International Journal of Imaging Systems and Technology

WILEY

RESEARCH ARTICLE

Multimodal Connectivity-Guided Glioma Segmentation From Magnetic Resonance Images via Cascaded 3D Residual U-Net

Xiaoyan Sun | Chuhan Hu | Wenhan He | Zhenming Yuan | Jian Zhang

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

Correspondence: Jian Zhang (jeyzhang@hznu.edu.cn)

Received: 14 June 2024 | **Revised:** 30 August 2024 | **Accepted:** 10 October 2024

Funding: This work was supported by National Natural Science Foundation of China (Grant number [61902103]).

Keywords: 3D U-Net | brain tumor segmentation | hierarchical structure | multimodal MRI | residual network

ABSTRACT

Glioma is a type of brain tumor with a high mortality rate. Magnetic resonance imaging (MRI) is commonly used for examination, and the accurate segmentation of tumor regions from MR images is essential to computer-aided diagnosis. However, due to the intrinsic heterogeneity of brain glioma, precise segmentation is very challenging, especially for tumor subregions. This article proposed a two-stage cascaded method for brain tumor segmentation that considers the hierarchical structure of the target tumor subregions. The first stage aims to identify the whole tumor (WT) from the background area; and the second stage aims to achieve fine-grained segmentation of the subregions, including enhanced tumor (ET) region and tumor core (TC) region. Both stages apply a deep neural network structure combining modified 3D U-Net with a residual connection scheme to tumor region and subregion segmentation. Moreover, in the training phase, the 3D masks generation of subregions with potential incomplete connectivity are guided by the completely connected regions. Experiments were performed to evaluate the performance of the methods on both area and boundary accuracy. The average dice score of the WT, TC, and ET regions on BraTS 2020 dataset is 0.9168, 0.0.8992, 0.8489, and the Hausdorff distance is 6.021, 9.203, 12.171, respectively. The proposed method outperforms current works, especially in segmenting fine-grained tumor subregions.

1 | Introduction

Glioma is a kind of commonly seen malignant brain tumor that occupies a high proportion of all nervous system tumors, and the clinical diagnosis usually relies on magnetic resonance imaging (MRI) scans of the brain. MRI scans provide valuable information about the glioma structures, which help doctors in identifying the whole tumor (WT) region and subregions including tumor core (TC) and enhancing tumor (ET) region that are essential to diagnosis. Accurately telling the tumor region (subregions) and healthy tissues apart from MRI scans is the key to diagnosing and treating. In computer-aided glioma diagnosis, this is referred to as glioma segmentation (from MR images). However, because the growth of glioma is diffuse and invasive, glioma segmentation remains a big challenge.

MRI examination usually generates different sequences, which have different abilities for imaging specific tissue characteristics and are usually referred to as multimodal MRI data. Mainly used sequences for tumor imaging include T1-weighted MRI (T1), T2-weighted MRI (T2), T1-weighted MRI with gadolinium contrast enhancement (T1-Gd or T1ce), and Fluid Attenuated Inversion Recovery (FLAIR). The multi–modal Brain Tumor Image Segmentation Benchmark (BRATS) provides preregistered multimodal MR images, including the four sequences mentioned above [1–3]. Among these, T1 images are

used for distinguishing healthy tissues, whereas T2 images are used to delineate the edema region [4]. Therefore, learning from multisequence (multimodal) MRI data has become the recent trend for brain tumor segmentation. Most works simply concatenate all four sequences as the input of the segmentation model [5, 6]. Dehghani et al. [7] explored the influence of single and multiple MRI sequences on glioma segmentation results and concluded that segmentation with multi- and dual-channel sequences outperforms the segmentation with single-channel MR sequences. Cinar et al. [8] selectively used only Flair, T2, and T1ce to segment the tumor area, considering that the T1 image resolution is lower. Tseng et al. [9] proposed cross-modality convolution layer after the encoder combines information from different modalities. Zhang et al. [10] treated T1-Tce and T2-FLAIR as modality pairs and learned features from cross-modality-pair for segmentation. Effectively fusing multimodal MRI data can boost the segmentation performance, but this is still an open problem.

Different strategies have been adopted for the target segmentation areas WT, TC, and ET. The basic one is segmenting the three areas with a single multiclassification network [6, 11]. Other typical strategies include cascaded or ensembled networks [12]. Many researchers use cascaded structures to segment different areas separately, with the same or different segmentation models. Lyu et al. [13] proposed a two-stage codec segmentation model. In the first stage, relatively rough segmentation was performed. In the second stage, the concatenation of the preliminary segmentation maps from the first stage and the MRI image were used as input to refine the sub-region prediction. Chinnam et al. [14] proposed a cascaded model which segments the WT with a nine layer U-Net model and the TC and ET later with a seven layer U-Net model. Li et al. [15] used a cascaded architecture to decompose the segmentation into three two-class classification subtasks. Different subnetworks were used for subtasks, where WT and TC were segmented with 3D U-Net, and ET was segmented with 3D U-Net++. Each submodel was trained on slices in three orthogonal planes. The cascaded structure has been commonly recognized as effective for WT, TC, and ET region segmentation.

U-Net [16], 3D U-Net [17], and their variants are mainly used networks for MRI-based segmentation tasks, but researchers may vary the network structures in specific works. In [14], different number of layers of U-Net were designed with attention gates. Squeeze-and-excitation (SE) and attention modules were added to the U-Net structures to improve the segmentation performance in [18]. In [8], a hybrid model was proposed using DenseNet121 instead of the encoder in U-Net. Similarly, a deep residual network is used as the encoder in the hybrid architecture with the decoder of the U-Net in [6]. Because the MR images contain 3D information, many researchers selected 3D U-Net as the baseline model and modified it to improve the effect. Fidon et al. [19] used a general 3D U-Net architecture, combined with three optimizations including the generalized Wasserstein Dice loss, etc. Qamar et al. [20] proposed a hyper-dense inception 3D U-Net with stacking factorization of 3D-weighted convolutional layers in the residual inception block to improve brain tumor segmentation. In both BraTS Challenge 2020 [21] and 2021 [22], the top solutions were based on nn UNet proposed by Isensee et al. [23, 24], which empirically chooses the best-performing configuration or ensemble of three different U-Net configurations with cross-validation. However, MRI-based glioma segmentation treats each sample as a 3D cube composed of MRI-scanned voxels, which occupies much memory resource and inevitably reduces the batch size. Accordingly, the batch normalization of 3D MRI data lacks stability and influences the training. In addition, the ReLU activation in 3D U-Net is susceptible to the gradient vanishing problem about negative input. More importantly, the layer-wise feature representation of original 3D U-Net is still unsatisfactory, which may diminish the ability to extract multiscale information.

To this end, this article proposes a multimodal glioma segmentation approach via cascaded 3D RU-Net with residual connection. The segmentation consists of two stages. We fuse T2 and FLAIR sequence (modality) through element-wise linear combination as the input of Stage I and yield the rough segmentation of the WT. Based on the rough segmentation, we concatenate T1, T2, T1-Gd or T1ce, and FLAIR as the input of Stage II and yield fine-grained segmentation of TC and ET region. In both stages, we improve traditional 3D U-Net by integrating Group Normalization (GN), Leaky ReLU, residual connection, and optional SE block to achieve segmentation.

The main contributions of this article include:

1. 3D RU-Net improves traditional 3D U-Net by replacing the batch normalization and ReLU activation with GN and Leaky ReLU, respectively, and introduces layer-wise residual connection into both encoder and decoder path of 3D U-Net to enhance the multiscale feature representation ability.
2. We propose a novel modality selection and fusion strategy to improve the quality of input data to 3D RU-Net. We choose preferable MRI sequences for each stage of the cascaded pipeline according to respective segmentation objectives and develop effective sequence fusion method to forge the input to neural network.
3. The ET region usually lacks complete spatial connectivity, which obstructs accurate annotation before training and degrades subsequent segmentation. We acquire accurate training annotation of ET by leveraging the difference of surrounding tumor regions with complete spatial connectivity to solve this problem.

2 | Method

Figure 1 introduces the overall design of our method. In data preparation, multisequence MR images are firstly processed and cropped along the three dimensions according to the statistical characteristics of the 3D image data to eliminate the impact of the background area. Then, we generate 3D binary masks for the tumor regions to be segmented following a spatial connectivity-guided mask annotation CMA strategy. These masks will be used as training labels and fed to the cascaded segmentation pipeline together with MRI samples. In addition, we select preferable MRI sequences for each stage of the cascaded pipeline according to respective segmentation objective and fuse them together as the input data for each stage. The two-stage cascaded segmentation first determines

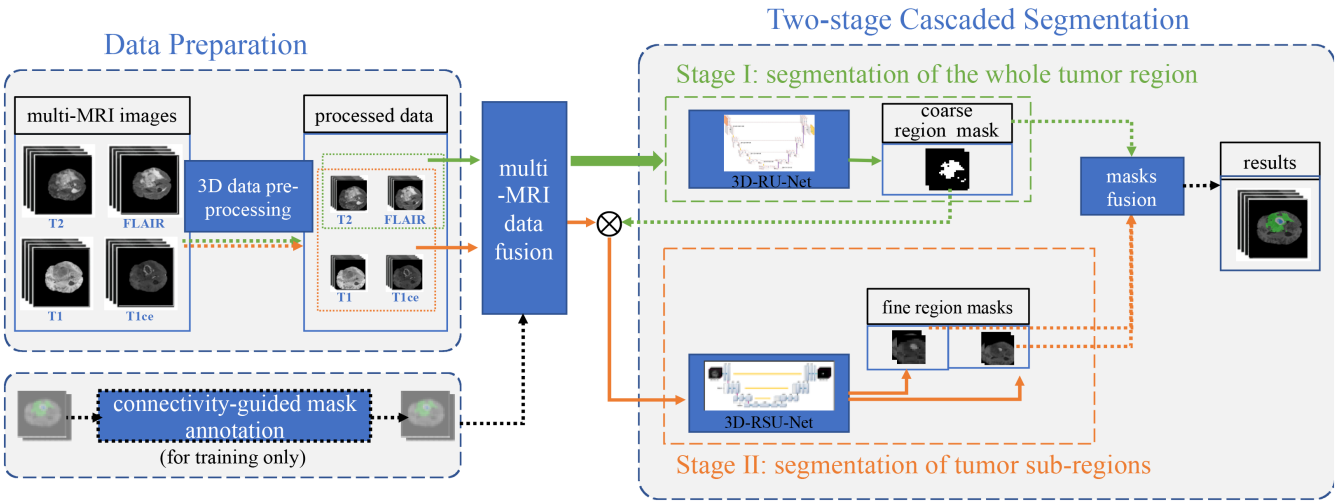

FIGURE 1 | Overview of the proposed methods.

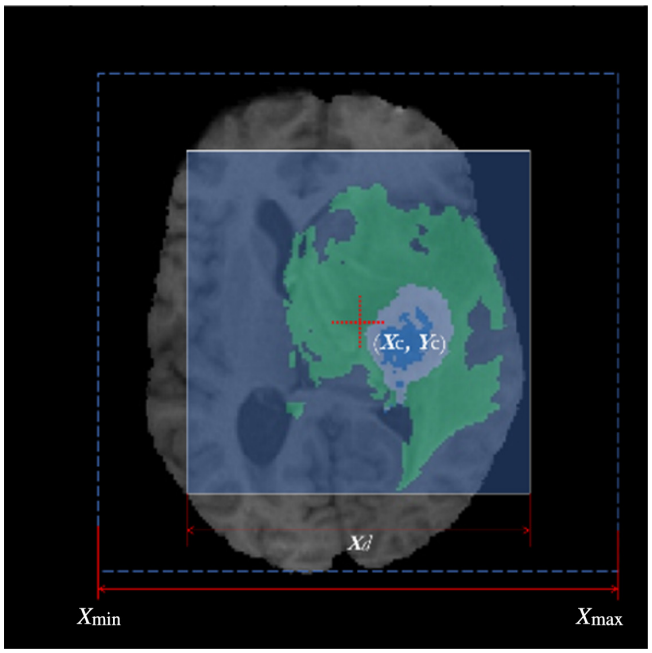

FIGURE 2 | 3D data clipping illustration.

the region where the tumor is located and then accurately segments the subregion of the tumor area, including WT, TC region, and enhanced tumor (ET) region. The final target segmentation regions are calculated by combining the results from both stages. Both stages exploit the 3D RU-Net model yielded by improving traditional 3D U-Net through GN, Leaky ReLU, and residual connection in each layer of the U-shaped structure.

2.1 | 3D Data Preprocessing

In head MRI scans, each MRI sequence file consists of three-dimensional data (which is $240 \times 240 \times 155$ voxels for samples in BraTS 2020 dataset). This substantial data volume significantly impacts computational efficiency and imposes high demands on GPU memory. Besides, a large portion of voxels captured are outside the brain and contain no diagnostic information, which is the background area for the segmentation task. Thus, a procedure to crop the original large-size data to some extent is designed. The clipping region is determined by considering both the statistics of MRI acquisition and patient-specific information.

To remove irrelevant background (nonbrain) areas while avoiding inadvertent exclusion of tumor areas during cropping, this study determined the outer boundary range for cropping based on the statistics of the dataset. Because the brain position when taking MRI scans is relatively constant, we assume that a bounding box covers most cases of patients' brain area. It is calculated by identifying the boundary of brain according to the segmentation label provided for each MRI scan in the dataset and calculating the minimum and maximum coordinates along the X, Y, and Z dimensions across all the scan, denoted by [Xmin, Xmax], [Ymin, Ymax], and [Zmin, Zmax], respectively. For each MRI data, the area outside of this 3D bounding box is first trimmed.

After N4 bias field correction [25] and Z-score normalization, the centroid part of the remaining data is further cropped with a preset dimension of $128 \times 128 \times 128$, whose centroid is aligned with the centroid of the 3D bounding box for the brain. Such preset values are chosen considering the input requirements of the deep learning model and have been experimentally validated to not alter the distribution of the tumor areas. An illustration of 3D data clipping in a single slice, i.e. along X-Y plane is shown in Figure 2, where the original image data (with black background) is first cropped into the rectangular area inside the blue dashed line and further cropped into the square area bounded with white border, both areas are centroid at (Xc, Yc, Zc).

2.2 | Region Annotation and Sequence Fusion

The annotation of WT and TC regions is easy because the voxels in these regions are spatially connected. The annotation of ET

region is much more difficult. It can be observed from Figure 2 that the ET region (with Label 4) usually encloses the region of the necrotic and nonenhancing TC (NET/NCR, with Label 1), which leads to possible incomplete connectivity of the ET region. To simplify the annotation of such regions, we annotate regions with complete connectivity, i.e., TC and NET/NCR, and subtract the NET/NCR from TC to obtain the annotation of ET. This is referred to as a spatial connectivity-guided mask annotation (CMA) approach, which provides better annotations to train the two-stage segmentation pipeline, which will be introduced in detail in Section 2.3.

The four sequences of MR images highlight different brain tissues and therefore play different roles in segmenting fine-grained parts of the tumor. It is necessary to select preferable sequences for each segmentation stage of the two-stage cascaded pipeline. The segmentation target is WT in the first stage (Stage I). Since T2 and FLAIR define the gross tumor volume according to [26], these two MRI sequences are considered as the input sequences to Stage I. We use an element-wise (voxel level) linear combination to merge T2 and FLAIR data with the weight equals 0.5 and 0.5, respectively. The second stage (Stage II) takes the concatenation of the four MRI sequences as input because we need as much information as possible to determine the fine-grained subregion of the glioma.

2.3 | Two-Stage Cascaded Segmentation

A two-stage cascaded segmentation method is designed to accurately segment glioma's WT, TC, and ET regions. As shown in Figure 3, a large portion of the preprocessed image still corresponds to nonbrain backgrounds and normal brain regions that contains no tumor inside. Therefore, Stage I of the method segments the WT from the background, and Stage II segments the subregions including TC and ET from the WT area. In Stage I, a 3D binary mask is estimated and applied to the cropped 3D image data of size $128\times128\times128$ by element-wise multiplication to identify the WT region. In Stage II, the subregions, including TC and ET, are further segmented in the same way as WT segmentation, based on the WT region obtained in Stage I. Both stages exploit the proposed 3D RU-Net to achieve segmentation, which will be depicted in detail in the next section. By combining the segmentation results from Stage I and Stage II, the three target regions can be obtained accordingly. The fused segmentation results from both stages are mapped back to the original data dimension to get the final output.

2.4 | The 3D RU-Net

The proposed model architecture is shown in Figure 4. The proposed model is a cascading of two 3D Residual U-Net-based networks (RU-Net), which share similar structures but with slight difference for each segmentation stage. The 3D RU-Net is a 3D U-Net-based neural network which combines Residual Block (RB) [19] and the optional SE block into a 3D U-Net. Each RU-Net is composed of an encoder and a decoder path. Both the encoder and decoder path consist of four layers. Each layer of the encoder extracts features with specific semantic levels and passes them to the upper layer through max pooling. The decoder layers deconvolve the features in a top-down way to generate the network output.

Similar to 3D U-Net, the RU-Net exploits 3D convolution in each encoding and decoding layer. However, we replace the subsequent ReLU activation with Leaky ReLU activation [27] and adopt GN [28] instead of batch normalization. Xu et al. [29] found that the Leaky ReLU and its variants outperform the ReLU activation function on small-scaled data. Our previous work on breast cancer classification also revealed similar results [30]. The Leaky ReLU activation function sets a slight slope for the area where the input data are negative, which solves the dead neuron problem of ReLU when it encounters a negative value. In the meanwhile, in the training procedure with multisequence MR image data, the batch size is usually smaller than that in the training with 2D images, which may lead to inaccurate estimation of the statistics of BN. The calculation of GN is independent of batch sizes and is more suitable for 3D data training.

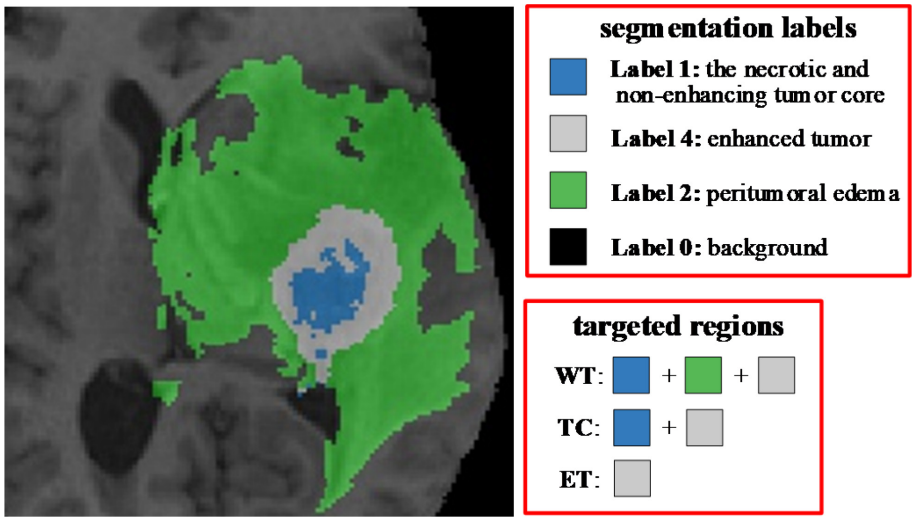

FIGURE 3 | Target segmentation regions and labels.

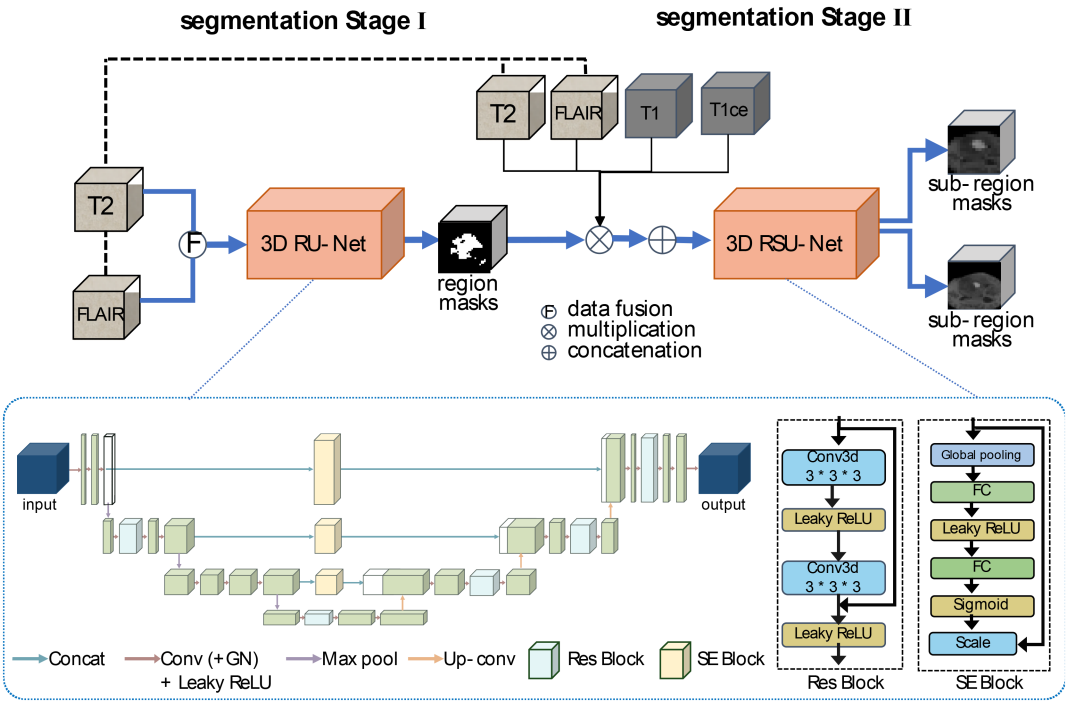

FIGURE 4 | The proposed cascaded RU-Net architecture.

More importantly, we insert a RB into each layer of encoder and decoder. An RB repeats a $3\times3\times3$ convolution layer followed by Leaky ReLU for twice, and the input of RB is concatenated by skip-connection to the output of the second 3D convolution. Another improvement of RU-Net is that we construct an SE block that can be optionally used to build extra connection between the encoder and decoder at each layer. An SE block is composed of global pooling, fully connected layers, Leaky ReLU, and sigmoid activation. We use SE block in the segmentation of Stage II. The detailed structures of the cascaded RU-Net architecture with RB and SE blocks are provided in Figure 4.

In implementation, the first 3D RU-Net in segmentation Stage I takes two of the MRI sequences (T2 and FLAIR) as the input and outputs the coarse segmentation results, i.e., the segmentation mask for WT areas. The second 3D RU-Net takes the concatenation of the four MRI sequences after multiplying them with the WT mask as its input and generates the fine-grained masks. With such cascaded design, the target tumor regions can be segmented from coarse to fine levels.

2.5 | The Training Loss

We adopt Dice function as the training loss function. The Dice function can be denoted as

$$\mathrm{Dice}\left(R_p^t, R_a^t\right) = 1 - \frac{2\times\left(R_p^t \cap R_a^t\right)}{R_p^t \cup R_a^t} \tag{1}$$

where R_p^t is the predicted tumor region or subregions denoted by the voxel set, R_a^t is the annotated ground truth tumor region or subregions. Dice function evaluates the degree of overlapping between predicted tumor region and annotated tumor region.

3 | Experiment

3.1 | Dataset and Implementation Details

The BraTS dataset is the publicly available benchmark for glioma segmentation. This article selects the BraTS 2020 [1–3] training dataset for model training and evaluation. The dataset contains MR image data of 369 patients. For each patient, the four MRI sequences (T1, T1ce, T2, and FLAIR) are coregistered with a size of (240, 240, 155), and segmentation labels for each image pixel are provided. Figure 5 shows the sample MRI data along the axial plane. The images, intensity range, and corresponding segmentation labels of the four sequences are displayed, respectively. The final segmentation goal is to segment three subregions, including WT, TC, and ET, as defined in Figure 5.

The proposed method is implemented in PyTorch1.10 of the GPU version with an NVIDIA GV100 [TITAN V] (12G). Adam optimizer is adopted to optimize the network, the learning rate is set as 1e-4, and the weight decay is set as 1e-5. The batch size is set as 2. We randomly split the dataset into training and evaluation (test) set, which occupy 80% and 20% of the total sample, respectively.

3.2 | Evaluating Metrics

Two commonly accepted metrics are used to evaluate the segmentation results, namely Dice score [31] and Hausdorff Distance [32]. The Dice score is computed using the evaluation samples. Formula (2) depicts the computation of Dice

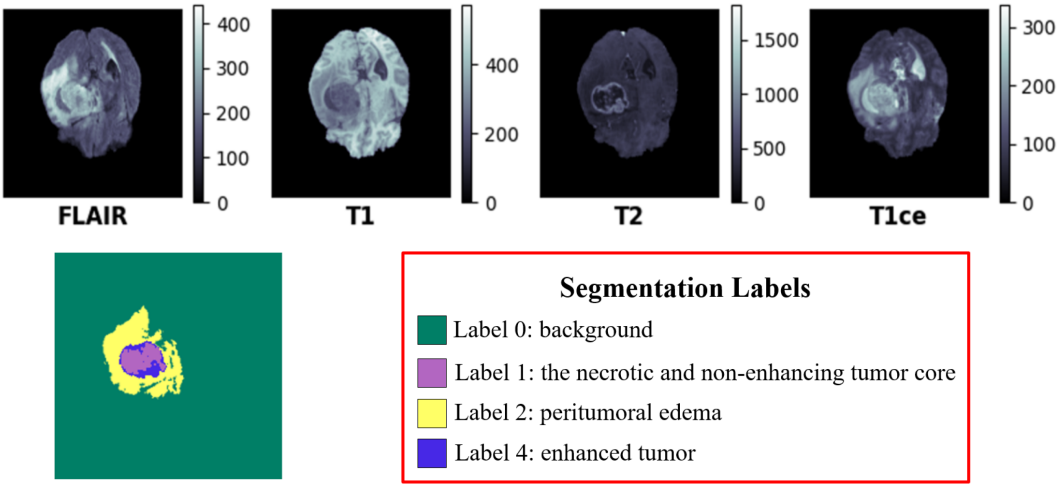

FIGURE 5 | Sample MRI data from BraTS 2020.

score, where R_p^s is the predicted tumor region or subregions of the evaluation sample and R_a^s represents the annotated region. Hausdorff Distance can be computed by formula (3), where $\|a-b\|$ denotes the Euclidean distance between the 3D coordinates of a and b. The Hausdorff distance is used to evaluate the distance between segmentation boundaries. Obviously, the bigger Dice score and smaller Hausdorff Distance represent better segmentation performances.

$$\mathrm{Dice}\left(R_p^s, R_a^s\right) = \frac{2 \times \left(R_p^s \cap R_a^s\right)}{R_p^s \cup R_a^s} \tag{2}$$

$$\begin{cases} H_d\left(R_p^s, R_a^s\right) = \max\left(h\left(R_p^s, R_a^s\right), h\left(R_a^s, R_p^s\right)\right) \\ h\left(R_p^s, R_a^s\right) = \max\limits_{a \in R_p^s}\left\{\min\limits_{b \in R_a^s}\|a-b\|\right\} \\ h\left(R_a^s, R_p^s\right) = \max\limits_{a \in R_a^s}\left\{\min\limits_{b \in R_p^s}\|a-b\|\right\} \end{cases} \tag{3}$$

3.3 | Ablation Experiments

Ablation experiments are performed to evaluate the effectiveness of different model configurations. Since cascaded method usually outperforms noncascaded method, cascaded 3D U-Net is used as the baseline, on which we gradually append CMA presented in Section 2.2, RB with optional SE block (RB/SE). The results in Dice score are shown in Table 1, where CMA beats the baseline in segmenting TC and ET regions, and our proposed method yields the highest Dice scores on all the regions, i.e., WT, TC, and ET. It is worth noting that CMA boosts the segmentation performance of ET region by 31 percentage points, which can be attributed to the proposed spatial CMA strategy that provides a more accurate mask to the ET regions with potential incomplete spatial connectivity. On this basis, our method adopts RB and SE block to improve the feature quality and yields the leading results.

We also compute the Hausdorff Distance using our proposed method, and the result with respect to WT, TC, and ET regions is 5.84, 8.87 and 11.74 mm, respectively. We will further compare both Dice score and Hausdorff Distance yielded by our method and other methods in Section 3.5.

TABLE 1 | Performance in dice score of different model configurations on BraTS 2020 dataset.

	Dice score (↑)		
Methods	**WT**	**TC**	**ET**
Baseline	0.93	0.87	0.50
Baseline + CMA	0.93	0.89	0.81
Baseline + RB/SE + CMA (ours)	0.94	0.92	0.87

3.4 | Results Visualization

The segmentation results under different model configurations depicted in Table 1 are shown in Figure 6, where the upper line shows the images of the four sequences of an MRI scan, and the lower line shows the multimodal segmentation results as well as the ground truth annotation, which is labeled by expert board-certified neuroradiologists and provided in the dataset. The necrotic and nonenhancing TC (NET/NCR) is annotated with blue color, the peritumoral edema region is annotated with green color, and the ET region is rendered in gray. Figure 6 also shows the segmentation result of noncascaded method (3D U-Net).

The noncascaded method noticeably underestimates the NET/NCR area. An important reason is that the NET/NCR area occupies only a small portion of the whole image, which requires fine-grained region segmentation ability that noncascaded method lacks. In comparison, cascaded segmentation obtains better results in NET/NCR and ET regions. However, both noncascaded and cascaded methods without CMA have unsatisfactory segmentation effect on the cavity part (ET region) due to the inaccurate target region annotation. This problem is well addressed by the proposed 3D RU-Net framework. In summary, Figure 6 reveals that the segmentation result with both the two-stage cascaded pipeline and CMA are closer to the ground truth.

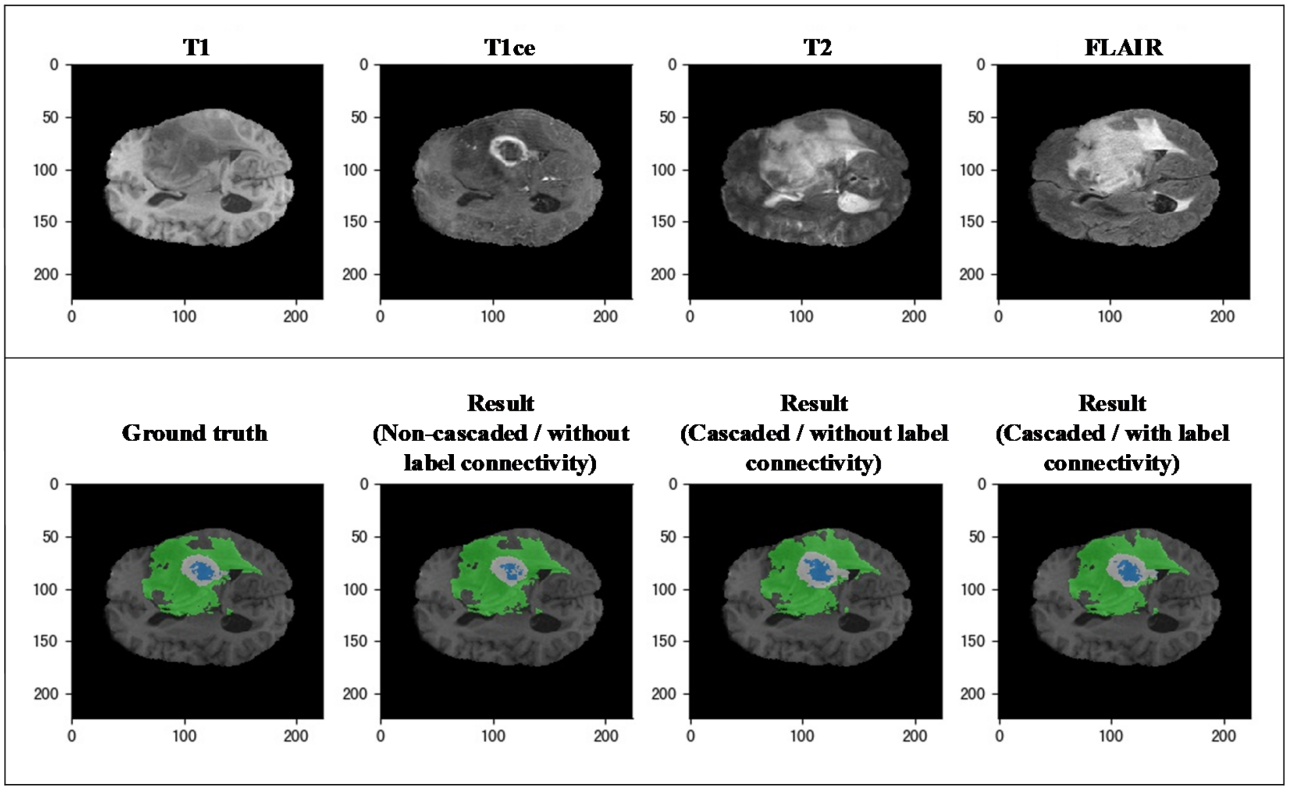

FIGURE 6 | Visualization of segmentation results (top: Multi-sequence MRI data; bottom: Ground truth annotation and segmentation results under different model configurations).

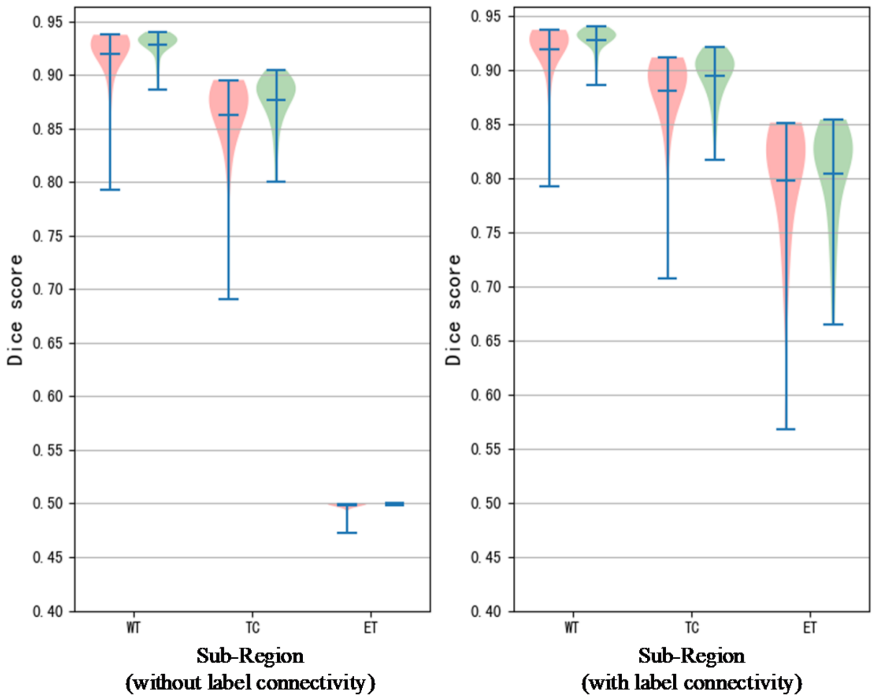

FIGURE 7 | Visualization of segmentation Dice scores with and without CMA.

The Dice scores about different target regions segmented by methods with and without CMA are further visualized in Figure 7, where the red and green part represent the result statistics of training set and evaluation (testing) set, respectively. It can be observed that the segmentation results for WT and TC regions have no obvious difference, regardless of whether CMA is adopted or not. This is because the two regions have complete connectivity, such that the accurate 3D mask can be easily

TABLE 2 | Comparison in Dice scores of our method with typical existing U-Net-based methods (the best and the second best results are highlighted in bold and underline).

Method	Framework of pipeline	Dice score		
		WT	TC	ET
dResU-Net [6]	3D	0.8660	0.8357	0.8004
Wasserstein Dice [19]	3D	0.8890	0.8410	0.8140
Hyperdense inception U-Net [20]	3D	0.8749	0.8371	0.7945
Asymmetric EM attention [33]	3D	0.8960	0.8390	0.7700
MAU-Net [34]	3D	0.9000	0.8163	0.7747
Bridged-U-Net-ASPP-EVO [35]	3D	0.9070	0.8159	0.7800
Asymmetrical U-Net [13]	3D cascaded	0.9041	0.8350	0.7958
Multimodal attention-gated U-Net [36]	2D cascaded	0.9045	0.8430	<u>0.8216</u>
nnU-Net [21]	2D/3D ensemble	<u>0.9187</u>	<u>0.8797</u>	0.8137
Ours	3D cascaded	**0.9425**	**0.9231**	**0.8698**

TABLE 3 | Comparison in Hausdorff distance of our method with typical existing U-Net-based methods (the best and the second best results are highlighted in bold and underline).

Method	Shape of U-Net	Hausdorff distance		
		WT	TC	ET
Asymmetrical U-Net [13]	3D cascaded	**4.953**	**6.299**	23.608
Wasserstein Dice [19]	3D	6.400	19.400	<u>15.800</u>
Asymmetric EM attention [33]	3D	7.7	11.7	32.4
Bridged-U-Net-ASPP-EVO [35]	3D	<u>5.3740</u>	15.9411	21.6850
Ours	3D cascaded	5.8365	<u>8.8737</u>	**11.7386**

annotated. However, for the ET region, the Dice score obtained with CMA outperforms that without CMA by 34 percentage points, from 0.5 to 0.84. This is because ET is wrapped around the NET/NCR region, resulting in a large hole in the middle of the 3D mask, and it is difficult to obtain accurate 3D annotation for ET region directly. The proposed spatial CMA method solves this problem by computing the difference between NET/NCR and TC, the completely connected regions.

3.5 | Comparison With Existing Methods

U-Net is the most commonly used neural network for medical image segmentation, which has many variants, including 3D U-Net, Res-U-Net, Dense Net, etc. We compare the proposed cascaded 3D RU-Net with some of the typical U-Net-based glioma segmentation methods on BraTS 2020 dataset. The results are listed in Tables 2 and 3.

Table 2 shows the Dice scores of our method and typical existing U-Net-based methods. The Dice score is the average over all the testing samples. The primary difference between these methods is the framework of the pipeline, i.e., whether cascaded strategy is used or not. Some methods like those proposed in [6, 19, 20] exploit 3D U-Net without a cascaded strategy, and they differ from one another in various loss functions, training pipelines, and network structures. For example, the model introduced in [6] also combines RDs with 3D U-Net, but it is not trained with cascaded scheme and sequence fusion. Compared with these methods, our method yields the highest Dice scores for all three target regions, especially in ET region segmentation. Specifically, we outperform dResU-Net [6], Hyperdense inception U-Net [20], Wasserstein Dice [19] by 6.94, 5.58, and 7.53 percentage points, respectively, in ET segmentation. Similar improvements can be observed over other noncascaded methods. Multimodal attention-gated U-Net adopts 2D cascaded pipeline, therefore achieves much better result in ET segmentation, with 0.8216 Dice score that is the second best result among all the comparative methods. This indicates that a cascaded pipeline is particularly effective for fine-grained tumor region segmentation. Our method beats all the comparative methods in that it benefits from the comprehensive use of CMA, multisequence fusion, and 3D RU-Net.

Table 3 shows the Hausdorff distance of our method and typical existing U-Net-based methods. The Hausdorff distance is the

average distance of all the testing samples. Our method achieves competitive results on WT and TC segmentation and outperforms all the comparative methods on ET segmentation by a large margin. For example, the distance yielded by our method approximately equals two-thirds of that yielded by the second-best method and is close to one-third of the biggest distance. The Asymmetrical U-Net with the VAE Branch proposed in [13] achieves slightly better results on WT and TC regions than ours. However, such results are achieved by using the segmentation results of multiple first-stage models to train the second-stage network and ensemble a total of 21 segmentation results. As stated in the paper, "it noticeably increases the training time as a trade-off." Moreover, the ET segmentation of the asymmetrical U-Net is far from satisfactory.

4 | Conclusion

This article proposed a Cascaded 3D RU-Net framework for glioma segmentation from multimodal MR images. We combine spatial CMA, target-oriented multisequence MRI fusion, and a two-stage cascaded pipeline for segmenting the WT, TC, and ET regions. Both stages insert layer-wise residual connections into 3D U-Net for fine-grained feature learning. Experiments on the BraTS 2020 training dataset resulted in an average Dice score of 0.9425, 0.9231, 0.8698 and an average Hausdorff distance of 5.8365, 8.8737, 11.7386 mm for WT, TC, and ET, respectively. The results indicate the efficacy of our method, which outperforms a group of mainstream methods.

Author Contributions

Xiaoyan Sun: conceptualization, methodology, manuscript review and revise, supervision, resources and funding acquisition; Chuhan Hu: programming and initial draft writing; Wenhan He: data collection and annotation; Zhenming Yuan: manuscript review; Jian Zhang: methodology, manuscript review and revise.

Acknowledgments

This work was partially supported by National Natural Science Foundation of China (Grant number [61902103]).

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are openly available in "BraTS 2020 data" at https://www.med.upenn.edu/cbica/brats2020/data.html.

References

1. B. H. Menze, A. Jakab, S. Bauer, et al., "The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)," *IEEE Transactions on Medical Imaging* 34 (2015): 1993–2024.

2. S. Bakas, H. Akbari, A. Sotiras, et al., "Advancing the Cancer Genome Atlas Glioma MRI Collections With Expert Segmentation Labels and Radiomic Features," *Scientific Data* 4 (2017): 1–13.

3. S. Bakas, M. Reyes, A. Jakab, et al., "arXiv," 2018, https://doi.org/10.48550/arXiv.1811.02629.

4. A. Işın, C. Direkoğlu, and M. Şah, "Review of MRI-Based Brain Tumor Image Segmentation Using Deep Learning Methods," *Procedia Computer Science* 102 (2016): 317–324.

5. M. Havaei, A. Davy, D. Warde-Farley, et al., "Brain Tumor Segmentation With Deep Neural Networks," *Medical Image Analysis* 35 (2017): 18–31.

6. R. Raza, U. I. Bajwa, Y. Mehmood, M. W. Anwar, and M. H. Jamal, "DResU-Net: 3D Deep Residual U-Net Based Brain Tumor Segmentation From Multimodal MRI," *Biomedical Signal Processing and Control* 79 (2022): 103861.

7. F. Dehghani, A. Karimian, and H. Arabi, "arXiv," 2022, https://doi.org/10.48550/arXiv.2203.03338.

8. N. Cinar, A. Ozcan, and M. Kaya, "A Hybrid DenseNet121-UNet Model for Brain Tumor Segmentation From MR Images," *Biomedical Signal Processing and Control* 76 (2022): 103647.

9. K.-L. Tseng, Y.-L. Lin, W. Hsu, and C.-Y. Huang, "Joint Sequence Learning and Cross-Modality Convolution for 3D Biomedical Segmentation," 2017, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 6393–6400.

10. D. W. Zhang, G. H. Huang, Q. Zhang, J. G. Han, J. W. Han, and Y. Z. Yu, "Cross-Modality Deep Feature Learning for Brain Tumor Segmentation," *Pattern Recognition* 110 (2021): 107562.

11. S. Sahayam, R. Nenavath, U. Jayaraman, and S. Prakash, "Brain Tumor Segmentation Using a Hybrid Multi Resolution U-Net With Residual Dual Attention and Deep Supervision on MR Images," *Biomedical Signal Processing and Control* 78 (2022): 103939.

12. Z. H. Liu, L. Tong, L. Chen, et al., "Deep Learning Based Brain Tumor Segmentation: A Survey," *Complex & Intelligent Systems* 9 (2022): 1001–1026.

13. C. Lyu and H. Shu, "A Two-Stage Cascade Model With Variational Autoencoders and Attention Gates for MRI Brain Tumor Segmentation," in *International MICCAI Brainlesion Workshop* (Berlin, Germany: Springer, 2021), 435–447.

14. S. K. R. Chinnam, V. Sistla, and V. K. K. Kolli, "Multimodal Attention-Gated Cascaded U-Net Model for Automatic Brain Tumor Detection and Segmentation," *Biomedical Signal Processing and Control* 78 (2022): 103907.

15. P. Li, W. Wu, L. Liu, F. M. Serry, J. Wang, and H. Han, "Automatic Brain Tumor Segmentation From Multiparametric MRI Based on Cascaded 3D U-Net and 3D U-Net++," *Biomedical Signal Processing and Control* 78 (2022): 103979.

16. O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," 2015, International Conference on Medical Image Computing and Computer-Assisted Intervention, 234–241: Springer.

17. Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-Net: Learning Dense Volumetric Segmentation From Sparse Annotation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (Berlin, Germany: Springer, 2016), 424–432.

18. J. Hu, L. Shen, and G. Sun, "Squeeze-and-Excitation Networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2018), 7132–7141.

19. L. Fidon, S. Ourselin, and T. Vercauteren, "Generalized Wasserstein Dice Score, Distributionally Robust Deep Learning, and Ranger for Brain Tumor Segmentation: BraTS 2020 Challenge," in *International MICCAI Brainlesion Workshop* (Berlin, Germany: Springer, 2021), 200–214.

20. S. Qamar, P. Ahmad, and L. Shen, "Hi-Net: Hyperdense Inception 3d Unet for Brain Tumor Segmentation," in *International MICCAI Brainlesion Workshop* (Berlin, Germany: Springer, 2021), 50–57.

21. F. Isensee, P. F. Jäger, P. M. Full, P. Vollmuth, and K. H. Maier-Hein, "nnU-Net for Brain Tumor Segmentation," in *International MICCAI Brainlesion Workshop* (Berlin, Germany: Springer, 2021), 118–132.

22. H. M. Luu and S.-H. Park, "Extending nn-UNet for Brain Tumor Segmentation," in *International MICCAI Brainlesion Workshop* (Berlin, Germany: Springer, 2022), 173–186.

23. F. Isensee, P. F. Jaeger, S. A. Kohl, J. Petersen, and K. H. Maier-Hein, "nnU-Net: A Self-Configuring Method for Deep Learning-Based Biomedical Image Segmentation," *Nature Methods* 18 (2021): 203–211.

24. F. Isensee, P. F. Jäger, S. A. Kohl, J. Petersen, and K. H. Maier-Hein, "arXiv," 2019, https://doi.org/10.1038/s41592-020-01008-z.

25. N. J. Tustison, B. B. Avants, P. A. Cook, et al., "N4ITK: Improved N3 Bias Correction, Ieee T Med," *IEEE Transactions on Medical Imaging* 29 (2010): 1310–1320.

26. M. Weller, M. van den Bent, M. Preusser, et al., "EANO Guidelines on the Diagnosis and Treatment of Diffuse Gliomas of Adulthood, Nature Reviews," *Nature Reviews. Clinical Oncology* 18 (2021): 170–186.

27. Y. Wu and K. He, "Group Normalization," in *Proceedings of the European Conference on Computer Vision (ECCV), Lecture Notes in Computer Science*, vol. 11217 (Munich, Germany: Springer, 2018), 3–19.

28. A. L. Maas, A. Y. Hannun, and A. Y. Ng, "Rectifier Nonlinearities Improve Neural Network Acoustic Models," *Proc. Icml, Citeseer* 30 (2013): 3.

29. B. Xu, N. Wang, T. Chen, and M. Li, "arXiv," 2015, https://doi.org/10.48550/arXiv.1505.00853.

30. C. Hu, X. Sun, Z. Yuan, and Y. Wu, "Classification of Breast Cancer Histopathological Image With Deep Residual Learning," *International Journal of Imaging Systems and Technology* 31 (2021): 1583–1594.

31. L. R. Dice, "Measures of the Amount of Ecologic Association Between Species," *Ecology* 26 (1945): 297–302.

32. F. Wang, "arXiv," 2020, https://doi.org/10.48550/arXiv.2005.06338.

33. J. Zhang, Z. Jiang, D. Liu, Q. Sun, Y. Hou, and B. Liu, "3D Asymmetric Expectation-Maximization Attention Network for Brain Tumor Segmentation," *NMR in Biomedicine* 35, no. 5 (2022): e4657.

34. Y. Zhang, Y. Han, and J. Zhang, "MAU-Net: Mixed Attention U-Net for MRI Brain Tumor Segmentation," *Mathematical Biosciences and Engineering* 20 (2023): 20510–20527.

35. R. Yousef, S. Khan, G. Gupta, B. M. Albahlal, S. A. Alajlan, and A. Ali, "Bridged-U-Net-ASPP-EVO and Deep Learning Optimization for Brain Tumor Segmentation," *Diagnostics* 13, no. 16 (2023): 2633.

36. S. K. R. Chinnam, V. Sistla, and V. K. K. Kolli, "Multimodal Attention-Gated Cascaded U-Net Model for Automatic Brain Tumor Detection and Segmentation[J]," *Biomedical Signal Processing and Control* 78 (2022): 103907.