The multimodal MRI brain tumor segmentation based on AD-Net ☆

Yanjun Peng [a,b], Jindong Sun [a,*]

[a] *College of Computer science and Engineering, Shandong University of Science and Technology, Qingdao, China*
[b] *Shandong Province Key Laboratory of Wisdom Mining Information Technology, China*

ARTICLE INFO

ABSTRACT

Multimodal glioma images provide different features of tumor boundaries based on magnetic resonance imaging (MRI), where multimodal features are often challenging to extract for deep learning segmentation methods. Disturbance between features of different modes is an important factor restricting multimodal learning. To efficiently extract multimodal features, we propose an automatic weighted dilated convolutional network (AD-Net) to learn multimodal brain tumor features through channel feature separation learning. Specifically, the auto-weight dilated convolutional unit (AD unit) utilizes dual-scale convolutional feature maps to acquire channel separation features. We employ two learnable parameters to fuse dual-scale convolutional feature maps in encoding layers, and the two learnable parameters are automatically adjusted with the back propagation of the gradient. We adopt the Jensen–Shannon divergence to constrain the distribution of its feature map, which in turn regularizes the weights of the entire down-sampling. In addition, we use the training technique of deep supervision to achieve fast fitting. Our proposed method got dice scores of 0.90, 0.80, and 0.76 for the whole tumor (WT), the tumor core (TC), and the enhancing tumor (ET) on the BraTS20 dataset. The experimental results showed good performance with the AD-Net network.

1. Introduction

Gliomas originate from intracranial tumors of glial cells and are highly lethal [1]. It is the occurrence of mutations that are sufficient for carcinogenesis at the level of the cell's genetic material (DNA) and epigenetic material (EPI) through the interaction of internal genetic predisposing factors with external environmental pathogenic factors. Gliomas are mainly classified into the following categories: Astrocytoma, Oligodendroglioma, Mixed gliomas (such as oligodendro astrocytomas, which contain mixed types of glial cells), and Ependymoma. The low-grade patients have a survival rate of months or even years, while the history of high-grade gliomas is often weeks to months.

Magnetic Resonance Imaging (MRI) has no ionizing radiation damage to the human body, imaging without injecting radioisotopes. The soft tissue structure appears clear on MRI images [2,3]. The multimodal glioma sequences, including T1, T1c, T2, and T2-Flair, are advantageous for assessing health risks and clinical diagnosis.

However, manual segmentation of multimodal brain tumors is time-consuming and expensive compared to the automatic methods [4,5]. It usually takes an expert radiologist about 3 h to perform pixel-level segmentation. The Dice Similarity Coefficient for manual segmentation is 74%–85% [6,7]. Therefore, accurate methods of tumor segmentation are of vital importance in clinical diagnostics and planning for treatment.

The automated segmented methods aim to segment these multimodal MRI images into four tissues, including the normal tissues, the whole tumor (WT), the tumor core (TC), and the enhancing tumor (ET). TC describes that most tumors need to be removed and represents the necrotic and non-enhancing parts. The WT contains peritumoral edema, ET, and TC.

Accurate automatic segmentation of gliomas remains a challenging task. In the 2010s, many methods were proposed for automated segmentation in brain tumors. These studies mainly include deep learning methods and traditional machine learning methods. Traditional machine learning mainly includes decision trees, random forests, artificial neural networks, and bayesian learning. Features, such as gradients and texture, used to segment images in most of these methods are hand-crafted. In the field of medical image segmentation, deep learning methods are popular topics by using a data-driven way.

Since the powerful generalization, the deep learning method has gained a significant advantage compared to other approaches. Deep learning puts forward a way to let computers learn the features automatically as data-driven to reduce the complexity of artificial design

features. The deep learning model advances segmentation performance, such as convolutional neural network (CNN), fully convolutional network (FCN), graph neural network (GNN), recurrent neural network (RNN), generative adversarial network (GAN), and other forms of network [8–11]. In brain tumor segmentation, the deeper neural network model is increasingly important to advance state-of-art performance.

Deep learning methods have attracted significant interest in segmenting brain tumors in recent years. The U-shaped model is an efficient and straightforward segmentation network in 3D medical images, learning features from deep and shallow neural units, especially in brain tumors. The UNet model consists of four encoder layers, and four decoder layers [12].

Most researchers applied U-shaped architecture to build deep learning model [13–15]. The Brain Tumor Segmentation Challenge (BraTS) dataset on Medical Image Computing and Computer Assisted Intervention Society (MICCAI) is the most popular dataset. The Nvidia team proposed a variational autoencoder branch to reconstruct the input image and gained the top rank in the BraTS 2018 (BraTS18) challenge [16]. The SA-LuT-Nets model developed a sample-adaptive intensity mapping network to transform the MRI intensity and applied the DMFNet to segmentation [17]. The DMFNet model realizes real-time segmentation through auto-weighted dilated Multi-Fiber unit [18]. Jiang et al. applied multiple networks from coarse to precise segmentation, in which the loss functions combine at different periods [19]. McKinley et al. applied a local attention mechanism and various postprocessing techniques to raise segmentation accuracy [20]. Agravat et al. discussed several deep learning methods and various open source tools [21]. The DeepMedic [22], UNet [12], VNet [15], SegNet [23], DenseNet [24], ResNet [25] give advantage segmentation results. U-shaped models have widely trusted architecture in medical image segmentation. UNet uses the skip-connection structure to fuse low-resolution information with high-resolution information. The skip connection between the peer layers is critical for a U-shaped model, and it can greatly help proper feature reconstruction. As the Transformer architectures have had great success in natural language processing (NLP), Transformer blocks have also been applied to medical image processing. The TransUNet model combines the advantages of the local feature extraction of UNet with the global feature extraction of Transformer, and achieves great success in medical image segmentation [26]. MedtNet adopted self-attention layers and gated axial attention layers to encode remote dependencies based on the Transformer architecture that learned global and local features, respectively [27].

There are three main challenges to getting stable and accurate multimodal segmentation.

- The most challenging task is segmenting the subregion's boundaries (WT, ET, TC, and normal tissue). Blood spots, strokes, and other complications are noisy regions that increase the difficulty of segmentation. We utilize dense residual blocks to build a deep learning model to solve this problem instead of simply overlaying convolutional layers. In this model, we combine the residual block with U-shaped architecture to maximize the feature of brain tumors and learn the edges of every lesion tissue. We employ the dilated convolution with dual-sized kernels to expand the receptive field of the convolutional kernel. Within each AD unit, we divide the feature channels into two groups, which are input into dual-sized kernel dilated convolution channels, respectively.
- The optimization strategy of a deep neural network is directly related to the final segmentation performance. Inspired by the idea of deep supervision, we map the most profound feature map directly to the output layer and adopt the Jensen–Shannon divergence to constrain the distribution of its feature map, which in turn regularize the weights of the entire downsampling.
- The last challenge is that the network is hard to train. Specifically, complex multimodal images provide rich organizational information but also increase the difficulty of network training.
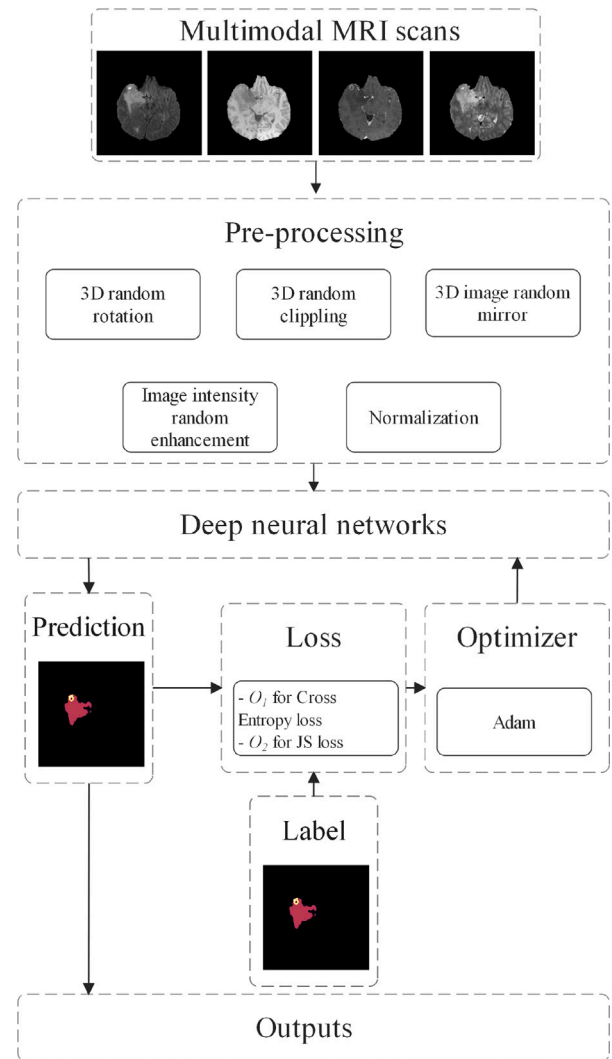

Fig. 1. The workflow of proposed method.

In ablation experiment, only using cross entropy loss in this task makes it difficult to reduce the loss to a reasonable range. We use the training technique of deep supervision, add additional loss calculation, and update the weight of the encoding layer with an additional gradient, to achieve the purpose of fast fitting.

Inspired by the groundbreaking proposal in the related work, our main task is to build a deep neural model to obtain the highest segmentation performance.

2. Method

Our workflow is shown in Fig. 1. Note that the pre-processing methods get the same results with different usage orders.

2.1. Pre-processing

We use the randomization strategy as image pre-processing, which could ensure that the deep learning model still maintains strong generalization performance after a large number of repeated training. Each batch of images uses the same random seeds so that the input images would be different in different epochs during training. It helps to learn the image features of different modes in the same brain while obtaining generalization. The figure shows the pre-processing image methods:
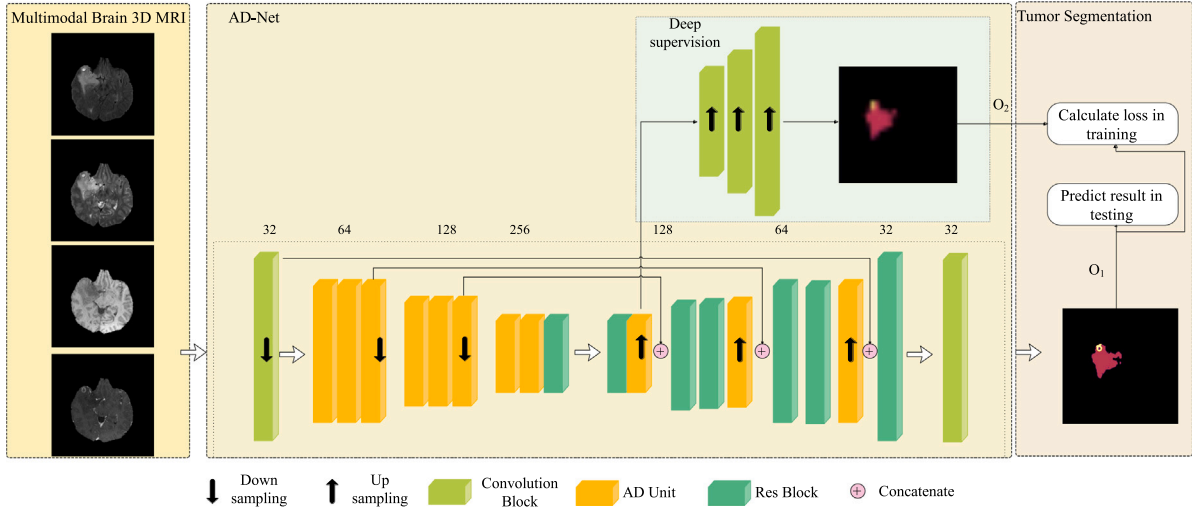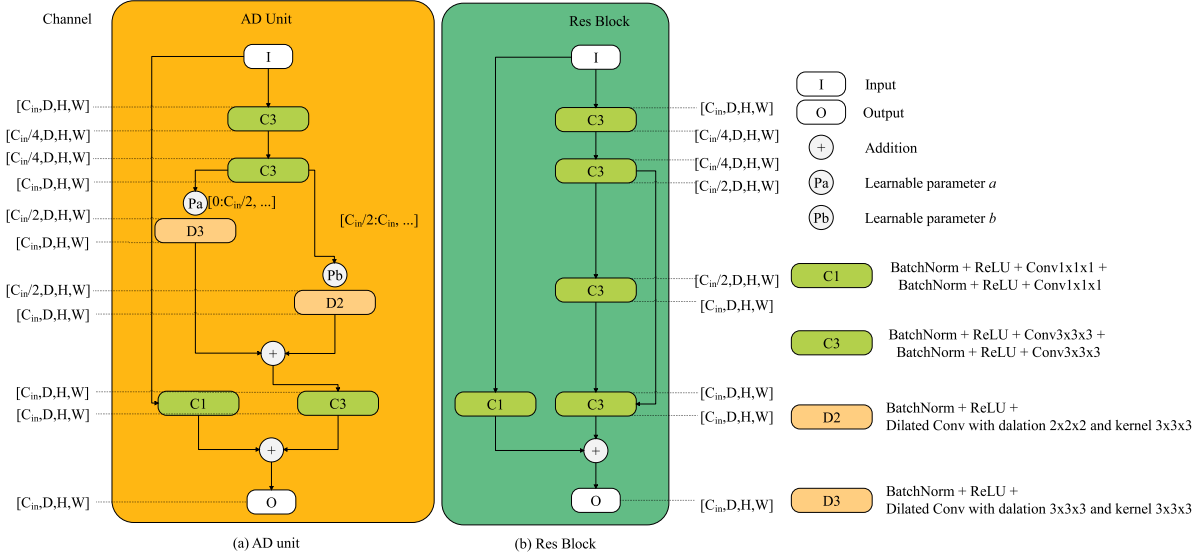
Fig. 2. The proposed AD-Net.

Fig. 3. The details of AD unit and Res block in Fig. 2. (a) The architecture of AD unit. (b) The architecture of Res block. [C_{in},D,H,W] represents input channel, deep, height, and width of feature map.

3D random clipping, 3D random rotation, 3D image intensity random enhancement, 3D image random mirror inversion, and normalization.

Image normalization is a widely used technique in computer vision, pattern recognition and other fields. The z-score normalization was applied in this work. It is defined as:

$$z = \frac{x - \mu}{\sigma} \tag{1}$$

where σ is the standard deviation, and μ is the mean value. Then, the 3D random clipping method randomly cuts the MRI image (240, 240, 155) into a matrix (144, 144, 128). The 3D random rotation method rotates the reduced image by the angle $U(-10, +10)$. The random intensity enhancement method of 3D image sets the image pixel value as:

$$x_{new} = x_{old} * U(0.9, 1.1) + U(-0.1, 0.1) \tag{2}$$

where U is the uniform distribution. Random mirror processing symmetrizes the image according to deep, height, and width directions. We apply these image enhancement routines to extend the training data set to improve the performance and generalization ability of the deep neural network.

2.2. Deep learning model

The proposed model is shown in Fig. 2, where C_{in} is the input channel, C_{out} is the output channel and + is the addition operation. In this model, the four channel model inputs correspond to the MRI images of four modes, respectively. The main body of the network is composed of auto-weight dilated (AD) unit, Residual (Res) block, linear upsampling, and the first and last convolution units. In the downsampling stage (feature encoding extraction), we use 8 AD units to obtain multi-scale feature maps. In the upsampling stage (feature decoding), we use the AD unit, Res unit, and a linear upsampling layer to form a primary decoding layer. Finally, a convolution unit with ReLU operation outputs the results of the network model. Moreover, each convolution unit, AD unit and Res block contains batch normalization [28] and ReLU functions. We use dilated convolution to extract fine-grained and multi-scale glioma features, and use residual structure to obtain long-dependent glioma features. All convolutional layer is with group parameter of 16.

As for the Res block in Fig. 3.(b), we use two convolution units to reduce and then enlarge the number of convolution kernels so as to realize feature learning and feature map reorganization. From an
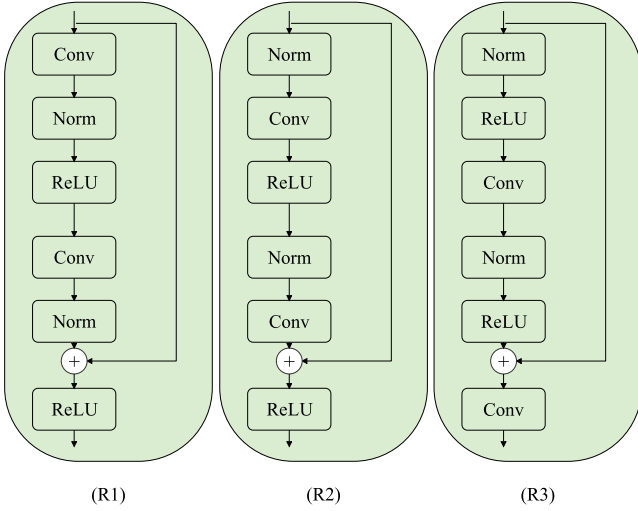
(R1)   (R2)   (R3)

Fig. 4. The different residual convolution block. The + is the addition operation.

experimental point of view, this is an efficient encoding method [17]. Then, we use two group convolution units with stripe 1 and groups 16, and the kernel size is $3 \times 3 \times 3$. Finally, we employ a convolution residual element to obtain the characteristic graph of long dependence.

As for AD Unit in Fig. 3.(a), we use two convolution units first (like Res unit). Then, we use two dilated convolution units and two learnable parameters to adjust and fused the feature. Finally, a group convolution unit is used to output the result of AD unit. We also set up residual calculations in the AD unit. The dilated convolution expands the receptive field of the convolutional kernel without sacrificing computational resources. We divide the feature maps into two groups according to the number of channels, and then use dilated convolutions of different strides to calculate the two groups of feature maps respectively, and learn two learnable weight factors (a and b) to adjust the weight of the two sets of feature maps in the AD unit. The fusion of the two types of convolutions strengthens the ability of the network to extract features.

In order to better evaluate the performance of the AD unit, we used three different residual units instead of the AD unit to construct three neural networks (Block-R1, Block-R2, and Block-R3), as shown in Fig. 4.

In the encoder stage, each residual block is a dual-pathway structure. In this stage, we set channel depth to 32, 64, 128, and 256. The residual block is the critical structure of down-sampling. In the decoder stage, we connect a convolutional unit and a de-convolutional unit for upsampling. The stride of the de-convolutional unit is $2 \times 2 \times 2$, and the kernel size is $3 \times 3 \times 3$. Similarly, we set the channel depth in the decoder stage to $32 \times 64 \times 128 \times 256$. Combined deep neural network with the residual blocks enables the network to obtain more significant gradients in deep layers. So, the phenomenon of gradient disappearance is relatively rare and gets more practical features of gliomas. The formula of the gradient propagation in the convolutional layer can be defined,

$$\delta_l = \sigma'(O_l) \cdot (w_{l+1})^T \delta_{l+1} O_l \tag{3}$$

where σ' means the first derivative of the loss function, w describes the weight, O indicates the output matrix vector, and l is the layer l. Then, the gradient in Block-R1, Block-R2 and Block-R3 can be defined,

$$O_{r1_{l+1}} = f(\delta_2(f(\delta_1(O_{r1_l})) + O_{r1_l})) \tag{4}$$

$$O_{r2_{l+1}} = f(\delta_2(f(\delta_1(O_{r2_l})) + O_{r2_l})) \tag{5}$$

$$O_{r3_{l+1}} = \delta_2(f(\delta_1(f(O_{r3_l})) + O_{r3_l})) \tag{6}$$

where f means the activation function, δ_1 and δ_2 represent the first and second convolution calculations, respectively. It is worth noting that the difference between Eqs. (4) and (5) lies in the order of normalization, which is not reflected in the equation

Multiplication is widely used in the calculation of series convolution, such as $\delta_2(f(\delta_1))$. The cumulative multiplication between $(-1, 1)$ makes it possible for the gradient to appear the result of approximate 0, so that the classical gradient disappears. The residual connection weakens this problem through weight addition and enhances the stability of the network. Obviously, it is a very effective way to use residual blocks to build architecture in very deep neural layers, especially in calculating the depth feature map.

We define the convolution block including Batch Normalization (BN), ReLU (RL), Convolution (Conv), BN, RL, and Conv in AD unit as follows,

$$\ell_{(c_i,k)} = w_{c_i}{}^T f(I_i) + b_{c_i} \tag{7}$$

where c_i is the convolution layer i, and k describes the kernel size, f is the activation function. The w and b represent the convolution weight and bias, I_i is the input data. Then, the AD block can be define simply,

$$\ell_{AD} = \ell_{(c_1,1)}\ell_{(c_2,1)}(a\ell_{(d_0,3)} + b\ell_{(c_3,3)})\ell_{(c_4,3)} + \ell_{(c_5,1)} \tag{8}$$

where ℓ_{d_0} is the dilated convolution. In the gradient back-propagation process, a and b can automatically adjust the weight ratio of the convolution integral branch in the main path. In addition, the channel parameters of the AD-Net were set to 32, 64,128, 256, and the AD-unit and Res-unit adopted the 3D matrix concatenate method.

2.3. Update strategy

The Update strategy consists of two parts, the optimizer and the loss function. The Adam method is adopted as optimizer. The loss function is divided into two parts. One part comes from the output O_1 and the other part comes from the deep supervision output O_2, as shown in the network Fig. 2. We compute the loss function separately using the cross entropy and the JS divergence, which can be defined as follow,

$$loss(y, x_{o_1}, x_{o_2}) = L_{CE}(y, x_{o_1}) + \alpha L_{JS}(y, x_{o_2}) \tag{9}$$

where y is the ground truth, x_{o_1} and x_{o_2} are the predicted output and the deeply supervised output, respectively. L_{CE} represents the Cross Entropy loss and L_{JS} represents the Jensen–Shannon (JS) divergence. The JS divergence can be represented by the Kullback–Leibler (KL) divergence as L_{JS}, Cross Entropy loss can be defined as L_{CE}, and KL divergence can be defined as L_{KL}.

$$L_{CE(y,x)} = -\sum_{i=1}^{n} y_i \log(x_i) \tag{10}$$

$$L_{JS}(y, x) = \frac{1}{2}(L_{KL}(y, \frac{y+x}{2}) + L_{KL}(x, \frac{x+y}{2})) \tag{11}$$

$$L_{KL}(y, x) = \sum_{i=1}^{n} y_i \log \frac{y_i}{x_i} \tag{12}$$

where n denotes the number of the class (n = 4), y and x denote the ground truth and the output matrix, respectively. Therefore, the loss function can be defined as follow,

$$loss_{total}(y, x_{o_1}, x_{o_2}) = -\sum_{i=1}^{4} y_i \log(x_{i,o_1})$$

$$+ \frac{\alpha}{2} \sum_{i=1}^{4} y_i \log \frac{2y_i}{y_i + x_{i,o_2}}$$

$$+ \frac{\alpha}{2} \sum_{i=1}^{4} x_{i,o_2} \log \frac{2x_{i,o_2}}{y_i + x_{i,o_2}}$$

Table 1
The parameters of the training.

	Parameter	Value
Convolution	Weights	$\mathcal{N}(0, \sqrt{2/kernels})$
	Bias	0
BatchNorm	Weight	1
	Bias	0
Training	GPU	$4 \times$ RTX 2080 Ti
	Batch size	8
	Max epoch	200
	Learning rate	0.001
	Learning decay	0.9

$$= \sum_{i=1}^{4} \log \frac{(2y_i)^{\frac{\alpha}{2}}}{x_{i,o_1}(y_i + x_{i,o_2})^{\frac{\alpha}{2}}}$$
$$+ \frac{\alpha}{2} \sum_{i=1}^{4} x_{i,o_2} \log \frac{2x_{i,o_2}}{y_i + x_{i,o_2}} \tag{13}$$

where α parameter is set to 0.5. The CE loss constrains the overall network architecture and therefore the advantage of CE loss over JS divergence needs to be maintained. Because of the need to make the overall loss in a reasonable magnitude, we multiply the loss of JS divergence by alpha.

To evaluate the performance of segmentation, we set ablation experiments including by using only the Cross Entropy loss (AD-Net$_{L_{CE}}$) and by using JS loss (AD-Net$_{L_{JS}}$) for both of O_1 and O_2.

2.4. Learning rate strategy

Our strategy to update the learning rate follows,

$$newLR = initLR * (1 - \frac{epoch}{maxepoch})^{Ld} \tag{14}$$

where $initLR$ is the initialization of learning rate, and Ld is the learning decay which is set in Table 1. Therefore, the learning rate is in the range of 1e−3 to 8e−6. An appropriate learning rate adjustment strategy can improve the performance of the model. When the model training tends to be stable, reducing the learning rate can further reduce the training loss. Our training strategies were supported by the PyTorch framework, which employed 4 GPUs of Nvidia RTX 2080 Ti and more than 33 GB of memory. It takes almost 40 h for one training scheduler. The parameter batch size was set to 8. To evaluate the performance of segmentation, we provide ablation experiment by using a fixed learning rate (AD-Net$_{FL}$).

2.5. Post-processing

In post-processing, we used detail processing to obtain higher scores. In the analysis of the output results of the previous deep learning model, it is found that there are no ET regions in some low-level brain gliomas. However, there may be few ET result predictions in the segmentation results, which leads to the dice score of ET being 0. Therefore, we replaced the ET prediction with TC prediction when the number of ET pixels in the segmentation result is less than 500. Finally, we submitted the prediction results of the validation set to the CBICA IPP platform and return the results through online evaluation https://ipp.cbica.upenn.edu.

3. Result

This section includes the data set, evaluation method, ablation results, and benchmarking.

3.1. Dataset

The MICCAI Brain Tumor Segmentation (BraTS) datasets consist of many pre-operative multimodal MR images from multi-institution [6, 29]. MICCAI is a comprehensive academic conference held by the international Medical Image Computing and Computer Assisted Intervention society. It is a top-level conference in the field of medical image computing (MIC) and computer assisted intervention (CAI). The BraTS datasets already applied various preprocessing steps by the organizers. BraTS challenge has always focused on evaluating advanced methods for multimodal brain tumor segmentation. The results of the validation set and test set required to be evaluated and returned by the online platform. In this study, we evaluated our method and made ablation study on BraTS20 dataset. Then, we evaluated AD-Net on BraTS19 and Federated Tumor Segmentation Challenge 2021 (FeTS21) dataset.

The BraTS19 training set consists of 335 multimodal MRI scans, including 259 high-grade gliomas (HGG) and 76 low-grade gliomas(LGG). In addition, the validation set includes 125 MRI scans. Note that, the test set can only be obtained by participating in the challenge (Closed). We participated in BraTS20.

The BraTS20 training set consists of 369 multimodal MRI scans, including HGG and LGG and the organizers pay no attention to dividing into HGG and LGG. In addition, the validation set contains 125 MRI scans and the testing set includes 166 MRI scans. The test set is for participants and can submit only once. Each scans contains five MRI images (T1, T1c, T2, Flair, and segmentation label).

The FeTS21 training set consists of 336 multimodal MRI scans. Each scans contains five MRI images (T1, T1c, T2, Flair and segmentation label). In addition, the validation set contains 111 MRI scans.

Each multimodal scan contains 5 images of T1, T1c, T2, Flair, and the ground truth for both training sets. For validation sets and test sets, each multimodal scan includes 4 images of T1, T1c, T2, and Flair. The results of the validation and test set come from the online evaluation platform (https://ipp.cbica.upenn.edu/).

3.2. Evaluation method

Normally, the results of brain tumor segmentation consist of four types of measures: Dice Similarity Coefficient (DSC), Sensitivity, Specificity, and Hausdorff. The DSC measures the spatial overlap between the label and the predicted result from the model. It can be defined as:

$$DSC = \frac{2TP}{FP + 2TP + FN} \tag{15}$$

where FP, FN, and TP are the false positive, the false negative, and the true positive, respectively. Sensitivity measures the proportion of positives, and that can be defined as such:

$$Sensitivity = \frac{TP}{TP + FN} \tag{16}$$

In general, higher sensitivity tends to predict better segmentation results. Finally, specificity measures the proportion of negatives. It is defined as:

$$Specificity = \frac{TN}{TN + FP} \tag{17}$$

where TN is true negative detections. DSC is more sensitive to the inner filling of the mask, while Hausdorff distance is more sensitive to the boundary. The Hausdorff distance is defined as:

$$d_H(X, Y) = \max(\max_{x \in X} \min_{y \in Y} d(x, y), \max_{y \in Y} \min_{x \in X}(d(x, y))) \tag{18}$$

where X, Y describe the predicted segmentation and the ground truth, d(x, y) describes the Euclidean Distance between points of the predicted and the ground truth.
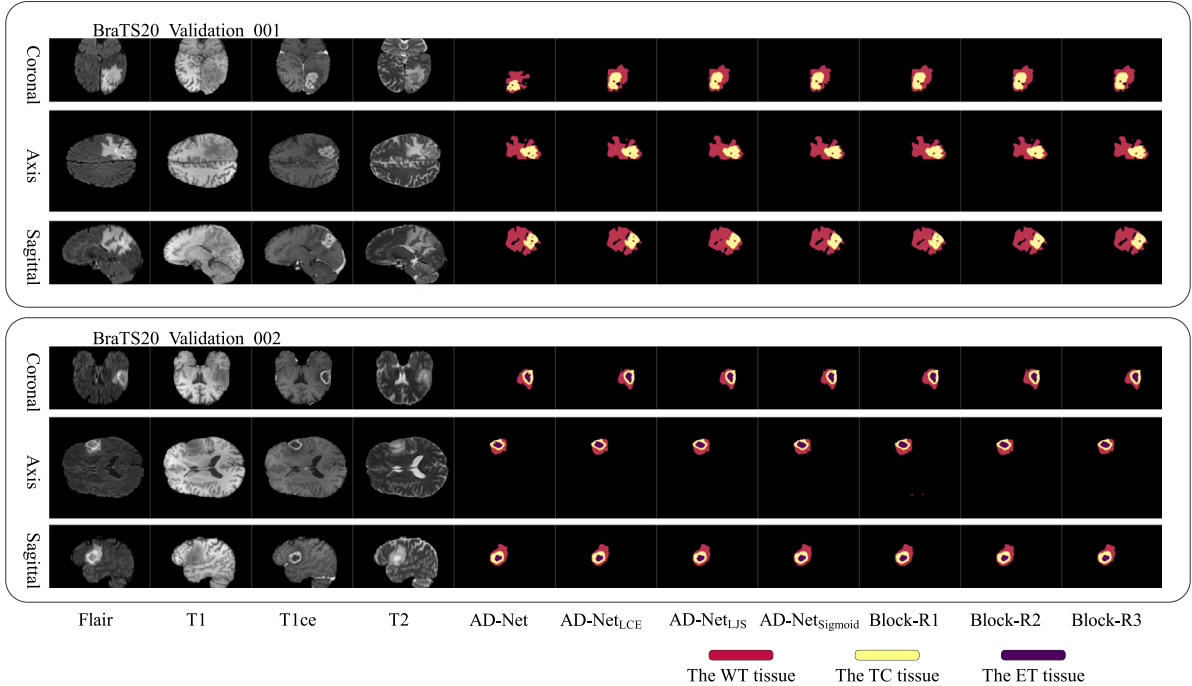
Table 2
All results evaluated on the BraTS20 validation set except the second line which is on BraTS19 validation set.

Methods	Dice			Sensitivity			Specificity			Hausdorff		
	ET	WT	TC	ET	WT	TC	ET	WT	TC	ET	WT	TC
AD-Net	0.76	0.90	0.80	0.78	0.91	0.78	0.99	0.99	0.99	35.2	7.22	15.3
AD-Net$_{Sigmoid}$	0.74	0.90	0.80	0.76	0.91	0.78	0.99	0.99	0.99	36.24	7.50	40.74
AD-Net$_{FL}$	0.77	0.89	0.78	0.77	0.89	0.76	0.99	0.99	0.99	3.66	6.58	8.40
AD-Net$_{L_{CE}}$	0.72	0.89	0.78	0.73	0.90	0.77	0.99	0.99	0.99	31.7	6.78	10.0
AD-Net$_{L_{JS}}$	0.64	0.85	0.73	0.61	0.83	0.70	0.99	0.99	0.99	42.8	12.1	16.7
Block-R1	0.67	0.86	0.72	0.75	0.87	0.74	0.99	0.99	0.99	43.6	22.5	26.4
Block-R2	0.69	0.87	0.69	0.71	0.85	0.65	0.99	0.99	0.99	43.5	12.5	26.4
Block-R3	0.70	0.89	0.77	0.75	0.91	0.77	0.99	0.99	0.99	41.6	6.40	15.0
AD-Net (BraTS19)	0.76	0.90	0.81	0.78	0.91	0.78	0.99	0.99	0.99	35.5	4.31	12.4
AD-Net (FeTS21)	0.80	0.91	0.85	0.82	0.92	0.83	0.99	0.99	0.99	23.32	3.89	4.81

Table 3
The compared results with state-of-art methods on BraTS20 validation set.

Methods	Dice				Hausdorff			
	ET	WT	TC	Mean	ET	WT	TC	Mean
Our method	0.76	0.90	0.80	0.82	35.2	7.22	15.3	19.24
Chen et al. [26]	0.60	0.81	0.61	0.67	56.9	22.3	25.7	34.97
Ding et al. [30]	0.66	0.88	0.73	0.76	46.5	7.95	16.0	23.48
Liu et al. [31]	0.68	0.88	0.76	0.77	38.7	5.37	15.3	19.79
Luo et al. [32]	0.67	0.84	0.73	0.75	53.2	34.2	33.5	40.30
Zhao et al. [33]	0.78	0.90	0.83	0.84	26.35	4.02	6.51	12.29
Ding et al. [34]	0.61	0.86	0.78	0.75	–	–	–	–
Agravat and Raval [35]	0.76	0.87	0.75	0.79	–	–	–	–
Ballestar and Vilaplana [36]	0.77	0.81	0.82	0.80	–	–	–	–
Nguyen et al. [37]	0.80	0.85	0.81	0.82	–	–	–	–
Ghaffari et al. [38]	0.78	0.90	0.83	0.83	–	–	–	–
Guan et al. [39]	0.68	0.85	0.69	0.74	47.4	8.44	31.6	29.14
Isensee et al. [40] (2020 Winner)	0.79	0.91	0.85	0.85	29.2	3.79	7.77	13.59
Jia et al. [41] (2020 2nd place)	0.78	0.91	0.85	0.85	26.57	4.97	11.91	14.48

Fig. 5. The visualization of ablation experiments.

3.3. Ablation results

As shown in Table 2, AD-Net obtained average dice scores of 0.90, 0.80, and 0.76 on the BraTS20 validation set. This proved that the AD unit is a very efficient feature extraction structure. This method obtains

35, 7, and 15 scores in Hausdorff and had achieved great success in ET and TC segmentation tasks.

The AD-Net$_{FL}$ shows the result by using fixed learning rate. The AD-Net$_{L_{CE}}$ shows the result by using only Cross Entropy loss for O_1 and O_2. The AD-Net$_{L_{JS}}$ shows the result by using JS loss for O_1 and O_2. The
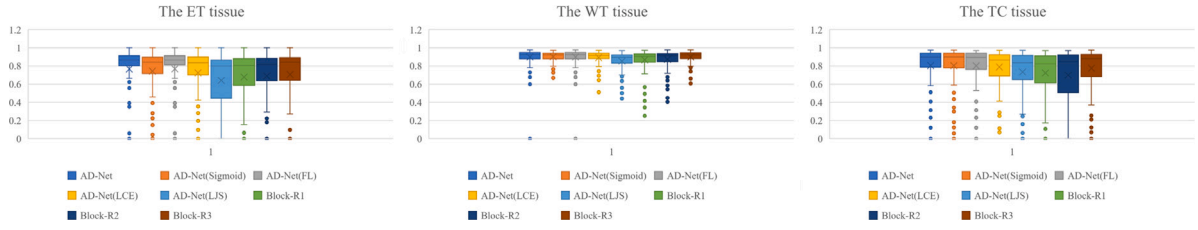
Fig. 6. The boxplots of ablation models in Dice metrics.

Table 4
The part results of Dice on BraTS20 challenge test set.
Source: The experimental results were from Agravat and Raval [42], Ballestar and Vilaplana [36], Colman et al. [43], Ghaffari et al. [44], Nguyen et al. [37] and Silva et al. [45].

Methods	Dice		
	ET	WT	TC
Our method (Block-R3)[a]	0.803	0.872	0.823
Agravat and Raval [42]	0.779	0.875	0.815
Ballestar and Vilaplana [36]	0.77	0.81	0.82
Colman et al. [43]	0.751	0.867	0.798
Ghaffari et al. [44]	0.81	0.88	0.82
Nguyen et al. [37]	0.805	0.856	0.819
Silva et al. [45]	0.790	0.885	0.829

[a]Since the network entrance of the online test platform has been closed and could submit only once, we only provide the test result of Block-R3.

Table 5
The compared results with state-of-art models for flops and parameters while inputting the array of $144 \times 144 \times 128$.

Methods	GFlops	Param.
Our method	134.98	**5.27M**
Ding et al. [30]	390.81	17.53M
Liu et al. [31]	156.21	7.49M
Luo et al. [32]	340.96	24.73M
Zhao et al. [33]	340.97	10.75M
Agravat and Raval [35]	169.25	9.38M
Ballestar and Vilaplana [36]	299.89	21.61M
Nguyen et al. [37]	370.31	17.18M
Ghaffari et al. [38]	590.76	20.42M
Ding et al. [34]	198.48	26.9M
Guan et al. [39]	148.12	6.48M
Isensee et al. [40]	1056.75	101.88M
Jia et al. [41]	301.77	45.59M

Block-R1, Block-R2, and Block-R3 show the results by alternating the AD Unit in the AD-Net. The AD-Net$_{Sigmoid}$ shows the result by using Sigmoid as an activation function instead of ReLU. Since the value range of JS is in the range of 0 to 1, the activation function should also be in the range of 0 to 1. The value range of ReLU and Sigmoid is from 0 to 1.

The Fig. 5 shows the visualization of ablation experiments and we provided four sets of samples from the T1ce mode.

3.4. Compare to state-of-art

The section contains two parts including results of reproducing the state-of-art model and results of challenging test set.

The Table 3 shows the results of the state-of-art model in BraTS20. The Fig. 7 shows the visualization of the state-of-art models. We reproduced deep learning models and evaluated these models by using the same preprocessing approaches.

Isensee et al. [40] got the best WT and TC Dice score. Nguyen et al. [37] proposal could achieve better segmentation on ET tissue than others. Luo et al. proposed method may make the segmentation failure in the complex 3D MRI images.

The result of the challenging test dataset is reported in Table 4. The network entrance of the online test platform has been closed and could submit only once.

3.5. Flops and params

We reproduced deep learning models and evaluated these models by using the same preprocessing approaches. We did a comparative study on calculation and parameters for an input image size of $144 \times 144 \times 128$ in Table 5.

Smaller GFlops mean less time will spend on a single training session. Our method achieved the multimodal glioma segmentation task with minimal model parameters contributing to densely stacked grouped convolutions. We used different dilated convolutions to obtain feature maps with larger receptive fields, compared to ordinary convolutions, with the same computational cost. Our method got 5.27M model parameter and 134.98 GFlops and had a great advantage over others. Ghaffari et al. [38] got 590.76 GFlops in a single calculation. Isensee et al. [40] proposed a nnUNet and got 1056.75 GFlops which model would take a lot of time to train.

4. Discussions

From Fig. 6, we can see the segmentation distribution of all validation samples using different deep neural networks, and make analysis and discussion in combination with the previous section.

4.1. Experimental discussions

Overall, AD-Net has an absolute advantage in ET tissue segmentation, while showing remarkable performance in TC tissue segmentation, but not enough advantage in WT tissue segmentation. In ablation experiments, the AD-Net$_{LJS}$ gets poor segmentation results, even worse than the segmentation results of the pure residual network. This enlightens us that JS divergence can be used as a constraint to regularize the parameters of neural networks, but if only relying on JS loss to optimize neural networks, it will bring great errors. Therefore, JS loss combined with cross entropy loss to optimize the neural network help to improve the segmentation performance of the neural network and reduce the error of the final results.

The neural network combined with the R3 block can bring better segmentation results in the current task. Therefore, we adopt the same convolution module to design the AD-Net. In addition, the proposed learning rate reduction strategy brings a slight improvement to the segmentation results.

Chen et al. proposed TransUNet model (2D model), which combined UNet and Transformer modules but results in considerable error segmentation. We speculate that this may be due to the complexity of 3D medical images, which is not well compatible with 2D segmentation methods because we modify 3D convolution, 3D normalization, and other layers in the reproduction model for qualitative analysis. In theory, the linear units in the Transformer module will destroy the spatial features of the feature map to a certain extent and may get good segmentation performance for 2D images. However, the computational complexity of the linear layers increases exponentially in the 3D computational graph.
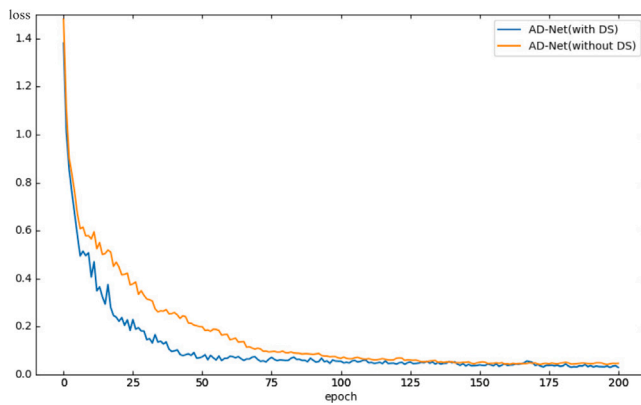
Fig. 7. The compared loss results during the training stage between the method with deep supervision and without deep supervision.

4.2. Deep supervision

Fig. 7 presents the performance of the model using deep supervision connection. Deep supervision technology can accelerate the decline of loss in the first 100 epochs of training, and finally, achieve the stability of loss in advance.

4.3. Limitation

In the comparing state-of-art experiment, we could not use the proposed loss optimization method to optimize the reproduced models, because their model architecture only gave one output O_1. Therefore, we only use the cross entropy loss, which may introduce some potential errors.

5. Conclusion

Our method achieved state-of-the-art segmentation performance with a small number of parameters. The pre-processing methods played an essential role in robust performance. We evaluated the performance of different residual blocks based on the U-shaped architecture. The network that applied Block-R3 had higher segmentation performance than the networks of Block-R1 and Block-R2. In the U-shaped network, feature extraction at the encoding stage is the most important component. Designing the network to extract the features of interest efficiently was crucial. Our proposed AD unit obtained multi-scale convolutional receptive fields to extract more detailed tumor features to achieve advanced segmentation effects. The advantages of our proposed method are quantitatively analyzed by reproducing four deep learning models using 3D processing methods.

We used the JS divergence as part of the loss function, mainly to constrain the convolution weights of the downsampling and thus accelerated the convergence of the model weights. The contribution of the cross entropy mainly lied in the update of the upsampling weight, and we used the JS divergence of 0.5 parameters to balance the contribution of the two loss functions. The code is available at github.com/JalexDooo/BrainstormTS20-Paper4.

CRediT authorship contribution statement

Yanjun Peng: Funding acquisition, Supervision, Writing – review & editing. **Jindong Sun:** Writing – original draft, Conceptualization, Methodology.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Yanjun Peng reports financial support was provided by National Natural Science Foundation of China, Shandong Natural Science Foundation.

Data availability

The code is available at https://github.com/JalexDooo/Brainstorm TS20-Paper4.

References

[1] S. Pereira, A. Pinto, V. Alves, C.A. Silva, Brain tumor segmentation using convolutional neural networks in MRI images, IEEE Trans. Med. Imaging 35 (5) (2016) 1240–1251.

[2] S. Bauer, R. Wiest, L.-P. Nolte, M. Reyes, A survey of MRI-based medical image analysis for brain tumor studies, Phys. Med. Biol. 58 (13) (2013) R97.

[3] Q.T. Ostrom, L. Bauchet, F.G. Davis, I. Deltour, J.L. Fisher, C.E. Langer, M. Pekmezci, J.A. Schwartzbaum, M.C. Turner, K.M. Walsh, et al., The epidemiology of glioma in adults: a "state of the science" review, Neuro-Oncology 16 (7) (2014) 896–913.

[4] Z. Akkus, A. Galimzianova, A. Hoogi, D.L. Rubin, B.J. Erickson, Deep learning for brain MRI segmentation: state of the art and future directions, J. Digit. Imaging 30 (4) (2017) 449–459.

[5] W. Yue, Z. Wang, B. Tian, M. Pook, X. Liu, A hybrid model-and memory-based collaborative filtering algorithm for baseline data prediction of Friedreich's Ataxia patients, IEEE Trans. Ind. Inf. (2020).

[6] B.H. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, Y. Burren, N. Porz, J. Slotboom, R. Wiest, et al., The multimodal brain tumor image segmentation benchmark (BRATS), IEEE Trans. Med. Imaging 34 (10) (2014) 1993–2024.

[7] L. Liu, H. Zhang, I. Rekik, X. Chen, Q. Wang, D. Shen, Outcome prediction for patient with high-grade gliomas from brain functional and structural networks, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2016, pp. 26–34.

[8] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, in: Advances in Neural Information Processing Systems, 2012, pp. 1097–1105.

[9] W. Dong, M. Wozniak, J. Wu, W. Li, Z. Bai, De-noising aggregation of graph neural networks by using principal component analysis, IEEE Trans. Ind. Inf. (2022).

[10] H. Basak, R. Kundu, P.K. Singh, M.F. Ijaz, M. Woźniak, R. Sarkar, A union of deep learning and swarm-based optimization for 3D human action recognition, Sci. Rep. 12 (1) (2022) 1–17.

[11] M. Wieczorek, J. Siłka, M. Woźniak, S. Garg, M.M. Hassan, Lightweight convolutional neural network model for human face detection in risk situations, IEEE Trans. Ind. Inf. 18 (7) (2021) 4820–4829.

[12] 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.

[13] F. Isensee, J. Petersen, A. Klein, D. Zimmerer, P.F. Jaeger, S. Kohl, J. Wasserthal, G. Koehler, T. Norajitra, S. Wirkert, et al., Nnu-net: Self-adapting framework for u-net-based medical image segmentation, 2018, arXiv preprint arXiv:1809.10486.

[14] X. Li, H. Chen, X. Qi, Q. Dou, C.-W. Fu, P.-A. Heng, H-DenseUNet: hybrid densely connected UNet for liver and tumor segmentation from CT volumes, IEEE Trans. Med. Imaging 37 (12) (2018) 2663–2674.

[15] F. Milletari, N. Navab, S.-A. Ahmadi, V-net: Fully convolutional neural networks for volumetric medical image segmentation, in: 2016 Fourth International Conference on 3D Vision (3DV), IEEE, 2016, pp. 565–571.

[16] A. Myronenko, 3D MRI brain tumor segmentation using autoencoder regularization, in: International MICCAI Brainlesion Workshop, Springer, 2018, pp. 311–320.

[17] B. Yu, L. Zhou, L. Wang, W. Yang, M. Yang, P. Bourgeat, J. Fripp, SA-LuT-Nets: Learning sample-adaptive intensity lookup tables for brain tumor segmentation, IEEE Trans. Med. Imaging 40 (5) (2021) 1417–1427.

[18] C. Chen, X. Liu, M. Ding, J. Zheng, J. Li, 3D dilated multi-fiber network for real-time brain tumor segmentation in MRI, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2019, pp. 184–192.

[19] Z. Jiang, C. Ding, M. Liu, D. Tao, Two-stage cascaded u-net: 1st place solution to brats challenge 2019 segmentation task, in: International MICCAI Brainlesion Workshop, Springer, 2019, pp. 231–241.

[20] R. McKinley, M. Rebsamen, R. Meier, R. Wiest, Triplanar ensemble of 3d-to-2d cnns with label-uncertainty for brain tumor segmentation, in: International MICCAI Brainlesion Workshop, Springer, 2019, pp. 379–387.

[21] R.R. Agravat, M.S. Raval, Deep learning for automated brain tumor segmentation in mri images, in: Soft Computing Based Medical Image Analysis, Elsevier, 2018, pp. 183–201.

[22] 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.

[23] V. Badrinarayanan, A. Kendall, R. Cipolla, Segnet: A deep convolutional encoder-decoder architecture for image segmentation, IEEE Trans. Pattern Anal. Mach. Intell. 39 (12) (2017) 2481–2495.

[24] F. Iandola, M. Moskewicz, S. Karayev, R. Girshick, T. Darrell, K. Keutzer, Densenet: Implementing efficient convnet descriptor pyramids, 2014, arXiv preprint arXiv:1404.1869.

[25] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.

[26] J. Chen, Y. Lu, Q. Yu, X. Luo, E. Adeli, Y. Wang, L. Lu, A.L. Yuille, Y. Zhou, Transunet: Transformers make strong encoders for medical image segmentation, 2021, arXiv preprint arXiv:2102.04306.

[27] J.M.J. Valanarasu, P. Oza, I. Hacihaliloglu, V.M. Patel, Medical transformer: Gated axial-attention for medical image segmentation, 2021, arXiv preprint arXiv:2102.10662.

[28] S. Ioffe, C. Szegedy, Batch normalization: Accelerating deep network training by reducing internal covariate shift, in: International Conference on Machine Learning, PMLR, 2015, pp. 448–456.

[29] 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.

[30] Y. Ding, L. Gong, M. Zhang, C. Li, Z. Qin, A multi-path adaptive fusion network for multimodal brain tumor segmentation, Neurocomputing (2020).

[31] L. Liu, J. Cheng, Q. Quan, F.-X. Wu, J. Wang, A survey on U-shaped networks in medical image segmentations, Neurocomputing (2020).

[32] Z. Luo, Z. Jia, Z. Yuan, J. Peng, HDC-Net: Hierarchical decoupled convolution network for brain tumor segmentation, IEEE J. Biomed. Health Inf. PP (99) (2020) 1.

[33] X. Zhao, P. Zhang, F. Song, C. Ma, G. Fan, Y. Sun, Y. Feng, G. Zhang, Prior attention network for MultiLesion segmentation in medical images, IEEE Trans. Med. Imaging (2022).

[34] Y. Ding, X. Yu, Y. Yang, RFNet: Regionaware fusion network for incomplete multi-modal brain tumor segmentation, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 3975–3984.

[35] R.R. Agravat, M.S. Raval, 3D semantic segmentation of brain tumor for overall survival prediction, in: International MICCAI Brainlesion Workshop, Springer, 2020, pp. 215–227.

[36] L.M. Ballestar, V. Vilaplana, MRI brain tumor segmentation and uncertainty estimation using 3D-UNet architectures, 2020, arXiv preprint arXiv:2012.15294.

[37] H.T. Nguyen, T.T. Le, T.V. Nguyen, N.T. Nguyen, Enhancing MRI brain tumor segmentation with an additional classification network, in: International MICCAI Brainlesion Workshop, Springer, 2020, pp. 503–513.

[38] M. Ghaffari, G. Samarasinghe, M. Jameson, F. Aly, L. Holloway, P. Chlap, E.-S. Koh, A. Sowmya, R. Oliver, Automated postoperative brain tumour segmentation: A deep learning model based on transfer learning from preoperative images, Magn. Reson. Imaging 86 (2022) 28–36.

[39] X. Guan, G. Yang, J. Ye, W. Yang, X. Xu, W. Jiang, X. Lai, 3D AGSEVNet: an automatic brain tumor MRI data segmentation framework, BMC Med. Imaging 22 (1) (2022) 1–18.

[40] F. Isensee, P.F. Jäger, P.M. Full, P. Vollmuth, K.H. MaierHein, NnUNet for brain tumor segmentation, in: International MICCAI Brainlesion Workshop, Springer, 2020, pp. 118–132.

[41] H. Jia, W. Cai, H. Huang, Y. Xia, H2NFNet for brain tumor segmentation using multimodal MR imaging: 2nd place solution to BraTS challenge 2020 segmentation task, in: Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries-6th International Workshop, Springer, 2021, pp. 58–68.

[42] R. Agravat, M.S. Raval, 3D semantic segmentation of brain tumor for overall survival prediction, 2020, arXiv preprint arXiv:2008.11576.

[43] J. Colman, L. Zhang, W. Duan, X. Ye, DR-unet104 for multimodal MRI brain tumor segmentation, 2020, arXiv preprint arXiv:2011.02840.

[44] M. Ghaffari, A. Sowmya, R. Oliver, Brain tumour segmentation using cascaded 3D densely-connected U-net, 2020, arXiv preprint arXiv:2009.07563.

[45] C.A. Silva, A. Pinto, S. Pereira, A. Lopes, Multi-stage deep layer aggregation for brain tumor segmentation, 2021, arXiv preprint arXiv:2101.00490.

Yanjun Peng received his Dr. degree in march, 2004. He joined the Department of Computer Science, Shandong University of science and technology, Qingdao, China, in 1996, where he was promoted to professor in 2010. His main research interests include image processing, deep learning and visualization.
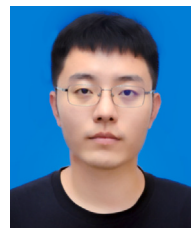
Jindong Sun is currently studying for Ph.D. at Shandong University of Science and Technology. He received the bachelor's degree for software engineering in Shandong University of Science and Technology in 2018. His research interests include deep learning, medical image processing and visualization.