = 4$ and $C_2 = 2$. The proposed M2GCM is considered for application at deeper levels (3rd, 4th, 5th, 6th) of the network to enhance feature extraction learning. From the comparison results, it is observed that applying the M2GCM at the 4th level of the network achieves the best Hausdorff Distance of 6.0 and 4.3 on tumor core and enhancing tumor regions, respectively. Applying it on the 5th level yields the best Hausdorff Distance of 5.7 on the whole tumor region. However, the best performance in terms of average Dice Similarity Coefficient and average Hausdorff Distance, respectively, is achieved when applying the M2GCM at the bottom level of the network. This observation suggests that the M2GCM excels in capturing intricate relationships among multi-modal information, particularly when applied to

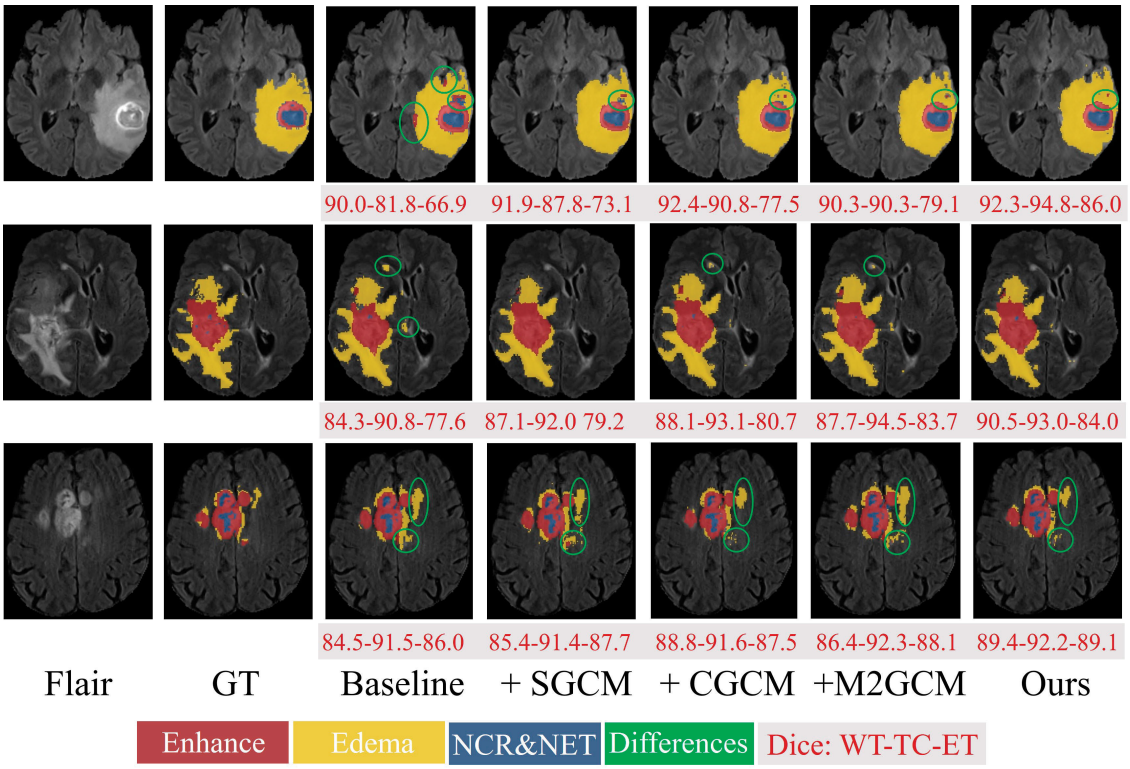

Fig. 6. Visualization of the segmentation results. From left to right: Flair modality, Ground-truth, Baseline, Baseline + SGCM, Baseline + CGCM, Baseline + M2GCM and the proposed method (Baseline + M2GCM + CLoss). M2GCM denotes the application of both SGCM and CGCM.

TABLE V

EXPERIMENT OF APPLYING THE PROPOSED M2GCM AT DIFFERENT LEVELS OF THE SEGMENTATION NETWORK, CONDUCTED ON THE BRATS 2018 DATASET IN TERMS OF DICE SIMILARITY COEFFICIENT (%) AND 95% HAUSDORFF DISTANCE (MM)

No. level	Dice Similarity Coefficient (%) ↑				Hausdorff Distance (mm) ↓			
	WT	TC	ET	Avg	WT	TC	ET	Avg
3rd	84.9	80.9	73.7	79.8	5.8	7.3	5.5	6.2
4th	84.6	83.4	75.0	81.0	5.8	**6.0**	**4.3**	5.4
5th	84.7	82.0	74.9	80.5	**5.7**	6.9	4.8	5.8
6th	**85.1**	**84.3**	**76.1**	**81.8**	**5.7**	**6.0**	**4.3**	**5.3**

high-level representations. To this end, the proposed method applies the M2GCM at the bottom level of the network.

B. Efficacy of the Proposed CLoss

To evaluate the impact of incorporating CLoss at various levels of the segmentation network, the experiments are conducted on BraTS 2018 dataset, and the results are illustrated in Table VI. It can be observed that applying the CLoss at the 2nd level of the network achieves the best Hausdorff Distance of 4.2 on the enhancing tumor region. Applying the CLoss at the 4th level of the network achieves the best Hausdorff Distance of 5.7 on the tumor core region. However, the best performance is obtained when considering the correlation of the features from the bottom level. The experiment results prove that incorporating the proposed CLoss at the last level of the segmentation network yields the best results with 85.1%, 84.3% 76.1%, and 81.8% on the whole tumor, tumor core, enhancing tumor and average result in terms of Dice Similarity Coefficient, and 5.7, 5.3 on the whole tumor and average result in terms of Hausdorff Distance. This observation indicates that the model benefits significantly from capturing higher-level contextual information in the final stages of feature extraction, leading to more precise boundary delineation of brain tumors.

TABLE VI

EXPERIMENT OF APPLYING THE PROPOSED CORRELATION LOSS (CLOSS) AT DIFFERENT LEVELS OF THE SEGMENTATION NETWORK, CONDUCTED ON THE BRATS 2018 DATASET IN TERMS OF DICE SIMILARITY COEFFICIENT (%) AND 95% HAUSDORFF DISTANCE (MM)

No. level	Dice Similarity Coefficient (%) ↑				Hausdorff Distance (mm) ↓			
	WT	TC	ET	Avg	WT	TC	ET	Avg
1st	83.8	82.6	74.0	80.1	6.6	6.4	4.8	5.9
2nd	83.9	80.9	74.5	79.8	6.1	6.1	**4.2**	5.5
3rd	83.1	81.0	73.9	79.3	5.9	6.9	4.6	5.8
4th	83.6	80.7	73.1	79.1	6.7	**5.7**	4.5	5.6
5th	84.1	81.2	74.0	79.8	6.7	6.7	5.7	6.4
6th	**85.1**	**84.3**	**76.1**	**81.8**	**5.7**	6.0	4.3	**5.3**

C. Efficacy of the Proposed SGCM and CGCM

To evaluate the design of the model, I compare the effects of applying concatenation or addition operations between the two sub-graph modules (from the two pairs of correlated modalities) in the proposed SGCM (refer to Section III-C.1) and CGCM (refer to Section III-C.2). For simplicity, the

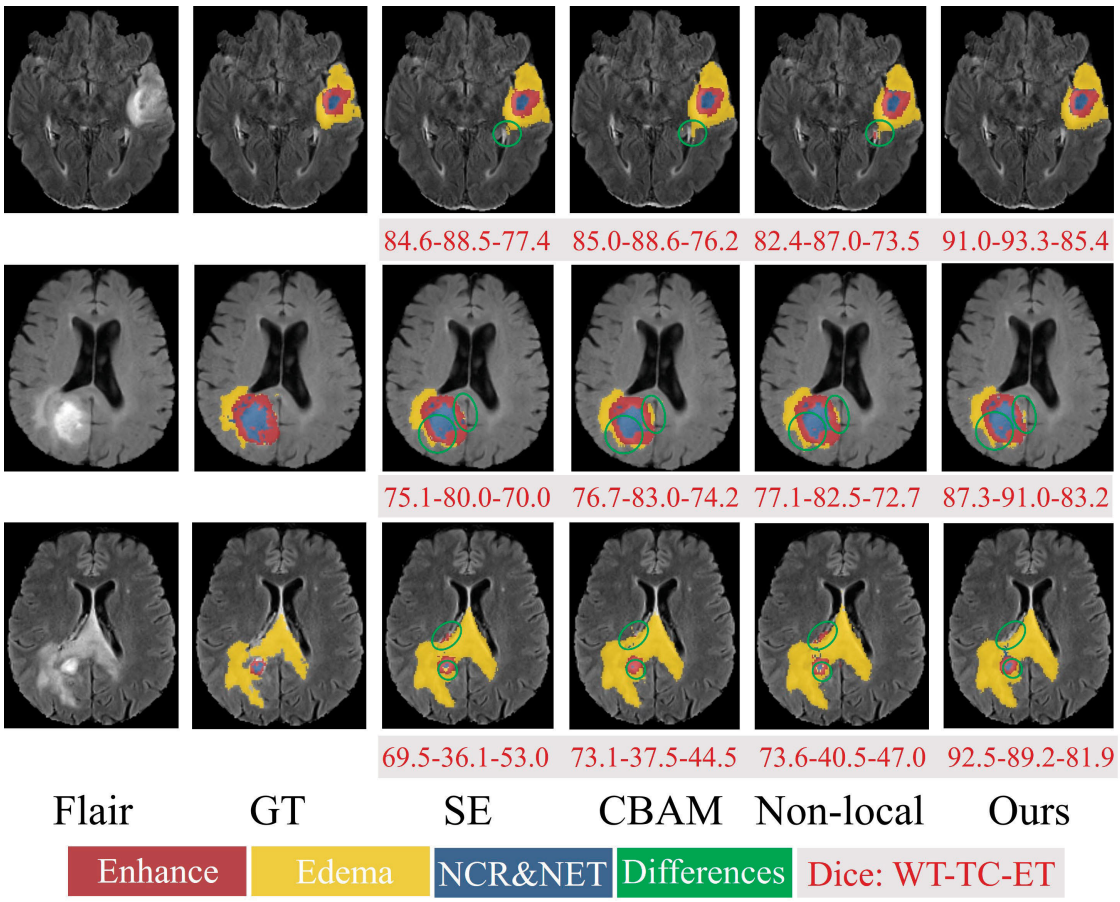

Fig. 7. Visualization of the segmentation results. From left to right: Flair modality, Ground-truth, SE, CBAM, Non-local and the proposed method.

experiments are conducted on BraTS 2018 dataset. The corresponding results are illustrated in Table VII. It can be observed that applying concatenation operation on SGCM and addition operation on CGCM can obtain the best Hausdorff Distance of 6.0 on the tumor core region. However, applying concatenation operation on both SGCM and CGCM obtains the best performance, with 85.1%, 84.3%, 76.1% and 81.8% in terms of Dice Similarity Coefficient, and 5.7, 6.0, 4.3, 5.3 in terms of Hausdorff Distance, on the whole tumor, tumor core, enhancing tumor and average result, respectively. This implies that concatenating features at the chosen layer allow the model to leverage more comprehensive contextual information, enhancing its ability to precisely identify tumor regions. Therefore, the proposed method adopts the concatenation operation due to its superior performance.

TABLE VII

EXPERIMENT OF APPLYING ADDITION OR CONCATENATION OPERATION ON SGCM AND CGCM, CONDUCTED ON THE BRATS 2018 DATASET IN TERMS OF DICE SIMILARITY COEFFICIENT (%) AND 95% HAUSDORFF DISTANCE (MM). $\|$ AND $\oplus$ DENOTE THE CONCATENATION AND ADDITION OPERATIONS, RESPECTIVELY

Method		Dice Similarity Coefficient (%) ↑				Hausdorff Distance (mm) ↓			
SGCM	CGCM	WT	TC	ET	Average	WT	TC	ET	Average
$\|$	$\oplus$	85.0	81.8	74.6	80.5	5.8	**6.0**	4.5	5.4
$\oplus$	$\|$	84.8	82.0	75.5	80.8	6.2	6.5	4.6	5.8
$\oplus$	$\oplus$	82.9	80.2	72.2	78.4	6.8	8.1	5.4	6.8
$\|$	$\|$	**85.1**	**84.3**	**76.1**	**81.8**	**5.7**	**6.0**	**4.3**	**5.3**

D. Limitation and Future Work

In evaluating the proposed segmentation method, it is crucial to comprehensively analyze its successes and cases where the segmentation process falls short. In this section, I visualize some failed segmentation cases for potential improvement of the proposed method, and the results are presented in Figure 8. It can be observed that, in the first case, the NET&NCR regions are characterized by their small and scattered nature, highlighted within green circles. The proposed method encounters difficulty in detecting these regions accurately. Additionally, in the second case, the segmentation of the left edema region proves to be challenging, which arises from the fact that this region is not clearly recognized in the original MR modalities, leading to a failure of the proposed method to detect it accurately. The analysis of the failed segmentation cases sheds light on the limitations of the proposed method. Strategies for improving the detection of small and scattered regions could be explored to refine the segmentation performance. For example, applying post-processing techniques, such as morphological operations, to refine segmented regions. This can be useful for dealing with scattered regions. Additionally, addressing challenges posed by inherently ambiguous areas, implementing uncertainty modelling

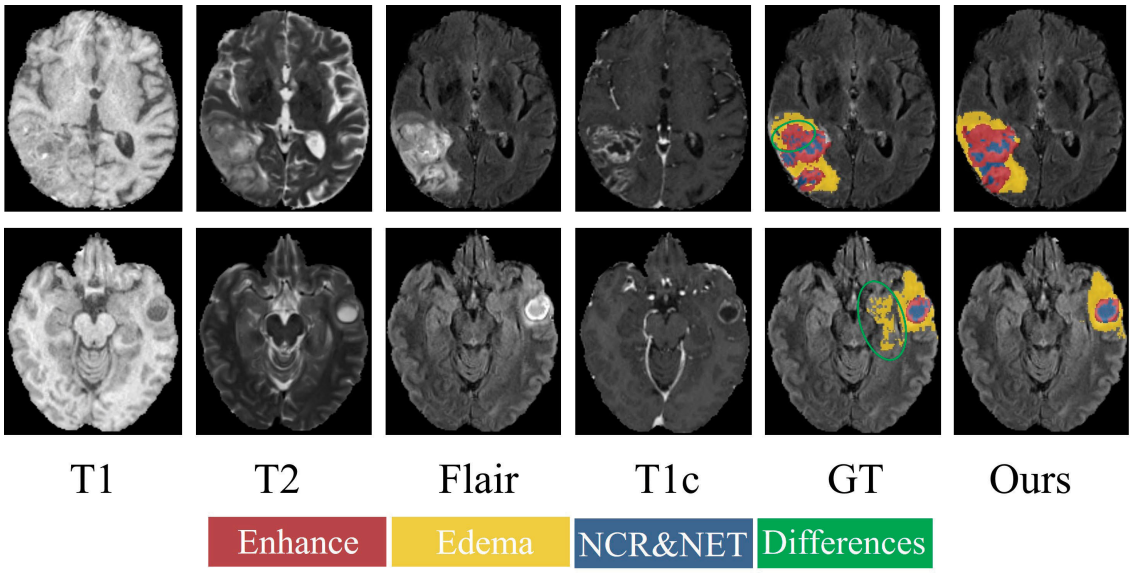

Fig. 8. Visualization of some failed segmentation cases. From left to right: T1 modality, T2 modality, Flair modality, T1c modality, Ground-truth and the proposed method.

techniques, such as Bayesian neural networks or Monte Carlo dropout, to quantify and account for uncertainty in predictions. This can be useful in ambiguous regions where the model is unsure about the correct segmentation. By addressing these limitations, the proposed method can better handle complex scenarios and contribute to advancements in brain tumor segmentation.

The complex nature of the proposed network, including multiple specialized components (SGCM, CGCM, and the multi-modal correlation loss), may pose challenges in terms of computational efficiency, scalability, and interpretability, which are crucial for clinical applications. To address these concerns, further investigations and optimizations are planned. This includes exploring techniques such as model pruning, knowledge distillation, constructing sparse graphs to reduce the number of connections, or focusing computations on regions of interest (ROIs) to improve computational efficiency. Additionally, efforts will be made to enhance the scalability of the network by considering strategies for distributed computing or parallel processing to handle larger datasets and more complex segmentation tasks. Furthermore, interpretability is vital in the clinical decision-making process, and methods will be explored to enhance it. This may involve incorporating techniques for visualizing and explaining model predictions, along with conducting comprehensive validation studies to ensure the reliability and trustworthiness of the segmentation results.

Moreover, integrating advanced AI-based tools like the proposed method into existing clinical workflows presents both opportunities and challenges. One key aspect is ensuring the interoperability of the AI-based tools with existing clinical systems, which involves standardizing data formats and protocols for data exchange. Additionally, addressing privacy and security concerns, such as patient data confidentiality, is important. Another challenge is ensuring the usability of the AI tool by healthcare professionals, requiring strategies for designing intuitive user interfaces and providing adequate training and support to clinical staff. Furthermore, it should address the potential impact of AI-based tools on clinical decision-making and patient care. Developing transparent and interpretable models using explainable AI techniques can ensure that clinicians can understand and trust AI predictions. Overall, discussing these considerations and potential solutions can offer valuable insights into effectively integrating the proposed method into healthcare settings, ultimately enhancing patient outcomes and advancing clinical practice.

VII. Conclusion

This paper proposes a multi-modal graph convolution network for brain tumor segmentation across multiple MRI sequences. The proposed M2GCNet can effectively capture spatial and contextual information via spatial-wise graph convolution network and channel-wise graph convolution network, enabling it to exploit the complementary information across different MR modalities. Fusing various MR modalities through the multi-modal graph convolution network allows for a more comprehensive understanding of the tumor boundaries and surrounding tissues, leading to more accurate segmentation results. In addition, a novel multi-modal correlation loss is proposed to capture latent relationships among features from different modalities and learn effective feature representations for brain tumor segmentation. The experiments conducted on publicly available datasets have consistently demonstrated the superiority of the proposed method in terms of Dice similarity coefficient and Hausdorff distance. Furthermore, the proposed method has highlighted the importance of multi-modal data fusion and deep learning techniques in the context of brain tumor analysis. As a part of future endeavours, there is a keen interest in exploring the integration of more advanced neural network architectures and leveraging larger datasets to further refine the performance of multi-modal brain tumor segmentation. Developing interpretable and explainable models will

also be crucial for gaining clinicians' trust and facilitating the clinical adoption of the advanced segmentation techniques.

REFERENCES

[1] T. A. Soomro et al., "Image segmentation for MR brain tumor detection using machine learning: A review," *IEEE Rev. Biomed. Eng.*, vol. 16, pp. 70–90, 2023.

[2] C. Kruchko, Q. T. Ostrom, H. Gittleman, and J. S. Barnholtz-Sloan, "The CBTRUS story: Providing accurate population-based statistics on brain and other central nervous system tumors for everyone," *Neuro-Oncol.*, vol. 20, no. 3, pp. 295–298, Feb. 2018.

[3] R. Ranjbarzadeh, A. Caputo, E. B. Tirkolaee, S. J. Ghoushchi, and M. Bendechache, "Brain tumor segmentation of MRI images: A comprehensive review on the application of artificial intelligence tools," *Comput. Biol. Med.*, vol. 152, Jan. 2023, Art. no. 106405.

[4] M. Ghaffari, A. Sowmya, and R. Oliver, "Automated brain tumor segmentation using multimodal brain scans: A survey based on models submitted to the BraTS 2012–2018 challenges," *IEEE Rev. Biomed. Eng.*, vol. 13, pp. 156–168, 2020.

[5] T. Zhou, S. Ruan, P. Vera, and S. Canu, "A tri-attention fusion guided multi-modal segmentation network," *Pattern Recognit.*, vol. 124, Apr. 2022, Art. no. 108417.

[6] T. Zhou, S. Ruan, and S. Canu, "A review: Deep learning for medical image segmentation using multi-modality fusion," *Array*, vols. 3–4, Sep. 2019, Art. no. 100004.

[7] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," 2016, *arXiv:1609.02907*.

[8] Y. Li, Y. Zhang, W. Cui, B. Lei, X. Kuang, and T. Zhang, "Dual encoder-based dynamic-channel graph convolutional network with edge enhancement for retinal vessel segmentation," *IEEE Trans. Med. Imag.*, vol. 41, no. 8, pp. 1975–1989, Aug. 2022.

[9] P. Xuan et al., "Dynamic graph convolutional autoencoder with node-attribute-wise attention for kidney and tumor segmentation from CT volumes," *Knowl.-Based Syst.*, vol. 236, Jan. 2022, Art. no. 107360.

[10] Y. Chen, M. Rohrbach, Z. Yan, Y. Shuicheng, J. Feng, and Y. Kalantidis, "Graph-based global reasoning networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 433–442.

[11] Z. Liu et al., "Deep learning based brain tumor segmentation: A survey," *Complex Intell. Syst.*, vol. 9, no. 1, pp. 1001–1026, 2023.

[12] S. Vadhnani and N. Singh, "Brain tumor segmentation and classification in MRI using SVM and its variants: A survey," *Multimedia Tools Appl.*, vol. 81, no. 22, pp. 31631–31656, Sep. 2022.

[13] A. S. Reddy and P. C. Reddy, "MRI brain tumor segmentation and prediction using modified region growing and adaptive SVM," *Soft Comput.*, vol. 25, no. 5, pp. 4135–4148, Mar. 2021.

[14] C. Ma, G. Luo, and K. Wang, "Concatenated and connected random forests with multiscale patch driven active contour model for automated brain tumor segmentation of MR images," *IEEE Trans. Med. Imag.*, vol. 37, no. 8, pp. 1943–1954, Aug. 2018.

[15] A. Phophalia and P. Maji, "Multimodal brain tumor segmentation using ensemble of forest method," in *Proc. Int. MICCAI Brainlesion Workshop*, Quebec City, QC, Canada. Cham, Switzerland: Springer, Sep. 14, 2017, pp. 159–168.

[16] M. Havaei, P.-M. Jodoin, and H. Larochelle, "Efficient interactive brain tumor segmentation as within-brain kNN classification," in *Proc. 22nd Int. Conf. Pattern Recognit.*, Aug. 2014, pp. 556–561.

[17] P. Vasuda and S. Satheesh, "Improved fuzzy c-means algorithm for MR brain image segmentation," *Int. J. Comput. Sci. Eng.*, vol. 2, no. 5, p. 1713, 2010.

[18] A. Bal, M. Banerjee, A. Chakrabarti, and P. Sharma, "MRI brain tumor segmentation and analysis using rough-fuzzy C-means and shape based properties," *J. King Saud Univ. Comput. Inf. Sci.*, vol. 34, no. 2, pp. 115–133, Feb. 2022.

[19] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Jan. 1998.

[20] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 3431–3440.

[21] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. 18th Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, vol. 9351. Cham, Switzerland: Springer, 2015, pp. 234–241.

[22] H.-Y. Yang and J. Yang, "Automatic brain tumor segmentation with contour aware residual network and adversarial training," in *Proc. Int. MICCAI Brainlesion Workshop*, Granada, Spain. Cham, Switzerland: Springer, Sep. 2018, pp. 267–278.

[23] K. Kamnitsas et al., "Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation," *Med. Image Anal.*, vol. 36, pp. 61–78, Feb. 2017.

[24] A. Myronenko, "3D MRI brain tumor segmentation using autoencoder regularization," in *Proc. Int. MICCAI Brainlesion Workshop*, 2018, pp. 311–320.

[25] F. Isensee, P. F. Jäger, P. M. Full, P. Vollmuth, and K. H. Maier-Hein, "nnU-Net for brain tumor segmentation," in *Proc. Int. MICCAI Brainlesion Workshop*, Oct. 2020, pp. 118–132.

[26] J. Zhang, Z. Jiang, J. Dong, Y. Hou, and B. Liu, "Attention gate ResU-Net for automatic MRI brain tumor segmentation," *IEEE Access*, vol. 8, pp. 58533–58545, 2020.

[27] T. Zhou and S. Zhu, "Uncertainty quantification and attention-aware fusion guided multi-modal MR brain tumor segmentation," *Comput. Biol. Med.*, vol. 163, Sep. 2023, Art. no. 107142.

[28] A. M. Gab Allah, A. M. Sarhan, and N. M. Elshennawy, "Edge U-Net: Brain tumor segmentation using MRI based on deep U-net model with boundary information," *Expert Syst. Appl.*, vol. 213, Mar. 2023, Art. no. 118833.

[29] M. U. Rehman, J. Ryu, I. F. Nizami, and K. T. Chong, "RAAGR2-Net: A brain tumor segmentation network using parallel processing of multiple spatial frames," *Comput. Biol. Med.*, vol. 152, Jan. 2023, Art. no. 106426.

[30] R. Zhang, S. Jia, M. J. Adamu, W. Nie, Q. Li, and T. Wu, "HMNet: Hierarchical multi-scale brain tumor segmentation network," *J. Clin. Med.*, vol. 12, no. 2, p. 538, Jan. 2023.

[31] Q. Li, H. Liu, W. Nie, and T. Wu, "Brain tumor image segmentation based on prior knowledge via transformer," *Int. J. Imag. Syst. Technol.*, vol. 33, no. 6, pp. 2073–2087, Nov. 2023.

[32] T. Zhou, "Multi-modal brain tumor segmentation via disentangled representation learning and region-aware contrastive learning," *Pattern Recognit.*, vol. 149, May 2024, Art. no. 110282.

[33] Z. Luo, L. Liu, J. Yin, Y. Li, and Z. Wu, "Deep learning of graphs with nGram convolutional neural networks," *IEEE Trans. Knowl. Data Eng.*, vol. 29, no. 10, pp. 2125–2139, Oct. 2017.

[34] J. Gao, T. Zhang, and C. Xu, "Graph convolutional tracking," in *Proc. EEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jul. 2019, pp. 4649–4659.

[35] C. R. Qi, L. Yi, H. Su, and L. J. Guibas, "PointNet++: Deep hierarchical feature learning on point sets in a metric space," in *Proc. Adv. neural Inf. Process. Syst.*, vol. 30, 2017, pp. 1–11.

[36] R. Wang, C. Huang, and X. Wang, "Global relation reasoning graph convolutional networks for human pose estimation," *IEEE Access*, vol. 8, pp. 38472–38480, 2020.

[37] H. Gao and S. Ji, "Graph U-Nets," in *Proc. Int. Conf. Mach. Learn.*, 2019, pp. 2083–2092.

[38] T. Zhao et al., "3D graph anatomy geometry-integrated network for pancreatic mass segmentation, diagnosis, and quantitative patient management," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 13738–13747.

[39] Y. Huang and A. C. S. Chung, "Disease prediction with edge-variational graph convolutional networks," *Med. Image Anal.*, vol. 77, Apr. 2022, Art. no. 102375.

[40] Q. Ma et al., "DGRUnit: Dual graph reasoning unit for brain tumor segmentation," *Comput. Biol. Med.*, vol. 149, Oct. 2022, Art. no. 106079.

[41] L. Mishra and S. Verma, "Graph attention autoencoder inspired CNN based brain tumor classification using MRI," *Neurocomputing*, vol. 503, pp. 236–247, Sep. 2022.

[42] Z. Liu et al., "CANet: Context aware network for brain glioma segmentation," *IEEE Trans. Med. Imag.*, vol. 40, no. 7, pp. 1763–1777, Jul. 2021.

[43] T. Zhou, S. Canu, P. Vera, and S. Ruan, "Latent correlation representation learning for brain tumor segmentation with missing MRI modalities," *IEEE Trans. Image Process.*, vol. 30, pp. 4263–4274, 2021.

[44] J. Lapuyade-Lahorgue, J.-H. Xue, and S. Ruan, "Segmenting multi-source images using hidden Markov fields with copula-based multivariate statistical distributions," *IEEE Trans. Image Process.*, vol. 26, no. 7, pp. 3187–3195, Jul. 2017.

[45] B. B. Avants, N. Tustison, and G. Song, "Advanced normalization tools (ANTS)," *Insight J.*, vol. 2, pp. 1–35, Jun. 2009.

[46] R. Mehta and T. Arbel, "3D U-Net for brain tumour segmentation," in *Proc. Int. MICCAI Brainlesion Workshop*, Granada, Spain. Cham, Switzerland: Springer, Sep. 16, 2018, pp. 254–266.

[47] N. Sheng, D. Liu, J. Zhang, C. Che, and J. Zhang, "Second-order ResU-Net for automatic MRI brain tumor segmentation," *Math. Biosciences Eng.*, vol. 18, no. 5, pp. 4943–4960, 2021.

[48] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2018, pp. 7132–7141.

[49] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis.*, Sep. 2018, pp. 3–19.

[50] X. Wang, R. Girshick, A. Gupta, and K. He, "Non-local neural networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7794–7803.

[51] Y. Hu, X. Liu, X. Wen, C. Niu, and Y. Xia, "Brain tumor segmentation on multimodal MR imaging using multi-level upsampling in decoder," in *Proc. Int. MICCAI Brainlesion Workshop*, Granada, Spain. Cham, Switzerland: Springer, Sep. 16, 2018, pp. 168–177.

[52] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: A nested U-Net architecture for medical image segmentation," 2018, *arXiv:1807.10165*.

[53] X. Hu, H. Li, Y. Zhao, C. Dong, B. H. Menze, and M. Piraud, "Hierarchical multi-class segmentation of glioma images using networks with multi-level activation function," in *Proc. Int. MICCAI Brainlesion Workshop*. Cham, Switzerland: Springer, 2018, pp. 116–127.

[54] A. Kermi, I. Mahmoudi, and M. T. Khadir, "Deep convolutional neural networks using U-Net for automatic brain tumor segmentation in multimodal MRI volumes," in *Proc. Int. MICCAI Brainlesion Workshop*. Cham, Switzerland: Springer, 2018, pp. 37–48.

[55] L. Weninger et al., "Segmentation of brain tumors in 3D-MRI data and patient survival prediction: Methods for the brats 2018 challenge," in *Proc. 4th Int. Workshop Brainlesion, Glioma, Multiple Sclerosis, Stroke Traumatic Brain Injuries*. Granada, Spain: Springer, Sep. 2018.

[56] H. Yao, X. Zhou, and X. Zhang, "Automatic segmentation of brain tumor using 3D se-inception networks with residual connections," in *Proc. Int. MICCAI Brainlesion Workshop*. Cham, Switzerland: Springer, 2018, pp. 346–357.

[57] L. Weninger, Q. Liu, and D. Merhof, "Multi-task learning for brain tumor segmentation," in *Proc. Int. MICCAI Brainlesion Workshop*. Cham, Switzerland: Springer, 2019, pp. 327–337.

[58] A. Di Ieva et al., "Application of deep learning for automatic segmentation of brain tumors on magnetic resonance imaging: A heuristic approach in the clinical scenario," *Neuroradiology*, vol. 63, no. 8, pp. 1253–1262, Aug. 2021.

[59] T. Zhou, S. Canu, P. Vera, and S. Ruan, "A dual supervision guided attentional network for multimodal MR brain tumor segmentation," in *Proc. Int. Conf. Med. Imag. Comput.-Aided Diagnosis*. Cham, Switzerland: Springer, 2021, pp. 3–11.

Tongxue Zhou received the Ph.D. degree in computer science from the National Institute of Applied Sciences of Rouen (INSA Rouen), France, in 2022. She is currently a Lecturer with the School of Information Science and Technology, Hangzhou Normal University, China. Her current research interests include medical image analysis, deep learning, and information fusion.