J Med Syst (2019) 43: 221
https://doi.org/10.1007/s10916-019-1358-6

IMAGE & SIGNAL PROCESSING

DRRNet: Dense Residual Refine Networks for Automatic Brain Tumor Segmentation

Jiawei Sun[1] · **Wei Chen**[1] · **Suting Peng**[1] · **Boqiang Liu**[1]

Received: 26 March 2019 / Accepted: 30 May 2019 / Published online: 8 June 2019

Abstract

Glioma is one of the most common and aggressive brain tumors. Segmentation and subsequent quantitative analysis of brain tumor MRI are routine and crucial for treatment. Due to the time-consuming and tedious manual segmentation, automatic segmentation methods are required for accurate and timely treatment. Recently, segmentation methods based on deep learning are popular because of their self-learning and generalization ability. Therefore, we propose a novel automatic 3D CNN-based method for brain tumor segmentation. In order to better capture the contextual information, we design the network architecture based on u-net and replace the simple skip connection with encoder adaptation blocks. To further improve the performance and reduce computational burden at the same time, we also use dense connected fusion blocks in decoder. We train our model with generalised dice loss function to alleviate the problem of class imbalance. The proposed model is evaluated on the BRATS 2015 testing dataset and obtains dice scores of 0.84, 0.72 and 0.62 for whole tumor, tumor core and enhancing tumor, respectively. Our model is accurate and efficient, achieving results that comparable to the reported state-of-the-art results.

Keywords Brain tumor · Deep learning · CNN · Segmentation · Multi-modal MRI

Introduction

Brain tumors are formed by cells that proliferate in an abnormal and uncontrolled manner. According to a report published by the NCC [1], brain tumors rank in the top ten incident-rate cancers among both men and women, with a low 5-year survival rate. Gliomas are one of the most common brain tumors, which can be divided into two grades: low grade gliomas (LGG) with benign tendency and high grade gliomas (HGG) with malignant tendency [2]. In view of the high incidence and poor prognosis of gliomas, early detection of tumors are of great significance for treatments. Segmenting the tumor images is the crucial method to solve the above problem because images demonstrate the physiological function and anatomical structure of tumors intuitively. Magnetic resonance imaging (MRI) is the main neuroimaging protocol used for brain tumors at the present stage. MRI can obtain multi-modal imaging of the same tissue. Multi-modal imaging makes it possible to comprehensively analyze anatomical and functional information of different modalities, including T1, T2, FLAIR, and T1c, and allows accurate visualization and delineation of the structures of interest [3].

Tumors vary in shape, size, and location, whose heterogeneous appearance makes segmentation a challenging task. Manual segmentation is the gold standard for in vivo images but is complained of tediousness and subjectivity. Especially a multitude of multi-modal information is a burden and even impractical for manual segmentation. Therefore, there is a need for automated segmentation methods to provide accuracy close to that of experts' with a high consistency [4].

Various deep learning algorithms, especially convolutional neural networks (CNNs), are widely used for automatic brain tumor segmentation. Thanks to their self-learning and generalization ability over large amounts of data, CNNs automatically learn a complex hierarchy of features from images, which obtain state-of-the-art results [4–8]. To reduce memory consumption, patch-wise architecture has been proposed. The method predicts the class

This article is part of the Topical Collection on *Image & Signal Processing*

✉ Boqiang Liu
bqliu@sdu.edu.cn

[1] School of Control Science and Engineering, Shandong University, Jinan, 250061, China

of central voxels by extracting small patches of 2D slices. To further boost the performance, Havaei et al. [9] put forward a dual-path architecture. The architecture captures both local and global contextual features through simultaneous extraction of two different sized patches. However, the 2D architectures fail to fully exploit volumetric medical image data. Thus [10] propose DeepMedic, which utilizes this dual-path structure in the 3D CNN-based method for multi-scale processing. However, patch-wise architectures lack global contextual information and produce redundancy, leading to degradation in segmentation [11]. Long et al. [12] propose a fully convolutional network (FCN) for pixel-wise prediction from arbitrary-sized inputs. U-Net [13] is a popular structure build upon FCN for biomedical image segmentation which contains a contracting path to capture context and a symmetric expanding path that enables precise localization [14, 15]. Dong et al. [16] develop a 2D segmentation network based on the U-Net architecture and proves the validity of pre-processing. Kayalibay et al. [17] modify this U-Net architecture and achieve excellent results.

Deeper networks learn more representative features, yet a limitation of the U-Net architecture is its scalability [18]. Therefore, it is difficult for U-Net to achieve better segmentation results. And stacking more layers not only causes the computation burden, but also leads to the gradient vanishing. He et al. [19] address the degradation problem by introducing a deep residual learning framework, ResNet, which inserts shortcut connections for the summation of two signals: a non-linear conversion of the input and its identity mapping. These connections boost gradient flows so as to facilitate the training [20]. Xie et al. [21] present its extended version, ResNeXt, which applies a split-transform-merge strategy and introduces group convolution for better performance. However, ResNet and ResNeXt cannot thoroughly solve the gradient vanishing. Thus DenseNet [22] establishes connections from early layers to later layers by using concatenation that is different from the ResNet's summation and achieves better accuracy with fewer parameters than ResNet. The dense connections are realized by using concatenations which require an enormous amount of memory. To achieve better results, [23] propose a deep encoder-decoder architecture for 2D natural images which incorporates the structures of U-Net, ResNeXt, and dense connections, and obtains state-of-the-art performance.

ResNeXt and DenseNet are mainly applied for natural image classification tasks. However, the low resolution of medical image and fuzzy boundary of lessions make it difficult to segment accurately. Especially, too many stride operations and stacked layers in ResNeXt and DenseNet make the image fuzzier, which leads to imprecise results. To extract effective and representative features, we adopt an encoder-decoder framework which is suitable for medical image segmentation. We also embed the ResNeXt blocks and dense connections to improve performance. In addition, the skip connections, residual connections and dense connections are used to share information, allowing shallow features and deep features to be combined for a more refined segmentation. Inspired by the above ideas, we propose a 3D encoder-decoder network containing ResNeXt blocks, dense connections and refine blocks, so we call it Dense-ResNeXt-Refine Net (DRRNet). The DRRNet is evaluated on BRATS 2015 testing dataset. Our main contributions are given as follows:

1. We propose a 3D encoder-decoder network which explore spatial and global information.
2. We thoroughly explore the advantages of skip connections, residual connections and dense connections. Their 3D modifications are proposed for volumetric medical image segmentation and concatenations are replaced with summation for computation and memory reduction.
3. We conduct ablation experiments to evaluate the performance of key components.
4. We achieve comparable results for multi-modal MRI brain tumor segmentation and compare our proposed model with state-of-the-art works in BRATS 2015.

The rest of our paper is organized as follows: Section "Methods" describes the network architecture and the functions of each module. Results, ablation experiments and discussions are included in "Results and discussion". Finally we conclude in Section "Conclusion".

Methods

In this part, we will start with a brief overview of our network architecture. Then a detailed description of the included functional modules will be given.

Figure 1 illustrates our proposed DRRNet architecture employing a classic encoder-decoder structure. Each convolutional layer is followed by instance normalization [24] and a LeakyRelu [25]. In the encoding phase, there are four stacked ResNeXt blocks, each made up of three ResNeXt units. Between the residual blocks, there are three convolutional layers to implement down-sampling operations. Different from the common down-sampling by pooling layers, DRRNet adopts convolution operations with a stride of 2 to reduce the resolution of feature maps, to expand the receptive field, and to obtain more contextual information [26]. In the decoding phase, we use a dense fusion structure (DF) to better restore high-resolution details. In addition, DRRNet uses encoder adaptation blocks (EA) similar to skip connections in u-net, which adapt feature maps of encoder for decoder. Thus the forward flow and the

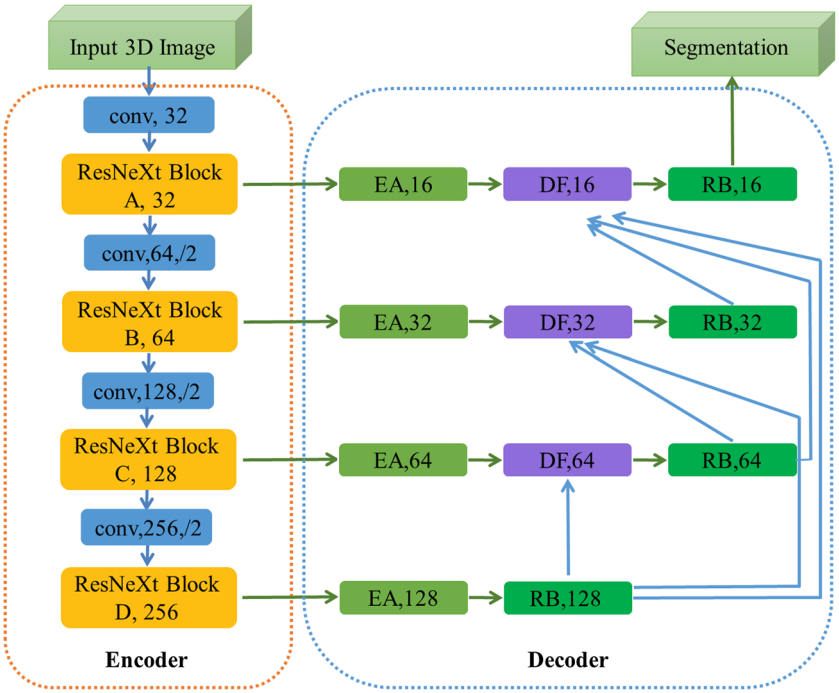

Fig. 1 Architecture of our proposed DRRNet. 'Conv' denotes a 3×3×3 convolution operation. '/2' stands for down-sampling. The number following each layer name represents the number of output channels. Each 'ResNeXt Block' consists of three ResNeXt units. 'EA' denotes skip connections. 'DF' denotes fusion layers. 'RB' represents convolutional layers with Inception structure

backward propagation of information have been facilitated [27]. Finally, refine blocks (RB) are used to further extract contextual information and optimize semantic segmentation results.

In order to fully explain the network structure, the ResNeXt blocks, EA block, DF blocks, and RB are described in detail below.

ResNeXt blocks

As one of the most successful architectures, ResNet uses shortcut connections to make inputs added to outputs of the stacked layers, which can facilitate the training of very deep networks [20]. ResNet has gained wide application in the field of computer vision, but there is little research on medical image segmentation. We explore its extended version, ResNeXt. Unlike the basic ResNet version, ResNeXt applies a split-transform-merge strategy and introduces group convolution to divide the feature maps into small groups [20]. Cardinality is put forward as a new dimension in addition to the network's depth and width, which denotes the number of groups of intermediate convolutional layer in the bottleneck block. Increasing cardinality can improve precision and is more effective than deepening or widening network structures [21]. Therefore, we choose ResNeXt blocks in this study.

However, the dataset adopted in this paper is volumetric images. In order to make full use of volume images and take advantage of deep residual learning, we extend 2D ResNeXt into a 3D variant. Each ResNeXt block consists of three ResNeXt units. The structure of a ResNeXt unit is shown in Fig. 2.

In each ResNeXt unit, the input feature maps of i-th residual unit x_i and residual function $F(x_i)$ are added together by a shortcut connection, which can be expressed as:

$$x_{i+1} = f(x_i) + x_i \tag{1}$$

Therefore the information can be propagated directly in the forward and backward directions. The first and third convolutional layers are bottleneck layers, whose kernel sizes are 1×1×1. The first bottleneck layers are applied to produce the low-dimensional embedding before the expensive 3×3×3 convolutions to reduce computations and parameters [21]. In our network, we set the input channels of the first layer as 2F, and the output channels as F. The second convolutional layer is the group convolutional layer which divides its input channels into groups. We employ small 3×3×3 convolutional kernels in the layer, which have demonstrated the state-of-the-art performance for 3D convolutional networks [28, 29]. In Fig. 2, input and output channels are divided into C groups, and convolutions are separately performed within each group, and then the group convolutional layer concatenates them as the outputs of the layer [21]. Normalization layers are inserted into the architecture intermediately for reducing internal covariate shift, hence help prevent vanishing gradient in networks with saturable nonlinearities and allow for higher learning rates [24, 28]. In our network, we choose instance normalization [24, 30] which is expressed as: let input $x \in$

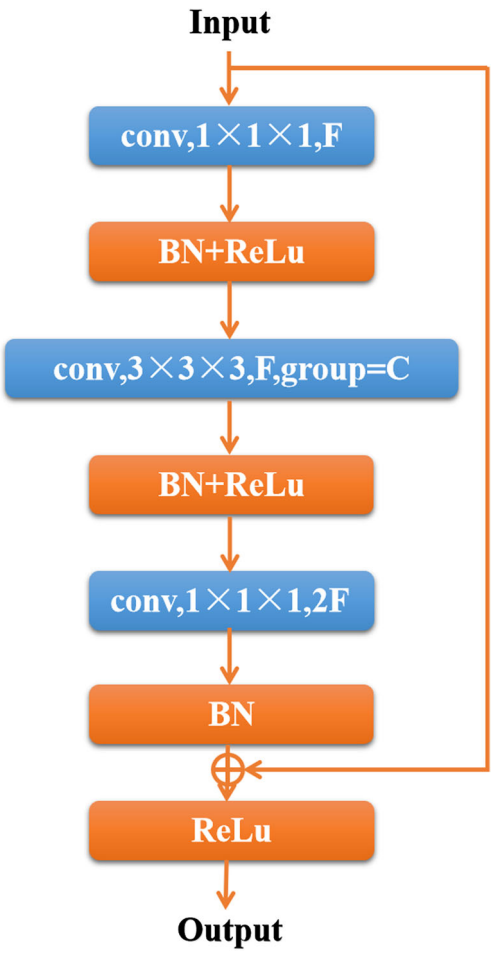

Fig. 2 Structure of each ResNeXt unit. We use conv, m×m×m to indicate the convolutional kernel size. 'F' represents the output channels. 'C' represents the number of groups, which divides the channels into small groups

$\mathbb{R}^{N\times C\times W\times H}$, x_{ncwh} denotes the element at width w, height h in c-th channel of the n-th image, then output

$$IN(x_{ncwh}) = \frac{x_{ncwh} - \mu_{nc}}{\sqrt{\sigma_{nc}^2 + \varepsilon}}, \tag{2}$$

where $\mu_{nc} = \frac{1}{HW}\sum_{w=1}^{W}\sum_{h=1}^{H} x_{ncwh}$ and $\sigma_{nc}^2 = \frac{1}{HW}\sum_{w=1}^{W}\sum_{h=1}^{H}(x_{ncwh} - h\mu_{nc})^2$. We can see that instance normalization and batch normalization have the same essence. The difference is that they apply normalization to different objects. Batch normalization is to calculate mean and variance of all images in each batch, and then to normalize; instance normalization is to calculate mean and variance of each single image in each batch [24, 30]. Therefore, instance normalization is more stable and conductive for image style transfer. And we utilize LeakyRelu as the activation function for non-linear transform because gradients of 0 in the negative part of ReLU would stop the learning, which is the "dead ReLU" problem. LeakyRelu solves the problem by using a little slope on the negative part to allow the gradients to flow on which helps to enhance the gradient flow and adjust the weights consequently [5, 25].

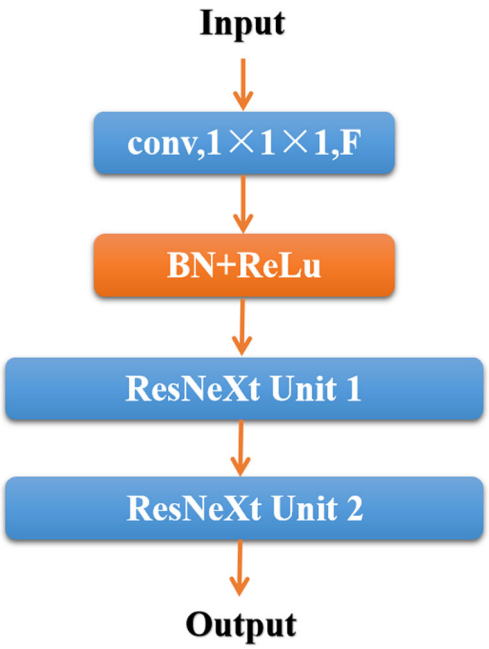

Fig. 3 Structure of an encoder adaptation block. 'F' denotes the number of output channels. 'ResNeXt unit' is detailed in Fig. 2

EA blocks

It is the initial stage of each decoder to adjust the feature size and channels for the DF blocks or the RB modules.

Its structure is shown in Fig. 3. The design of this encoder adaptation block is motivated by [23] and modified to 3D version. First, we use a 3×3×3 convolution to reduce the number of input channels. We set input and output channels are 2F and F, respectively. Two stacked residual units are then utilized to adjust feature maps for the subsequent stages.

Dense fusion blocks

It is the second stage of decoder. It fuses features from the corresponding encoder adaptations and the previous refine blocks.

Many theoretical studies have shown that increasing the depth of networks will hinder the propagation of gradients, leading to gradient vanishing. Various solutions have been proposed and share the same core idea: to create shortcut connections between layers [22]. Densenet establishes connections from early layers to later layers by using concatenation and connects all layers (with matching feature-map sizes) with each other directly [22]. The network has three advantages. First, each layer can obtain supervision information from the loss function because of

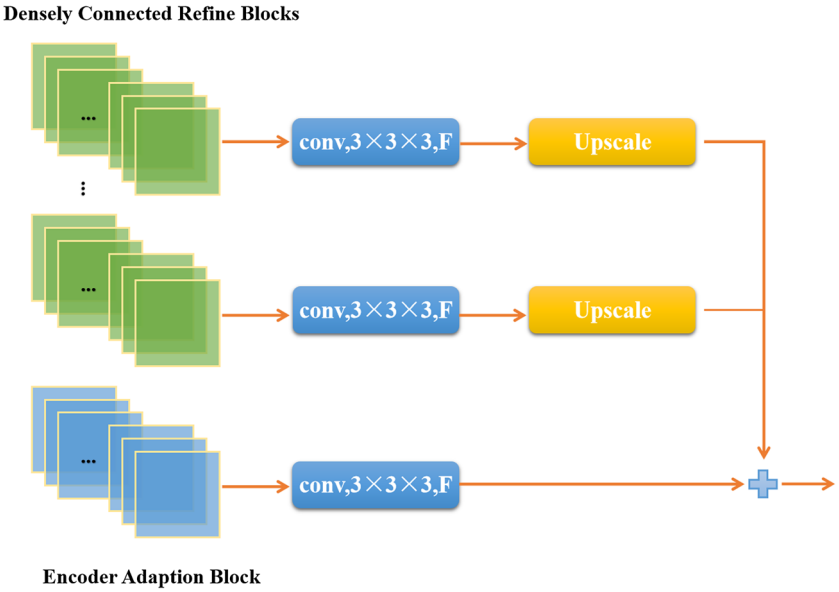

Fig. 4 Structure of a dense fusion block. Green layers denote the densely connected refine blocks and blue layers denote the encoder adaptation block

the short connections. Thus densenet can be understood as an implicit deep supervision. Second, the propagation of information and gradients in the network can be improved, which makes the network easier optimized. Third, dense connections have a certain regularization effect, which can help to avoid over-fitting with small training set.

In order to better enhance the information flow and reuse features, we use dense connection in the decoder. To further reduce the memory consumption, we replace the concatenation operation of densenet with a summation operation. Before the final fusion, we need to make sure that the feature maps have the same spatial resolution and number of channels [23]. First of all, we use a $3\times3\times3$ convolution, and simply set the output channels to F, which is the same as F in EA blocks in Fig. 3. A trilinear interpolation is employed later to upscale the small sized features. Densely connected features could be fused by pixel-wise summation eventually. The structure can be seen in Fig. 4.

Refine blocks

Refine blocks are responsible for capturing contextual information and refining semantic features. To increase the depth and width of the network and to better capture contextual information, we use Inception structures [31], which mainly consider that different sizes of convolutional kernels can enhance the adaptability of the network and merge the branches through a concatenation at the end. The split-transform-merge strategy adopted in ResNeXt is learned from the Inception module exactly. This strategy has the representative capability comparable to large and dense layers, yet at a lower computational complexity [21]. After the Inception structure, we input it into a ResNeXt unit to refine semantic features. Finally the number of output channels is adjusted by a $3\times3\times3$ convolution. The structure we designed is shown in Fig. 5.

Results and discussion

Dataset

We evaluated our model on BRATS 2015 testing dataset, which is a brain tumor image segmentation challenge at the MICCAI conference to compare the state-of-the-art methods [32]. The training set includes 220 cases with high grade (HGG) and 54 cases with low grade (LGG), and the testing set includes 110 cases with mixed grades [4]. Each case consists of four MRI sequences: FLAIR, T1, T1c, and T2. According to the results of manual segmentation, brain tumors are labeled with four labels: label 1 for necrosis, label 2 for edema, label 3 for non-enhancing tumor, and label 4 for enhancing tumor. However, the segmentation results are evaluated for the following three categories: whole tumor (all four labels), tumor core (label 1,3,4), and the enhancing tumor (label 4).

Evaluation metrics

To thoroughly evaluate the proposed model, we consider three segmentation metrics: Dice similarity coefficient (DSC), positive predictive value (PPV) and sensitivity [5]. DSC measures the overlap between the ground truth and the automatic segmentation. It is defined as,

$$DSC = \frac{2TP}{FP + 2TP + FN} \tag{3}$$

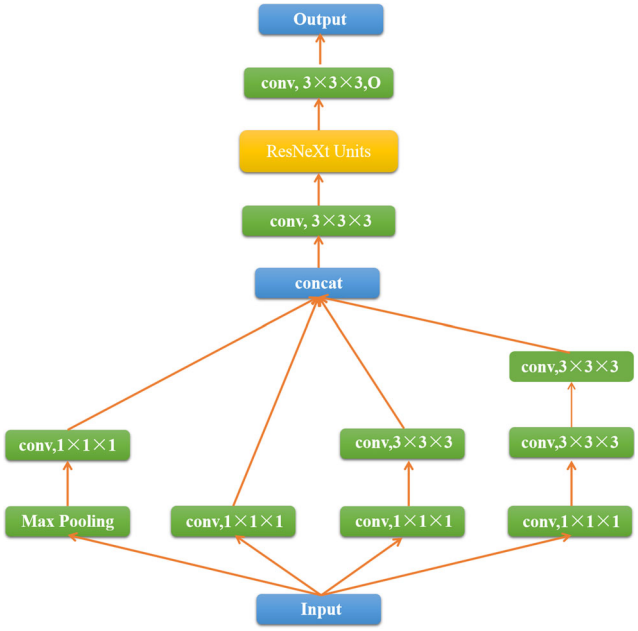

Fig. 5 Structure of a refine block. 'O' denotes the number of output channels. Except for the final refine block, 'O' is equal to the number of input channels. In the final refine block, 'O' is equal to the number of labels

where TP, FP and FN are the numbers of true positive, false positive and false negative voxels respectively. PPV is the proportion of positive results in the automatic segmentation that are true positive results, defined as,

$$PPV = \frac{TP}{TP + FP} \tag{4}$$

And sensitivity measures the proportion of true positives in the ground truth, defined as,

$$Sensitivity = \frac{TP}{TP + FN} \tag{5}$$

Experimental setting

Loss function CNN models are popular for the capability to learn distinctive features. Nevertheless, there are still some critical challenges during training the CNN models. Especially brain tumor dataset mainly bring two problems, one is the limitation of training dataset size, the other is class imbalance. In this article, we focus on alleviating class imbalance. Two methods are usually used to overcome the problem. One is undersampling of the majority class by extracting patches, and the other is using weighted loss functions [33–35]. To fully exploit volumetric image data, we utilize weighed loss function to address the problem. [36] propose Generalized Dice Loss (GDL) which is the excellent loss function for training deep convolutional neural networks. We employ this loss function and extend it into multi-class to train our Model. It takes the form:

$$GDL = 1 - \frac{2}{K}\sum_{k\in K}\frac{w_k \cdot \sum_n p_{nk} r_{nk}}{w_k \cdot \left(\sum_n p_{nk} + \sum_n r_{nk}\right)} \tag{6}$$

where p is the softmax output of the network and r is the one-hot encoding of the ground truth segmentation maps, each with K classes and N voxels. $w_k = 1/\left(\sum_{n=1}^{N} r_{kn}\right)^2$ is the weight to provide invariance for different label set properties.

Network configuration and training First we remove as many zero voxels as possible to crop the images. Then the images are resized to 128×128×128. We randomly select 28 images as a validation set each epoch and set the maximum of training epoch is 100. For each epoch, we use a batch size of 1 and draw input images in random order. Adam optimization algorithm is applied with an initial learning rate of 1e-3 and a l2 weight decay of 1e-5. The learning rate will be reduced by factor 5 if the validation loss does not decrease within 10 epochs. The training will be stopped early if the validation loss does not decrease within 20 epochs. Our model is implemented using the PyTorch deep learning framework on a NVIDIA GeForce GTX 1080Ti GPU.

Results and comparison

We randomly select three subjects in BRATS 2015 as examples and visualize the predictions in Fig. 6. The 2D slices of the axial view of MRI scan of three subjects are shown at the top, middle and bottom, respectively. From the first two rows, we notice that our model has a promising performance. Validated from the ground truth, we can see that our model could pinpoint subregions of tumors, particularly the whole tumor and enhancing tumor. However, one limitation of our model, from our observation (bottom row), is over segmentation. Our future goal is to address this problem by using pre- and post-processing.

Table 1 gives all comparison results including the proposals by [10, 37, 38] and us. From the results in Table 1, we conclude that our DRRNet compares favorably to the existing state-of-the-art on BRATS 2015. We have obtained relatively high dice scores and got a certain balance between PPV and sensitivity. Both [10] and [38] apply 3D patch-wise CNN models which are significantly inefficient due to repetitive computation. 3D conditional random field (CRF) is further applied in their post-processing to refine the predictions. Since our method does not apply post-processing, the evaluation results of these models without post-processing are listed. Xue et al. [37] use multi-scale L1 loss function to train the model. Although obtaining a

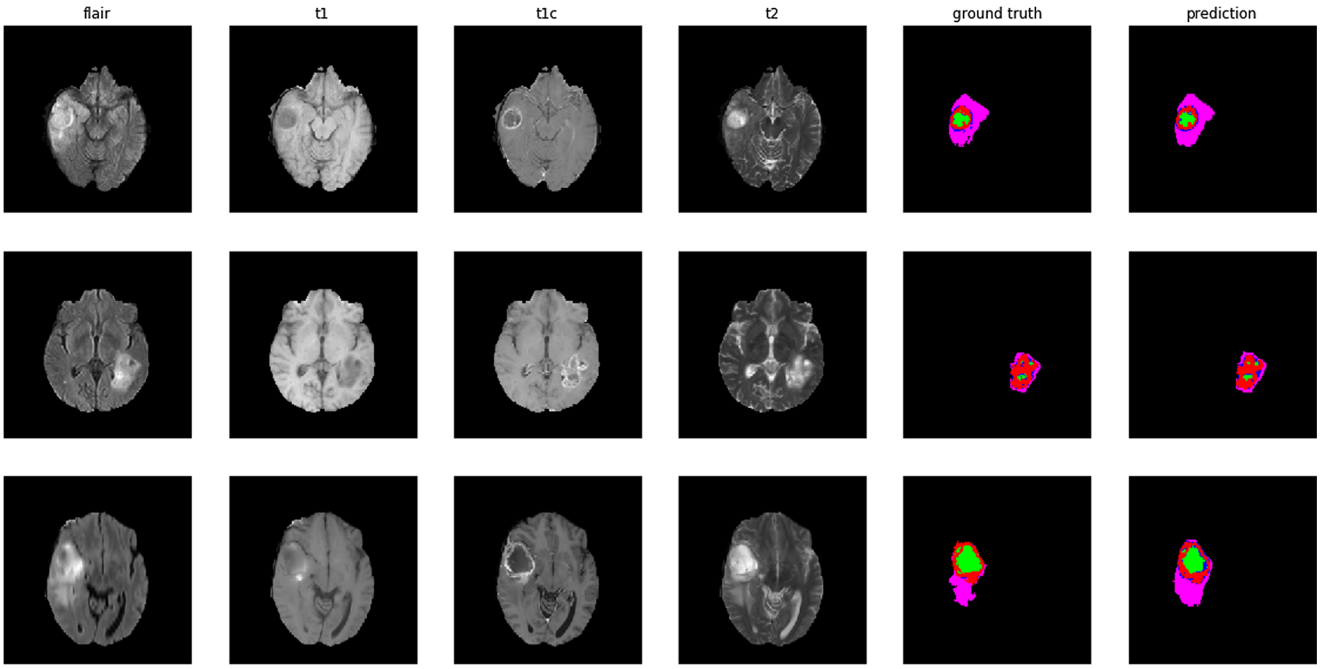

Fig. 6 Randomly selected segmentation results of three MRI subjects. The 2D slices of the axial view of MRI scan of three subjects are presented at the top, middle and bottom, respectively. From left to right: FLAIR, T1, T1c, T2, ground truth and predictions. Color: green for necrosis; purple for edema; blue for non-enhancing tumor; red for enhancing tumor

higher dice score of enhancing tumor, they fail to balance PPV and sensitivity. In our model, we incorporate multi-scale contextual information by skip connections and dense connections, which make our model more efficient and improves the performance. The proposed DRRNet achieves Dice scores of 0.84, 0.72 and 0.62 for the whole tumor, tumor core and enhancing tumor, respectively. Compared with previous models, our method achieves the highest Dice score in tumor core segmentation and the highest sensitivity scores. These results demonstrate the potential of our model in 3D multi-modal MRI segmentation tasks. The reasons for our model's good performance are three-fold. First, we directly use 3D multi-modal MRI as input, which enhances spatial, global and modal information sharing. Second, we apply skip connections, residual connections and dense connections to boost information and gradient flows. Third, instance normalization, LeakyRelu and the generalized dice loss function are employed to train the model more effectively.

Ablation study

We evaluate the effect of each key component, ResNeXt, dense connections, and refine blocks, by studying the improvement in performance. This is evaluated by the metrics (DSC, PPV and Sensitivity). The results are shown in Table 2.

With regard to the key components mentioned above, we respectively conduct comparison tests whether to use group convolution, whether to use dense connections, and whether to include Inception module. As we can see from Table 2, the addition of key components contribute to the

Table 1 The results of our proposed method and comparisons with other methods on BRATS 2015 testing dataset

	Dice			PPV			Sensitivity		
	whole	core	enhance	whole	core	enhance	whole	core	enhance
Xue et al. [37]	**0.85**	0.70	**0.66**	**0.92**	0.80	**0.69**	0.80	0.65	0.62
DeepMedic [10]	0.84	0.67	0.63	0.82	**0.85**	0.64	0.89	0.62	0.66
Zhao et al. [38]	0.82	0.72	0.62	0.84	0.78	0.60	0.83	0.73	0.69
Proposed DRRNet	0.84	**0.72**	0.62	0.82	0.77	0.60	**0.89**	**0.73**	**0.69**

The bold symbols represent the best score in each item

Table 2 The results of our ablation study on BRATS 2015 testing dataset

	Dice			PPV			Sensitivity		
	whole	core	enhance	whole	core	enhance	whole	core	enhance
Non-Group-Convolution	0.83	0.71	0.61	**0.83**	0.77	0.60	0.87	0.72	0.67
Non-Dense	0.83	0.72	0.61	0.81	0.77	0.60	0.87	0.72	0.67
Non-Inception	0.83	0.72	0.61	0.80	0.75	0.57	**0.90**	**0.76**	**0.70**
Proposed DRRNet	**0.84**	**0.72**	**0.62**	0.82	**0.77**	**0.60**	0.89	0.73	0.69

The bold symbols represent the best score in each item

improvement of dice scores, especially the whole tumor and enhancing tumor. Moreover, none-group-convolution and none-Inception gain highest scores in PPV and sensitivity respectively. Obviously, combination of group convolution and Inception module slightly reduces the value of PPV and sensitivity, but successfully reach a balance. These ablation studies suggest that the combination of key components in our network increases scores without compromising the balance between three metrics. Therefore, they all play an irreplaceable role and contribute to the improvement of performance.

Conclusion

In this paper, we present a densely connected 3D CNN-based model, DRRNet, to meet the challenge of brain tumor segmentation in multi-modal MRI. Our network is based on u-net to combine deep and shallow features. Encoder is made up of ResNeXt blocks which are verified effectiveness by ablation study. Decoder has dense fusion blocks and refine blocks. Skip connections are performed by encoder adaptation blocks. Finally, we insert instance normalization and LeakyRelu between each convolutional layer to accelerate convergence and train the model more effectively. In addition, we optimize the model by using the generalized dice loss function to solve the classification imbalance problem.

We evaluate the proposed model on the BRATS 2015 testing dataset. Our model has achieved promising performance without pre- and post-processing compared with other top methods. our ongoing study is to add pre- and post-processing in our approach, and to further fine-tune the modules so that it could have a great potential in clinical practice and serve as a preliminary step in a computer-aided diagnosis system.

Acknowledgements This work was supported by the Department of Science and Technology of Shandong Province (Grant No.2017CXGC1502).

Compliance with Ethical Standards

Conflict of interests The authors declare that they have no conflict of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

1. Zeng, H., Chen, W., Zheng, R., Zhang, S., Ji, J. S., Zou, X., Xia, C., Sun, K., Yang, Z., Li, H., and et al, Changing cancer survival in China during 2003–15: a pooled analysis of 17 population-based cancer registries. *Lancet Glob. Health* 6(5):e555–e567, 2018.
2. Wang, G., Li, W., Ourselin, S., and Vercauteren, T., Automatic brain tumor segmentation using cascaded anisotropic convolutional neural networks. In: *International MICCAI Brainlesion Workshop*, pp. 178–190: Springer, 2017.
3. Dolz, J., Gopinath, K., Yuan, J., Lombaert, H., Desrosiers, C., and Ayed, I. B., Hyperdense-net: a hyper-densely connected cnn for multi-modal image segmentation. arXiv:180402967, 2018.
4. Akkus, Z., Galimzianova, A., Hoogi, A., Rubin, D. L., and Erickson, B. J., Deep learning for brain mri segmentation: state of the art and future directions. *J. Digit. Imaging* 30(4):449–459, 2017.
5. Pereira, S., Pinto, A., Alves, V., and Silva, C. A., Brain tumor segmentation using convolutional neural networks in mri images. *IEEE Trans. Med. Imaging* 35(5):1240–1251, 2016.
6. Zhou, C., Ding, C., Lu, Z., Wang, X., and Tao, D., One-pass multi-task convolutional neural networks for efficient brain tumor segmentation. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 637–645: Springer, 2018.
7. Wang, S. H., Tang, C., Sun, J., Yang, J., Huang, C., Phillips, P., and Zhang, Y. D., Multiple sclerosis identification by 14-layer convolutional neural network with batch normalization, dropout, and stochastic pooling. *Front. Neurosci.* 12:818, 2018a.
8. Wang, S. H., Sun, J., Phillips, P., Zhao, G., and Zhang, Y. D., Polarimetric synthetic aperture radar image segmentation by convolutional neural network using graphical processing units. *J. Real-Time Image Proc.* 15(3):631–642, 2018b.
9. Havaei, M., Davy, A., Warde-Farley, D., Biard, A., Courville, A., Bengio, Y., Pal, C., Jodoin, P. M., and Larochelle, H., Brain tumor segmentation with deep neural networks. *Med. Image Anal.* 35:18–31, 2017.

10. Kamnitsas, K., Ledig, C., Newcombe, V. F., Simpson, J. P., Kane, A. D., Menon, D. K., Rueckert, D., and Glocker, B., Efficient multi-scale 3d cnn with fully connected crf for accurate brain lesion segmentation. *Med. Image Anal.* 36:61–78, 2017.
11. Imai, H., Matzek, S., Le, T. D., Negishi, Y., and Kawachiya, K., Fast and accurate 3d medical image segmentation with data-swapping method. arXiv:181207816, 2018.
12. Long, J., Shelhamer, E., and Darrell, T., Fully convolutional networks for semantic segmentation. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3431–3440, 2015.
13. Ronneberger, O., Fischer, P., and Brox, T., U-net: convolutional networks for biomedical image segmentation. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 234–241: Springer, 2015.
14. Feng, X., Tustison, N., and Meyer, C., Brain tumor segmentation using an ensemble of 3d u-nets and overall survival prediction using radiomic features. In: *International MICCAI Brainlesion Workshop*, pp. 279–288: Springer, 2018.
15. Çiçek, Ö., Abdulkadir, A., Lienkamp, S. S., Brox, T., and Ronneberger, O., 3d u-net: learning dense volumetric segmentation from sparse annotation. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 424–432: Springer, 2016.
16. Dong, H., Yang, G., Liu, F., Mo, Y., and Guo, Y., Automatic brain tumor detection and segmentation using u-net based fully convolutional networks. In: *Annual Conference on Medical Image Understanding and Analysis*, pp. 506–517: Springer, 2017.
17. Kayalibay, B., Jensen, G., and van der Smagt, P., Cnn-based segmentation of medical imaging data. arXiv:170103056, 2017.
18. Chen, L., Bentley, P., Mori, K., Misawa, K., Fujiwara, M., and Rueckert, D., Drinet for medical image segmentation. *IEEE Trans. Med. Imaging* 37(11):2453–2462, 2018.
19. He, K., Zhang, X., Ren, S., and Sun, J., Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778, 2016.
20. Hara, K., Kataoka, H., and Satoh, Y., Can spatiotemporal 3d cnns retrace the history of 2d cnns and imagenet? In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 6546–6555, 2018.
21. Xie, S., Girshick, R., Dollár, P., Tu, Z., and He, K., Aggregated residual transformations for deep neural networks. In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5987-5995: IEEE, 2017.
22. Huang, G., Liu, Z., Van Der Maaten, L., and Weinberger, K. Q., Densely connected convolutional networks. In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2261–2269: IEEE, 2017.
23. Bilinski, P., and Prisacariu, V., Dense decoder shortcut connections for single-pass semantic segmentation. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 6596–6605, 2018.
24. Ulyanov, D., Vedaldi, A., and Lempitsky, V. S., Instance normalization: the missing ingredient for fast stylization. arXiv:160708022, 2016.
25. Maas, A. L., Hannun, A. Y., and Ng, A. Y., Rectifier nonlinearities improve neural network acoustic models. In: *Proceedings icml*, Vol. 30, p. 3, 2013.
26. Zhang, R., Zhao, L., Lou, W., Abrigo, J. M., Mok, V. C., Chu, W. C., Wang, D., and Shi, L., Automatic segmentation of acute ischemic stroke from dwi using 3d fully convolutional densenets. *IEEE Trans. Med. Imaging* 37:2149–2160, 2018.
27. Kang, H., and Chen, D., Multi-scale fully convolutional network for cardiac left ventricle segmentation. arXiv:180910203, 2018.
28. Chen, H., Dou, Q., Yu, L., and Heng, P. A., Voxresnet: deep voxelwise residual networks for volumetric brain segmentation. arXiv:160805895, 2016.
29. Tran, D., Bourdev, L., Fergus, R., Torresani, L., and Paluri, M., Learning spatiotemporal features with 3d convolutional networks. In: *Proceedings of the IEEE International Conference on Computer Vision*, pp. 4489–4497, 2015.
30. Xu, Z., Yang, X., Li, X., Sun, X., and Harbin, P., Strong baseline for single image dehazing with deep features and instance normalization. In: *BMVC*, Vol. 2, p. 5, 2018.
31. Szegedy, C., Ioffe, S., Vanhoucke, V., and Alemi, A. A., Inception-v4, inception-resnet and the impact of residual connections on learning. In: *AAAI*, Vol. 4, p. 12, 2017.
32. Menze, B. H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R. et al., The multimodal brain tumor image segmentation benchmark (brats). *IEEE Trans. Med. Imaging* 34(10):1993–2024, 2015.
33. Wong, K. C., Moradi, M., Tang, H., and Syeda-Mahmood, T., 3d segmentation with exponential logarithmic loss for highly unbalanced object sizes. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 612–619: Springer, 2018.
34. Bernal, J., Kushibar, K., Asfaw, D. S., Valverde, S., Oliver, A., Martí, R., and Lladó, X., Deep convolutional neural networks for brain image analysis on magnetic resonance imaging: a review. *Artif. Intell. Med.* 95:64–81, 2018.
35. Fidon, L., Li, W., Garcia-Peraza-Herrera, L. C., Ekanayake, J., Kitchen, N., Ourselin, S., and Vercauteren, T., Generalised wasserstein dice score for imbalanced multi-class segmentation using holistic convolutional networks. In: *International MICCAI Brainlesion Workshop*, pp. 64–76: Springer, 2017.
36. Sudre, C. H., Li, W., Vercauteren, T., Ourselin, S., and Cardoso, M. J., Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. In: *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, pp. 240–248: Springer, 2017.
37. Xue, Y., Xu, T., Zhang, H., Long, L. R., and Huang, X., Segan: adversarial network with multi-scale l 1 loss for medical image segmentation. *Neuroinformatics* 16(3-4):383–392, 2018.
38. Zhao, X., Wu, Y., Song, G., Li, Z., Zhang, Y., and Fan, Y., A deep learning model integrating fcnns and crfs for brain tumor segmentation. *Med. Image Anal.* 43:98–111, 2018.