Biomedical Signal Processing and Control 81 (2023) 104524

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc

Modality-level cross-connection and attentional feature fusion based deep neural network for multi-modal brain tumor segmentation

Tongxue Zhou

School of Information Science and Technology, Hangzhou Normal University, Hangzhou 311121, China

ARTICLE INFO

Keywords:
Brain tumor segmentation
Cross-connection
Feature fusion
Multi-modality
Deep learning

ABSTRACT

Brain tumor segmentation from Magnetic Resonance Imaging is essential for early diagnosis and treatment planning for brain cancers in clinical practice. However, existing brain tumor segmentation methods cannot sufficiently learn high-quality feature information for segmentation. To address this issue, a modality-level cross-connection and attentional feature fusion based deep neural network is proposed for multi-modal brain tumor segmentation. The proposed method can not only locate the whole tumor region but also can accurately segment the sub-tumor regions. The proposed network architecture is a multi-encoder based 3D U-Net. Inspired by the characteristics of multi-modalities, a modality-level cross-connection (MCC) is first proposed to take advantage of the complementary information between the related modalities. Moreover, to enhance the feature learning capacity of the network, the attentional feature fusion module (AFFM) is proposed to fuse the multi-modalities as well as to extract the useful feature representation for segmentation. It consists of two components: multi-scale spatial feature fusion (MSFF) block and dual-path channel feature fusion (DCFF) block. They aim at learning multi-scale spatial contextual information and the channel-wise feature information to improve the segmentation accuracy. Also, the proposed fusion module can be easily integrated into other fusion models and deep neural network architectures. Comprehensive experiments evaluated on the BraTS 2018 dataset demonstrate that the proposed network architecture can effectively improve the brain tumor segmentation performance when compared with the baseline methods and the state-of-the-art methods.

1. Introduction

Brain tumors is the mass or growth of abnormal cells in the brain, which are one of the most life-threatening cancers in the world [1]. Brain tumors take up 85% to 90% of all primary central nervous system (CNS) tumors. According to the cell characteristics of the abnormal regions and the degree of harm to the body, brain tumors are mainly divided into two categories: benign and malignant. Based on the statistics from the American Cancer Society, it is estimated that around 25,050 malignant tumors of the brain or spinal cord will be diagnosed in 2022. In addition, the 5-year survival rate for the malignant brain or CNS tumor is only about 36%, and the 10-year survival rate is only 31%.[1] Therefore, effective brain tumor segmentation approaches are urgently required in clinical diagnosis, assessment and following radiotherapy treatment planning.

Many imaging techniques can be applied for brain tumor screening, such as Positron Emission Tomography (PET), Computed Tomography (CT) and Magnetic Resonance Imaging (MRI). MRI is the most prominent medical imaging technology to present detailed brain structures as well as obtain multimodal images [2]. The commonly used multimodal MR images include T1-weighted, contrast-enhanced T1-weighted (T1c), T2-weighted, and fluid-attenuated inversion recovery (FLAIR), shown in Fig. 1. All of these modalities are available in the sagittal, coronal, and axial planes. And they can provide complementary information to help distinguish brain tumors from the normal brain tissues, and further identify the intra-tumoral structures. Concretely, FLAIR and T2 can highlight the whole tumor region, especially, the edema region, T1 and T1c can distinguish the tumor core region, including necrotic tumor, non-enhancing tumor core and enhancing tumor regions [3].

Manual brain tumor segmentation requires a lot of expertise, and the segmentation is often achieved in a slice-by-slice manner. Although we can take advantage of the expert knowledge, this method is very tedious and time-consuming, and the results are very subjective [4,5]. In the last few years, a large number of automated approaches have been proposed to segment the brain tumors, which can be generally categorized into two groups:

(1) Machine learning-based methods. These methods use little prior knowledge of the brain and depend on hand-crafted features, such as local histograms, shape difference, inter-image gradient and symmetry

E-mail address: txzhou@hznu.edu.cn.

[1] https://www.cancer.net/cancer-types/brain-tumor/statistics.

https://doi.org/10.1016/j.bspc.2022.104524
Received 12 July 2022; Received in revised form 25 October 2022; Accepted 18 December 2022
Available online 21 December 2022
1746-8094/

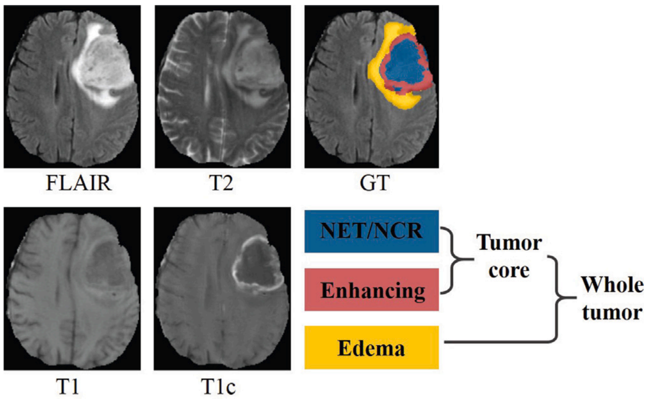

Fig. 1. The four MR modalities FLAIR, T2, T1, T1c and the corresponding tumor sub-region ground truth, consisting of NET/NCR, enhancing tumor and edema. NET denotes non-enhancing tumor core, NCR denotes necrotic tumor.

analysis [6]. For example, Lapuyade-Lahorgue et al. [7] proposed to use probability theory to segment brain tumors. Lian et al. [8] applied belief function on brain tumor segmentation. Bonte et al. [9] and Chen et al. [10] adopted random forests to address the brain tumor segmentation issue. Hu et al. [11] introduced conditional random fields to the field of brain tumor segmentation. Krishnakumar et al. [12] applied the K-means algorithm and multi-kernel support vector machines (SVM) for brain tumor segmentation. Nie et al. [13] proposed to use hidden markov random field for brain tumor segmentation. However, brain tumor segmentation is a very challenging task due to the large variety in tumor shape, position, texture, scanning modalities and scanning parameters.

(2) Deep learning-based methods. Recently, deep learning starts to impact medical research field recently, such as in the field of brain tumor segmentation [14], detection of breast cancer on screening mammograms [15], diagnosis of Attention Deficit Hyperactivity Disorder (ADHD) [16] and prediction of COVID fatality in Alzheimer's patients. The reason is that deep learning-based methods have powerful ability to integrate vast datasets, learn arbitrarily complex relationships and incorporate existing knowledge [17]. Lately, several deep learning-based network architectures have been employed in brain tumor segmentation [18], such as Convolutional Neural Networks (CNN) [19], Fully Convolutional Networks (FCN) [20], Residual Net [21] and U-net [22]. For example, Hu et al. [23] proposed an encoder–decoder based brain tumor segmentation network called MU-Net. They designed a module named global attention (GA) to combine high-level features and low-level features instead of concatenating them directly. Weninger et al. [24] proposed a U-Net based two-step network by first locating the whole tumor, then segmenting the whole tumor into tumor core, enhanced tumor, and edema regions. Yang et al. [25] introduced a contour-aware Residual U-Net to segment brain tumors. Besides, the adversarial training is embedded in the proposed network to improve the segmentation results. Yao et al. [26] proposed a cascaded network framework by combining the Inception model, 3D Squeeze and Excitation structures, and dilated convolution filters to achieve the brain tumor segmentation. Ghaffari et al. [27] proposed an automated post-operative brain tumor segmentation method. Ma et al. [28] proposed a dual graph reasoning unit for brain tumor segmentation.

Although the aforementioned methods can achieve brain tumor segmentation, they did not consider exploring and utilizing the potential correlation between multi-modalities, which is essential for the subsequent multi-modal fusion stage. To this end, a novel modality-level cross-connection is proposed to explore and utilize the multi-modalities complementary information. In addition, the proposed attentional feature fusion can selectively extract the important features to improve the segmentation performance. The main contributions of this paper are concluded as follows:

(1) A modality-level cross-connection (MCC) is proposed to exploit the complementary information from multi-modalities and to perform effective feature aggregation for segmentation.

(2) A novel feature fusion module, named attentional feature fusion module (AFFM) is developed. It consists of a multi-scale spatial feature fusion (MSFF) block and a dual-path channel feature fusion (DCFF) block. It can fuse different modalities and selectively extract useful feature information for segmentation.

(3) It is proved that combining the multi-scale spatial feature fusion block and the dual-path channel feature fusion block can significantly strengthen the ability of feature learning and improve the segmentation performance.

(4) Comprehensive experiments conducted on BraTS 2018 demonstrate the effectiveness of the proposed method, and the superior performance compared with the state-of-the-art methods.

The rest of this paper is organized as follows: Section 2 introduces the proposed network and describe the details of each proposed components. Section 3 provides the experimental setup. Section 4 describes and discusses the experimental results. Section 5 concludes the proposed work.

2. Method

In this section, I first introduce the proposed network architecture. Then, I elaborate on the modality-level cross-connection for exploiting the complementary information from the related MR modalities. Finally, I detail the attentional feature fusion module to fuse different MR modalities and extract the important feature information for segmentation.

2.1. Network architecture

The proposed network architecture is a multi-encoder based 3D U-Net, which is depicted in Fig. 2. First, four MR modalities are fed to four individual encoders to learn the modality-specific feature representations. Besides, modality-level cross-connections (MCC) are proposed to exploit the complementary information from the related modalities (FLAIR-T2, T1-T1c). Then, the attentional feature fusion module (AFFM) is proposed to fuse the multi-modalities, as well as to extract the useful feature representation for segmentation at each level. Finally, the fused feature representation is passed to the label space by a sigmoid activation function to obtain the segmentation result for each label. The detailed network architecture is depicted in Table 1.

2.2. Modality-level cross-connection

As is shown in Fig. 1, we can observe that FLAIR and T2 modalities are suitable to detect the whole tumor region, especially the edema region. T1 and T1c modalities are suitable to detect the tumor core regions, consisting of NET/NCR, and enhancing tumor regions. Therefore, it can be assumed that there is a strong potential relationship between these two pairs of modalities. To take advantage of the complementary information between the related modalities, the modality-level cross-connection (colorful arrows in Fig. 2) is proposed. It is applied in the four encoder paths between the two pairs of modalities: FLAIR and T2, T1 and T1c. Specifically, in each level, after the convolution operation, the extracted feature from one modality will be connected (concatenated) to the corresponding related modality, and as the input in the encoder path of the related modality at the next level. To make it clear, let x^i_{T1} be the convolution output of T1 modality at the ith level in the encoder path, y^i_{T1c} be the convolution output of T1c modality at the ith level in the encoder path. Then, the cross-connection will perform on the two outputs x^i_{T1} and y^i_{T1c} via concatenation operation ([...]). Following that, $\left[x^i_{T1}, y^i_{T1c}\right]$ will be passed as the input of T1c modality

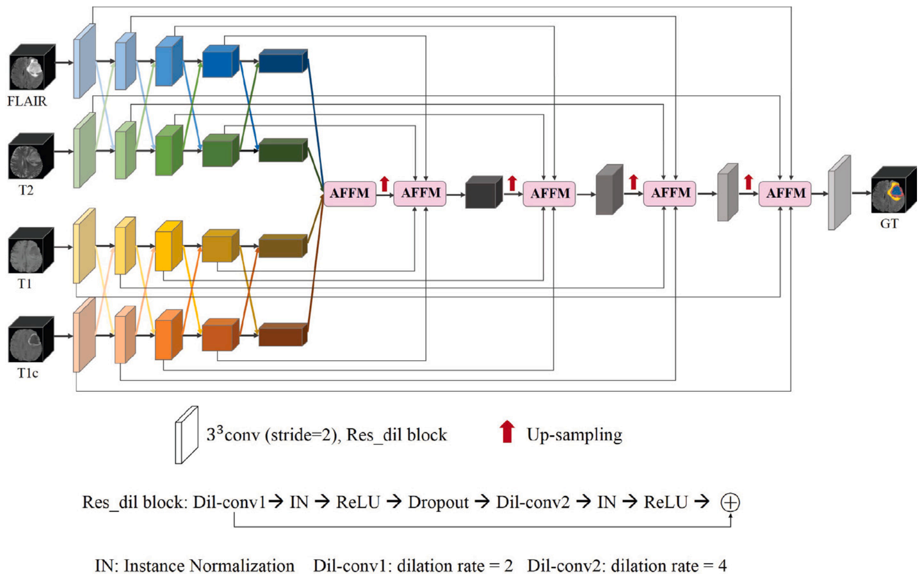

Fig. 2. The overview of the proposed network architecture. Four MR modalities are fed to four individual encoders to learn the modality-specific feature representations. In addition, multi-level cross-connections (indicated by colorful arrows) are employed to exploit the complementary information from the related modalities. Subsequently, an attentional feature fusion module is proposed to fuse the multi-modalities and extract the useful feature representation for segmentation at each level. The colorful arrows represent the concatenation.

Table 1
The detailed network architecture of the proposed network architecture. For simplicity, only the encoders part are presented, the decoders are symmetrical. Conv. is short for the 3D convolution, IN is short for instance normalization, and # filters indicate the number of filters.

Level	Input size	Type	Conv. kernel	#Kernel	Stride	Output size
1	$1 \times 128 \times 128 \times 128$	Conv.	$3 \times 3 \times 3$	16	$1 \times 1 \times 1$	$16 \times 128 \times 128 \times 128$
		IN	–	–	–	–
		LeakyReLU	–	–	–	–
		Res_dil	$3 \times 3 \times 3$	16	$1 \times 1 \times 1$	$16 \times 128 \times 128 \times 128$
2	$16 \times 128 \times 128 \times 128$	Conv.	$3 \times 3 \times 3$	32	$2 \times 2 \times 2$	$32 \times 64 \times 64 \times 64$
		IN	–	–	–	–
		LeakyReLU	–	–	–	–
		Res_dil	$3 \times 3 \times 3$	32	$1 \times 1 \times 1$	$32 \times 64 \times 64 \times 64$
3	$32 \times 64 \times 64 \times 64$	Conv.	$3 \times 3 \times 3$	64	$2 \times 2 \times 2$	$64 \times 32 \times 32 \times 32$
		IN	–	–	–	–
		LeakyReLU	–	–	–	–
		Res_dil	$3 \times 3 \times 3$	64	$1 \times 1 \times 1$	$64 \times 32 \times 32 \times 32$
4	$64 \times 32 \times 32 \times 32$	Conv.	$3 \times 3 \times 3$	128	$2 \times 2 \times 2$	$128 \times 16 \times 16 \times 16$
		IN	–	–	–	–
		LeakyReLU	–	–	–	–
		Res_dil	$3 \times 3 \times 3$	128	$1 \times 1 \times 1$	$128 \times 16 \times 16 \times 16$
5	$128 \times 16 \times 16 \times 16$	Conv.	$3 \times 3 \times 3$	256	$2 \times 2 \times 2$	$256 \times 8 \times 8 \times 8$
		IN	–	–	–	–
		LeakyReLU	–	–	–	–
		Res_dil	$3 \times 3 \times 3$	256	$1 \times 1 \times 1$	$256 \times 8 \times 8 \times 8$

at the $(i+1)$th level in the encoder path. It is the same for the encoder path of the T1 modality. In conclusion, by introducing the cross-connection structure, the network can learn the complimentary feature information from related modalities. And the learned complimentary feature can guide the following feature fusion stage to achieve effective feature aggregation for segmentation.

2.3. Attentional feature fusion module

It is known that attention plays a significant role in human perception [29–31]. For multi-modal brain MR images, different MR modalities can highlight different tissue structures and underlying anatomy. Therefore, designing an effective feature fusion method to stand out the important features from different modalities is crucial for multi-modal segmentation tasks. Many recent works have been proposed to achieve the feature fusion, such as Squeeze and Excitation (SE) block [32] and Convolutional Block Attention Module (CBAM) [29]. SE method only aims at recalibrating the channel-wise feature. CBAM method combines the channel attention mechanism with the spatial attention mechanism. However, these two methods were proposed to achieve image classification and object detection tasks, which are not designed for image segmentation tasks. In addition, since different scale features can provide different receptive fields for the network, therefore, capturing the multi-scale context feature information is particularly essential for the segmentation task. To this end, a special feature fusion method, named attentional feature fusion module (AFFM), is designed to emphasize the important information and suppress the weak ones, the architecture is depicted in Fig. 3. Instead of directly fusing the multi-modalities in the input space [33–36], the network adopts the multi-encoder framework. The individual encoders can obtain the modality-specific feature representations. Then, the AFFM is proposed to not only fuse the multi-modal feature representations but also selectively learn the important features and discard the useless ones. It is noted that the AAFM is utilized at each level of the network. AFFM is composed of two blocks: a multi-scale spatial feature fusion (MSFF) block and a dual-path channel feature fusion (DCFF) block.

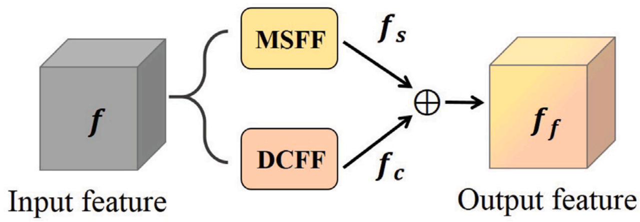

Fig. 3. The architecture of the proposed attentional feature fusion module, consisting of a multi-scale spatial feature fusion (MSFF) block and a dual-path channel feature fusion (DCFF) block.

2.3.1. Multi-scale spatial feature fusion

The architecture of the multi-scale spatial feature fusion block is depicted in Fig. 4. Let (f_1, f_2, f_3, f_4, …, f_n) represent the independent features obtained from the individual encoders. They are first concatenated $f = [f_1, f_2, f_3, f_4, \dots, f_n]$, $n = 4$ Then, three parallel convolution operations with different kernel sizes ($1 \times 1 \times 1$, $3 \times 3 \times 3$ and $5 \times 5 \times 5$) are utilized to learn the multi-scale feature information: $s_1 = W_{s1} * f$, $s_2 = W_{s2} * f$ and $s_3 = W_{s3} * f$, where W_{s1}, W_{s2} and W_{s3} are the convolution weights. Subsequently, a sigmoid activation function is applied to obtain the spatial attention $\sigma(s_1)$, $\sigma(s_2)$ and $\sigma(s_3)$, respectively. Following that, the multi-scale features can be obtained by multiplying the three spatial attention $\sigma(s_1)$ and $\sigma(s_2)$ with the input feature f: $f_{s1} = \sigma(s_1) \otimes f$, $f_{s2} = \sigma(s_2) \otimes f$ and $f_{s3} = \sigma(s_3) \otimes f$. The final fused feature can be obtained by merging the three multi-scale features $f_s = f_{s1} + f_{s2} + f_{s3}$ via element-wise summation. In conclusion, the proposed multi-scale spatial feature fusion block can take advantage of different scale information, and combine them to obtain rich feature information. Besides, the obtained spatial contextual feature can guide the network to focus on the brain tumor regions from different views in order to improve the segmentation accuracy.

2.3.2. Dual-path channel feature fusion

The architecture of the dual-path channel feature fusion block is depicted in Fig. 5. Here, I take modality f_1 as an example. In the first path, a 3D global average pooling is first utilized to obtain the global spatial information f_{1c}^a. Simultaneously, in the second path, a 3D global max pooling is utilized to capture another important feature f_{1c}^m.

$$f_{1c}^a = GAP(f_{1c}(i,j)) \tag{1}$$

$$f_{1c}^m = GMP(f_{1c}(i,j)) \tag{2}$$

Then, the two features f_{1c}^a and f_{1c}^m are forwarded to two fully-connected layers, respectively, to learn the individual channel attentions.

$$\hat{f}_{1c}^a = (W_{c2}(\delta(W_{c1} f_{1c}^a))) \tag{3}$$

$$\hat{f}_{1c}^m = (W_{c2}(\delta(W_{c1} f_{1c}^m))) \tag{4}$$

where W_{c1} and W_{c2} are the weights of two fully-connected layers, and $\delta(\cdot)$ the ReLU operator.

Subsequently, $\hat{f}_{1c}^a$ and $\hat{f}_{1c}^m$ are merged via element-wise summation, and followed by a sigmoid activation function to achieve the dual-path channel attention. Finally, they are multiplied with the input feature f_1 to obtain the channel-wise feature representation.

$$f_{1c} = (\sigma(\hat{f}_{1c}^a + \hat{f}_{1c}^m)) \otimes f_1 \tag{5}$$

All the modalities will follow the same operation like this, and they are finally fused via concatenation operation.

$$f_c = [f_{1c}, f_{2c}, f_{3c}, f_{4c}] \tag{6}$$

Table 2
Hyper-parameter settings of the proposed network architecture.

Hyper-parameter	Value
Learning rate	5e−4
Patience	10
Early stopping	50
Learning rate drop	0.5
Batch size	1
Layer	5
Initial filter	16

Therefore, the dual-path channel feature fusion block can selectively learn channel-wise feature by considering global average feature and global maximum feature. Finally, the multi-scale spatial feature and dual-path channel feature are merged via element-wise summation to achieve the final fused feature representation. In this way, the proposed attentional fusion module can guide the network to learn multi-scale spatial contextual information and the channel-wise attentional feature information to improve the segmentation accuracy.

$$f_f = f_c + f_s \tag{7}$$

3. Experimental settings

3.1. Implementation details

The proposed network is implemented by Keras using a single Nvidia Tesla V100 (32G). Nadam is used as the optimizer with an initial learning rate of 0.0005. It will be halved with the patience of 10 epochs. To avoid over-fitting, early stopping is adopted when the validation loss is not improved by 50 epochs. The batch size is set as 1. The dataset is randomly split into 0.8 training and 0.2 testing. The network is trained by using Dice loss function, defined in Eq. (8). Table 2 presents more details about the hyper-parameter settings.

$$L_{Dice} = 1 - 2\frac{\sum_{i=1}^{C}\sum_{j=1}^{N} p_{ij} g_{ij} + \epsilon}{\sum_{i=1}^{C}\sum_{j=1}^{N} (p_{ij} + g_{ij}) + \epsilon} \tag{8}$$

where N is the number of the examples, C is the number of the classes, p_{ij} indicates the probability that pixel i belongs to the tumor class j, which is the same for g_{ij}, and ϵ is a small constant to avoid dividing by 0.

3.2. Dataset and pre-processing

The proposed method is evaluated on the largest public multimodal brain tumor segmentation dataset, BraTS 2018 [37], which is widely used to evaluate the model's performance. This dataset provides 285 training samples, each sample consists of four pre-operative MRI modalities, T1, T1c, T2, FLAIR, and ground truth annotated by experienced neuro-radiologists following the same annotation protocol. The ground truth annotation comprises enhancing tumor (ET), edema (ED), the necrotic and non-enhancing tumor core (NCR/NET).

The original image size is 155 × 240 × 240. These images have been pre-processed by the provider, including the co-registering to the same anatomical template, interpolating to the same resolution (1 mm^3), and skull-stripping. In the experiment, N4ITK [38] is first employed to correct bias field and intensity inhomogeneities in MRI images. Then, the images are resized to 128 × 128 × 128. Finally, Z-score normalization is conducted on each modality of each sample by subtracting the mean and dividing by the standard deviation to a zero-mean, unit-variance space.

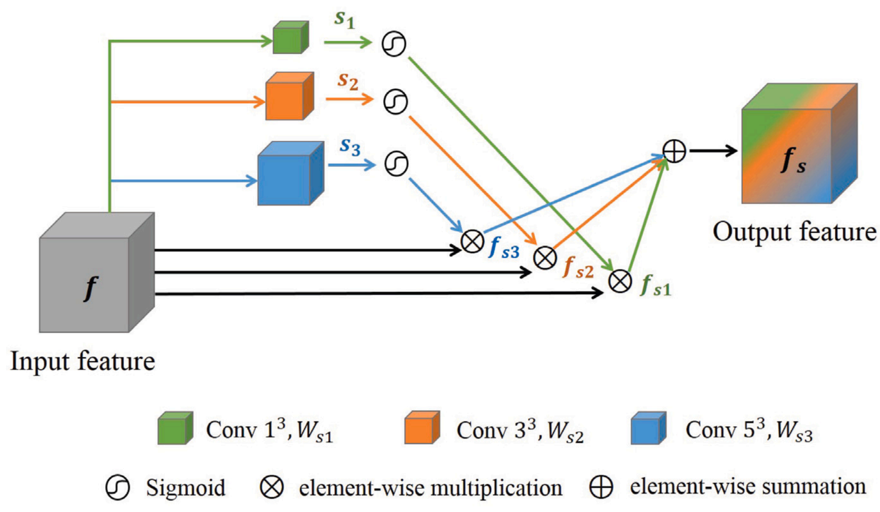

Fig. 4. The architecture of the proposed multi-scale spatial feature fusion block.

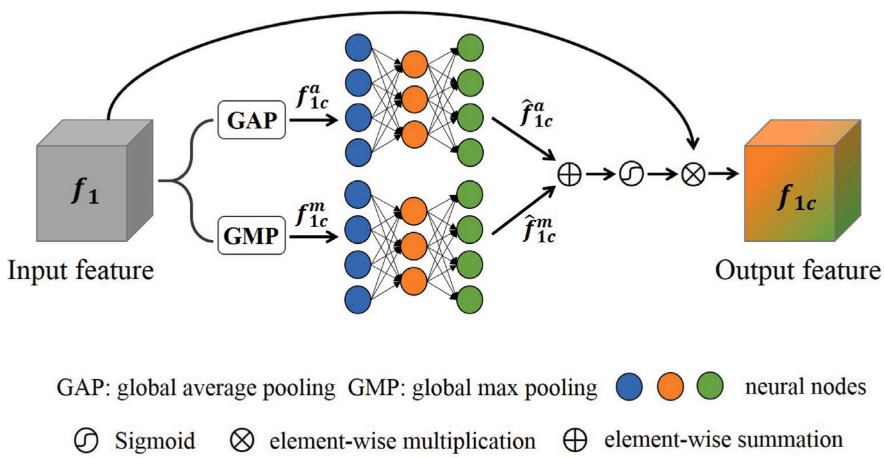

Fig. 5. The architecture of the proposed dual-path channel feature fusion block.

3.3. Evaluation metrics

To evaluate the performance of the proposed network architecture, two popular evaluation metrics are adopted in semantic segmentation, Dice Score and Hausdorff Distance.

Dice Score is used to measure the overlap between the prediction region and the ground truth region. A higher Dice Score value indicates a better segmentation performance, which is defined as follows:

$$Dice\ Score = \frac{2|P \cap T|}{|P| + |T|} \tag{9}$$

where P and T denote the volumes of the prediction and ground truth, respectively.

Hausdorff Distance is an indicator of the largest segmentation error, it is calculated between the boundaries of prediction and ground truth. The lower value indicates a better segmentation performance.

$$Hausdorff\ Distance = \max\{sup_{r \in \partial R} d_m(s,r), sup_{s \in \partial S} d_m(r,s)\} \tag{10}$$

where ∂S and ∂R are the sets of tumor border voxels for the predicted and the ground truth, respectively, and $d_m(v,v)$ is the minimum of the Euclidean distances between a voxel v and voxels in a set V.

4. Experiment results

In this section, First, an ablation analysis is conducted on the proposed method to demonstrate the effectiveness of the proposed components in Section 4.1.1. Then, to prove the superior performance of the proposed attentional feature fusion module, it is compared with the state-of-the-art attention modules in Section 4.1.2. Following that, the proposed method is compared with the state-of-the-art brain tumor segmentation methods in Section 4.1.3. Finally, the qualitative analysis is implemented to further prove the advantages of the proposed method in Section 4.2.

4.1. Quantitative analysis

4.1.1. Ablation analysis of the proposed components

First, an ablation analysis is performed to quantify the effects of the proposed components, including modality-level cross-connection (MCC), multi-scale feature fusion (MSFF) block, and dual-path channel feature fusion (DCFF) block. The comparison results are reported in Table 3. It can be observed that compared with the baseline method, the proposed modality-level cross-connection (MCC) can significantly improve the segmentation performance, with a 1.7% improvement on tumor core region, 2.2% on enhancing tumor region in terms of Dice Score, and a 15.9% improvement on tumor core region, 14.3% on enhancing tumor region in terms of Hausdorff Distance, respectively. The reason is that the proposed modality-level cross-connection can perform effective feature aggregation. Based on it, by incorporating a multi-scale feature fusion (MSFF) block, compared with the baseline method, a significant improvement of 2.1% on tumor core region,

Table 3
Comparison results of the ablation analysis in terms of Dice Score and Hausdorff Distance on BraTS 2018 dataset. WT, TC and ET denote whole tumor, tumor core and enhancing tumor, respectively, bold results denote the best score.

Method	Dice Score				Hausdorff Distance			
	WT	TC	ET	Mean	WT	TC	ET	Mean
Baseline	85.8	84.0	76.6	82.1	4.9	4.4	2.8	4.0
Baseline + MCC	85.9	85.4	78.3	83.2	5.0	3.7	**2.4**	3.7
Baseline + MCC + MSFF	86.0	85.8	79.3	83.7	5.5	3.7	2.5	3.9
Baseline + MCC + DCFF	86.3	85.4	78.1	83.3	**4.5**	3.7	2.6	**3.6**
Baseline + MCC + MSFF + DCFF	**86.5**	**87.0**	**79.4**	**84.3**	4.6	**3.6**	2.5	**3.6**

Table 4
Comparison results with the state-of-the-art attention modules in terms of Dice Score and Hausdorff Distance on the BraTS 2018 dataset. WT, TC, ET denote whole tumor, tumor core and enhancing tumor, respectively, bold results denote the best score.

Method	Dice Score				Hausdorff Distance			
	WT	TC	ET	Mean	WT	TC	ET	Mean
CBAM [29]	82.7	75.2	68.2	75.4	6.4	8.5	6.7	7.2
SE [32]	85.4	84.5	77.4	82.4	**4.6**	4.6	3.0	4.1
Ours	**86.5**	**87.0**	**79.4**	**84.3**	**4.6**	**3.6**	**2.5**	**3.6**

Table 5
Comparison results with the state-of-the-art methods in terms of Dice Score and Hausdorff Distance on the BraTS 2018 dataset. WT, TC, ET denote whole tumor, tumor core and enhancing tumor, respectively, bold results denote the best score.

Method	Dice Score				Hausdorff Distance			
	WT	TC	ET	Mean	WT	TC	ET	Mean
Hu et al. [40]	83.0	73.0	61.0	72.3	47.2	41.1	41.5	43.3
Serrano. et al. [41]	84.0	73.3	59.5	72.3	9.0	14.7	11.7	11.8
Zhang et al. [42]	87.6	77.2	72.0	78.9	–	–	–	–
Mehta et al. [43]	88.8	79.3	69.0	79.0	6.6	7.9	7.3	7.3
Kermi et al. [44]	86.7	79.8	71.7	79.4	8.7	6.4	4.7	6.6
Yang et al. [25]	83.8	78.6	78.1	80.2	5.9	7.1	4.7	5.9
Aboeleneinet al. [45]	86.5	80.8	74.5	80.6	7.5	8.8	4.4	6.9
Weninger et al. [24]	86.0	81.7	76.3	81.3	7.0	7.9	5.6	6.8
Yao et al. [26]	88.5	83.4	72.9	81.6	17.0	7.1	5.7	9.9
Chen et al. [46]	88.8	84.4	74.0	82.4	5.9	5.7	4.6	5.4
Popli et al. [39]	**90.7**	86.9	72.7	83.4	5.3	5.6	4.8	5.2
Ours	86.5	**87.0**	**79.4**	**84.3**	**4.6**	**3.6**	**2.5**	**3.6**

and 3.5% improvement on enhancing tumor region in terms of Dice Score can be observed. With the assistance of the dual-path channel feature fusion (DCFF) block, the segmentation performance is improved by 1.7% on the tumor core region, and 2.0% on enhancing tumor region in terms of Dice Score. A similar improvement can be seen in terms of Hausdorff Distance. In addition, by comparing MSFF, DCFF and the proposed method, it can be observed that using only one single feature fusion block can only obtain poor performance due to the inadequate feature information. Using the proposed method by considering both spatial and channel-wise features can significantly improve the segmentation accuracy. Therefore, the comparison results Table 3 prove the effectiveness of the proposed method. In addition, the computation analysis is performed, the proposed method uses 36M parameters, and takes 36 h for training the model, 2.8 s for testing a new subject, implying that the proposed method has a good real-time performance.

4.1.2. Comparison with the state-of-the-art attention methods

To further verify the importance of the proposed attentional feature fusion module (AFFM), a comparative experiment is performed with two suggested state-of-the-art attention modules. As reported in Table 4, we can observe that the proposed attentional feature fusion module (AFFM) outperforms both Squeeze and Excitation (SE) block [32] and CBAM [29]. Especially, compared with CBAM, the proposed method can achieve an significant 11.8% improvement in terms of average Dice Score, and a 50% improvement in terms of average Hausdorff Distance, respectively. This implies the advantages of the proposed feature fusion module on brain tumor segmentation.

4.1.3. Comparison with the state-of-the-art segmentation methods

The proposed method is also compared with several state-of-the-art methods designed for brain tumor segmentation. The results are illustrated in Table 5, it can be observed that the proposed method can outperform all the methods in terms of average Dice Score and average Hausdorff Distance. In addition, the method from [39] can achieve a higher Dice Score on the whole tumor region than the proposed method, however, the performance on enhancing tumor region is poor. It is worth noting that among the three sub-tumor regions, the enhancing tumor usually locates between the edema and necrosis regions, which is easily to be falsely predicted into the neighbor regions, and it is the most challenging sub-tumor region. For the enhancing tumor region, compared with the method from [39], the proposed method can obtain a 9.2% improvement in terms of Dice Score and 30.8% improvement in terms of Hausdorff Distance, respectively. It demonstrates the superior performance of the proposed method. It is noticed that the method from [26] uses 3D Squeeze and Excitation (SE) block. Compared with it, the proposed method can obtain a 3.3% improvement in terms of average Dice Score and 63.6% in terms of Hausdorff Distance, respectively. It indicates that the proposed method can further learn more informative features to obtain better segmentation results than the SE block. Therefore, the above experiments prove the effectiveness of the proposed method.

4.2. Qualitative analysis

To obtain a better understanding of the segmentation results and the effectiveness of the proposed method, some examples of the proposed method are presented in Figs. 6 and 7. The comparison results are consistent with the previous analysis in Tables 3 and 4. Specifically, from Fig. 6, we can observe that, by incorporating the proposed components, it can progressively obtain the competitive segmentation performance. Especially, in the first example, the baseline method fails to detect the isolated edema region around the left bottom, while the proposed components can gradually ameliorate the segmentation result. To verify the remarkable feature learning of the proposed attention module, I also compare it with the existing suggested attention methods: CBAM and SE methods. The results are presented in Fig. 7. We can observe the superior performance of the proposed method. For example, in the second example, the CBAM method detects a lot of false enhancing tumor regions. Also, the SE method produces some false enhancing tumor regions and NET/NCR regions. On the contrary, the proposed attention feature fusion module obtains a better performance. It can be explained that the other two methods are not designed specially for segmentation tasks, also they did not take the multi-scale feature representations into account. Thus, the qualitative results further demonstrate the effectiveness of the proposed feature fusion module.

Moreover, some examples are selected to visualize the feature maps from different methods, which are presented in Fig. 8. From Fig. 8, on the one hand, we can observe that the tumor regions are not obvious in the original feature maps (before applying the AFFM). However, the proposed multi-scale spatial feature fusion (MSFF) block and dual-path channel feature fusion (DCFF) block can help the network focus on

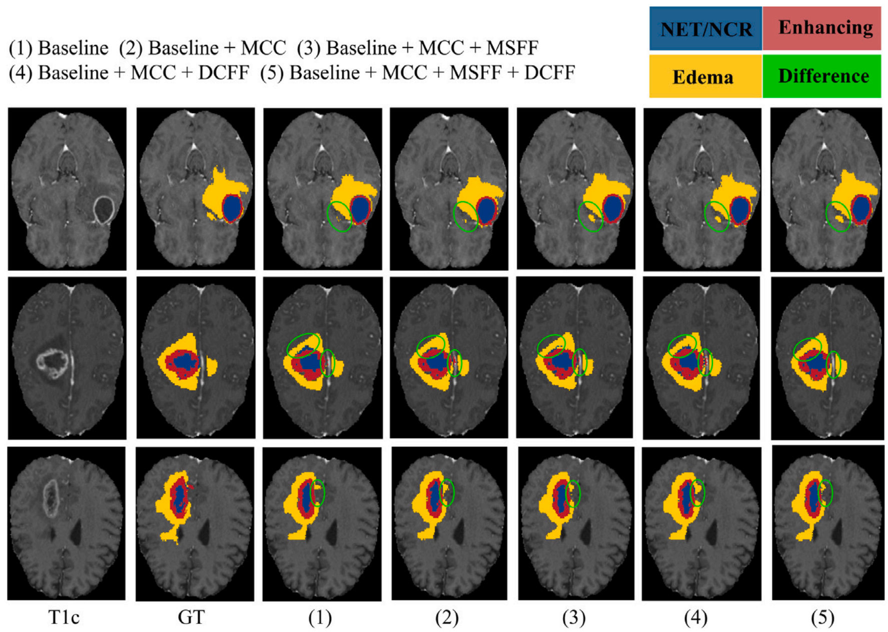

Fig. 6. Visualization of the segmentation results in the proposed method. The first column shows the MR modality. For simplicity, only the T1c modality is presented. The second column presents the ground truth. From the third to seventh columns present the predictions of (1) baseline (2) baseline + MCC (3) Baseline + MCC + MSFF (4) Baseline + MCC + DCFF (5) the proposed method. The green circle highlights the segmentation differences.

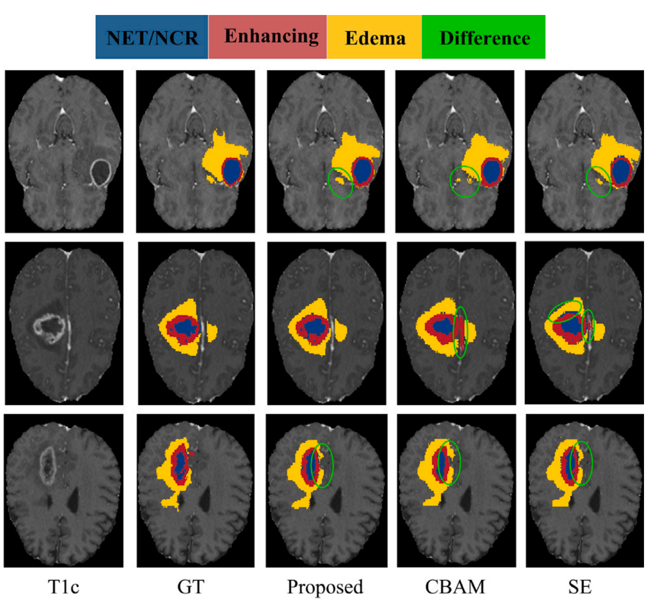

Fig. 7. Visualization of the segmentation results from different attention methods. The first column shows the MR modality. For simplicity, only the T1c modality is presented. The second column presents the ground truth. The third to fifth columns present the predictions of the proposed method, CBAM and SE. The green circle highlights the segmentation differences.

the interested target regions to improve the segmentation performance. Specifically, the MSFF can highlight the whole tumor region, and the DCFF can further distinguish the sub-tumor regions. With the assistance of the two blocks, we can finally obtain the discriminative fused features. On the other hand, from Fig. 8, it can be also seen that the proposed attention method achieves better feature representations than the two compared methods: CBAM and SE. For example, in the first and third cases, we can distinguish the sub-tumor regions from the fused feature. However, in the other two methods, only the whole tumor is detected. And in the second case, the proposed fused feature has clearer contour than the others. The reason is that the fusion module considers not only the multi-scale feature representations but also the channel-wise ones. It can learn more effective and important features to ensure the network can achieve better segmentation performance.

5. Conclusion

In this paper, an attentional deep neural network for brain tumor segmentation is proposed. First, considering the characteristics of multiple MR modalities, a modality-level cross-connection (MCC) is proposed to utilize the complementary information between the related modalities. In addition, to sufficiently learn high-quality feature information for segmentation, an attentional feature fusion module (AFFM) is proposed, consisting of two fusion blocks: a multi-scale spatial feature fusion (MSFF) block and a dual-path channel feature fusion (DCFF) block. The former can learn the multi-scale spatial contextual feature, the latter one aims to learn the channel-wise feature. The combination of these two blocks can enhance the feature learning ability of the network to improve the segmentation performance. The comprehensive experiments conducted on the BraTS 2018 dataset demonstrate the effectiveness of the proposed network. The proposed method achieves 86.5 in terms of Dice Score and 4.6 in terms of Hausdorff Distance on whole tumor, 87.0 in terms of Dice Score and 3.6 in terms of Hausdorff Distance on tumor core, 79.4 in terms of Dice Score and 2.5 in terms of Hausdorff Distance on enhancing tumor. In addition, it can achieve superior segmentation results and feature learning ability compared with state-of-the-art methods. Furthermore, the proposed fusion module can be easily integrated into other data fusion models and deep neural network architectures.

To our knowledge, there are some limitations of the proposed work. First, the segmentation accuracy can be further improved by exploiting the feature fusion method, for example, the recent multi-head attention mechanism can be exploited to enhance the ability of feature learning to improve the segmentation accuracy. Second, the proposed modality-level cross connections are utilized on two pairs of related modalities (T1-T1c, Flair-T2), in the future work, we plan to conduct more experiments to find other suitable combination modes. Third, I would like to develop a more efficient network to reduce computational complexity and improve segmentation accuracy.

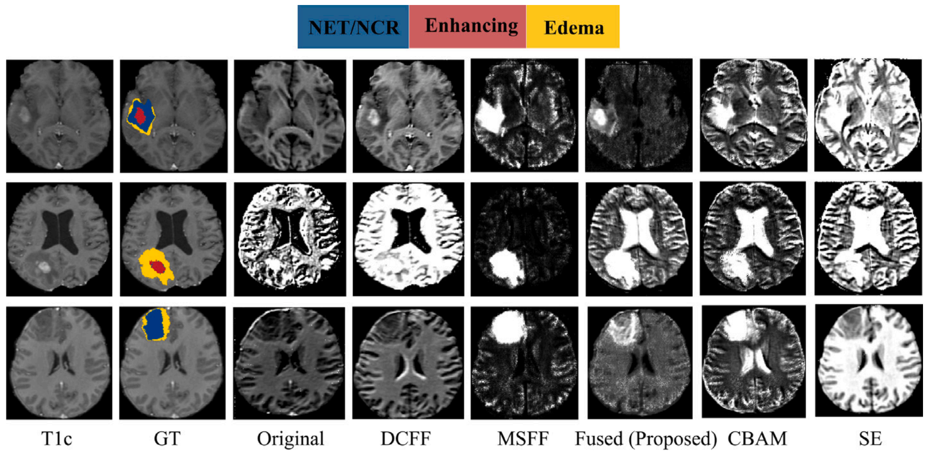

Fig. 8. Visualization of the feature maps from different methods. The first column shows the MR modality. For simplicity, only the T1c modality is presented. The second column presents the ground truth. From the third to sixth columns present the feature maps of the proposed method: original feature (before using AFFM), DCFF, MSFF and fused feature (before using AFFM). The last two columns present the feature maps after applying the CBAM, and SE methods.

CRediT authorship contribution statement

Tongxue Zhou: Conceptualization, Methodology, Writing and Editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgment

This work is supported by National Natural Science Foundation of China (No. 6220071321).

References

[1] T. Zhou, S. Canu, S. Ruan, Fusion based on attention mechanism and context constraint for multi-modal brain tumor segmentation, Comput. Med. Imaging Graph. 86 (2020) 101811.
[2] J. Liu, J. Zheng, G. Jiao, Transition Net: 2D backbone to segment 3D brain tumor, Biomed. Signal Process. Control 75 (2022) 103622.
[3] D.E. Cahall, G. Rasool, N.C. Bouaynaya, H.M. Fathallah-Shaykh, Inception modules enhance brain tumor segmentation, Front. Comput. Neurosci. 13 (2019) 44.
[4] M.P. Starmans, S.R. van der Voort, J.M.C. Tovar, J.F. Veenland, S. Klein, W.J. Niessen, Radiomics: data mining using quantitative medical image features, in: Handbook of Medical Image Computing and Computer Assisted Intervention, Elsevier, 2020, pp. 429–456.
[5] R. Raza, U.I. Bajwa, Y. Mehmood, M.W. Anwar, M.H. Jamal, dResU-Net: 3D deep residual U-Net based brain tumor segmentation from multimodal MRI, Biomed. Signal Process. Control (2022) 103861.
[6] Y. Wang, Z. Zhao, S. Hu, F. Chang, CLCU-Net: Cross-level connected U-shaped network with selective feature aggregation attention module for brain tumor segmentation, Comput. Methods Programs Biomed. 207 (2021) 106154.
[7] J. Lapuyade-Lahorgue, J.-H. Xue, S. Ruan, Segmenting multi-source images using hidden Markov fields with copula-based multivariate statistical distributions, IEEE Trans. Image Process. 26 (7) (2017) 3187–3195.
[8] C. Lian, S. Ruan, T. Denœux, H. Li, P. Vera, Joint tumor segmentation in PET-CT images using co-clustering and fusion based on belief functions, IEEE Trans. Image Process. 28 (2) (2018) 755–766.
[9] S. Bonte, I. Goethals, R. Van Holen, Machine learning based brain tumour segmentation on limited data using local texture and abnormality, Comput. Biol. Med. 98 (2018) 39–47.
[10] G. Chen, Q. Li, F. Shi, I. Rekik, Z. Pan, RFDCR: Automated brain lesion segmentation using cascaded random forests with dense conditional random fields, NeuroImage 211 (2020) 116620.
[11] K. Hu, Q. Gan, Y. Zhang, S. Deng, F. Xiao, W. Huang, C. Cao, X. Gao, Brain tumor segmentation using multi-cascaded convolutional neural networks and conditional random field, IEEE Access 7 (2019) 92615–92629.
[12] S. Krishnakumar, K. Manivannan, Effective segmentation and classification of brain tumor using rough K means algorithm and multi kernel SVM in MR images, J. Ambient Intell. Humaniz. Comput. 12 (6) (2021) 6751–6760.
[13] J. Nie, Z. Xue, T. Liu, G.S. Young, K. Setayesh, L. Guo, S.T. Wong, Automated brain tumor segmentation using spatial accuracy-weighted hidden Markov Random Field, Comput. Med. Imaging Graph. 33 (6) (2009) 431–441.
[14] T. Zhou, S. Canu, P. Vera, S. Ruan, Latent correlation representation learning for brain tumor segmentation with missing MRI modalities, IEEE Trans. Image Process. 30 (2021) 4263–4274.
[15] Y. Shen, M. Gao, Brain tumor segmentation on MRI with missing modalities, in: International Conference on Information Processing in Medical Imaging, Springer, 2019, pp. 417–428.
[16] A. Ahmadi, M. Kashefi, H. Shahrokhi, M.A. Nazari, Computer aided diagnosis system using deep convolutional neural networks for ADHD subtypes, Biomed. Signal Process. Control 63 (2021) 102227.
[17] M. Wainberg, D. Merico, A. Delong, B.J. Frey, Deep learning in biomedicine, Nature Biotechnol. 36 (9) (2018) 829–838.
[18] Z. Liu, L. Chen, L. Tong, F. Zhou, Z. Jiang, Q. Zhang, C. Shan, Y. Wang, X. Zhang, L. Li, et al., Deep learning based brain tumor segmentation: a survey, 2020, arXiv preprint arXiv:2007.09479.
[19] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, Proc. IEEE 86 (11) (1998) 2278–2324.
[20] J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 3431–3440.
[21] K. He, X. Zhang, S. Ren, J. Sun, Identity mappings in deep residual networks, in: European Conference on Computer Vision, Springer, 2016, pp. 630–645.
[22] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2015, pp. 234–241.
[23] Y. Hu, X. Liu, X. Wen, C. Niu, Y. Xia, Brain tumor segmentation on multimodal MR imaging using multi-level upsampling in decoder, in: International MICCAI Brainlesion Workshop, Springer, 2018, pp. 168–177.
[24] L. Weninger, O. Rippel, S. Koppers, D. Merhof, Segmentation of Brain Tumors in 3D-MRI Data and Patient Survival Prediction: Methods for the BraTS 2018 Challenge.
[25] H.-Y. Yang, J. Yang, Automatic brain tumor segmentation with contour aware residual network and adversarial training, in: International MICCAI Brainlesion Workshop, Springer, 2018, pp. 267–278.
[26] H. Yao, X. Zhou, X. Zhang, Automatic segmentation of brain tumor using 3D SE-inception networks with residual connections, in: International MICCAI Brainlesion Workshop, Springer, 2018, pp. 346–357.
[27] M. Ghaffari, G. Samarasinghe, M. Jameson, F. Aly, L. Holloway, P. Chlap, E.-S. Koh, A. Sowmya, R. Oliver, Automated post-operative brain tumour segmentation: A deep learning model based on transfer learning from pre-operative images, Magn. Reson. Imaging 86 (2022) 28–36.
[28] Q. Ma, S. Zhou, C. Li, F. Liu, Y. Liu, M. Hou, Y. Zhang, DGRUnit: Dual graph reasoning unit for brain tumor segmentation, Comput. Biol. Med. 149 (2022) 106079.

[29] S. Woo, J. Park, J.-Y. Lee, I.S. Kweon, Cbam: Convolutional block attention module, in: Proceedings of the European Conference on Computer Vision (ECCV), 2018, pp. 3–19.
[30] L. Itti, C. Koch, E. Niebur, A model of saliency-based visual attention for rapid scene analysis, IEEE Trans. Pattern Anal. Mach. Intell. 20 (11) (1998) 1254–1259.
[31] R.A. Rensink, The dynamic representation of scenes, Vis. Cogn. 7 (1–3) (2000) 17–42.
[32] J. Hu, L. Shen, G. Sun, Squeeze-and-excitation networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 7132–7141.
[33] F. Isensee, P. Kickingereder, W. Wick, M. Bendszus, K.H. Maier-Hein, Brain tumor segmentation and radiomics survival prediction: Contribution to the brats 2017 challenge, in: International MICCAI Brainlesion Workshop, Springer, 2017, pp. 287–297.
[34] G. Wang, W. Li, S. Ourselin, T. Vercauteren, Automatic brain tumor segmentation using cascaded anisotropic convolutional neural networks, in: International MICCAI Brainlesion Workshop, Springer, 2017, pp. 178–190.
[35] K. Kamnitsas, C. Ledig, V.F. Newcombe, J.P. Simpson, A.D. Kane, D.K. Menon, D. Rueckert, B. Glocker, Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation, Med. Image Anal. 36 (2017) 61–78.
[36] K. Kamnitsas, W. Bai, E. Ferrante, S. McDonagh, M. Sinclair, N. Pawlowski, M. Rajchl, M. Lee, B. Kainz, D. Rueckert, et al., Ensembles of multiple models and architectures for robust brain tumour segmentation, in: International MICCAI Brainlesion Workshop, Springer, 2017, pp. 450–462.
[37] S. Bakas, M. Reyes, A. Jakab, S. Bauer, M. Rempfler, A. Crimi, R.T. Shinohara, C. Berger, S.M. Ha, M. Rozycki, et al., Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the BRATS challenge, 2018, arXiv preprint arXiv:1811.02629.
[38] B.B. Avants, N. Tustison, G. Song, Advanced normalization tools (ANTS), Insight J. 2 (2009) 1–35.
[39] A. Popli, M. Agarwal, G. Pillai, Automatic brain tumor segmentation using u-net based 3d fully convolutional network, in: Pre-Conference Proceedings of the 7th MICCAI BraTS Challenge, 2018, pp. 374–382.
[40] X. Hu, H. Li, Y. Zhao, C. Dong, B.H. Menze, M. Piraud, Hierarchical multi-class segmentation of glioma images using networks with multi-level activation function, in: International MICCAI Brainlesion Workshop, Springer, 2018, pp. 116–127.
[41] J. Serrano-Rubio, R. Everson, H. Hutt, Brain tumour segmentation method based on sparse feature vectors, in: Pre-Conference Proceedings of the 7th MICCAI BraTS Challenge, Granada, Spain, 2018, pp. 420–427.
[42] J. Zhang, Z. Jiang, J. Dong, Y. Hou, B. Liu, Attention gate resU-Net for automatic MRI brain tumor segmentation, IEEE Access 8 (2020) 58533–58545.
[43] R. Mehta, T. Arbel, 3D U-Net for brain tumour segmentation, in: International MICCAI Brainlesion Workshop, Springer, 2018, pp. 254–266.
[44] A. Kermi, I. Mahmoudi, M.T. Khadir, Deep convolutional neural networks using U-Net for automatic brain tumor segmentation in multimodal MRI volumes, in: International MICCAI Brainlesion Workshop, Springer, 2018, pp. 37–48.
[45] N.M. Aboelenein, P. Songhao, A. Koubaa, A. Noor, A. Afifi, HTTU-Net: Hybrid Two Track U-Net for automatic brain tumor segmentation, IEEE Access 8 (2020) 101406–101415.
[46] W. Chen, B. Liu, S. Peng, J. Sun, X. Qiao, S3D-UNet: separable 3D U-Net for brain tumor segmentation, in: International MICCAI Brainlesion Workshop, Springer, 2018, pp. 358–368.