symmetry

MDPI

Article

AResU-Net: Attention Residual U-Net for Brain Tumor Segmentation

Jianxin Zhang [1,2,†], Xiaogang Lv [1,†], Hengbo Zhang [2] and Bin Liu [3,4,*]

1 Key Lab of Advanced Design and Intelligent Computing (Ministry of Education), Dalian University, Dalian 116622, China; jxzhang@dlnu.edu.cn (J.Z.); lvxiaogang0428@163.com (X.L.)
2 School of Computer Science and Engineering, Dalian Minzu University, Dalian 116600, China; zhanghengbo@dlnu.edu.cn
3 International School of Information Science and Engineering (DUT-RUISE), Dalian University of Technology, Dalian 116620, China
4 Key Lab of Ubiquitous Network and Service Software of Liaoning Province, Dalian University of Technology, Dalian 116620, China
* Correspondence: liubin@dlut.edu.cn
† These authors contributed equally to this work.

Received: 3 March 2020; Accepted: 7 April 2020; Published: 2 May 2020

Abstract: Automatic segmentation of brain tumors from magnetic resonance imaging (MRI) is a challenging task due to the uneven, irregular and unstructured size and shape of tumors. Recently, brain tumor segmentation methods based on the symmetric U-Net architecture have achieved favorable performance. Meanwhile, the effectiveness of enhancing local responses for feature extraction and restoration has also been shown in recent works, which may encourage the better performance of the brain tumor segmentation problem. Inspired by this, we try to introduce the attention mechanism into the existing U-Net architecture to explore the effects of local important responses on this task. More specifically, we propose an end-to-end 2D brain tumor segmentation network, i.e., attention residual U-Net (AResU-Net), which simultaneously embeds attention mechanism and residual units into U-Net for the further performance improvement of brain tumor segmentation. AResU-Net adds a series of attention units among corresponding down-sampling and up-sampling processes, and it adaptively rescales features to effectively enhance local responses of down-sampling residual features utilized for the feature recovery of the following up-sampling process. We extensively evaluate AResU-Net on two MRI brain tumor segmentation benchmarks of BraTS 2017 and BraTS 2018 datasets. Experiment results illustrate that the proposed AResU-Net outperforms its baselines and achieves comparable performance with typical brain tumor segmentation methods.

Keywords: brain tumor segmentation; MRI; deep learning; attention mechanism; AResU-Net

1. Introduction

Brain tumors are abnormal cells growing in human brains, regarded as a type of common neurological disease, which is harmful to human health extremely [1]. As an important way to assist in the diagnosis and treatment of brain tumors, automatic brain tumor segmentation performed on brain magnetic resonance images is of great significance in clinical medicine [2]. The most common malignant brain tumor is glioma, which can be further divided into high-grade glioma (HGG) and low-grade glioma (LGG) [3]. Magnetic resonance imaging (MRI) is a typical non-invasive imaging technology, which can produce high-quality brain images without damage and skull artifacts, and is regarded as the main technical means for the diagnosis and treatment of brain tumors. With multimodal brain images, doctors can perform quantitative analysis of brain tumors to develop the best diagnosis and

treatment plan for patients [4]. However, due to changes in brain tumor size, shape, and structure, as well as the influence of neighboring tissues and device noise, it is very challenging to localize and segment tumors from MRI brain images accurately. Fortunately, with the continuous breakthrough of the deep learning technology, automatic image segmentation methods based on deep learning have also achieved great development. In 2014, Long et al. [5] proposed a novel end-to-end fully convolution network (FCN) for natural image segmentation, which injected vitality into the natural image segmentation field [6–13] and was quickly introduced to resolve the brain tumor segmentation problem. Motivated by the FCN model, Ronneberger et al. [14] further presented a symmetric fully convolutional network named U-Net for medical image segmentation. U-Net consists of a contracting path that contains several convolutional layers for down-sampling input images, an expanding path for up-sampling deep feature maps, and a skip connection to merge cropped feature maps from the encoder-decoder network, largely improving segmentation performance of medical images. Nowadays, U-Net has already become a milestone of resolving the brain tumor segmentation task. Meanwhile, a variety of improved U-Net methods, such as ResU-Net [15] and Ensemble Net [16], have also been put forward to gain superior performance for the brain tumor segmentation problem.

Besides, attention mechanisms have been proved effective in capturing long-range dependencies and important responses in the field of computer vision. Wang et al. [17] proposed non-local neural networks (NL-Nets) to capture long-range dependencies through aggregating query-specific global context to each query position, achieving favorable results on video classification and image recognition applications. As a concurrent work, Hu et al. presented squeeze-and-excitation networks (SENets) [18] to address global spatial information of various channels in a soft-attention manner, i.e., learning and re-scaling scaling factors for channels. As an effectiveness and efficiency model, SENet was quickly introduced to boost the image segmentation performance [19–21]. Additionally, attention mechanisms have also been successfully introduced into the field of medical image segmentation [22,23]. Since U-Net has to gradually recover down-sampling image caused by the inherent pooling and stridden convolution, attention mechanisms can better bridge information flow from deep layers to shallow layers and guide the learning of up-sampling. Therefore, for the existing U-Net model on brain tumor segmentation task, it may lead to a higher segmentation accuracy by enhancing its capacity of capturing local responses. Motivated by this, we try to embed the attention mechanism to explore the effects of local responses for brain tumor segmentation in this work. Specifically, an end-to-end 2D brain tumor segmentation network is presented via simultaneously embedding attention mechanism and residual units into U-Net for further performance improvement. The main contributions of the work are summarized as follows: (1) In this work, an end-to-end 2D attention residual U-Net (AResU-Net) is proposed to address the brain tumor segmentation task, which successfully embeds the attention and squeeze-excitation (ASE) unit and residual block into the U-Net network structure. The overall architecture of AResU-Net can be shown in Figure 1. (2) AResU-Net adds a series of ASE units on skip connections to adaptively enhance local responses of down-sampling features utilized for the feature recovery of the following up-sampling process, which helps to reduce the semantic gap between down-sampling and up-sampling stages. (3) Experimental results on two commonly used brain tumor segmentation datasets, i.e., BraTS 2017 and BraTS 2018, illustrate that AResU-Net can gain better performance than its baseline, as well as several typical 2D and 3D brain tumor segmentation methods.

The rest of the paper is organized as follows: Section 2 provides an overview of related works. Section 3 introduces the details of the proposed AResU-Net method. Section 4 presents the experiments and analysis, and conclusions are given in Section 5.

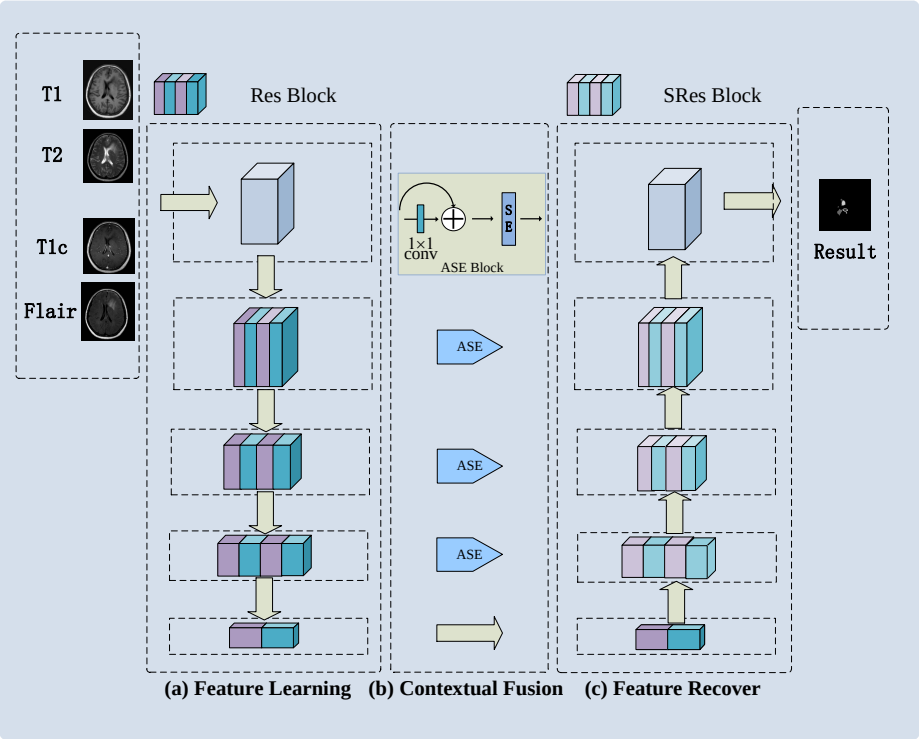

Figure 1. The overall architecture of the proposed attention residual U-Net (AResU-Net), which mainly consists of three modules, i.e., the feature learning module, the contextual fusion module and the feature recovery module. The intermediate features among corresponding down-sampling layers in feature learning module and up-sampling layers in feature recovery module are adaptively refined to enhance the local responses of brain tumors, resulting in more favorable segmentation result for the whole network.

2. Related Work

In this section, we review literature in three domains related to our AResU-Net model, including patch-wise brain tumor segmentation networks, brain tumor semantic segmentation networks and attention mechanisms.

Patch-wise Segmentation. Recently, a number of deep networks have been developed in the field of brain tumor segmentation, which has achieved significant performance improvement over traditional methods. Among them, patch-wise based brain tumor segmentation networks, as representative works proposed early, are trained on small patches with labels to correctly distinguish brain tumors from normal tissues. To achieve favorable performance, researchers have designed various modules to introduce more contextual contact information among different slices into networks. Havaei et al. [24] embedded multi-scale and multi-path modules into a 2D network structure to capture richer contextual information. Instead of utilizing 2D convolutional neural networks (CNNs) as the backbone, Urban et al. presented a patch-wise based brain tumor segmentation network by using 3D CNN [25] architecture. Pereira et al. [26] explored small 3×3 convolutional kernels to design a deeper network, achieving more non-linearities and effectively reducing over-fitting problem. To further boost the segmentation performance, Kamnitsas et al. [27] adopted a dual pathway 3D CNN model with the dense structure for the brain tumor segmentation task, which also performed multi-scale processing on input images and post-processing on result images by using conditional random field (CRF). This work finally gained the first place in the BraTS 2015 competition. In addition, Zhao et al. [28] integrated a fully convolutional neural network and CRF, training three 2D patch-wise models from axial, sagittal and coronal views with a voting-based fusion strategy to finish the brain tumor segmentation.

Semantic-wise Segmentation. The semantic segmentation model classifies each pixel of the whole brain image into an assigned label to complete brain tumor segmentation. Most of the semantic segmentation models for the brain tumor segmentation task are based on U-Net architecture proposed by Ronneberger et al. [14], which has also been widely applied to other medical image segmentation tasks. U-Net contains a contracting path to capture context information and an expanding path that ensures accurate location, largely improving the performance of medical image segmentation task. Dong et al. [29] developed a 2D U-Net based brain tumor segmentation network and employed real-time data augmentation to refine its segmentation performance. Kong et al. [30] embedded a feature pyramid module into U-Net architecture to integrate multi-scale semantic and location information, which effectively improved the segmentation accuracy. Additionally, cascade strategy, dense block, dilated convolution and up skip connection have also been introduced into U-Net architecture [31–35], continuously optimizing network structures to pursue more accurate tumor segmentation results.

Attention mechanism. Attention mechanisms have been increasingly applied to a variety of computer vision tasks, which can be roughly divided into two categories in terms of purposes. The first purpose is to focus on long-range dependencies. NL-Net, as a representative work presented by Wang et al. [17], can generate new spatial feature responses through the weighted sum of all responses to capture long-range dependency of spatial dimension. Based on the NL-Net model, Zhao et al. [21] designed location-sensitive NL to learn long-range context, which achieved impressive segmentation results. Fu et al. [36] proposed dual attention modules that consisted of spatial and channel attention for semantic segmentation, in which the spatial attention was similar to the non-local(NL) operation in NL-Net and channel attention followed the same idea. Moreover, Zhang et al. [37] extended NL with a prior distribution and built an ensemble of NLs with weights to further improve segmentation performance. Another purpose of attention mechanism is to learn the scaling factors of each channel for feature maps. A typical work is SENet [18] that focuses on the channel relationship and performs dynamic channel-wise feature recalibration to enhance feature expression. Instead of using simple global average pooling to summarize statistics of features, EncNet [19] employed VLAD [38] encoder to collect them, and the output of the encoder was also passed through fully connected layers to get channel-wise factors. In the up-sampling stage of image segmentation, DFN [20] and PAN [39] fed the features of deep layers with stronger semantics into SE-like attention block to provide high-level category information used to precisely recovery details. With such a block, the features from deep and shallow layers were well combined to enhance the learning of features with larger resolution but weaker semantic, helpful for restructuring original image resolution. Inspired by the success of attention, we explore it to learn channel-wise factors to augment channel responses selectively.

3. Method

In this section, we firstly give a brief introduction to the data preprocessing utilized in this work. Then, the details of AResU-Net, including the feature learning module, the contextual fusion module and feature recovery module, are further described. Finally, we introduce the loss function adopted in AResU-Net.

3.1. Data Preprocessing

MRI brain tumor segmentation is a challenging task in the field of medical image analysis due to the complexity of brain structure and biological tissue, as well as the influence of imaging quality. Though deep learning-based models are robust to noise, data processing is still an important step to improve the brain tumor segmentation performance. In this work, we mainly utilize two multi-modality MRI brain scan datasets, i.e., BraTS 2017 and BraTS 2018 datasets. To better fit for the network, we also perform data processing on the original MRI brain tumor images, whose overall steps can be shown in Figure 2. As demonstrated in this figure, most invalid pixels are firstly removed from the original 3D brain image data, and then each refined 3D image data is sliced into a number

of 2D images. Then, several patches with size of 128 × 128 are extracted from each 2D slice image, followed by a z-score normalization, i.e., zero-mean normalization, operation performed on each 128 × 128 patch. The z-score normalization is calculated as following:

$$z' = \frac{z - \mu}{\delta}, \tag{1}$$

where z and z' are the input image and normalized output image, respectively. μ is the mean value of the input image, and δ is the standard deviation of the input image. After these steps, images with normalized multi-center and inhomogeneous intensity can be achieved. Finally, to mitigate the effects of overfitting problem, the Gaussian noise reduction [40] is further performed on each normalized patch, whose result can be taken as the input of the segmentation network.

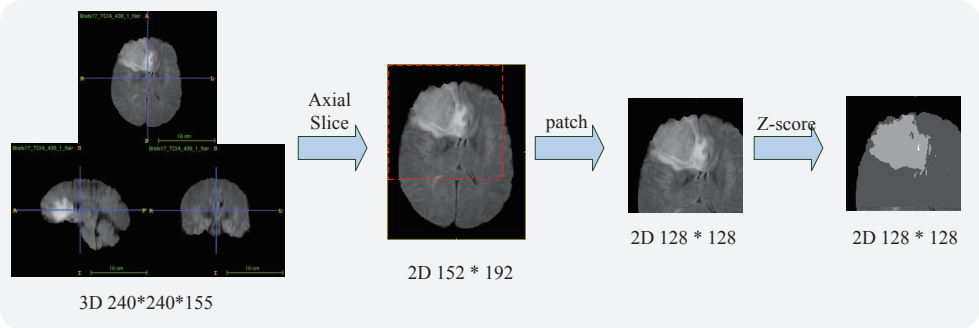

Figure 2. Overall steps of brain tumor data preprocessing.

3.2. AResU-Net

The AResU-Net model is an end-to-end brain tumor segmentation network through simultaneously embedding attention mechanism and residual block into the existing U-Net architecture, which is suitable for processing medical image stabilization with limited training data. AResU-Net utilizes the training strategy of 2D convolutional neural network to achieve the multi-modal pixel-level classification. The size of input image data for the network is 128 × 128 × 4, i.e., the image size is 128 and 128, and the number of channels is 4.

As shown in Figure 1, our brain tumor segmentation network can be seen as a classical encoder-decoder structure that preserves high-level information in deep layers by combining shallow features and deep features. The feature learning (encoder) module includes three residual (Res) blocks and a bottom convolutional layer with dropout function to obtain high-level features of context semantic information. The feature recovery (decoder) module adopts three similar up-sampling residual (SRes) blocks for precisely positioning and feature recovery. In order to gain richer low-level and high-level information, the attention and squeeze excitation (ASE) block is added as the horizontal connection, effectively enhancing feature information representations between down-sampling features and up-sampling information. Finally, we integrate the softmax layer to obtain the final segmentation results of multiple classes. By integrating these blocks into a unified architecture, the AResU-Net network will capture richer information and achieve stable segmentation results.

3.2.1. Feature Learning Module

Feature learning module consists of the down-sampling process through a variety of residual(Res) blocks whose basic architecture can be illustrated in Figure 3. As given in Figure 3, the residual block is achieved by a shortcut connection element-wise addition operation, which greatly improves the training speed and accuracy without any extra parameters. Each block of encoder contains two convolution layers and one pooling layer. In our work, we mainly adopt ResNet-34 [41] for the feature

learning module, which includes two same units, and each unit is composed of a regularization, an activation function, and a 3×3 convolutional layer. The residual operation [41] can be denoted as

$$y = F(x, W_i) + x, \tag{2}$$

where x and y are the input and output vectors of related layers, and $F(x, W_i)$ is the mapping function for the residual path. The result of $F(x, W_i)$ should have the same dimension as x. Residual block adds a shortcut mechanism to avoid the gradient vanishing and accelerate the network convergence, integrating the rough local and global features. Considering the size of brain tumors, we adopt three down-sampling units to achieve feature map learning, and the final image size is 16×16 pixels after the whole down-sampling process.

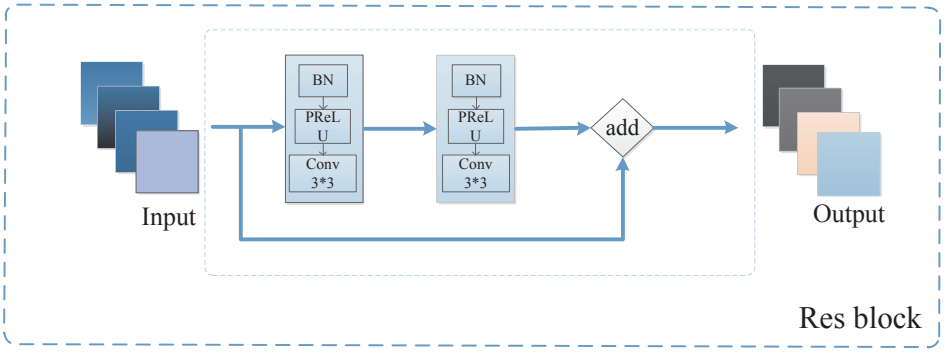

Figure 3. The structure of Res block. An individual Res block includes two same convolutional units, a shortcut connection and an element-wise addition operation, where each convolutional unit is composed of a regularization, an activation function and a 3×3 convolutional layer.

3.2.2. Contextual Fusion Module

Here, we combine the attention mechanism with the shortcut idea to construct a novel contextual fusion module, instead of the direct connection given in the original U-Net architecture. The contextual fusion module takes advantage of preserving some detailed information from an encoder to a decoder that helps to recovery the feature information loss. As shown in Figure 1, the contextual fusion module is mainly composed of a series of attention and squeeze-excitation (ASE) blocks. The ASE block encodes dependencies of each channel through a fully connected operation, embedding global spatial information into each channel vector. The details of ASE block can be illustrated in Figure 4. The ASE block allows the network to give different attention to various channels based on the importance of feature maps. To achieve more robust feature information, we firstly exploit a regularization, PReLU activation, 1×1 convolution layer and sum operation on corresponding input maps, which can be computed as

$$F_{tr} = F(\beta, \theta, x) + x, \tag{3}$$

where x and F_{tr} are the input and output of related layers, respectively. β is the normalized parameter, and θ is the activation function PReLU.

Then, the corresponding output feature information F_{tr} will pass through the squeeze operation implemented by a global average pooling layer to aggregate global information for each channel of the whole image. To fully capture channel-wise dependencies, the output feature maps via squeeze operation are fed to an excitation operation that consists of two fully connected layers around the non-linearity operations interaction between channels, i.e., the ReLU and the sigmoid activation function. Finally, the weight vector is reshaped into $(1, 1, 1, C)$ size, where C is the number of feature

maps, applied to each feature map by multiplication operation returning to the channel dimension of the transformation output $F_{scale}(F_{tr}, F_{eq})$. The corresponding process [18] can be denoted as follows:

$$F_{sq}(F_{tr}) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} F_{tr}(i, j), \tag{4}$$

$$F_{eq}(F_{sq}, W, r) = \rho(W_2 \sigma(W_1, \frac{F_{sq}}{r})), \tag{5}$$

$$F_{scale}(F_{tr}, F_{eq}) = F_{tr} \cdot F_{eq}. \tag{6}$$

In the above equations, F_{sq}, F_{eq} and F_{scale} denote squeeze global spatial information, channel-wise dependencies information and output vectors of related layers, respectively. W_1 and W_2 are the dimensionality-reduction layer and the dimensionality-increasing layer, respectively. r represents the reduction ratio, ρ and σ are parameters of non-linearity activation. In addition, H and W are spatial dimensions.

Therefore, ASE block emphasizes useful feature information and suppresses redundant feature information through variable weights obtained by attention mechanism, focusing on local details for achieving better results. By integrating several ASE blocks, the contextual fusion module well utilizes the attention mechanism and shortcut idea in the architecture, which will pay more attention to the important feature maps.

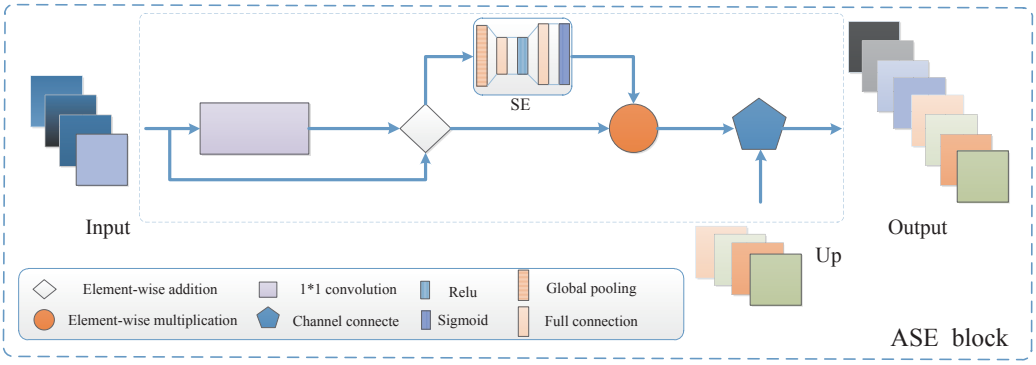

Figure 4. The structure of attention and squeeze-excitation (ASE) block. The ASE block first exploits the PReLU activation, 1×1 convolution and sum operation on the corresponding input information, then the output feature information will be squeezed, which is implemented by a global average pooling layer to aggregate global information for each channel of the entire image.

3.2.3. Feature Recovery Module

The feature recovery module is designed to restore high-level image features extracted from the feature learning module and the contextual fusion module. The bilinear interpolation and deconvolution are two common operations for decoder structures, where the bilinear interpolation operation increases the image size with linear interpolation while the deconvolution operation employs convolution operation to enlarge the image information. We adopt an efficient spatial residual (SRes) block that is similar to the residual structure [41] for enhancing the decoding performance, and its basic structure can be shown in Figure 5. The SRes block mainly includes two convolutional units as given in Res block, a 1×1 convolution concatenation with input feature maps as shortcut connection, and element-wise addition operation. In the SRes block, the feature decoder module outputs a mask whose size is the same as that of the original input. As shown in Figure 1, we can achieve the richer feature information recovery through deconvolution based on the combination of several SRes blocks.

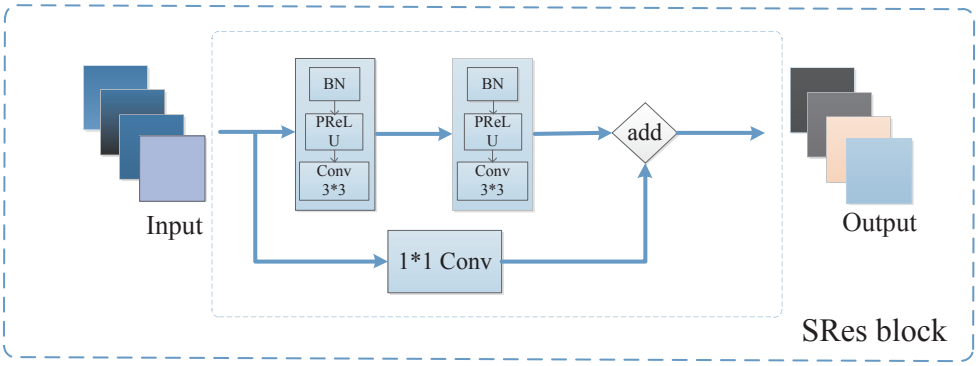

Figure 5. The structure of SRes block. The up-sampling residual (SRes) block mainly includes two convolutional units as given in Res block, a 1×1 concatenation with input feature maps as shortcut connection, and an element-wise addition operation.

3.3. Loss Function

The MRI brain tumor segmentation task usually exhibits a severe class imbalance problem. Table 1 illustrates the distribution of sub-classes in the training data of BraTS dataset, approximately 98.46% of voxels belong to either the healthy tissue or the black surrounding area, labeled as background. However, the edema and the enhancing tumor only cover 1.02% and 0.29% voxels of the whole data, respectively. Moreover, the necrotic and non-enhancing tumors occupy the lowest volume among all categories, which has a rate of only 0.23%. Although the data pre-processing alleviates this problem to some extent, it still severely affects the segmentation accuracy. Here, we employ a combined loss function [15] that integrates weight cross-entropy (WCE) and generalized dice loss(GDL) to address this class balance problem as below.

$$Loss = L_{GDL} + L_{WCE}, \tag{7}$$

where L_{GDL} and L_{WCE} respectively represent the generalized dice loss and the weighted cross entropy loss, which are correspondingly defined as Equations (8) and (9).

$$L_{GDL} = 1 - 2\frac{\sum_i^L w_i g_i p_i}{\sum_i^L w_i \left(g_i + p_i\right)} \tag{8}$$

$$L_{WCE} = -\sum_i^L w_i g_i \log\left(p_i\right), \tag{9}$$

where L is the total number of labels, and w_i denotes the weight assigned to the ith label. As for the generalized dice loss, p_i and g_i represent the pixel value of the segmented binary image and the binary ground truth image, respectively.

Table 1. The distribution of sub-classes in the training data of BraTS dataset.

Class	Rate %
Background	98.46
edema	1.02
enhancing tumor	0.29
necrotic and non-enhancing tumor	0.23

4. Experiments and Results

In this section, we execute experiments on two brain tumor benchmarks to evaluate the effectiveness of AResU-Net. We firstly describe the details of the employed datasets for model

evaluation. Then, experimental settings are further introduced, followed by a simple description of evaluation metric. Finally, compared experiment results on two brain tumor benchmarks are given and analyzed.

4.1. Datasets

In this work, we mainly adopt the public BraTS 2017 and BraTS 2018 [4,42] brain tumor segmentation datasets for the performance evaluation. These two datasets are released by the Multimodal Brain Tumor Segmentation Challenge (BraTS) that run in conjunction with the International Conference On Medical Image Computing and Computer-Assisted Intervention (MICCAI) 2017 and 2018. The BraTS 2017 dataset consists of training dataset, validation dataset and test dataset, and each sample has four different modalities, i.e., fluid-attenuated inversion recovery (FLAIR), T1 weighting (T1), T1 weighted contrast enhancement (T1ce), and T2 weighting (T2). However, only the training dataset is available to the public among the three datasets. The BraTS 2017 training dataset includes 285 patient samples, in which 210 samples are from high-grade glioma (HGG) patients and the remaining 75 samples belong to low-grade glioma (LGG) patients. All the images are skull-striped and re-sampled to an isotropic $1 \times 1 \times 1\,\mathrm{mm}^3$ resolution with image size of $240 \times 240 \times 155$, and the four sequences from the same patient have been co-registered. The ground truth of each image is labeled based on manual segmentation results given by experts. The basic labels include four types named as the GD-enhanced tumor (labeled as 4), the peritumoral edema (labeled as 2), the necrotic and non-enhancing tumor (labeled as 1), and the healthy pixel (labeled as 0). According to these labels, the whole tumor (Whole, the combined areas of labels 1, 2, 4), core tumor (Core, the combined areas of labels 1, 4), and Enhancing tumor (Enhancing, the area of label 4) are further constructed. Two typical samples from this dataset can be illustrated as. Figure 6. Besides, BraTS 2018 dataset shares the same training data with BraTS 2017 dataset but has different validation and testing datasets.

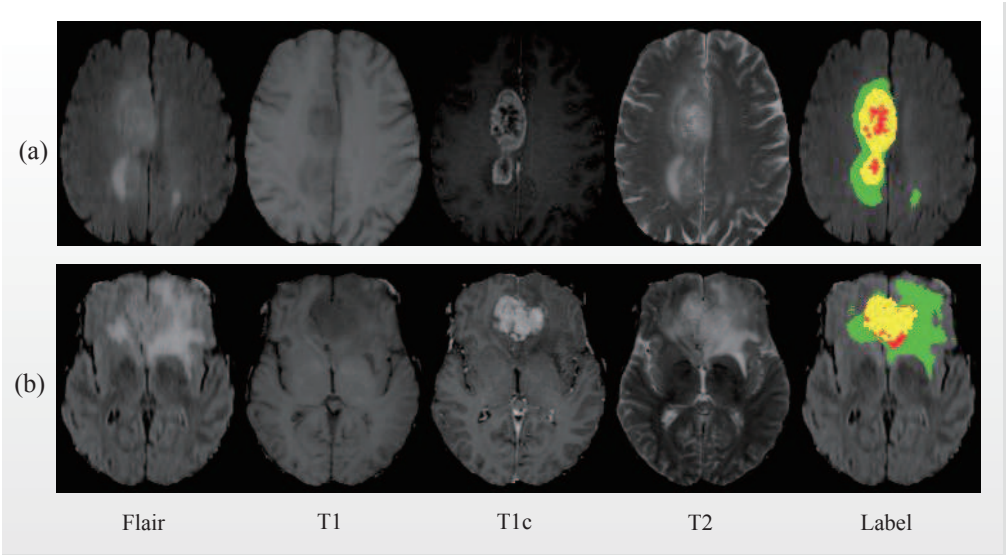

Figure 6. The legend shows the multimodal and label plots for two cases (**a**,**b**), from left to right are Flair, T1, T1ce(T1c), T2 and the Ground truth(Label). For the two rightmost images, each color represents a tumor class: red-necrosis and non-enhancing, green-edema and yellow-enhancing tumor.

4.2. Experimental Settings

The AResU-Net is conducted in Keras 2.2.4 using the Tensorflow [43] 1.5.0 backend, running on a PC equipped with 64 GB RAM and a single NVIDIA GTX 1080 GPU. Stochastic gradient descent (SGD) algorithm is employed as an optimizer with an initial learning rate of 0.085, a momentum of 0.95, and a weight decay of $5e^{-6}$. Besides, we utilize the patch-wise and the weight dice loss function to obtain a superior network model, and the size of each input patch is $128 \times 128 \times 4$ pixels. All networks

are trained from scratch with a batch size of 10 for 5 epochs. For fair comparisons, we report results of methods marked with * through publicly released codes of authors and try our best to fine-tune their parameters.

4.3. Evaluation Metric

We adopt the commonly used Dice score as the evaluation metric to estimate the given AResU-Net model. Dice score measures the rate of overlap area between the manual segmentation result and the automatic segmentation result, which can be computed by the following equation:

$$Dice = \frac{2TP}{FN + FP + 2TP}. \tag{10}$$

In the above equation, TP, FP and FN represent true positive, false positive and false negative prediction, respectively.

4.4. Experiment Results

4.4.1. Experiment Results on the BraTS 2017 Dataset

The first experiment is implemented on the BraTS 2017 HGG dataset. In this experiment, 80% of the BraTS 2017 HGG training cases (168 HGG cases) are used to train brain tumor segmentation networks, and the remaining 42 HGG cases are for testing. We compare AResU-Net with its baseline of U-Net [29], as well as four recently proposed networks of FCNN, ResU-Net, Densely CNN and the CNN [15,26,28,44]. The compared experiment results are listed in Table 2. As shown in Table 2, for the CNN and the FCNN models, they employ 2D CNN models with 33×33 patches as inputs to predict center voxel. In addition, the FCNN model additionally applies conditional random field as post-process to improve the prediction performance. Without utilizing any post-processing strategy, AResU-Net respectively achieves mean Dice scores of 0.892, 0.853 and 0.825 on the whole tumor, core tumor and enhancing tumor. It obtains 6.1%, 5.2% and 7.5% gains over its baseline of U-Net on the segmentation of these three areas, illuminating a significant accuracy improvement. Meanwhile, it respectively outperforms ResU-Net 1.2%, 0.3% and 0.5% on the whole tumor, core tumor and enhancing tumor. Additionally, compared with the FCNN, the CNN and Densely CNN, AResU-Net also obtains the highest Dice score on the whole tumor and enhancing tumor segmentation. These results demonstrate the effectiveness of our model for the brain tumor segmentation task.

Table 2. Compared experiment results on the BraTS 2017 HGG dataset.

Methods	Whole	Core	Enhancing
U-Net [29]	0.831	0.801	0.750
ResU-Net * [15]	0.880	0.850	0.820
FCNN [28]	0.865	**0.864**	0.816
Densely CNN [44]	0.720	0.830	0.810
CNN [26]	0.840	0.720	0.620
AResU-Net (ours)	**0.892**	0.853	**0.825**

The second experiment is conducted on the whole BraTS 2017 training data. In this experiment, we choose 80% of images for training and the rest images are used for testing, which means that brain tumor images from 228 patients construct the training set and images from the rest 57 cases constitute the testing set. We compare AResU-Net with its baseline of U-Net [29], as well as four recently proposed networks including SegNet, PSP-Net, NovelNet and ResU-Net [15,34,45]. The compared experiment results are reported in Table 3. As given in Table 3, AResU-Net achieves mean dice coefficients of 0.881, 0.780 and 0.719 for the whole tumor, core tumor and enhancing tumor segmentation, respectively. Compared with the basic U-Net, AResU-Net outperforms it by 1.10%, 1.80% and 1.60% gains on the

whole tumor, core tumor and enhancing tumor, respectively. AResU-Net is also superior to the other networks in all of the three areas. Moreover, some examples of visual comparison are also given in Figure 7. Overall, these results demonstrate that the attention mechanism helps improve the accuracy of the brain tumor segmentation task.

Table 3. Compared experiment results on the BraTS 2017 dataset (57 MRI scans).

Methods	Whole	Core	Enhancing
U-Net	0.870	0.762	0.700
SegNet [45]	0.833	0.703	0.496
PSPNet [34]	0.809	0.701	0.554
NovelNet [34]	0.876	0.763	0.642
ResU-Net [15]	0.873	0.768	0.716
AResU-Net (ours)	**0.881**	**0.780**	**0.719**

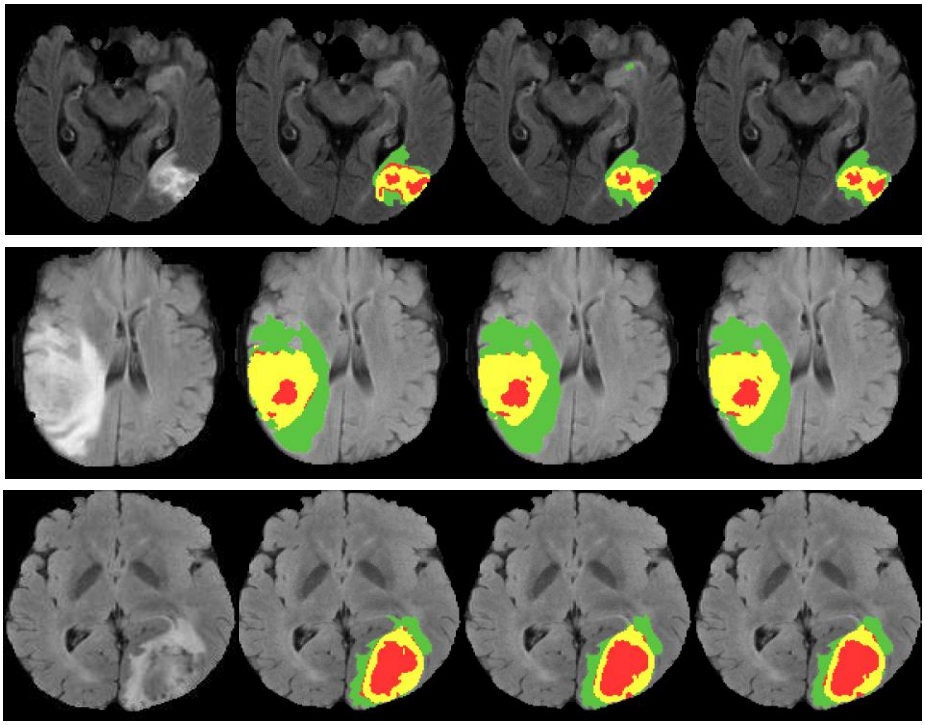

Figure 7. Qualitative comparison results of three sample images. From left to right: image, ground truth (label), U-Net and AResU-Net segmentation method.

4.4.2. Experiment Results on the BraTS 2018 Dataset

To further evaluate AResU-Net, we also perform an experiment on the BraTS 2018 dataset. In this experiment, 285 samples from the 2018 BraTS training dataset are adopted for training and 66 subjects from the validation dataset are utilized for testing. We compare AResU-Net with the baseline of U-Net and other networks including ResU-Net, Ensemble Net 3DU-Net, S3DU-Net, TTA and MMC [15,16,46–49]. Table 4 shows the compared experiment results. From Table 4 we can see that AResU-Net achieves average dice scores of 0.876, 0.810 and 0.773 on the whole tumor, core tumor and enhancing tumor segmentation, respectively. Compared with the baseline of U-Net, AResU-Net gains 1.60%, 2.00% and 0.60% performance improvement on the whole tumor, core tumor and enhancing tumor segmentation, respectively. In comparison with some recent methods of [16,46–49], AResU-Net also achieves the best performance on the enhancing tumor segmentation. Meanwhile, it obtains the second place among the compared methods on the core tumor segmentation, which is only slightly

inferior to S3DU-Net [47]. However, its dice score is not favorable on the whole tumor segmentation. The reason lies in that these methods utilize either 3D networks or more complicated 2D network structures. Nevertheless, these compared results still prove the effectiveness of the given AResU-Net method for brain tumor segmentation application.

Table 4. Compared experiment results on the BraTS 2018 validation dataset (66 MRI cases).

Methods	Whole	Core	Enhancing
U-Net	0.860	0.790	0.767
ResU-Net [15]	0.867	0.803	0.768
Ensemble Net [16]	0.881	0.777	0.773
3DU-Net [46]	0.885	0.718	0.760
S3DU-Net [47]	**0.894**	**0.831**	0.749
TTA [48]	0.873	0.783	0.754
MCC [49]	0.882	0.748	0.718
AResU-Net (ours)	0.876	0.810	**0.773**

5. Conclusions

In this paper, we presented a novel AResU-Net model for the MRI brain tumor segmentation task, which simultaneously embedded attention mechanism and residual units into U-Net to improve the segmentation performance of brain tumors. By adding a series of attention units among corresponding down-sampling and up-sampling processes, AResU-Net adaptively rescaled features to enhance local responses of down-sampling residual features, as well as the recovery effects of the up-sampling process. Experiment results on two brain tumor segmentation benchmarks demonstrated that AResU-Net outperformed U-Net by a large margin and gained comparable performance with other typical brain tumor segmentation methods. However, due to the limitation of computational resources, our model utilized 2D slices as inputs to construct the segmentation network, which led to more loss of context information among different slices of brain tumors to a certain extent. Therefore, in the future, we will extend our AResU-Net to 3D network to purse better segmentation results and apply it to other medical image segmentation tasks for further evaluation. In addition, some more powerful feature extraction modules will also be explored to gain the extra performance improvement.

Author Contributions: J.Z., X.L. and B.L. designed the methods and concepts of this article. J.Z., X.L. and H.Z. completed the experimental analysis section. J.Z., X.L., H.Z. and B.L. finished writing this article. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Key R&D Program of China (Grant No. 2018YFC0910506), the National Natural Science Foundation of China (Grant No. 61972062), the Natural Science Foundation of Liaoning Province (Grant No. 2019-MS-011), the Key R&D Program of Liaoning Province (Grant No. 2019 JH2/10100030) and the Liaoning BaiQianWan Talents Program.

Conflicts of Interest: The authors declare no conflict of interest.

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] [PubMed]
2. Cui, S.; Mao, L.; Jiang, J.; Liu, C.; Xiong, S. Automatic Semantic Segmentation of Brain Gliomas from MRI Images Using a Deep Cascaded Neural Network. *J. Healthc. Eng.* **2018**, *1*, 1–14. [CrossRef] [PubMed]
3. Işın, A.; Direkoğlu, C.; Şah, M. Review of MRI-based brain tumor image segmentation using deep learning methods. *Procedia Comput. Sci.* **2016**, *102*, 317–324. [CrossRef]
4. Menze, B.H.; Jakab, A.; Bauer, S.; Kalpathy-Cramer, J.; Farahani, K.; Kirby, J.; Burren, Y.; Porz, N.; Slotboom, J.; Wiest, R. The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). *IEEE Trans. Med. Imaging* **2015**, *34*, 1993–2024. [CrossRef] [PubMed]

5. Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015.
6. 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.* **2017**, *40*, 834–848. [CrossRef] [PubMed]
7. Kayalibay, B.; Jensen, G.; van der Smagt, P. CNN-based segmentation of medical imaging data. *arXiv* **2017**, arXiv:1701.03056.
8. Abbas, N.; Saba, T.; Mohamad, D.; Rehman, A.; Almazyad, A.S.; Al-Ghamdi, J.S. Machine aided malaria parasitemia detection in Giemsa-stained thin blood smears. *Neural Comput. Appl.* **2018**, *29*, 803–818. [CrossRef]
9. Esteva, A.; Kuprel, B.; Novoa, R.A.; Ko, J.; Swetter, S.M.; Blau, H.M.; Thrun, S. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* **2017**, *542*, 115. [CrossRef]
10. Gulshan, V.; Peng, L.; Coram, M.; Stumpe, M.C.; Wu, D.; Narayanaswamy, A.; Venugopalan, S.; Widner, K.; Madams, T.; Cuadros, J. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *J. Am. Med. Assoc.* **2016**, *316*, 2402–2410. [CrossRef]
11. Liu, Y.; Gadepalli, K.; Norouzi, M.; Dahl G.E.; Kohlberger, T.; Boyko, A; Venugopalan, S.; Timofeev, A.; Nelson, H.Q.; Corrado, G.S. Detecting cancer metastases on gigapixel pathology images. *arXiv* **2017**, arXiv:1703.02442.
12. Wang, D.; Khosla, A.; Gargeya, R.; Irshad, H.; Beck, A.H. Deep learning for identifying metastatic breast cancer. *arXiv* **2016**, arXiv:1606.05718.
13. Chen, H.; Dou, Q.; Yu, L.Q.; Qin, J.; Heng, P.A. VoxResNet: Deep voxelwise residual networks for brain segmentation from 3D MR images. *NeuroImage* **2017**, *170*, 446–455. [CrossRef]
14. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, 5–9 October 2015.
15. Kermi, A.; Mahmoudi, I.; Khadir, M.T. Deep Convolutional Neural Networks Using U-Net for Automatic Brain Tumor Segmentation in Multimodal MRI Volumes . In Proceedings of the International MICCAI Brainlesion Workshop, Granada, Spain, 16 September 2018.
16. Albiol, A.; Albiol, A.; Albiol, F. Extending 2D Deep Learning Architectures to 3D Image Segmentation Problems. In Proceedings of the International MICCAI Brainlesion Workshop, Granada, Spain, 16 September 2018.
17. Wang, X.; Girshick, R.; Gupta, A.; He, K. Non-local neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018.
18. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018.
19. Zhang, H.; Dana, K.; Shi, J.; Zhang, Z.; Wang, X.; Tyagi, A.; Agrawal, A. Context encoding for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018.
20. Yu, C.; Wang, J.; Peng, C.; Gao, C.; Yu, G.; Sang, N. Learning a discriminative feature network for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018.
21. Zhao, H.S.; Zhang, Y.; Liu, S.; Jia, J.P.; Loy, C.C.; Lin, D.H.; Jia, J.Y. Psanet: Point-wise spatial attention network for scene parsing. In Proceedings of the European Conference on Computer Vision, Munich, Germany, 8–14 September 2018.
22. Zhou, C.H.; Chen, S.C.; Ding, C.X.; Tao, D.C. Learning contextual and attentive information for brain tumor segmentation. In Proceedings of the International MICCAI Brainlesion Workshop, Granada, Spain, 16 September 2018.
23. Qi, K.H.; Yang, H.; Li, C.; Liu, Z.Y.; Wang, M.Y.; Liu, Q.G.; Wang, S.S. X-Net: Brain Stroke Lesion Segmentation Based on Depthwise Separable Convolution and Long-Range Dependencies. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Shenzhen, China, 13–17 October 2019.
24. 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]

25. Urban, G.; Bendszus, M.; Hamprecht, F. Multi-modal brain tumor segmentation using deep convolutional neural networks. In Proceedings of the MICCAI Multimodal Brain Tumor Segmentation Challenge, Boston, MA, USA, 14–18 September 2014.
26. Pereira, S.; Pinto, A.; Alves, V.; Silva, C.A. Brain Tumor Segmentation Using Convolutional Neural Networks in MRI Images. *IEEE Trans. Med. Imaging* **2016**, *35*, 1240–1251. [CrossRef]
27. Kamnitsas, K.; Ledig, C.; Newcombe, V.F.J.; Simpson, J.P.; Kane, A.D.; Menon, D.K.; Rueckert, D.; Glocker, B. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. *Med. Image Anal.* **2017**, *36*, 61–78. [CrossRef]
28. Zhao, X.M.; Wu, Y.H.; Song, G.D.; Li, Z.Y.; Zhang, Y.Z.; Fan, Y. A deep learning model integrating FCNNs and CRFs for brain tumor segmentation. *Med. Image Anal.* **2018**, *43*, 98–111. [CrossRef]
29. Dong, H.; Yang, G.; Liu, F.; Mo, Y.; Guo, Y. Automatic brain tumor detection and segmentation using u-net based fully convolutional networks. In Proceedings of the Annual Conference on Medical Image Understanding and Analysis, Edinburgh, UK, 11–13 July 2017.
30. Kong, X.G.; Sun, G.X.; Wu, Q.; Liu, J.; Lin, F.M. Hybrid Pyramid U-Net Model for Brain Tumor Segmentation. In Proceedings of the International Conference on Intelligent Information Processing, Nanning, China, 19–22 October 2018.
31. Wang, G.; Li, W.; Ourselin, S.; Vercauteren, T. Automatic brain tumor segmentation using cascaded anisotropic convolutional neural networks. In Proceedings of the International MICCAI Brainlesion Workshop, Quebec City, QC, Canada, 14 September 2017.
32. Tseng, K.L.; Lin, Y.L.; Hsu, W.; Huang, C.Y. Joint sequence learning and cross-modality convolution for 3d biomedical segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017.
33. Liu, D.; Zhang, H.; Zhao, M.M.; Yu, X.J.; Yao, S.W.; Zhou, W. Brain Tumor Segmention Based on Dilated Convolution Refine Network. In Proceedings of the International Conference on Software Engineering Research, Management and Applications, Kunming, China, 13–15 June 2018.
34. Li, H.; Li, A.; Wang, M. A novel end-to-end brain tumor segmentation method using improved fully convolutional networks. *Comput. Biol. Med.* **2019**, *108*, 150–160. [CrossRef]
35. Jin, Q.G.; Meng, Z.P.; Sun C.M.; Wei, L.Y.; Su, R. RA-UNet: A hybrid deep attention-aware network to extract liver and tumor in CT scans. *arXiv* **2018**, arXiv:1811.01328.
36. Fu, J.; Liu, J.; Tian, H.; Li, Y.; Bao, Y.J.; Fang, Z.W.; Lu, H.Q. Dual attention network for scene segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 16–20 June 2019.
37. Zhang, H.; Zhang, H.; Wang, C.; Xie, J. Co-occurrent Features in Semantic Segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 16–20 June 2019.
38. Jegou, H.; Douze, M.; Schmid, C.; Perez, P. Aggregating local descriptors into a compact image representation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, San Francisco, CA, USA, 13–18 June 2010.
39. Li, H.; Xiong, P.; An J.; Wang, L.X. Pyramid attention network for semantic segmentation. *arXiv* **2018**, arXiv:1805.10180.
40. Kolnogorov, A.V. Gaussian Two-Armed Bandit and Optimization of Batch Data Processing. *Probl. Inf. Transm.* **2018**, *54*, 84–100. [CrossRef]
41. He, K.M.; Zhang, X.; Ren, S.Q.; Sun, J. Deep Residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016.
42. Bakas, S.; Reyes, M.; Jakab, A.; Bauer, S.; Rempfler, M.; Crimi, A.; Shinohara, R.T.; Berger, C.; Ha, S.M.; Rozycki, M. 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.
43. Abadi, M.; Barham, P.; Chen, J.M.; Chen, Z.F.; Davis, A.; Dean, J.; Devin, M.; Ghemawat, S.; Irving, G.; Isard, M. Tensorflow: A system for large-scale machine learning. In Proceedings of the USENIX Conference on Operating Systems Design and Implementation, Savannah, GA, USA, 2–4 November 2016.
44. Chen, L.; Wu, Y.; DSouza, A.M.; Abidin, A.Z.; Wismüller, A.; Xu, C. MRI tumor segmentation with densely connected 3D CNN. In Proceedings of the International Society for Optics and Photonics, San Diego, CA, USA, 19–23 August 2018.

45. Badrinarayanan, V.; Kendall, A.; Cipolla, R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 2481–2495. [CrossRef] [PubMed]
46. 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 International Conference on Medical Image Computing and Computer-Assisted Intervention, Athens, Greece, 17–21 October 2016.
47. Chen, W.; Liu, B.; Peng, S.; Sun, J.; Qiao, X. S3d-unet: Separable 3du-net for brain tumor segmentation. In Proceedings of the International MICCAI Brainlesion Workshop, Granada, Spain, 16 September 2018.
48. Wang, G.; Li, W.; Ourselin, S.; Vercauteren, T. Automatic brain tumor segmentation using convolutional neural networks with test-time augmentation. In Proceedings of the International MICCAI Brainlesion Workshop, Granada, Spain, 16 September 2018.
49. Hu, K.; Gan, Q.; Zhang, Y.; Deng, S.H.; Xiao, F.; Huang, W.; Cao, C.H.; Gao, X.P. Brain tumor segmentation using multi-cascaded convolutional neural networks and conditional random field. *IEEE Access* **2019**, *7*, 92615–92629. [CrossRef]

© 2020 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 (http://creativecommons.org/licenses/by/4.0/).