

Article

TACA-RNet: Tri-Axis Based Context-Aware Reverse Network for Multimodal Brain Tumor Segmentation

Hyunjin Kim [†], Youngwan Jo [†], Hyojeong Lee  and Sanghyun Park ^{*}

Department of Computer Science, Yonsei University, Seoul 03722, Republic of Korea; hjkim99@yonsei.ac.kr (H.K.); jyy1551@yonsei.ac.kr (Y.J.); hyojoy@yonsei.ac.kr (H.L.)

* Correspondence: sanghyun@yonsei.ac.kr

[†] These authors contributed equally to this work.

Abstract: Brain tumor segmentation using Magnetic Resonance Imaging (MRI) is vital for clinical decision making. Traditional deep learning-based studies using convolutional neural networks have predominantly processed MRI data as two-dimensional slices, leading to the loss of contextual information. While three-dimensional (3D) convolutional layers represent an advancement, they have not fully exploited pathological information according to the three-axis nature of 3D MRI data—axial, coronal, and sagittal. Recognizing these limitations, we introduce a Tri-Axis based Context-Aware Reverse Network (TACA-RNet). This innovative approach leverages the unique 3D spatial orientations of MRI, learning crucial information on brain anatomy and pathology. We incorporated three specialized modules: a Tri-Axis Channel Reduction module for optimizing feature dimensions, a MultiScale Contextual Fusion module for aggregating multi-scale features and enhancing spatial discernment, and a 3D Axis Reverse Attention module for the precise delineation of tumor boundaries. The TACA-RNet leverages three specialized modules to enhance the understanding of tumor characteristics and spatial relationships within MRI data by fully utilizing its tri-axial structure. Validated on the Brain Tumor Segmentation Challenge 2018 and 2020 datasets, the TACA-RNet demonstrated superior performances over contemporary methodologies. This underscores the critical role of leveraging the three-axis structure of MRI to enhance segmentation accuracy.

Keywords: medical image segmentation; brain tumor; three-axis based network; multi-scale fusion; axis reverse attention



Citation: Kim, H.; Jo, Y.; Lee, H.; Park, S. TACA-RNet: Tri-Axis Based Context-Aware Reverse Network for Multimodal Brain Tumor Segmentation. *Electronics* **2024**, *13*, 1997. <https://doi.org/10.3390/electronics13101997>

Received: 9 April 2024
Revised: 15 May 2024
Accepted: 17 May 2024
Published: 20 May 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Gliomas are the most common primary brain tumors in adults, accounting for 70% of all malignant primary brain tumors [1]. They can be classified as high-grade gliomas (HGGs) and low-grade gliomas (LGGs), with HGGs being more aggressive and invasive than LGGs [2]. Magnetic Resonance Imaging (MRI) is commonly used for the diagnosis and treatment planning of brain tumors owing to its high resolution, soft tissue contrast, and non-invasive nature [3]. For gliomas, four MRI modalities (T1, T1ce, T2, and FLAIR) are typically employed. Multimodal imaging facilitates the capture of a wide range of histopathological parameters, effectively reducing informational uncertainty and enhancing the precision of clinical diagnoses [4].

Identifying gliomas in MRI is crucial for the clinical diagnoses and formulation of treatment plans. However, the traditional segmentation process for diagnosis, which involves manually inspecting MRI volumes slice-by-slice, is time consuming and depends significantly on the experience of the radiologists [5]. Moreover, the significant individual variability in tumors' location, size, shape, margins, and density makes it challenging to manually distinguish gliomas in MRI data [2]. Additionally, morphological uncertainty complicates the process, as the outer layers of brain tumors consist of edematous tissue, making the edges of the tissue surrounding the tumor ambiguous and the tumor contours difficult to define [1].

These challenges have prompted research on automatic brain tumor segmentation. With the advancement of deep learning, numerous studies have been conducted utilizing deep learning-based methods. In particular, the introduction of U-Net [6], a U-shaped convolutional network, has aided various brain tumor segmentation studies using convolutional neural networks (CNNs). Initially, two-dimensional (2D) convolution was predominantly used, thus employing three-dimensional (3D) MRI data as sliced 2D data. Previous studies have proposed cascade structures or multitask methods that leverage 2D convolution for precise brain tumor segmentation [7–11]. However, these 2D-based models often lose crucial contextual information present in volumetric data and face challenges in accurately distinguishing pathological features as the volume of data increases. Recently, numerous methods that employ 3D data have been explored. These approaches have been augmented by leveraging 3D convolution to enhance the segmentation performance [12–14]. Nonetheless, the reliance on 3D CNNs alone has limitations due to variability in the size and location of brain tumors. To address these limitations, studies have proposed the use of atrous convolution [15] and hierarchical feature pyramid structures to detect tumors of varying sizes and utilize multi-scale features [16,17]. Attention mechanisms have been integrated to concentrate on areas where tumors are present [18–21].

However, despite significant advancements in automatic brain tumor segmentation research, researchers have often overlooked the full potential of leveraging the unique characteristics of 3D MRI data. While 3D-based models align more closely with the intrinsic volumetric nature of MRI data, processing 3D data with convolution layers that utilize $3 \times 3 \times 3$ kernels fails to fully exploit the unique characteristics of MRI. MRI inherently comprises three dimensions—axial, coronal, and sagittal—each providing unique and critical perspectives on brain anatomy and pathology, as illustrated in Figure 1. Focusing on the axial, coronal, and sagittal dimensions is essential because each dimension offers a different view of the brain's anatomy, revealing various aspects of tumors, such as their spread, volume, and interaction with surrounding tissues [22,23]. This highlights the need for a novel approach that not only preserves the rich contextual and spatial nuances inherent in MRI data but also enhances tumor segmentation precision by deeply prompting the understanding of the multi-axis structure of MRI data.

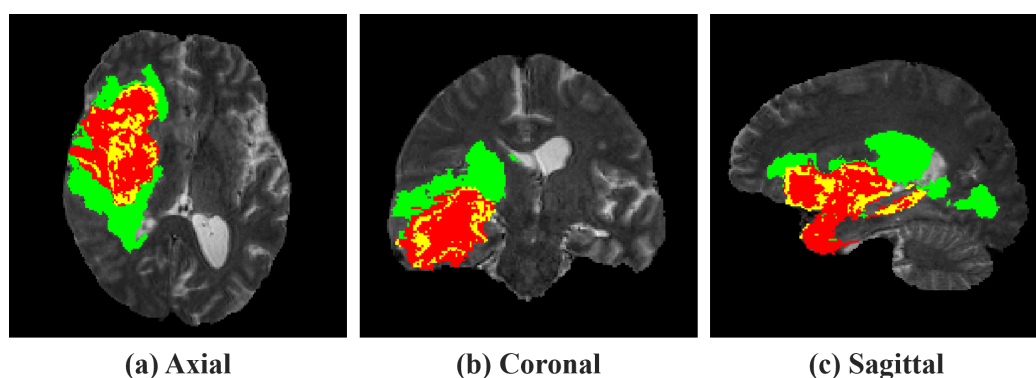


Figure 1. Three-axis visualization of Magnetic Resonance Imaging (MRI) data. (a) Axial, (b) coronal, (c) sagittal. The colors indicate regions of tumors. Red: necrosis and non-enhancing tumor (NCR/NET); yellow: enhancing tumor (ET); green: edema (ED).

In this paper, we propose a Tri-Axis based Context-Aware Reverse Network (TACARNet), which is a novel approach that carefully considers the axial, coronal, and sagittal perspectives of MRI data. This method was designed to address the aforementioned challenges by leveraging the unique 3D spatial orientations inherent in MRI data, aiming to significantly enhance the accuracy and precision of brain tumor segmentation through a comprehensive understanding and utilization of the volumetric information provided by MRI technology. Our approach was validated using the Brain Tumor Segmentation

Challenge (BraTS) 2018 and 2020 datasets [24–26]. The experimental results demonstrate that the proposed TACA-RNet outperforms other recent networks.

The main contributions of this research are summarized as follows:

- We introduce the TACA-RNet, a novel framework specifically designed to leverage the axial, coronal, and sagittal MRI directions. This approach enables a deeper understanding of the complex spatial relationships inherent in volumetric MRI data.
- Our approach integrates three specialized modules: a Tri-Axis Channel Reduction module (TACR), which targets dimension reduction and feature enhancement across MRI's axial, coronal, and sagittal planes; a MultiScale Contextual Fusion module (MSCF), which integrates features from multiple scales to enhance spatial discernment; and a 3D Axis Reverse Attention module (ARA), which concentrates on essential details for precise tumor segmentation.
- We evaluated the efficiency of the proposed network using the BraTS 2018 and 2020 datasets. The results demonstrate that our approach generates a superior segmentation performance and outperforms other recent CNN methodologies.

The remainder of this paper is organized as follows. Section 2 provides a discussion of existing research related to our study. Section 3 details the method and design of this study. In Section 4, we outline the datasets utilized, describe the preprocessing steps, present the evaluation metrics, and detail the experimental configurations; this is then followed by comparison and ablation experiments. Finally, conclusions are presented in Section 5.

2. Related Work

2.1. 2D MRI Data-Based Brain Tumor Segmentation Methods

Typically, 2D MRI data-based brain tumor segmentation methods leverage 2D MRI data generated by slicing the 3D data along the axial axis. Ranjbarzadeh et al. [7] proposed preprocessing methods and utilized a cascade CNN to emphasize the smaller regions of brain images for brain tumor segmentation. Havaei et al. [8] employed a cascade CNN with a two-pathway architecture designed to learn both the local details and the broader context of brain tumors. MTAU [9] implemented a U-Net-based model with three identical structures for multitask brain tumor segmentation based on 2D data. Shen et al. [10] used a multitask framework based on a fully convolutional network to predict tumor boundaries and regions. AGResU-Net [11] utilizes attention gates to enhance local feature expression, addressing the issue of failing to detect small-scale tumors owing to dimension reduction caused by downsampling.

However, models utilizing 2D MRI data often fail to accurately learn voxel-level features and relationships within the 3D data. Consequently, there is growing interest in research utilizing 3D MRI data to exploit volumetric information more effectively and improve the segmentation accuracy.

2.2. 3D MRI Data-Based Brain Tumor Segmentation Methods

Existing 3D MRI data-based brain tumor segmentation methods utilize 3D MRI data, which are resized or cropped to create inputs of a uniform size.

CNN based network. The 3D U-Net [12] employs 3D data as input and adapts the architecture of the 2D U-Net by replacing 2D operations, such as convolution, max pooling, and up-convolution, with their corresponding 3D operations, thereby enabling 3D brain tumor segmentation. Huang et al. [13] utilized a multitasking approach that incorporates segmentation and distance decoders for accurate brain tumor segmentation. CANet [14] employs a feature interaction graph between glioma cells and their surrounding pixels to resolve inter-class ambiguity among different tumor types.

Multi-scale based network. Utilizing a CNN alone poses challenges in identifying brain tumors of various sizes and leveraging multi-scale features effectively. AFPNet [16] overcomes the problems of reduced feature map resolution and loss of information on small tissues by implementing 3D atrous convolution [15]. This method enables the detection of tumors of various sizes utilizing multi-scale features through atrous convolution. DenseAF-

PNet [17], which aims at efficient learning in deep networks, utilizes a densely connected CNN and employs a 3D hierarchical feature pyramid to learn multi-scale features, further enhancing its ability to capture the complexity of brain tumors.

Attention based network. Attention mechanisms are employed to focus on features within the regions of interest where tumors are present. AMMGS [18] and Single-Level U-Net3D [19] implement channel and spatial attention blocks in their respective 3D U-Net and atrous convolution-based models to concentrate on pixels related to tumors. NLCA-VNet [20], based on V-Net [27], utilizes a nonlocal block module and a convolutional block attention module [28] to focus on brain tumors. The scSE-NLV-Net [21] leverages a Squeeze-and-Excitation (SE) block [29] to exploit the spatial dependencies among features, further emphasizing critical areas for accurate tumor segmentation.

These 3D data-based approaches aim to enhance the brain tumor segmentation performance utilizing techniques such as 3D convolution, atrous convolution, and attention mechanisms. However, they do not exploit information according to the three axes of 3D medical data. In particular, studies employing 3D convolution typically use a $3 \times 3 \times 3$ kernel size and overlook the distinct characteristics of each axis. Therefore, we propose the TACA-RNet, which was designed to leverage the unique information present along each axis of 3D medical data, addressing this gap and aiming for more accurate and comprehensive brain tumor segmentation.

3. Proposed Method

In this section, we present the overall network framework. We then introduce the designed components: the TACR, partial decoder (PD), Multi-Resolution Fusion (MRF), MSCF, and 3D ARA modules.

3.1. Overview of Network

As shown in Figure 2, the proposed network consists of six main components: an encoder composed of convolution blocks, a TACR module, a PD, MRF, an MSCF module, and a 3D ARA module.

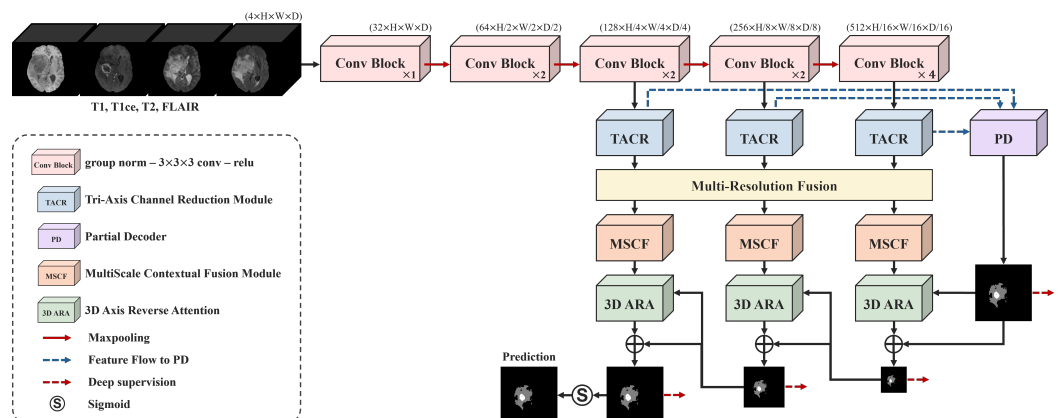


Figure 2. Overview of the proposed TACA-RNet for automatic brain tumor segmentation.

The encoder, which is composed of convolution blocks, extracts high-dimensional semantic features related to gliomas. Each convolution block within the encoder is structured with group normalization, a $3 \times 3 \times 3$ convolution layer, and a ReLU activation function. This setup is used to classify and analyze the sub-level local pixel values of gliomas on MRI. We consider only high-level features $\{f_i, i = 3, 4, 5\}$ among the features $\{f_i, i = 1, \dots, 5\}$ extracted from the encoder because low-level features demand more computational resources owing to their large spatial resolution compared to high-level features, yet they contribute less to the performance [30]. To optimally leverage the unique 3D characteristics of the MRI data, including their axial, coronal, and sagittal orientations, we integrated the TACR module. This strategic addition supplements the conventional Receptive Field Block,

previously situated before the PD [30], with a more sophisticated mechanism tailored to the three-axis structure of the MRI. Subsequently, we present a PD method that generates a global map to serve as an initial guide. Simultaneously, to preserve the details of tumors of varying sizes while minimizing irrelevant information, our approach incorporates MRF. Subsequently, the MSCF module is employed to integrate the global contextual information and accommodate the diverse resolutions inherent in the volumetric data. Finally, guided by the features generated by the PD, the 3D ARA module transforms the suppressed details from the downsampling process into emphasized features, thereby enhancing the focus on locally important information. The network details of the TACA-RNet are provided in Appendix A.

3.2. Tri-Axis Channel Reduction Module

The TACR module refines the high-level feature maps $\{f_i, i = 3, 4, 5\}$ extracted by the encoder, directly addressing the unique 3D spatial orientations (axial, coronal, and sagittal) found in MRI data. This module decreases dimensionality while accentuating the salient features that are vital for accurate tumor segmentation.

As illustrated in Figure 3, the TACR module comprises four branches, each initially using a $1 \times 1 \times 1$ convolution to adjust the number of channels. The first branch of the module, denoted as B_0 , utilizes a $3 \times 3 \times 3$ convolution to capture a broad spectrum of features across the data, establishing a baseline for feature extraction. The subsequent branches, denoted as B_1 for axial, B_2 for coronal, and B_3 for sagittal, employ specialized convolutions with varying kernel sizes, such as $3 \times 3 \times 1$, $1 \times 3 \times 3$, and $3 \times 1 \times 3$, to capture features pertinent to each orientation. After processing through branches B_1 , B_2 , and B_3 , the outputs are concatenated, merging the unique spatial features captured in each direction (axial, coronal, and sagittal). The specific expression for the concatenation process is as follows:

$$B_{123}(f_i) = C(B_1(f_i), B_2(f_i), B_3(f_i)) \tag{1}$$

where C denotes the concatenation operation. Following feature integration, an SE block [29], denoted as SE , dynamically recalibrates channel-specific responses, enhancing important features and diminishing less relevant ones through global information analysis. Subsequently, the outputs from B_0 and the outputs enhanced by the SE block are unified, creating a feature map that incorporates both wide-ranging and orientation-specific characteristics. Finally, shortcut connections seamlessly integrate the initial input with the output of the model, enhancing learning and feature representation while preventing information loss and gradient dissipation. The specific implementation can be mathematically defined as follows:

$$t_i = C(B_0(f_i), SE(B_{123}(f_i))) + f_i \tag{2}$$

Additionally, to reduce model complexity, the channel count for the output t_i of each TACR module was reduced to 32.

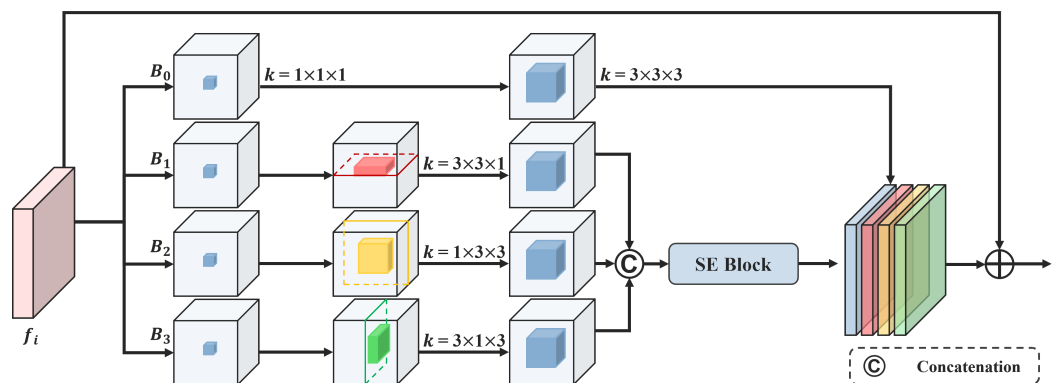


Figure 3. Tri-Axis Channel Reduction module. Here, k denotes the kernel size.

3.3. Partial Decoder

As mentioned in Section 3.1, we strategically incorporated a PD [30] within our model to significantly reduce information loss by enhancing the detail capture capability, particularly in the context of the encoder's downsampling process. This decoder was designed to process only high-level features $\{t_i, i = 3, 4, 5\}$ obtained from the TACR module. For the highest feature layer $i = 5$, we directly use the feature from the corresponding layer, setting $t_{c_5} = t_5$. For features where $i < 5$, each feature t_i is updated by multiplying it element-wise with deeper layer features. The specific expression for the updating process is as follows:

$$t_{c_i} = t_i \odot \prod_{j=i+1}^5 \text{Conv}(\text{Up}(t_j)), \quad i \in [3, 4] \quad (3)$$

where Up represents the upsampled feature by a factor 2^{j-i} , Conv denotes a $3 \times 3 \times 3$ convolutional layer, and \odot denotes element-wise multiplication. To integrate multi-level features $\{t_{c_i}, i = 3, 4, 5\}$, we employ an upsampling and concatenating strategy. The specific expression for the integration process is as follows:

$$t_{c_{\text{integrated}}} = C(t_{c_4}, \text{Conv}(\text{Up}(t_{c_5}))) \quad (4)$$

$$P_g = \text{Conv}_{1 \times 1 \times 1}(C(t_{c_3}, \text{Conv}(\text{Up}(t_{c_{\text{integrated}}}))))) \quad (5)$$

where Up denotes an upsampling operation that doubles the spatial dimensions through transposed convolution, Conv is a $3 \times 3 \times 3$ convolutional layer, C represents the concatenation operation, and $\text{Conv}_{1 \times 1 \times 1}$ is a $1 \times 1 \times 1$ convolutional layer. This strategic use of high-level features via parallel connections in the partial decoder efficiently generates a global map P_g . This map effectively guides the accurate determination of tumor shape, location, and size, enhancing our model's precision with optimized computational efficiency.

3.4. Multi-Resolution Fusion

The MRF process utilizes the high-level features $\{t_i, i = 3, 4, 5\}$ obtained from the TACR module, scaling them up or down to match their unique resolutions. Following this adjustment, the features of the three distinct resolutions are made compatible for concatenation by aligning them with the scale of each layer. The specific implementation can be mathematically defined as follows:

$$\text{MRF}_3 = C(t_3, \text{Up}(t_4), \text{Up}(\text{Up}(t_5))) \quad (6)$$

$$\text{MRF}_4 = C(\text{Down}(t_3), t_4, \text{Up}(t_5)) \quad (7)$$

$$\text{MRF}_5 = C(\text{Down}(\text{Down}(t_3)), \text{Down}(t_4), t_5) \quad (8)$$

In these equations, Up represents an upsampling operation that doubles the spatial dimensions using transposed convolution, Down denotes a downscaling operation that halves the spatial dimensions using convolution, and C indicates the concatenation operation. By harmonizing the features across different resolution layers, MRF enables the integration of diverse spatial information, thereby enhancing the segmentation capabilities of the network.

3.5. MultiScale Contextual Fusion Module

To enhance the network's ability to accurately segment brain tumors, we introduced the MSCF module to reflect the complexity and variability of tumor sizes and their spatial distribution in the MRI data. This module, inspired by Atrous Spatial Pyramid Pooling (ASPP) [31] and the Contextual Feature Pyramid (CFP) [32], aims to innovatively capture and integrate contextual information at multiple scales through a hybrid approach.

As shown in Figure 4, the MSCF module employs a combination of ASPP and CFP to comprehensively represent the spatial and contextual details necessary for accurately

identifying tumor boundaries. The ASPP component within the MSCF employs atrous convolution operations with a set of dilation rates $r_{ASPP} \in \{1, 6, 12, 18\}$. This enables the network to extract features from a wide range of receptive fields, capturing both local and global contextual information without a loss of resolution. Concurrently, the CFP component employs a series of convolutions with an increasing sequence of dilation rates $r_{CFP} \in \{1, 2, 4\}$. This setup progressively captures larger contextual features, thereby enhancing the network's ability to discern spatial relationships at various scales. Varying levels of 3D padding are employed as necessary in the ASPP and CFP components to ensure that the outputs have compatible resolutions for concatenation. After applying ASPP and CFP, the features are concatenated to form a comprehensive, multi-level feature map. This map is then adjusted to the appropriate channel count using a $1 \times 1 \times 1$ convolution.

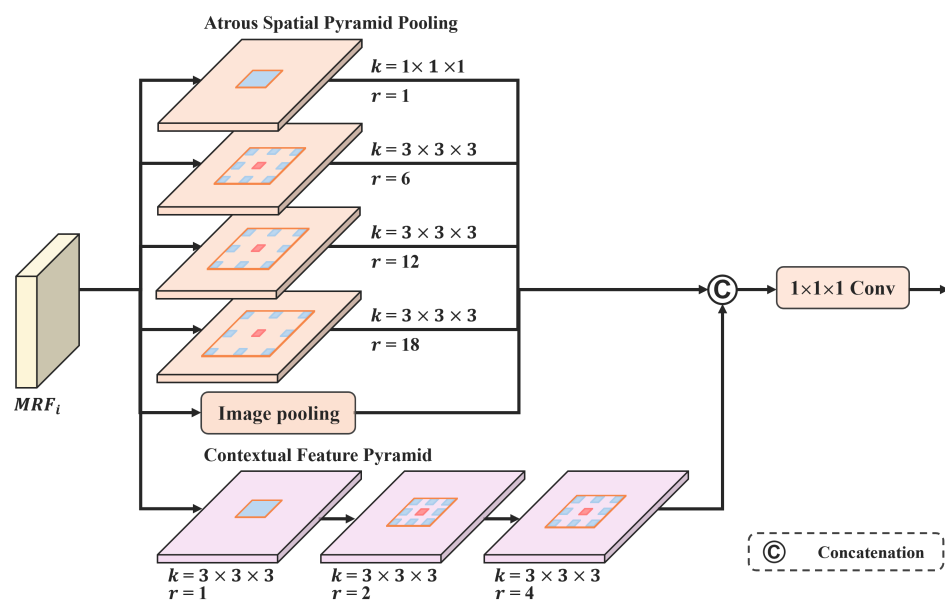


Figure 4. MultiScale Contextual Fusion Module. Here, k denotes the kernel size, and r represents the dilation rate.

3.6. 3D Axis Reverse Attention Module

The 3D ARA module, designed to capture high-resolution details critical for delineating tumor boundaries, is closely aligned with the intrinsic characteristics of the MRI data, particularly considering its axial, coronal, and sagittal orientations. While the MSCF module effectively identifies tumor regions across various scales, it may lack the precision needed for the fine-grained delineation of tumor margins. The 3D ARA module complements this by focusing on the critical details to ensure more accurate segmentation outputs.

As shown in Figure 5, 3D ARA components comprise bifurcated complementary mechanisms: axis attention (AA) and reverse attention (RA). The AA mechanism divides the 3D space into 2D planes, applying three separate 2D attentions across the dimensions of height, width, and depth. To facilitate this, the input is restructured into a 2D format where one dimension corresponds to the space of interest (height, width, or depth) and the other dimension combines the remaining two spatial dimensions. For instance, in the case of $Attention_h$, width and depth are merged into a single dimension, while height is kept separate, thus tailoring the input to suit 2D attention requirements. This approach mirrors the unique anatomical orientations found in MRI data: the height dimension within the axial plane, the width dimension in the coronal plane, and the depth dimension through the sagittal plane. The AA mechanism is defined by the equation $AA_i = Attention_d(Attention_w(Attention_h(m_i))) + m_i$. Here, $Attention_h$, $Attention_w$, and $Attention_d$ were specifically designed to process features corresponding to height, width, and depth, respectively. For instance, $Attention_h$ focuses attention within each

individual height plane, functioning across the 2D planes of width and depth. The variable m_i represents the output of the MSCF module for $i = 3, 4, 5$. Following the AA phase, the RA mechanism reclaims and emphasizes features that may have been overshadowed during the initial focusing process. The RA mechanism is defined by the equation $R_i = \ominus(S(UP(P_{i+1})))$, where $UP(\cdot)$ denotes an upsampling function that enhances the resolution of the feature map, $S(\cdot)$ applies sigmoid activation for a non-linear effect, and \ominus represents the subtraction of this activated output from a unitary matrix. The variable P_{i+1} indicates the output generated from a preceding processing stage in the cascade structure, with $i = 3, 4$, and 5 indicating the sequence of each layer. Notably, P_6 is designated as P_g . The culmination of the 3D ARA process involves the integration of the AA and RA mechanisms to produce a refined feature map that accurately delineates the tumor boundaries. The final representation of this process, which combines the focused and refocused features, is expressed as

$$ARA_i = AA_i \odot R_i \tag{9}$$

In this equation, \odot denotes element-wise multiplication, merging the AA and RA maps to form a feature map that is rich in detail and effectively captures the true edges of the tumors. This sophisticated mechanism of the 3D ARA module, emphasizing the height, width, and depth attentions in alignment with MRI's axial, coronal, and sagittal orientations of the MRI, not only identifies the general area of the tumor but also maps its precise contours.

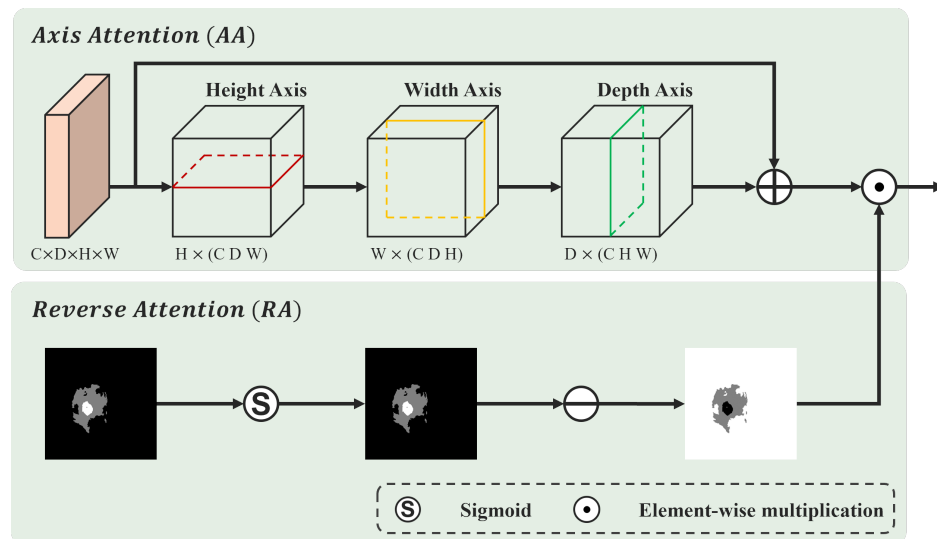


Figure 5. 3D Axis Reverse Attention module.

3.7. Deep Supervision

To ensure thorough learning across various abstraction levels and improve segmentation reliability, our network employs a deep supervision strategy. This strategy entails applying deep supervision to both the global map P_g and the outputs of the i -th layer, $\{P_i, i = 3, 4, 5\}$. To calculate the loss, we used upsampling to adjust the output of each layer to the same size as that of the ground truth. Therefore, the final loss can be expressed as

$$\mathcal{L}_{total} = W_g \cdot \mathcal{L}(G, P_g) + \sum_{i=3}^5 W_i \cdot \mathcal{L}(G, P_i) \tag{10}$$

where W_g and W_i are the weights for the output of each layer, being a constant of 0.3 for W_g , and 1, 0.5, and 0.4 for W_i , respectively. While deep supervision is employed during training to enhance learning at multiple levels, the final output of the network is specifically P_3 , which is optimized for the most detailed and accurate segmentation results.

In building on this foundation, our loss function is defined as the sum of the soft Dice loss function [27] and binary cross entropy (BCE) loss [33], calculated in a voxel-wise manner as follows:

$$\mathcal{L}_{Dice}(G, P) = 1 - \frac{2}{J} \sum_{j=1}^J \frac{2 \sum_{i=1}^I G_{i,j} \cdot P_{i,j}}{\sum_{i=1}^I G_{i,j}^2 + \sum_{i=1}^I P_{i,j}^2} \quad (11)$$

$$\mathcal{L}_{BCE}(G, P) = -\frac{1}{I} \sum_{i=1}^I \sum_{j=1}^J G_{i,j} \cdot \log P_{i,j} + (1 - G_{i,j}) \cdot \log(1 - P_{i,j}) \quad (12)$$

$$\mathcal{L}(G, P) = \mathcal{L}_{Dice}(G, P) + \mathcal{L}_{BCE}(G, P) \quad (13)$$

where G and P are the ground truth and predicted values, respectively, and I and J are the numbers of voxels and classes, respectively.

4. Experimental Results

4.1. Datasets

The datasets used in this study were sourced from the BraTS datasets for 2018 and 2020 [24–26]. In the BraTS 2018 and 2020 datasets, the training sets consisted of 285 and 369 cases, and the validation sets consisted of 66 and 125 cases, respectively. The training sets from both datasets consisted of 3D MRI data from four modalities (T1, T1ce, T2, and FLAIR), including voxel-wise ground truth labels. The labels comprised distinct categories for different tumor regions: necrotic and non-enhancing tumor core (NCR/NET), peritumoral edema (ED), enhanced tumors (ET), and non-tumor regions. Commonly used modalities (T1, T1ce, T2, and FLAIR) play important roles in segmenting different types of brain tumor regions, such as ED, NCR/NET, and ET. An example of multimodal MRI for brain tumor segmentation is shown in Figure 6. The segmentation accuracy was assessed using the Dice score and 95% Hausdorff Distance metrics for specific tumor regions: enhanced tumor (ET), whole tumor (WT, comprising NCR/NET, ED, and ET), and tumor core (TC, involving NCR/NET and ET) regions. The volume of each MRI data is $240 \times 240 \times 155$. Labels for the validation data were not provided; hence, all segmentation results were evaluated using the Center for Biomedical Image Computing and Analytics (CBICA) Image Processing Portal (<https://ipp.cbica.upenn.edu/> (accessed on 20 January 2024)).

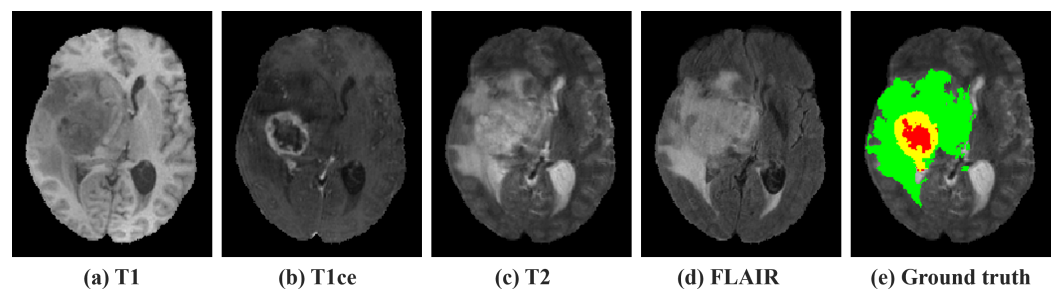


Figure 6. Examples of multimodal MRI. The colors indicate regions of tumors. Red: necrosis and non-enhancing tumor (NCR/NET); yellow: enhancing tumor (ET); green: edema (ED).

4.2. Implementation Details

The proposed model was implemented using PyTorch and trained on four GeForce RTX 3090 GPUs for 500 epochs with a batch size of 16. We utilized the AdamW optimizer, setting the learning rate at 0.001 and a weight decay of 1×10^{-2} . The input data for our network consisted of multiple channels (number of modalities), spatial resolution $H \times W$, and the depth dimension D (number of slices), represented as $C \times H \times W \times D$. The raw input resolution of the images was $240 \times 240 \times 155$. These images were resized to $128 \times 128 \times 128$ via random cropping. We also applied z-score normalization to standardize the input data. Additionally, we applied two data augmentation techniques: random mirror

flipping and random intensity shifting. For inference, we used a sliding window approach with an overlap of 0.5 for neighboring voxels.

4.3. Evaluation Metrics

Two widely used evaluation metrics in brain tumor segmentation, the Dice score and 95% Hausdorff Distance, were employed to evaluate the superiority of the proposed model. Both metrics assess the similarity between the predicted results and the ground truth. A higher Dice score and lower Hausdorff Distance indicate a higher similarity. The definitions of these two evaluation metrics are as follows:

$$Dice(G, P) = \frac{2 \sum_{i=1}^I G_i \cdot P_i}{\sum_{i=1}^I G_i + \sum_{i=1}^I P_i} \quad (14)$$

$$HD(G', P') = \max\{\max_{g' \in G'} \min_{p' \in P'} \|g' - p'\|, \max_{p' \in P'} \min_{g' \in G'} \|p' - g'\|\} \quad (15)$$

where G_i and P_i represent the ground truth and predicted values of voxel i , and G' and P' denote the set of surface points of the ground truth and predicted values, respectively.

4.4. Comparison with Other Methods

To validate the efficiency of our method for brain tumor segmentation, we conducted extensive comparisons with recent 2D and 3D segmentation methodologies. For 2D-based models, our comparative analysis included MTAU [9], Probabilistic U-Net [34], and AGResU-Net [11], all of which employ attention mechanisms to enhance the segmentation accuracy. For 3D-based models, we compared our method with several advanced architectures. These include the 3D U-Net [12], Huang [13], CANet [14], and Deep Supervision CNN [35]; models leveraging multi-scale information such as AFPNet [16], DenseAFPNet [17], and an MR Encoder–Decoder [36]; and models utilizing attention mechanisms, including NLCA-VNet [20], scSE-NLV-Net [21], Single-Level U-Net3D [19], and AMMGs [18].

Tables 1 and 2 present the average Dice scores and Hausdorff Distance metrics for these methods and our model for the BraTS 2018 and 2020 validation sets, respectively. The performance evaluation results for all the models were obtained either by citing the respective papers or through online validation available on official websites, with bold font indicating the highest scores achieved for each category, and underlined text marking the second highest scores.

Table 1. Comparison of different methods on BraTS 2018 validation set. A higher Dice score (\uparrow) and a lower Hausdorff Distance (\downarrow) indicate better performance. Bold font indicates the highest scores achieved for each category, while underlined text marks the second highest scores.

Methods	Dice Score (%) \uparrow				Hausdorff Dist. (mm) \downarrow			
	ET	WT	TC	AVG.	ET	WT	TC	AVG.
3D U-Net [12]	73.44	86.38	76.58	78.80	9.370	12.000	10.370	10.580
AFPNet [16]	72.55	84.94	75.00	77.50	–	–	–	–
DenseAFPNet [17]	75.25	86.42	77.38	79.68	–	–	–	–
AGResU-Net [11]	<u>77.20</u>	87.20	80.80	81.73	<u>3.570</u>	<u>5.620</u>	8.360	<u>5.850</u>
Huang et al. [13]	71.70	80.10	75.90	75.90	9.900	13.200	15.200	12.767
CANet [14]	76.70	<u>89.80</u>	<u>83.40</u>	<u>83.30</u>	3.859	6.685	<u>7.674</u>	6.073
NLCA-VNet [20]	75.00	87.00	78.00	80.00	5.390	7.060	9.890	7.447
Single-Level U-Net3D [19]	74.20	88.48	80.98	81.22	6.670	10.830	10.250	9.250
Ours	77.32	90.21	84.66	84.06	3.441	5.101	5.872	4.805

For the BraTS 2018 validation set (Table 1), our model achieved Dice scores of 77.32%, 90.21%, and 84.66% for the ET, WT, and TC regions, respectively, complemented by Hausdorff Distances of 3.441 mm, 5.101 mm, and 5.872 mm, respectively. These results had an average Dice score of 84.06% and an average Hausdorff Distance of 4.805 mm, positioning our model at the forefront of the performance metrics across all evaluated regions on

the BraTS 2018 dataset. In the realm of ET, AGResU-Net was identified as the leading contender in the literature. Our proposed model exhibited incremental improvements in ET, with a 0.12% increase in the Dice score and a 0.129 mm reduction in the Hausdorff Distance. For the WT and TC scenarios, CANet delivered optimal outcomes. Additionally, our model demonstrated improvements in the Dice score of 0.41% and 1.26% and in the Hausdorff Distance of 1.584 mm and 1.802 mm for WT and TC, respectively, underscoring its proficiency in delineating small tumor regions.

Table 2. Comparison of different methods on BraTS 2020 validation set. A higher Dice score (\uparrow) and a lower Hausdorff Distance (\downarrow) indicate better performance. Bold font indicates the highest scores achieved for each category, while underlined text marks the second highest scores.

Methods	Dice Score (%) \uparrow				Hausdorff Dist. (mm) \downarrow			
	ET	WT	TC	AVG.	ET	WT	TC	AVG.
3D U-Net [12]	68.76	84.11	79.06	77.31	50.983	13.366	13.607	25.985
MTAU [9]	57.00	73.00	61.00	63.67	47.220	24.030	31.530	34.260
Huang et al. [13]	75.00	86.00	77.20	79.40	34.600	<u>6.700</u>	15.100	18.800
Deep Supervision CNN [35]	70.40	87.94	77.31	78.55	–	–	–	–
MR Encoder–Decoder [36]	66.00	87.00	80.00	77.67	47.330	6.910	<u>7.800</u>	20.680
Probabilistic U-Net [34]	68.89	81.90	71.68	74.16	36.886	41.524	26.275	34.895
scSE-NLV-Net [21]	64.70	81.80	75.90	74.13	44.400	10.000	14.600	23.000
NLCA-VNet [20]	67.00	87.60	76.90	77.17	50.800	9.400	12.500	24.233
Single-Level U-Net3D [19]	72.91	<u>88.57</u>	80.19	80.56	31.970	10.260	13.580	18.603
AMMGS [18]	78.03	88.31	<u>81.72</u>	<u>82.69</u>	23.615	7.169	7.988	<u>12.924</u>
Ours	<u>75.52</u>	90.43	84.51	83.49	<u>25.104</u>	5.047	5.410	11.854

For the BraTS 2020 validation set (Table 2), our methodology had Dice scores of 75.52%, 90.43%, and 84.51% for ET, WT, and TC, respectively, with Hausdorff Distances of 25.104 mm, 5.047 mm, and 5.410 mm, respectively. Consequently, the proposed model attained an average Dice score of 83.49% and an average Hausdorff Distance of 11.854 mm across these regions. Aside from the ET region, our model demonstrated superior performances in terms of both Dice scores and Hausdorff Distances for WT and TC regions in the BraTS 2020 dataset compared to other models. In the ET region, the AMMGS model attained the highest performance. Our model showed a slight decline in performance for ET of 2.51% in Dice score and 1.489 mm in Hausdorff Distance but exhibited notable improvements in the WT and TC regions, with increases in the Dice score of 2.12% and 2.79% and decreases in the Hausdorff Distance of 2.122 mm and 2.578 mm, respectively. Our model exhibited the best performance when averaged across all considered regions for both the Dice scores and Hausdorff Distances.

The demonstrated efficacy of the TACA-RNet on the BraTS 2018 and 2020 validation sets underscores its superior capability to harness the full spectrum of inherent 3D spatial orientations of MRI. By meticulously integrating the axial, coronal, and sagittal perspectives, the TACA-RNet provides a deeper understanding of complex spatial relationships within volumetric MRI data. This comprehensive approach allows for the precise identification of tumor boundaries across various tumor types and scales, facilitated by innovative modules of the model designed for multi-scale contextual information processing and focused dimensional reduction. Moreover, the remarkable performance of the TACA-RNet in distinguishing small tumor regions and its ability to effectively integrate multi-scale contextual information affirm its advancements over conventional methodologies. Our comparative experiments validated not only the model’s efficiency in enhancing the segmentation accuracy but also its potential in setting a new benchmark for brain tumor segmentation research, demonstrating its superiority by surpassing existing approaches across all evaluated regions.

4.5. Ablation Study

In this section, we describe an in-depth ablation study conducted to evaluate various modules in the segmentation task using the BraTS 2020 dataset. Due to the absence of

ground truth labels in the official validation dataset, we divided the training dataset into a randomized 9:1 split. This split produced a validation set, which constituted 10% of the training dataset and was used primarily for visualization purposes. All visualizations were exclusively performed with this validation set. The quantitative evaluation results of the ablation studies, presented in the tables, were validated using an official website to confirm their robustness and ensure an accurate performance comparison.

4.5.1. Analysis of the Impact of Tri-Axis Integration within the TACR

The first ablation study was conducted to validate the efficacy of utilizing all the three axis orientations within the TACR module of the proposed TACA-RNet. This experiment was structured to assess the performance variations when employing the TACR module aligned exclusively along the axial, coronal, and sagittal axes, in addition to a configuration in which all three axes were integrated. We conducted four experiments under consistent conditions, where the B_0 branch of the TACR module was consistently used, and the branches designated as B_1 for axial, B_2 for coronal, and B_3 for sagittal were selectively applied. For instance, a TACR module using only the axial orientation employs only B_0 and B_1 . The results are presented in Table 3, in which the highest performance is highlighted in bold.

Table 3. Evaluation results of the TACR module across individual and combined anatomical axes (axial, coronal, and sagittal) on the BraTS 2020 validation dataset. A higher Dice score (\uparrow) and a lower Hausdorff Distance (\downarrow) indicate better performance. Bold font indicates the highest scores achieved for each category.

Method	Dice Score (%) \uparrow			Hausdorff Dist. (mm) \downarrow		
	ET	WT	TC	ET	WT	TC
TACR (Axial)	74.69	89.56	82.63	30.229	7.807	9.554
TACR (Coronal)	72.46	88.78	80.17	33.343	8.520	15.090
TACR (Sagittal)	74.34	89.23	81.02	33.280	6.467	11.413
TACR (Axial + Coronal + Sagittal)	75.86	90.07	83.09	25.242	5.239	6.351

In our experiments on the BraTS 2020 dataset, integrating the axial, coronal, and sagittal orientations within the TACR module of the TACA-RNet resulted in an enhanced segmentation performance compared to configurations where the TACR module was aligned exclusively along a single axis. While the axial orientation typically used in 2D models showed a higher performance than the other individual axial, coronal, and sagittal orientations, it was surpassed by the integrated module that leveraged all three orientations. This improvement underscores the significance of TACR in the model architecture, particularly during the process of channel reduction across the encoder's high-level features. In meticulously designing TACR to emphasize the unique 3D spatial orientations inherent in MRI data, our approach not only deepens the understanding of complex spatial relationships within volumetric MRI data but also significantly boosts the accuracy and precision of brain tumor segmentation. This advancement demonstrates the need to consider the three-axis structure of MRI to enhance the segmentation performance, validating the effectiveness of our method in leveraging the full spectrum of MRI volumetric information.

Figure 7 illustrates the tumor segmentation outputs of the TACR module across various anatomical orientations on the BraTS 2020 dataset. From left to right, each column shows the segmentation results with the TACR module aligned separately in the axial, coronal, and sagittal orientations, followed by the integrated approach using all three orientations, and finally, the ground truth for comparison. In the visualization results, the performance of the NCR/NET regions, marked in red, and the ED region, highlighted in green, was inaccurately predicted by all models except for the one integrating all three orientations. Specifically, in the first row, models exclusively employing axial, coronal, or sagittal orientations incorrectly identified normal tissue as ED in the same slice, and a review of the 3D-rendered results revealed a failure to predict within NCR/NET regions.

Furthermore, in the second row, all models, except the one integrating all three orientations, incorrectly identified normal tissue as the ED region.

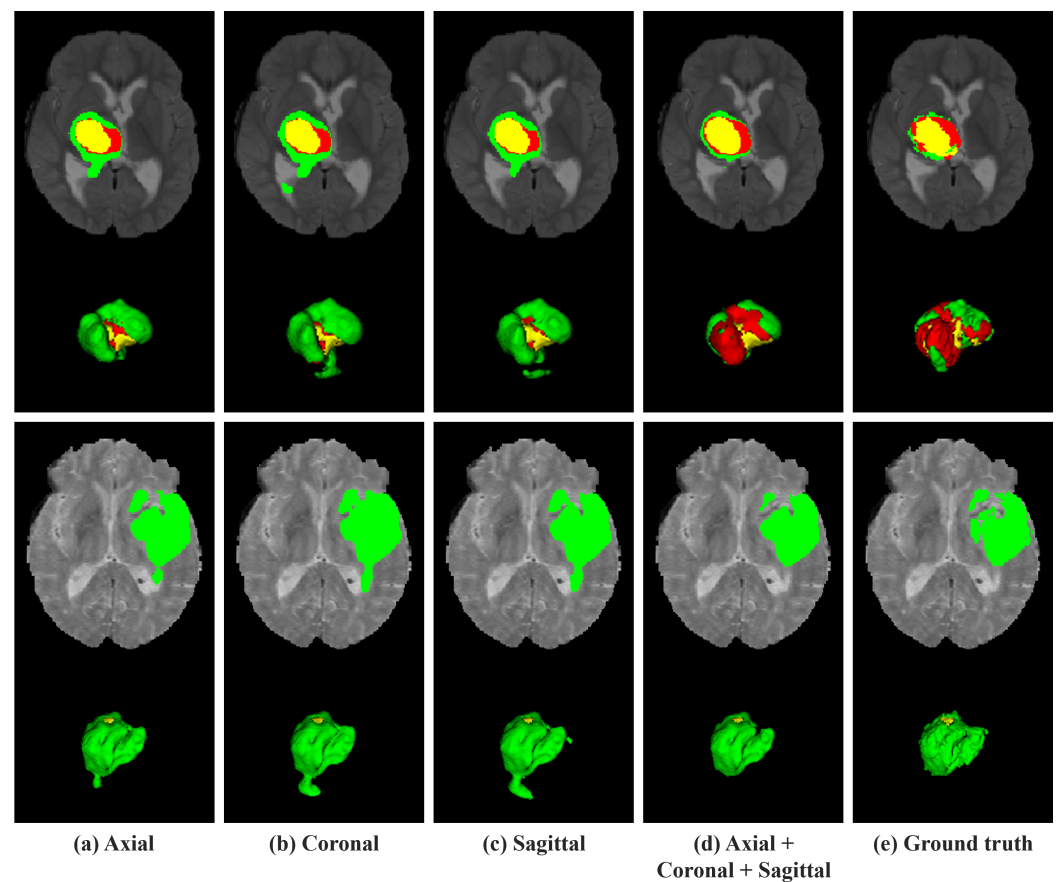


Figure 7. Visual comparison of tumor segmentation outputs from the TACR module across different anatomical orientations on the BraTS 2020 dataset. The colors indicate regions of tumors. Red: NCR/NET; yellow: ET; green: ED.

4.5.2. Assessment of Impact of Proposed Modules

In the second ablation study, meticulous evaluations were performed to discern the individual contributions of each component module to the overall performance of the proposed model. This comparative analysis involved the base network (BN), which comprised a CNN encoder and a PD, as well as variations in the base model enhanced with additional modules: the TACR, MSCF, and 3D ARA modules. These modules were methodically integrated to examine their individual and collective impacts on segmentation accuracy. The results are presented in Table 4, with the best performances highlighted in bold.

In our experiments on the BraTS 2020 dataset, the introduction of our designed modules led to notable improvements in the segmentation performance compared to the BN. Specifically, the Dice scores increased by 1.05%, 1.27%, and 3.09% for the ET, WT, and TC tumor regions, respectively. Moreover, the Hausdorff Distances showed significant reductions of 5.791 mm, 3.11 mm, and 15.925 mm across these regions. These enhancements attest to the effectiveness of our modules in brain tumor segmentation tasks. The implementation of the TACR module led to improved performance in all tumor regions, both in terms of Dice scores and Hausdorff Distances. This improvement can be attributed to TACR's capability to handle unique 3D spatial orientations of MRI data (axial, coronal, and sagittal), enabling the model to accurately capture and emphasize critical features. Further integration of the MSCF module yielded higher Dice scores and lower Hausdorff Distances. This outcome demonstrates the capacity of the MSCF module to reflect the complexity and variability of tumor sizes and their spatial distribution within the MRI data. When the 3D ARA module

was added, without the MSCF module, there was a slight decrease in Dice scores compared to configurations using the MSCF module. However, the 3D ARA module's focus on capturing essential details for accurate tumor boundary delineation resulted in improved accuracy in terms of Hausdorff Distances. This highlights the importance of high-resolution feature capture for enhancing segmentation precision. Ultimately, integrating all modules resulted in the highest segmentation accuracy, confirming the synergistic benefits of our comprehensive module design.

Table 4. Evaluation results of assessing the impact of proposed modules on the BraTS 2020 validation set. BN: base network (consists of a CNN encoder), PD: partial decoder, TACR: Tri-Axis Channel Reduction module, MSCF: MultiScale Contextual Fusion module, 3D ARA: 3D Axis Reverse Attention module. The “+” in “BN + PD” signifies that the Conv Block has replaced the receptive field block between the encoder and PD, and another plus sign indicates that the corresponding components were directly added to the original model. A higher Dice score (↑) and a lower Hausdorff Distance (↓) indicate better performance. Bold font indicates the highest scores achieved for each category.

Method	Dice Score (%) ↑			Hausdorff Dist. (mm) ↓		
	ET	WT	TC	ET	WT	TC
BN + PD	74.81	88.80	79.19	31.033	8.349	22.276
BN + PD + TACR	74.82	89.32	81.44	30.253	6.798	21.103
BN + PD + TACR + MSCF	75.25	90.03	82.19	28.007	6.075	12.733
BN + PD + TACR + 3D ARA	75.18	89.51	82.01	25.551	5.890	8.625
BN + PD + TACR + MSCF + 3D ARA	75.86	90.07	83.09	25.242	5.239	6.351

Figure 8 presents visualization results of the ablation study for the configuration modules on the BraTS 2020 dataset. From left to right, each column represents the base model, the model incorporating the TACR module, the model introducing both the TACR and MSCF modules, the model introducing both the TACR and 3D ARA modules, the model with all the modules integrated, and ground truth. In the visualization results, the performance in the ED region, marked in green, was inaccurately predicted across all models, excluding the model incorporating all modules. Specifically, in the first row, there was a failure to predict the ED region located on the right side, which aligned with the ground truth. Conversely, in the second row, normal tissue was erroneously identified as the ED region.

4.5.3. Effectiveness of MultiScale Contextual Fusion Module

The third ablation study was conducted to demonstrate the efficacy of MSCF in harnessing the complexity and variability of tumor sizes, as well as their spatial distribution across MRI data. Inspired by the ASPP and CFP, the MSCF module implements a hybrid approach that encapsulates contextual information across various scales and achieves a synergistic effect. The ASPP aspect of MSCF utilizes a series of dilation rates to extract features from a broad range of receptive fields without compromising resolution. Concurrently, CFP introduces a sequence of dilation rates designed to progressively capture larger contextual features, thereby enhancing the network's spatial discernment across scales. Table 5 presents the comparative analysis results for the BraTS 2020 validation dataset, focusing on the performance of the MSCF module, with the best scores highlighted in bold. The superiority of Full MSCF, which achieved the highest Dice scores and lowest Hausdorff distances when all components were integrated, is clearly demonstrated. These improvements are crucial because the Full MSCF module ensures a comprehensive representation of the spatial and contextual nuances essential for accurately delineating tumor boundaries.

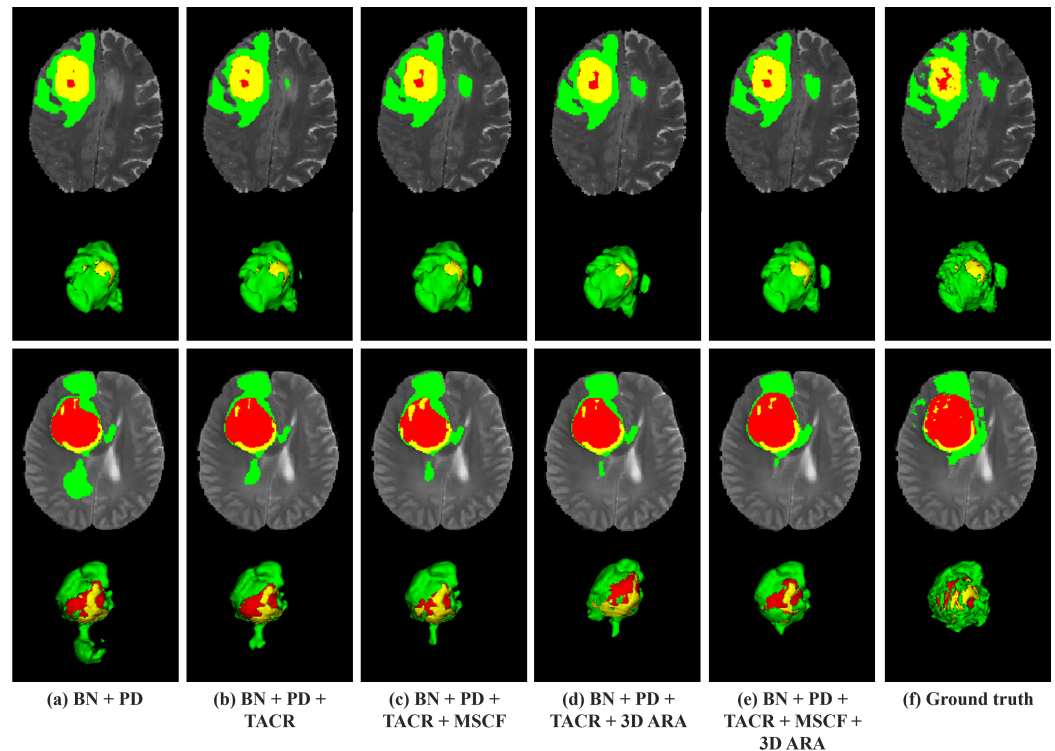


Figure 8. Visual comparison of tumor segmentation outputs from different configurations of the proposed model on the BraTS 2020 dataset. The colors indicate regions of tumors. Red: NCR/NET; yellow: ET; green: ED.

Table 5. Evaluation results of the effectiveness of the MSCF module on the BraTS 2020 validation set. ASPP and CFP components, integral to the MSCF structure, were individually assessed to discern their contribution to the module's efficacy. A higher Dice score (\uparrow) and a lower Hausdorff Distance (\downarrow) indicate better performance. Bold font indicates the highest scores achieved for each category.

Method	Dice Score (%) \uparrow			Hausdorff Dist. (mm) \downarrow		
	ET	WT	TC	ET	WT	TC
MSCF w/o ASPP	74.13	88.69	81.52	34.032	5.759	12.957
MSCF w/o CFP	75.67	89.38	82.21	32.811	5.347	11.636
Full MSCF	75.86	90.07	83.09	25.242	5.239	6.351

Figure 9 presents a visual comparison of the tumor segmentation outputs across various MSCF configurations on the BraTS 2020 dataset. From left to right, each column represents the MSCF without the ASPP component (MSCF w/o ASPP), MSCF without the CFP component (MSCF w/o CFP), Full MSCF with all components integrated, and the ground truth for comparison. In the visualization results, the performance in the NCR/NET regions, marked in red, was inaccurately predicted by all models except Full MSCF. Specifically, in the first row, the MSCF without the ASPP component failed to accurately predict the NCR/NET regions that matched the actual values. Moreover, the MSCF without the CFP component predicted only a small portion of this region. Additionally, in the second row, the same slices reveal that both MSCF without the ASPP component and MSCF without CFP component failed to predict the NCR/NET regions. These outcomes were further confirmed through 3D-rendered visualization results. While the Full MSCF predicts NCR/NET regions that closely resemble that of the ground truth, both the MSCF without the ASPP component and the MSCF without the CFP component failed to accurately predict the NCR/NET regions.

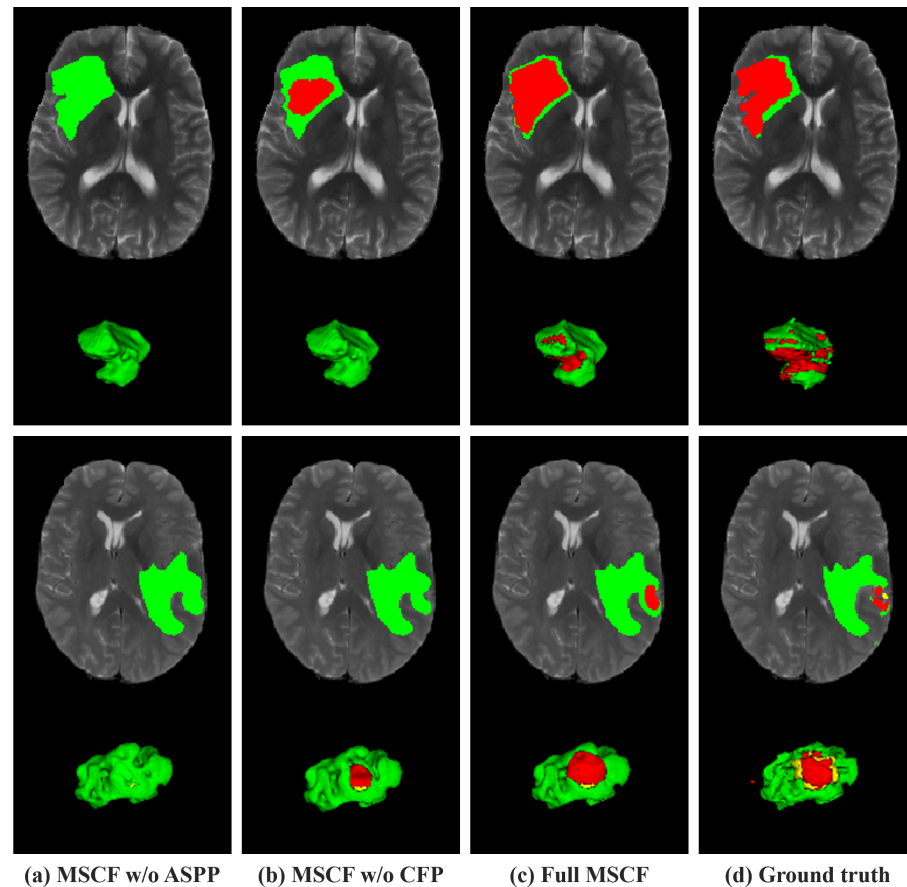


Figure 9. Visual comparison of MSCF module variations for tumor segmentation on the BraTS 2020 dataset. The colors indicate regions of tumors. Red: NCR/NET; yellow: ET; green: ED.

5. Conclusions

In this study, we introduced the TACA-RNet, which was designed to automate the segmentation of brain tumors from multimodal MRI data. The TACA-RNet uniquely incorporates the TACR module to leverage the distinct spatial orientations of MRI data (axial, coronal, and sagittal), thereby enhancing the model's understanding and processing of volumetric information. This was further complemented by the integration of PD, MSCF, and 3D ARA modules, each contributing to the model's ability to accurately delineate brain tumors. Our comprehensive evaluations using the BraTS 2018 and 2020 datasets demonstrated the superiority of the TACA-RNet over existing segmentation techniques, with significant advancements in the segmentation accuracy and precision. Ablation studies further validated the significance of the TACR module in the TACA-RNet for brain tumor segmentation using MRI data. This evaluation demonstrated the central role of TACR in harnessing the unique 3D spatial orientations of MRI data. Subsequently, the contribution of individual modules within the model was examined, emphasizing their collective impact on achieving precise segmentation across various tumor ranges and sizes. Finally, the analysis of the impact of the MSCF module in multi-scale fusion settings revealed its effectiveness in capturing the variability and complexity of tumor sizes and their spatial distribution.

Our approach not only aids the precise localization and diagnosis of brain tumors but also shows promise in addressing the challenges of multimodal brain tumor MRI segmentation. However, the issue of missing modalities in MRI data is often encountered in clinical practice, which significantly affects segmentation tasks. Therefore, it is crucial to develop segmentation methods capable of handling missing modalities. Future research could explore strategies to enhance the segmentation performance in the presence of missing modalities, aiming to bolster the robustness and clinical applicability of the TACA-RNet. This direction promises to refine the utility of the TACA-RNet for clinical applications,

offering valuable tools for medical professionals in the diagnosis of brain tumors. Furthermore, TACA-RNet’s potential extends beyond diagnostic imaging. In the context of drug development for brain tumors, TACA-RNet could play a pivotal role in clinical trials by monitoring tumor responses to new therapeutic agents. This capability facilitates quicker assessment of drug efficacy, enhancing the overall efficiency of drug development.

Author Contributions: Conceptualization, H.K.; data curation, H.K.; methodology, H.K.; supervision, S.P.; validation, H.K. and Y.J.; visualization, H.K.; writing—original draft, H.K. and Y.J.; Writing—review and editing, H.K., Y.J. and H.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the National Research Foundation (NRF) funded by the Korean government (MSIT) (No. RS-2023-00229822).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The datasets were provided by BraTS 2018 Challenge and BraTS 2020 Challenge and are allowed for personal academic research. The specific link to the datasets is <https://ipp.cbica.upenn.edu/> (accessed on 20 January 2024).

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Appendix A

In this Appendix, we provide the network details of the TACA-RNet in Table A1.

Table A1. The design details of our proposed TACA-RNet. Conv3 denotes a $3 \times 3 \times 3$ convolutional layer with specified number of filters; GN denotes group normalization (groups = 8); i denotes the i -th layer. Note that the size of the input image is $4 \times 128 \times 128 \times 128$.

Stage	Block Name	Details	Output Size
CNN Encoder	EnBlock ($i = 1$)	[GN, ReLU, Conv3 (32 filters)] $\times 1$	$32 \times 128 \times 128 \times 128$
	DownSample1	Max pooling 3d	$32 \times 64 \times 64 \times 64$
	EnBlock ($i = 2$)	[GN, ReLU, Conv3 (64 filters) GN, ReLU, Conv3 (64 filters)] $\times 2$	$64 \times 32 \times 32 \times 32$
	DownSample2	Max pooling 3d	$128 \times 16 \times 16 \times 16$
	EnBlock ($i = 3$)	[GN, ReLU, Conv3 (128 filters) GN, ReLU, Conv3 (128 filters)] $\times 2$	$128 \times 16 \times 16 \times 16$
	DownSample3	Max pooling 3d	$256 \times 8 \times 8 \times 8$
	EnBlock ($i = 4$)	[GN, ReLU, Conv3 (256 filters) GN, ReLU, Conv3 (256 filters)] $\times 2$	$256 \times 8 \times 8 \times 8$
	DownSample4	Max pooling 3d	$512 \times 4 \times 4 \times 4$
	EnBlock ($i = 5$)	[GN, ReLU, Conv3 (512 filters) GN, ReLU, Conv3 (512 filters)] $\times 4$	$512 \times 4 \times 4 \times 4$
TACR Module	TACR		$32 \times 128/2^{i-1} \times 128/2^{i-1} \times 128/2^{i-1}$
PD	PD		$96 \times 32 \times 32 \times 32$
MRF	MRF		$96 \times 128/2^{i-1} \times 128/2^{i-1} \times 128/2^{i-1}$
MSCF Module	MSCF		$32 \times 128/2^{i-1} \times 128/2^{i-1} \times 128/2^{i-1}$
3D ARA Module	3D ARA		$3 \times 128/2^{i-1} \times 128/2^{i-1} \times 128/2^{i-1}$
Deep Supervision	Final Layer	Trilinear interpolation	$3 \times 128 \times 128 \times 128$

References

1. Bauer, S.; Wiest, R.; Nolte, L.P.; Reyes, M. A survey of MRI-based medical image analysis for brain tumor studies. *Phys. Med. Biol.* **2013**, *58*, R97. [\[CrossRef\]](#)
2. Mohammed, Y.M.A.; El Garouani, S.; Jellouli, I. A survey of methods for brain tumor segmentation-based MRI images. *J. Comput. Des. Eng.* **2023**, *10*, 266–293. [\[CrossRef\]](#)
3. Magadza, T.; Viriri, S. Deep learning for brain tumor segmentation: A survey of state-of-the-art. *J. Imaging* **2021**, *7*, 19. [\[CrossRef\]](#)
4. Zhou, T.; Ruan, S.; Canu, S. A review: Deep learning for medical image segmentation using multi-modality fusion. *Array* **2019**, *3–4*, 100004. [\[CrossRef\]](#)
5. Soomro, T.A.; Zheng, L.; Afifi, A.J.; Ali, A.; Soomro, S.; Yin, M.; Gao, J. Image Segmentation for MR Brain Tumor Detection Using Machine Learning: A Review. *IEEE Rev. Biomed. Eng.* **2023**, *16*, 70–90. [\[CrossRef\]](#)
6. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, 5–9 October 2015; pp. 234–241. [\[CrossRef\]](#)
7. Ranjbarzadeh, R.; Bagherian Kasgari, A.; Jafarzadeh Ghoushchi, S.; Anari, S.; Naseri, M.; Bendechache, M. Brain tumor segmentation based on deep learning and an attention mechanism using MRI multi-modalities brain images. *Sci. Rep.* **2021**, *11*, 10930. [\[CrossRef\]](#)
8. Havaei, M.; Davy, A.; Warde-Farley, D.; Biard, A.; Courville, A.; Bengio, Y.; Pal, C.; Jodoin, P.M.; Larochelle, H. Brain tumor segmentation with Deep Neural Networks. *Med. Image Anal.* **2017**, *35*, 18–31. [\[CrossRef\]](#)
9. Awasthi, N.; Pardasani, R.; Gupta, S. Multi-Threshold Attention U-Net (MTAU) Based Model for Multimodal Brain Tumor Segmentation in Mri Scans. In *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries, Proceedings of the 6th International Workshop, BrainLes 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, 4 October 2020*; Springer: Cham, Switzerland, 2021; pp. 168–178. [\[CrossRef\]](#)
10. Shen, H.; Wang, R.; Zhang, J.; McKenna, S.J. Boundary-aware fully convolutional network for brain tumor segmentation. In Proceedings of the Medical Image Computing and Computer-Assisted Intervention- MICCAI 2017: 20th International Conference, Quebec City, QC, Canada, 11–13 September 2017; pp. 433–441. [\[CrossRef\]](#)
11. Zhang, J.; Jiang, Z.; Dong, J.; Hou, Y.; Liu, B. Attention Gate ResU-Net for Automatic MRI Brain Tumor Segmentation. *IEEE Access* **2020**, *8*, 58533–58545. [\[CrossRef\]](#)
12. Çiçek, Ö.; Abdulkadir, A.; Lienkamp, S.S.; Brox, T.; Ronneberger, O. 3D U-Net: Learning dense volumetric segmentation from sparse annotation. In Proceedings of the Medical Image Computing and Computer-Assisted Intervention–MICCAI 2016: 19th International Conference, Athens, Greece, 17–21 October 2016; pp. 424–432. [\[CrossRef\]](#)
13. Huang, H.; Yang, G.; Zhang, W.; Xu, X.; Yang, W.; Jiang, W.; Lai, X. A Deep Multi-Task Learning Framework for Brain Tumor Segmentation. *Front. Oncol.* **2021**, *11*, 690244. [\[CrossRef\]](#)
14. Liu, Z.; Tong, L.; Chen, L.; Zhou, F.; Jiang, Z.; Zhang, Q.; Wang, Y.; Shan, C.; Li, L.; Zhou, H. CANet: Context Aware Network for Brain Glioma Segmentation. *IEEE Trans. Med. Imaging* **2021**, *40*, 1763–1777. [\[CrossRef\]](#)
15. Chen, L.C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *40*, 834–848. [\[CrossRef\]](#)
16. Zhou, Z.; He, Z.; Jia, Y. AFPNet: A 3D fully convolutional neural network with atrous-convolution feature pyramid for brain tumor segmentation via MRI images. *Neurocomputing* **2020**, *402*, 235–244. [\[CrossRef\]](#)
17. Zhou, Z.; He, Z.; Shi, M.; Du, J.; Chen, D. 3D dense connectivity network with atrous convolutional feature pyramid for brain tumor segmentation in magnetic resonance imaging of human heads. *Comput. Biol. Med.* **2020**, *121*, 103766. [\[CrossRef\]](#)
18. Liu, X.; Hou, S.; Liu, S.; Ding, W.; Zhang, Y. Attention-based multimodal glioma segmentation with multi-attention layers for small-intensity dissimilarity. *J. King Saud Univ.-Comput. Inf. Sci.* **2023**, *35*, 183–195. [\[CrossRef\]](#)
19. Akbar, A.S.; Fatichah, C.; Suciati, N. Single level UNet3D with multipath residual attention block for brain tumor segmentation. *J. King Saud Univ.-Comput. Inf. Sci.* **2022**, *34*, 3247–3258. [\[CrossRef\]](#)
20. Fang, Y.; Huang, H.; Yang, W.; Xu, X.; Jiang, W.; Lai, X. Nonlocal convolutional block attention module VNet for gliomas automatic segmentation. *Int. J. Imaging Syst. Technol.* **2022**, *32*, 528–543. [\[CrossRef\]](#)
21. Zhou, J.; Ye, J.; Liang, Y.; Zhao, J.; Wu, Y.; Luo, S.; Lai, X.; Wang, J. scse-nl v-net: A brain tumor automatic segmentation method based on spatial and channel “squeeze-and-excitation” network with non-local block. *Front. Neurosci.* **2022**, *16*, 916818. [\[CrossRef\]](#)
22. Dekeyser, S.; De Kock, I.; Nikoubashman, O.; Vanden Bossche, S.; Van Eetvelde, R.; De Groote, J.; Acou, M.; Wiesmann, M.; Deblaere, K.; Achten, E. “Unforgettable”—a pictorial essay on anatomy and pathology of the hippocampus. *Insights Imaging* **2017**, *8*, 199–212. [\[CrossRef\]](#)
23. Lemaire, J.J.; De Salles, A.; Coll, G.; El Ouadih, Y.; Chaix, R.; Coste, J.; Durif, F.; Makris, N.; Kikinis, R. MRI atlas of the human deep brain. *Front. Neurol.* **2019**, *10*, 851. [\[CrossRef\]](#)
24. Menze, B.H.; Jakab, A.; Bauer, S.; Kalpathy-Cramer, J.; Farahani, K.; Kirby, J.; Burren, Y.; Porz, N.; Slotboom, J.; Wiest, R.; et al. The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). *IEEE Trans. Med. Imaging* **2015**, *34*, 1993–2024. [\[CrossRef\]](#) [\[PubMed\]](#)

25. Bakas, S.; Akbari, H.; Sotiras, A.; Bilello, M.; Rozycki, M.; Kirby, J.S.; Freymann, J.B.; Farahani, K.; Davatzikos, C. Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features. *Sci. Data* **2017**, *4*, 170117. [[CrossRef](#)] [[PubMed](#)]
26. Bakas, S.; Reyes, M.; Jakab, A.; Bauer, S.; Rempfler, M.; Crimi, A.; Shinohara, R.T.; Berger, C.; Ha, S.M.; Rozycki, M.; et al. Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the BRATS challenge. *arXiv* **2018**, arXiv:1811.02629. [[CrossRef](#)]
27. Milletari, F.; Navab, N.; Ahmadi, S.A. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In Proceedings of the 2016 Fourth International Conference on 3D Vision (3DV), Stanford, CA, USA, 25–28 October 2016; pp. 565–571. [[CrossRef](#)]
28. Woo, S.; Park, J.; Lee, J.Y.; Kweon, I.S. Cbam: Convolutional block attention module. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 3–19. [[CrossRef](#)]
29. Hu, J.; Shen, L.; Sun, G. Squeeze-and-Excitation Networks. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 7132–7141. [[CrossRef](#)]
30. Wu, Z.; Su, L.; Huang, Q. Cascaded partial decoder for fast and accurate salient object detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 3907–3916. [[CrossRef](#)]
31. Chen, L.C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-decoder with atrous separable convolution for semantic image segmentation. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 801–818. [[CrossRef](#)]
32. Lou, A.; Loew, M. CFPNET: Channel-Wise Feature Pyramid For Real-Time Semantic Segmentation. In Proceedings of the 2021 IEEE International Conference on Image Processing (ICIP), Anchorage, AK, USA, 19–22 September 2021; pp. 1894–1898. [[CrossRef](#)]
33. van Beers, F.; Lindström, A.; Okafor, E.; Wiering, M. Deep neural networks with intersection over union loss for binary image segmentation. In Proceedings of the 8th International Conference on Pattern Recognition Applications and Methods, Prague, Czech Republic, 19–21 February 2019; pp. 438–445. [[CrossRef](#)]
34. Savadikar, C.; Kulhalli, R.; Garware, B. Brain tumour segmentation using probabilistic u-net. In Proceedings of the Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 6th International Workshop, BrainLes 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, 4 October 2020; pp. 255–264. [[CrossRef](#)]
35. Ma, S.; Zhang, Z.; Ding, J.; Li, X.; Tang, J.; Guo, F. A deep supervision CNN network for brain tumor segmentation. In Proceedings of the Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 6th International Workshop, BrainLes 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, 4 October 2020; pp. 158–167. [[CrossRef](#)]
36. Soltaninejad, M.; Pridmore, T.; Pound, M. Efficient MRI brain tumor segmentation using multi-resolution encoder-decoder networks. In Proceedings of the Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 6th International Workshop, BrainLes 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, 4 October 2020; pp. 30–39. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.