Nested Dilation Networks for Brain Tumor Segmentation Based on Magnetic Resonance Imaging

Liansheng Wang[1,2], Shuxin Wang[2], Rongzhen Chen[2], Xiaobo Qu[3], Yiping Chen[1,2], Shaohui Huang[2]* and Changhua Liu[4]*

[1] Fujian Key Laboratory of Sensing and Computing for Smart City, School of Information Science and Engineering, Xiamen University, Xiamen, China, [2] Department of Computer Science, School of Information Science and Engineering, Xiamen University, Xiamen, China, [3] Department of Electronic Science, Fujian Provincial Key Laboratory of Plasma and Magnetic Resonance, School of Electronic Science and Engineering (National Model Microelectronics College), Xiamen University, Xiamen, China, [4] Department of Medical Imaging, Chenggong Hospital Affiliated to Xiamen University, Xiamen, China

Aim: Brain tumors are among the most fatal cancers worldwide. Diagnosing and manually segmenting tumors are time-consuming clinical tasks, and success strongly depends on the doctor's experience. Automatic quantitative analysis and accurate segmentation of brain tumors are greatly needed for cancer diagnosis.

Methods: This paper presents an advanced three-dimensional multimodal segmentation algorithm called nested dilation networks (NDNs). It is inspired by the U-Net architecture, a convolutional neural network (CNN) developed for biomedical image segmentation and is modified to achieve better performance for brain tumor segmentation. Thus, we propose residual blocks nested with dilations (RnD) in the encoding part to enrich the low-level features and use squeeze-and-excitation (SE) blocks in both the encoding and decoding parts to boost significant features. To prove the reliability of the network structure, we compare our results with those of the standard U-Net and its transmutation networks. Different loss functions are considered to cope with class imbalance problems to maximize the brain tumor segmentation results. A cascade training strategy is employed to run NDNs for coarse-to-fine tumor segmentation. This strategy decomposes the multiclass segmentation problem into three binary segmentation problems and trains each task sequentially. Various augmentation techniques are utilized to increase the diversity of the data to avoid overfitting.

Results: This approach achieves Dice similarity scores of 0.6652, 0.5880, and 0.6682 for edema, non-enhancing tumors, and enhancing tumors, respectively, in which the Dice loss is used for single-pass training. After cascade training, the Dice similarity scores rise to 0.7043, 0.5889, and 0.7206, respectively.

Conclusion: Experiments show that the proposed deep learning algorithm outperforms other U-Net transmutation networks for brain tumor segmentation. Moreover, applying cascade training to NDNs facilitates better performance than other methods. The findings of this study provide considerable insight into the automatic and accurate segmentation of brain tumors.

Keywords: brain tumor segmentation, nested dilation networks, residual blocks nested with dilations, squeeze-and-excitation blocks, coarse-to-fine

1. INTRODUCTION

Brain tumors are one of the deadliest cancers worldwide. Gliomas are the most common primary craniocerebral tumor and are caused by the carcinogenesis of glial cells in the brain and spinal cord (Bauer et al., 2013). In pathology, gliomas can be classified as low-grade or high-grade according to the malignant degree of the tumor cells (Cho and Park, 2017; Wang et al., 2018b). Low-grade gliomas are mainly represented by low-speed cell division and proliferation, whereas high-level gliomas are characterized by rapid cell division and proliferation accompanied by angiogenesis, hypoxia, and necrosis (Gerlee and Nelander, 2012; Bogdańska et al., 2017). Although significant advances have been made in healthcare so far, the vast majority of gliomas are incurable, except for a small number of low-grade gliomas, which can be completely resected surgically. Gliomas can be further divided into different tumor sub-regions according to the severity of the tumor cells, such as edemas, non-enhancing tumors, and enhancing tumors. Magnetic resonance imaging (MRI) is the most frequently used and most effective noninvasive auxiliary diagnostic tool (Wen et al., 2010; Yang et al., 2018), providing a reference for the formulation of treatment programs (Mazzara et al., 2004). Brain tumors are usually imaged with different MRI modalities, and these images are interpreted by image analysis methods (Bauer et al., 2013). The MRI sequence usually includes four different modalities: T1-weighted, T2-weighted, post-contrast T1-weighted, and fluid-attenuated inversion-recovery (FLAIR). Different MRI modalities are employed for different diagnosis tasks in clinical diagnosis and treatment. However, it is still a daunting task for clinicians to diagnose diseases with MRI, because there is a wide variation in the size, shape, regularity, location, and heterogeneous appearance of brain tumors (Dong et al., 2017). Therefore, automatic quantitative analysis and accurate segmentation of brain tumors are greatly needed clinically to help doctors make accurate diagnoses.

CNNs have become a prominent deep learning method and have been used to make a series of breakthroughs in different tasks, including computer vision (Krizhevsky et al., 2012; Long et al., 2015; Ren et al., 2015). The success of CNNs is credited to their ability to independently learn deep features instead of relying on manual features. With historical opportunities provided by a strong calculation capability and large numbers of annotations, the development of CNNs has been explosive. The original LeNet5 (LeCun et al., 1998) was proposed in 1998 with five layers, establishing the modern structure of CNNs. Krizhevsky et al. (2012) presented a classical CNN structure called "AlexNet, and made a historic breakthrough. The great success of AlexNet stimulated new research on CNNs. ZFNet (Zeiler and Fergus, 2014), VGGNet (Simonyan and Zisserman, 2014), GoogLeNet (Szegedy et al., 2015), and ResNet (He et al., 2016) were successively presented with more layers and better performances. Huang et al. (2017) used a more radical dense connection mechanism to maximize the flow of information. Hu et al. (2017) proposed an SE network that modeled the interdependencies between feature channels, adaptively learning important information. All of these CNN studies made it possible to apply neural networks to medical image processing.

Recent reports have shown that CNNs outperform state-of-the-art medical image analyses (Li et al., 2017; Lin et al., 2018). MRI-based brain tumor segmentation is a task that still requires extensive attention. Extant methods for automatic brain tumor segmentation are diverse. DeepMedic (Kamnitsas et al., 2016b) was designed as a dual-pathway three-dimensional (3D) network with 11 layers, to simultaneously process images at different scales and combine the results with fully connected layers. Kamnitsas et al. (2016a) and Castillo et al. (2017) further improved the architecture of DeepMedic by adding residual connections and parallel pathways. U-Net (Ronneberger et al., 2015) was proposed to train an end-to-end network with few images for the accurate segmentation of biomedical images. Many architectures similar to U-Net have been widely adopted for brain tumor segmentation. Kayalibay et al. (2017) and Isensee et al. (2017) employed deep supervision by combining segmentation layers from different levels in the localization pathway. Iqbal et al. (2018) increased the number of U-Net layers and trained the network with the Dice loss. Le and Pham (2018) used the U-Net architecture to extract features and put them into an ExtraTrees classifier. Zhao et al. (2018) integrated fully convolutional neural network (FCNN) and conditional random field (CRF) and trained three models using two-dimensional (2D) image patches obtained from axial, coronal, and sagittal views. A voting-based fusion strategy was used to obtain segmentation results. To deal with the class imbalance problem, Wang et al. (2017) proposed a triple-cascaded framework for brain tumor segmentation. Three similar networks were used to segment the entire tumor (all lesions, including edema, non-enhancing tumors, and enhancing tumors), and the tumor core (all lesions except edema). They then sequentially enhanced tumor core. Zhou et al. (2018) drew upon lesions with coarse-to-fine medical image segmentation methods and proposed a single multitask CNN that could learn correlations between different categories. Partial model parameters can be shared when different tasks are being trained according to different sets of training data to utilize the underlying correlation among classes.

We propose a CNN-based 3D segmentation algorithm, the NDN, which can handle multimodal images. Instead of simple convolution layers, residual blocks are stacked in the U-Net architecture to simplify optimization. The SE blocks used in NDNs fuse the global information and adaptively learn important information from each channel. A new block i.e., residual blocks nested with dilations (RnD) enlarges the receptive fields and avoids the gridding effect. RnD blocks can enrich information in shallow layers by using dilation convolutions while retaining detailed information during the rapid expansion of receptive fields by using residual connections. The cascade training strategy is adopted to train three tasks individually to deal with the class imbalance problem.

2. MATERIALS AND METHODS

This section describes the proposed NDNs algorithm for detailed brain tumor segmentation, including the data preprocessing,

network architecture, training strategy, and post-processing methods. We also concisely describe the experimental design.

2.1. Data Acquisition and Preprocessing

2.1.1. Data Acquisition
Most of the data used in this work are downloaded from the Medical Segmentation Decathlon (MSD) organized by the 21st Annual Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI) 2018. A small number of low-grade glioma data are abstained from MICCAIs Multimodal Brain Tumor Segmentation (BraTS) Challenge of the same year. These are used to test the stability of the proposed algorithm. The images for each patient comprise four scanning sequences: T1-weighted, T2-weighted, post-contrast T1-weighted, and FLAIR. Every scan is aligned to the same anatomical template space and interpolated into $1 \times 1 \times 1mm^3$ with an image size of $240 \times 240 \times 155$ voxels. The purpose of the study is to segment brain tumors (i.e., gliomas) into three different classes: edemas, non-enhancing tumors, and enhancing tumors. All data are labeled and verified by an expert human rater. Efforts were made to mimic the accuracy required for clinical use.

2.1.2. Data Preprocessing and Augmentation
Training an effective neural network requires thousands or even tens of thousands of data. However, the quantity of available medical images is usually well short of that. To avoid overfitting, more training data need to be generated from the limited images and annotations. Our method applies the following data augmentation techniques to make reasonable changes to the image shapes: flip the x-, y-, or z-axis with a probability of 50%; rotate the images with a rotation angle of $-15°$ to $15°$; apply gamma correction with the gamma value varied randomly from 0.4 to 1.6; and apply elastic distortion. **Figure 1** shows the data augmentation.

Images from multiple modalities may have varying intensity ranges. When the intensity values are not standardized, it is detrimental to the training of the neural network. Normalization is critical to allow images from different modalities to be trained with one algorithm. In our study, each modality is normalized individually by subtracting the mean from the value for each patient and dividing it by the standard deviation. The useless black borders in the images along the x- and y-axes are also removed. On the z-axis, we note that the head and tail of the image slices are uninformative. Therefore, 70% of the slices used for the network input are captured from the middle.

2.2. Residual Blocks
He et al. (2016) reformulated the layers as residual blocks and yielded unusually brilliant results in the 2015 ImageNet competition. Instead of simply stacking convolution layers to fit a desired underlying mapping, they added identity mapping, which was easier to optimize. The residual blocks depicted in **Figure 2A** are achieved by a shortcut connection and element-wise addition operation, performed on the input and output feature maps of the blocks, channel-by-channel. The operating principle of the residual blocks can be defined as

$$y = F(x, W_i) + x, \tag{1}$$

where x and y are the input and output vectors of the relevant layers; and $F(x, W_i)$ is the mapping function for the residual path. The results of $F(x, W_i)$ should have the same dimensions as x. Otherwise, we can perform linear mapping on the shortcut connection (**Figure 2B**). This simple algorithm does not add additional parameters or computations to the network, but it greatly increases the training speed of the model and improves the training effect.

The standard convolutional layers of a U-Net are replaced by the residual structure shown in **Figure 2A**. The residual path comprises two convolution layers with a kernel size of 3, followed by a batch normalization (BN) operation (Ioffe and Szegedy, 2015) and a rectified linear unit (ReLU). The input and output of the residual path are added element by element. The results of the residual blocks are fed directly into subsequent network layers.
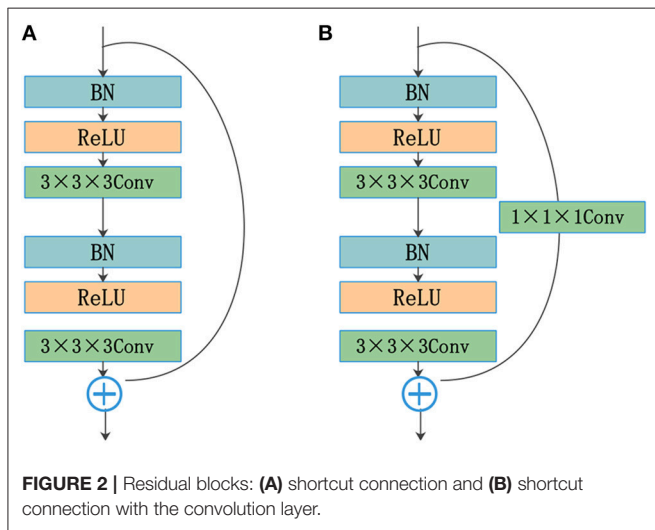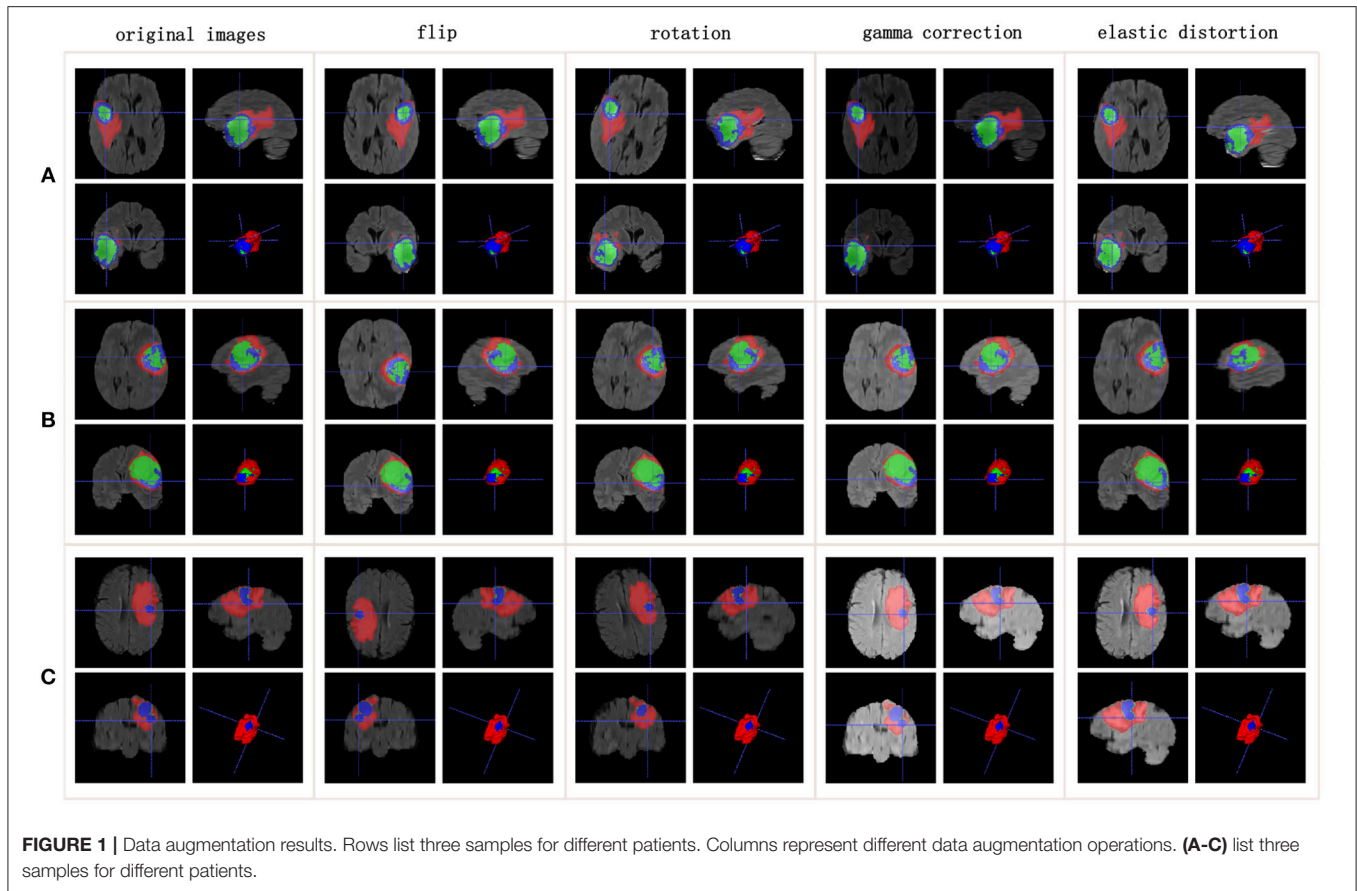
2.3. SE Blocks
A lot of research has recently been accomplished to strengthen the learning power of CNNs and to improve their performance. Hu et al. (2017) introduced the SE blocks to enhance the representations of features produced by a convolutional network. SE blocks embed the global spatial information into the channel vector by encoding each channel dependency with a fully connected operation. It allows the network to pay different amounts of attention to each channel according to the importance of the feature maps. **Figure 3A** illustrates the structure of SE blocks. The features are first passed through a squeeze operation achieved by a global average pooling layer to aggregate global information per channel for the whole image. Then, the outputs are fed into an excitation operation to get the final weights for each channel. The excitation operation is achieved by using two fully connected layers: one with ReLU activation and another with a sigmoid. Finally, the weight vectors are reshaped to $(1, 1, 1, C)$, where C is the number of the feature maps and are applied to each feature map by the multiply operation. The SE blocks emphasize useful features and suppress useless features through weights like an attention mechanism.

SE blocks have a simple structure and can be used directly in existing state-of-the-art architectures. We draw on experience with the attention mechanism and nested SE blocks in the architecture to help the network focus on important feature maps. As shown in **Figure 3B**, feature maps with size (X, Y, Z, C) are put into SE blocks. Then, the blocks generate a significant coefficient for each channel, finally gaining outputs with different weights and the same size as the inputs.

2.4. RnD Blocks
The traditional up-sampling and down-sampling structures lead to a loss of internal structure, and the information of small objects cannot be reconstructed. To solve this problem, Yu and Koltun (2015) presented a model with dilated convolutions that can increase the receptive fields without reducing the resolution or increasing the parameters. Chen et al. (2014,

FIGURE 1 | Data augmentation results. Rows list three samples for different patients. Columns represent different data augmentation operations. **(A-C)** list three samples for different patients.

FIGURE 2 | Residual blocks: **(A)** shortcut connection and **(B)** shortcut connection with the convolution layer.
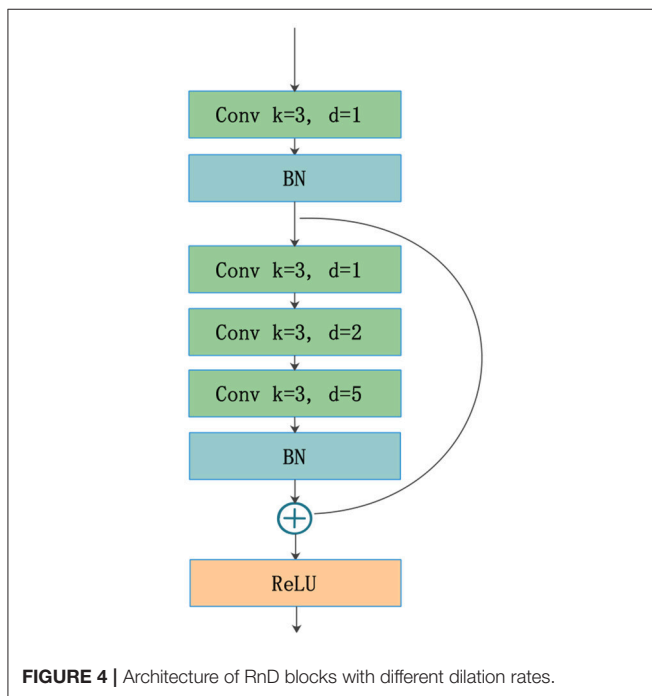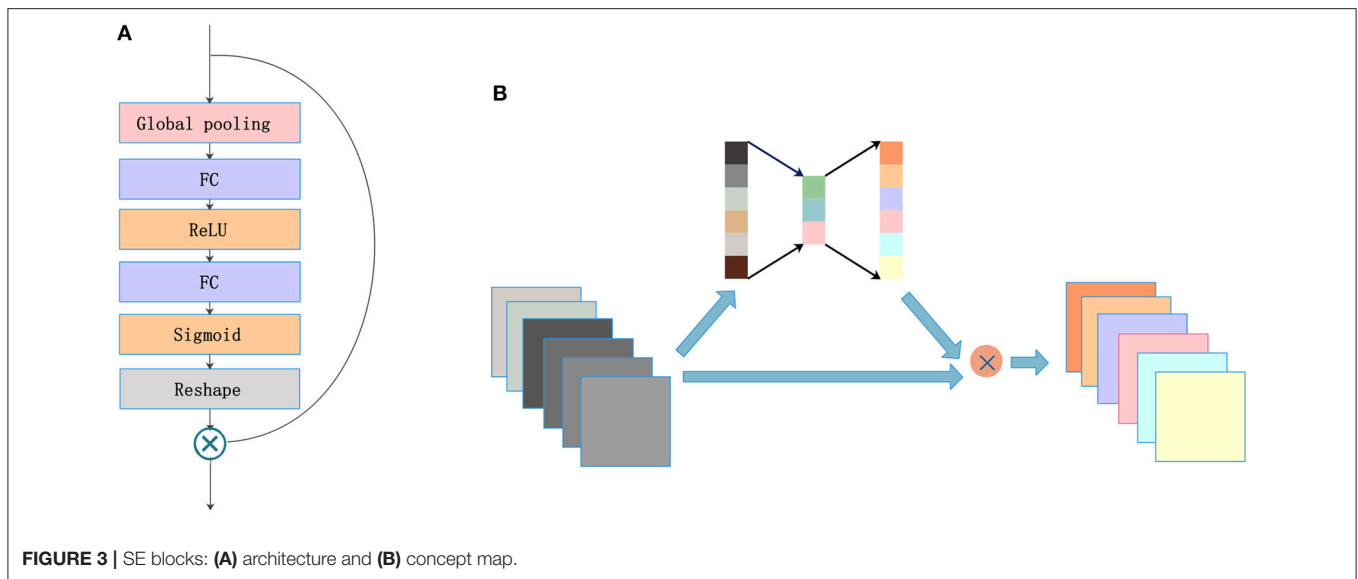
2017, 2018) used dilated convolutions in their networks and achieved good performance for dense prediction tasks. However, standard dilated convolution causes a gridding issue that will harm small objects. Wang et al. (2018a) proposed a hybrid dilated convolution (HDC) framework, which can not only expand receptive fields but also mitigate the gridding issue.

Implementing the HDC framework requires two conditions to be met. First, the dilation rates of a groups dilated convolutions should not have a common divisor > 1. The maximum distance between two nonzero values is defined as follows:

$$M_i = max[M_{i+1} - 2r_i, M_{i+1} - 2(M_{i+1} - r_i), r_i], \quad (2)$$

where r_i is the dilation rate in layer i, and M_i is the maximum dilation rate from layer 0 to layer i. The second condition requires satisfying $M_i < K$, where K is the kernel size.

The standard U-Net architecture does not get enough semantic information in the shallow layers because of the limited receptive fields. This is harmful to feature fusion in the first few cross-layer connections. To resolve this issue and avoid the gridding effect, we draw on an idea from the HDC framework. RnD blocks (**Figure 4**) are built to enlarge receptive fields in the first two layers of the network. This new type of block can obtain more extensive local information via 3 convolution layers with different dilation rates (e.g., 1, 2, 5). The kernel size is 3 for all dilated convolutions, which are followed by a ReLU activation. The residual connection in RnD blocks helps retain information and fill details during the rapid expansion of receptive fields.

FIGURE 3 | SE blocks: **(A)** architecture and **(B)** concept map.

FIGURE 4 | Architecture of RnD blocks with different dilation rates.

2.5. NDNs

The structure of our proposed NDNs is shown in **Figure 5**. The architecture is inspired by U-Net, which is a stable encoder–decoder network designed for limited data training, especially for medical images. Here, we carefully modify the standard U-Net to make it perform better for the brain tumor segmentation task. First, we use 3D convolution layers rather than 2D to adapt images from multiple modalities. The classic encoder–decoder structure that fuses the lower features in the shallow layers and higher features in the deep layers is retained to ensure the stability

of the proposed network. The architecture comprises three max-pooling layers to capture context and three up-sampling layers to enable precise localization. To obtain enough receptive fields, the first two encoder modules adopt RnD blocks to enrich the low-level features. This is followed by an SE block and a max-pooling layer. In the decoder part, each module comprises a stack of residual blocks, an SE block, and an up-sampling layer. The BN is employed immediately after each convolution and before activation. As shown in **Figure 5**, the network can obtain rich information to boost essential features and achieve a stable effect.

2.6. Cascade Training

The cascade strategy trains different models for each category sequentially, showing ideal results. Coarse-to-fine medical image segmentation is becoming increasingly popular because of the class imbalance problem. Cascaded models decompose complex problems into simple ones and capitalize on the hierarchical structure of tumor sub-regions. A single model is trained repeatedly to segment substructures of brain tumors hierarchically and sequentially. Each sequence is handled as a binary segmentation problem. The first task is to segment the entire tumor including edemas, enhancing tumors, and non-enhancing tumors. These three classes are regarded as a binary segmentation problem. Then, NDNs are trained to crop the target. After the first stage of training, the entire tumor region is segmented in the 3D volumes of a patient. A cuboid sub-region, based on the entire tumor, is used as inputs to the network to segment the enhancing and non-enhancing tumors together. Similarly, the third training differentiates enhancing tumors from non-enhancing ones by using the cuboid sub-region produced by the second stage as input.

In the training, the input of the network is generated based on the ground truth, as shown in **Figure 6A**. In the testing, the results of the previous stage are extended by 32 pixels on the x- and y-axes, and 8 slices on the z-axis as the input for the

FIGURE 5 | Architecture of the proposed NDNs with SE blocks and RnD blocks.
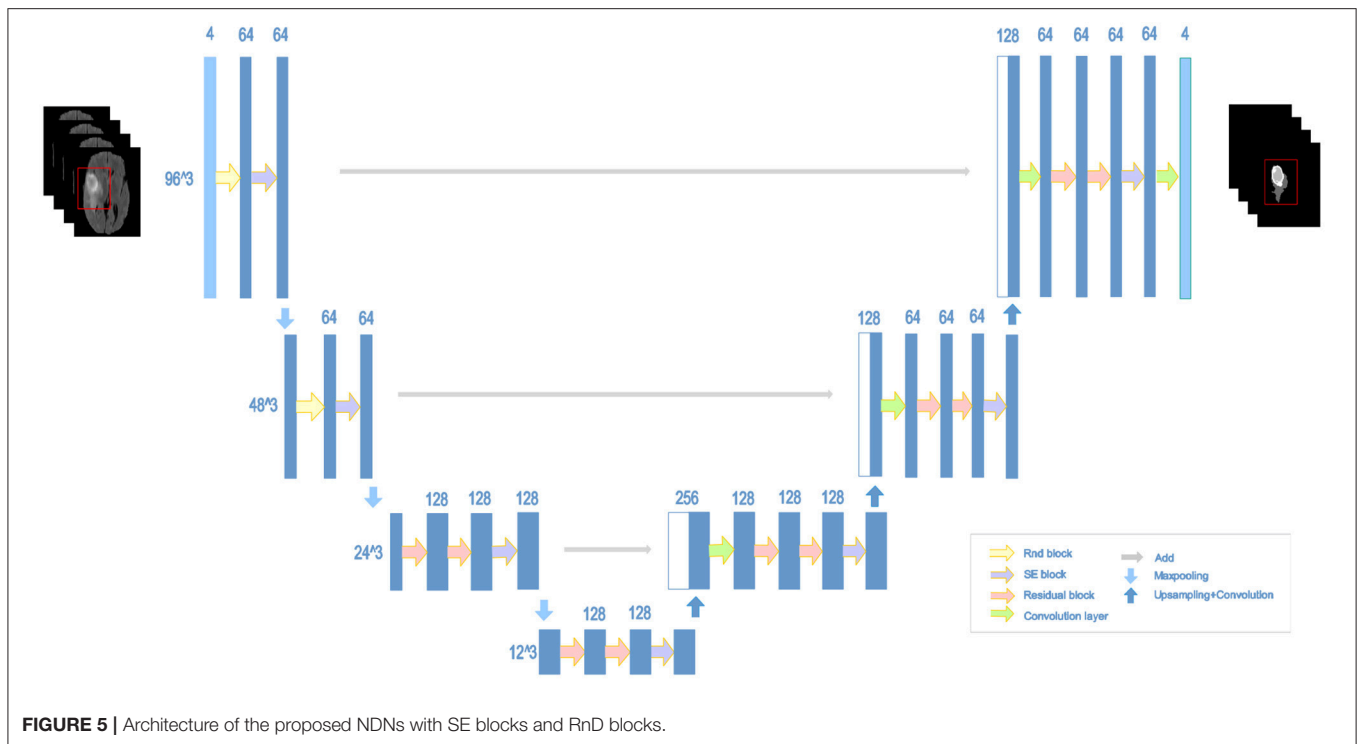
next stage. The process is described in **Figure 6B**. Finally, we integrate the three binary segmentation tasks to obtain the final segmentation results of multiple classes. Cascade training offers a way to adaptively alleviate the class imbalance problem of brain tumor segmentation.

2.7. Post-processing

Post-processing is further used to improve the segmentation results of NDNs. During data processing, we noticed that the brain tumors for all patients in the 3D volumes were of a single connected domain. Thus, isolated small clusters should be removed from the results. More specifically, connected domain analysis should be performed to retain the maximal region and remove other smaller clusters to better fit the ground truth. Moreover, some patients are observed to have benign tumors, which means that the gliomas only comprise edemas and non-enhancing tumors. Some small clusters are erroneously classified as enhancing tumors in our task instead of benign tumors, which harms the segmentation results. To deal with this issue, we impose volumetric constraints by removing enhancing tumor clusters in the segmentation that are smaller than a predefined threshold.

2.8. Dice Similarity Score

In our work, the Dice similarity score is calculated for quantitative evaluation. This performance metric measures the similarity between the ground truth and predicted results. The Dice similarity score is defined as follows:

$$DSC = \frac{2TP}{(FP + 2TP + FN)}, \qquad (3)$$

where TP, FP, and FN are the numbers of true positives, false positives, and false negatives, respectively.

MSD and BraTS 2018 provide three different tumor regions that can be described as edemas, enhancing tumors, and non-enhancing tumors. The Dice similarity scores are calculated for each tumor region to evaluate the segmentation results, and the scores are compared with those of other methods.

2.9. Experimental Design

We conduct three groups of experiments according to different requirements, which we describe in this section.

Experiment 1: We explored the effects of different network structures on brain tumor segmentation. Ronneberger et al. (2015) developed a U-Net architecture based on the fully convolutional network (FCN) (Long et al., 2015), which can work with very few training images and yield more precise segmentation. Some new architectures derived from U-Net have appeared and have been applied to the field of medical image processing. In Experiment 1, the standard Conv + BN + ReLU module in U-Net was replaced by frequently used blocks, such as residual blocks and dense blocks separately for comparison with the proposed NDNs.

Experiment 2: Different loss functions were attempted with NDNs to improve segmentation results. The loss function quantifies the amount by which the predicted value deviates from the actual value. Choosing a suitable loss function benefits both the training process and the results of brain tumor segmentation. In Experiment 2, different loss functions were applied to the brain tumor segmentation task: the categorical cross-entropy loss, Dice loss, and focal loss. The Dice similarity scores are calculated for each task.

Experiment 3: The proposed method was compared with other state-of-the-art methods. We implemented several previously published algorithms and trained the networks with the same datasets. The brain tumors comprise of edemas, enhancing tumors, and non-enhancing tumors with very different volumes, resulting in an imbalanced number of samples in each class. This category imbalance problem impairs the performance of a deep network. In Experiment 3, a cascade strategy was used to train NDNs, which decomposed a multiple classification problem into multiple binary classification problems. The segmentation results of the cascaded NDNs were compared with several state-of-the-art methods according to the Dice similarity score.

2.10. Implementation Details

All networks were implemented in Keras (Chollet et al., 2015) 2.1.2 using the Tensorflow (Abadi et al., 2016) 1.4.0 backend. Adaptive moment estimation (Kingma and Ba, 2014) was used as an optimizer with an initial learning rate of 0.0001, a momentum of 0.9, and a weight decay of 0.00001. Training was implemented on an NVIDIA 1080 Ti GPU with a version of CUDA 8.0 for 300 epochs. We did not use a dropout (Hinton et al., 2012) but rather L2 regularization and BN for the whole network structure. We cropped $96 \times 96 \times 48$ patches as inputs close to the ground truth from images and annotations. All networks were trained from scratch with a batch size of 4.

3. EXPERIMENTS AND RESULTS

In this section, we explain the advantages of the proposed algorithm with regard to brain tumor segmentation. The Dice similarity score is adopted as the evaluation criterion for each model. Edemas, non-enhancing tumors, and enhancing tumors were trained together with single NDNs in Experiments 1 and 2 for the sake of fairness. In Experiment 3, however, the cascade training strategy was used to train NDNs for each class, which was then compared with the state-of-the-art methods.

3.1. Experiment 1

To prove the effectiveness of the NDNs structure, different U-Net-like networks were trained with the same brain tumor dataset for comparison. A traditional 3D U-Net with three down-sampling layers and three symmetric up-sampling layers was trained first. It consisted of two convolution layers used repeatedly with a kernel size of 3, similar to the standard 2D U-Net structure presented by Ronneberger et al. (2015). The

TABLE 1 | Comparison of different U-Net-like architectures: (A) standard 3D U-Net; (B) U-Net with residual blocks; (C) U-Net with dense blocks; (D) NDNs without SE blocks; (E) NDNs without RnD blocks; and (F) NDNs network.

Dice Methods		Edema	Non-enhancing tumor	Enhancing tumor
A		0.6686	0.4734	0.6169
	ResNet18	0.6645	0.5022	0.6455
B	ResNet50	**0.6792**	0.5314	0.6289
	ResNet101	0.6752	0.5617	0.6342
C		0.6734	0.5527	0.6287
D		0.6590	0.5612	0.6305
E		0.6725	0.536	0.638
F		0.6652	**0.5880**	**0.6682**

The bold values mean the best result (the highest dice value) for each class.
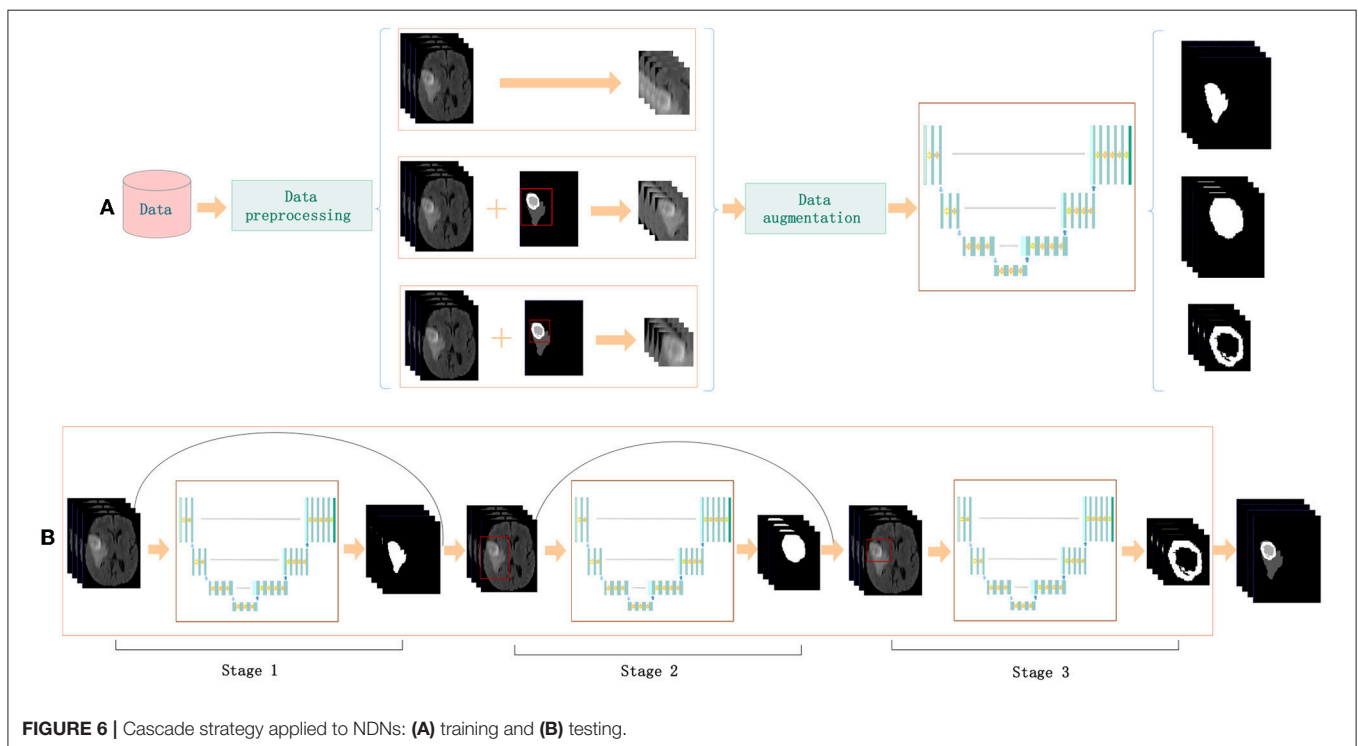
FIGURE 6 | Cascade strategy applied to NDNs: **(A)** training and **(B)** testing.

FIGURE 7 | Boxplots for each method in **Table 1**. Dice similarity scores for **(A)** edema, **(B)** non-enhancing tumors, and **(C)** enhancing tumors. The symbol "×" marks the mean.

FIGURE 8 | Brain tumor segmentation results predicted by different U-Net-like networks. The rows represent three samples from different patients, and the columns represent results predicted by each U-Net-like network. The organizers provided the ground truth images. **(A–C)** list three samples for different patients.

filter number was doubled at the end of each down-sampling layer and halved after each up-sampling layer. Then, the repeated convolution layers were replaced by residual blocks and dense blocks to be trained. ResNet18, ResNet50, and ResNet101 were each employed as an encoder path, and the decoder path was consistent with the expanding path in 3D U-Net. For the dense U-Net, dense blocks were used as substitutes for the two repeated convolution layers, and each dense block had four dense connected convolution layers. Finally, we studied the effect of the NDNs architecture with SE blocks or RnD blocks only. **Table 1** lists the Dice similarity scores calculated for brain tumor segmentation with these networks, and **Figure 7** presents the boxplots for each class. Note that all networks were trained with the Dice loss in Experiment 1.

We achieved better results for non-enhancing tumor segmentation and enhancing tumor segmentation with NDNs than with the other U-Net-like architectures. According to

TABLE 2 | Comparison with different losses: (A) categorical cross-entropy; (B) weighted categorical cross-entropy loss; (C) focal loss; and (D) Dice loss.

Dice Methods	Edema	Non-enhancing tumor	Enhancing tumor
A	**0.6708**	0.5321	0.6579
B	0.6634	0.5774	0.6604
C	0.5905	0.5721	0.6445
D	0.6652	**0.5880**	**0.6682**

The bold values mean the best result (the highest dice value) for each class.

Table 1, the non-enhancing tumor results segmented by NDNs are about 2.6% better than U-Net with ResNet101, and the enhancing tumor segmentation results are at least 2.0% better than the other methods. However, the proposed algorithm lacked the ability to segment the edema part with a result of 0.6652, which is worse than the other U-Net-like algorithms. **Figure 8**

presents the ground truth and prediction results for different U-Net-like architectures from different perspectives.

3.2. Experiment 2

Class imbalance is a severe issue in medical image segmentation and needs to be carefully tackled. The data provided by MSD and BraTS 2018 are heavily imbalanced, especially the classes of the non-enhancing tumors and enhancing tumors. To alleviate the class imbalance, we use a Dice loss function. We also explore the effects of other loss functions on NDNs for comparison. The categorical cross-entropy is used as a base loss function:

$$Crossentropy(p, q) = -\frac{1}{N} \sum_{x,y,z} \sum_{k} p_{x,y,z}^{k} \log q_{x,y,z}^{k}, \quad (4)$$

where $p_{x,y,z}^{k}$ and $q_{x,y,z}^{k}$ correspond to the ground truth and predicted results for class k, and N is the total number of samples. Based on previous experience, the class imbalance can be addressed by associating different weights with individual classes. Therefore, the weighted categorical cross-entropy is also used:

$$W\_Crossentropy(p, q) = -\frac{1}{N} \sum_{x,y,z} \sum_{k} w^{k} p_{x,y,z}^{k} \log q_{x,y,z}^{k}, \quad (5)$$

where w^{k} is the weight for class k. Here the weights for the background, edema, non-enhancing tumors, and enhancing tumors are defined as $(1, 1, 2, 1)$ respectively. The focal loss function described by Lin et al. (2017) for dense object detection is a modified version of binary cross-entropy and is aimed toward low-confidence labels. We adopt a multiclass focal loss for the segmentation task:

$$Focal(p, q) = -\frac{\sum_{x,y,z} \sum_{k} p_{x,y,z}^{k}(1 - q_{x,y,z}^{k})^{\gamma} \log q_{x,y,z}^{k}}{\sum_{x,y,z} \sum_{k} p_{x,y,z}^{k}}, \quad (6)$$

where $(1 - q_{x,y,z}^{k})^{\gamma}$ is a modulating factor and the value of γ is set to 2.0 in our algorithm. Finally, our proposed model is trained with the following Dice loss to segment different parts of the brain tumors:

$$Dice(p, q) = 1 - \frac{1}{N} \frac{2 \sum_{x,y,z} \sum_{k} p_{x,y,z}^{k} * q_{x,y,z}^{k}}{\sum_{x,y,z} \sum_{k} p_{x,y,z}^{k} + \sum_{x,y,z} \sum_{k} q_{x,y,z}^{k}}. \quad (7)$$

The Dice similarity scores for the different loss functions used in NDNs are presented in **Table 2** and **Figure 9**. We obtain final scores of 0.6652, 0.5880, and 0.6682 for edemas, non-enhancing tumors, and enhancing tumors, respectively, using the Dice loss. Normal loss functions like the categorical cross-entropy may achieve good results for balanced datasets, but datasets with a massive imbalance among classes require special attention. We avoid weighted categorical cross-entropy as much as possible, because it needs additional hyperparameters that may introduce another difficult problem for network optimization. The results show that the focal loss may be good for binary

TABLE 3 | Comparison of methods for the same dataset: (A) Isensee et al. (2017); (B) Iqbal et al. (2018); (C) Wang et al. (2017); (D) Zhou et al. (2018); and (E) our proposed method training with the cascade strategy.

Dice Methods	Edema	Non-enhancing tumor	Enhancing tumor
A	0.6574	0.5418	0.6943
B	0.6808	0.5727	0.6661
C	0.6919	0.5504	0.6793
D	0.6894	0.5376	0.6861
E	**0.7043**	**0.5889**	**0.7206**

The bold values mean the best result (the highest dice value) for each class.

classification problems to solve intra-class imbalance. However, it is less helpful for inter-class imbalance. The Dice loss is calculated based on the Dice coefficient and can deal with situations with large amounts of class imbalance. **Figure 10** shows the ground truth and prediction results for the different loss functions used in NDNs.

3.3. Experiment 3

We reproduced several state-of-the-art methods for brain tumor segmentation for comparison with our algorithm. Isensee et al. (2017) achieved a high Dice score in the BraTS 2017 Challenge by using a U-Net-like architecture. They employed deep supervision in the localization pathway to integrate segmentation layers at different levels of the network and combined them via element-wise summation to form the final network output. Iqbal et al. (2018) adopted SE blocks at the end of the decoder part and fused its output with the output of encoder blocks. These two methods were chosen for comparison, because they have similarities with our network structure. Wang et al. (2017) proposed a triple-cascaded framework to segment the entire tumor, tumor core, and enhancing tumor core sequentially. They used dilated convolutions after the down-sampling layers and set the dilation parameter from 1 to 3. Zhou et al. (2018) presented a one-single multitask CNN that can learn the correlations between different categories. These two methods used a cascade or cascade-like training strategy like our training process, and they both obtained high Dice scores in the brain tumor segmentation task. In this experiment, a multiclass segmentation problem was decomposed into three binary segmentation problems by repeated training of NDNs with the coarse-to-fine method just like (Wang et al., 2017). **Table 3** and **Figure 11** present the quantitative evaluation according to the Dice similarity scores for the same datasets.
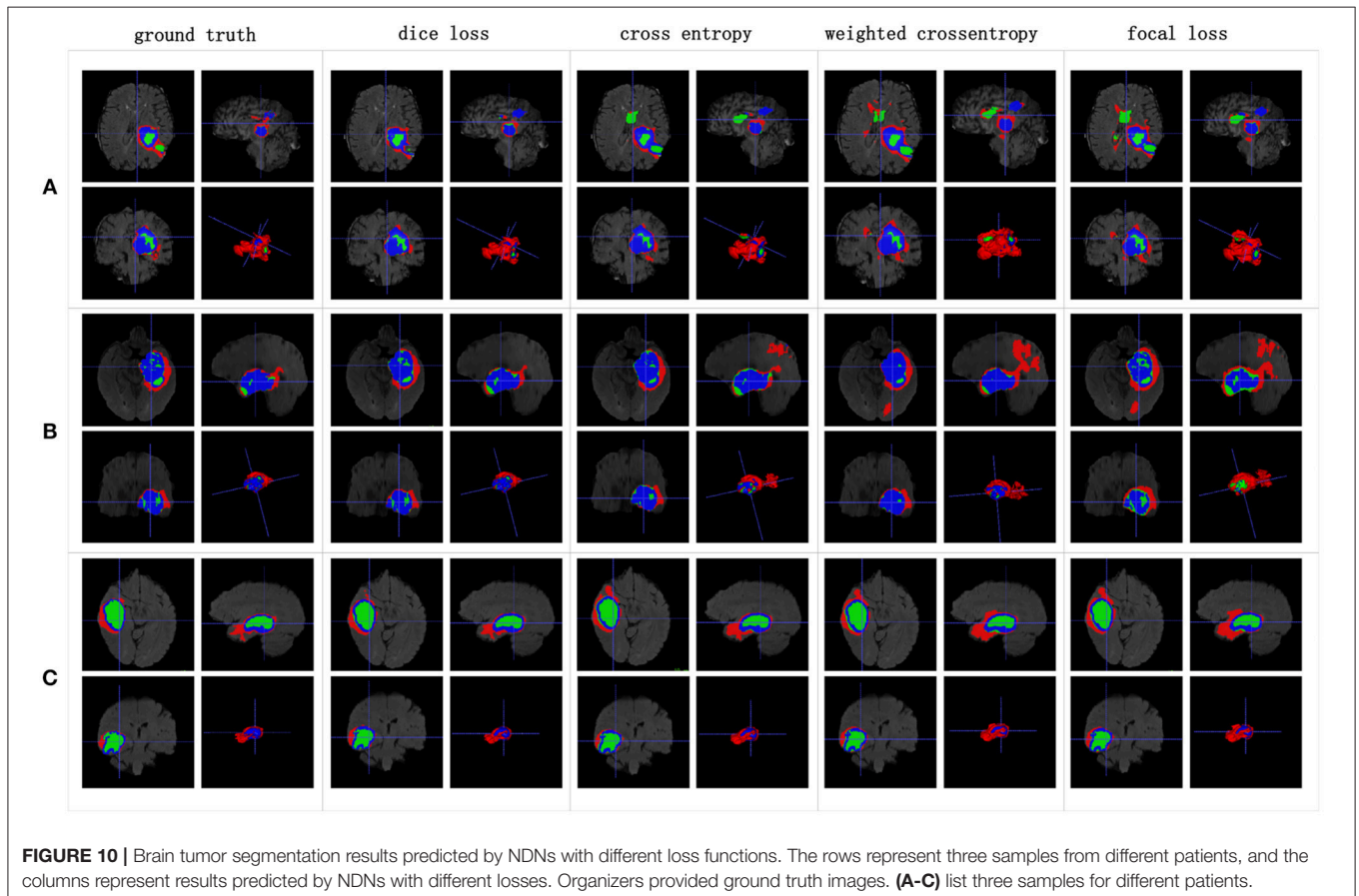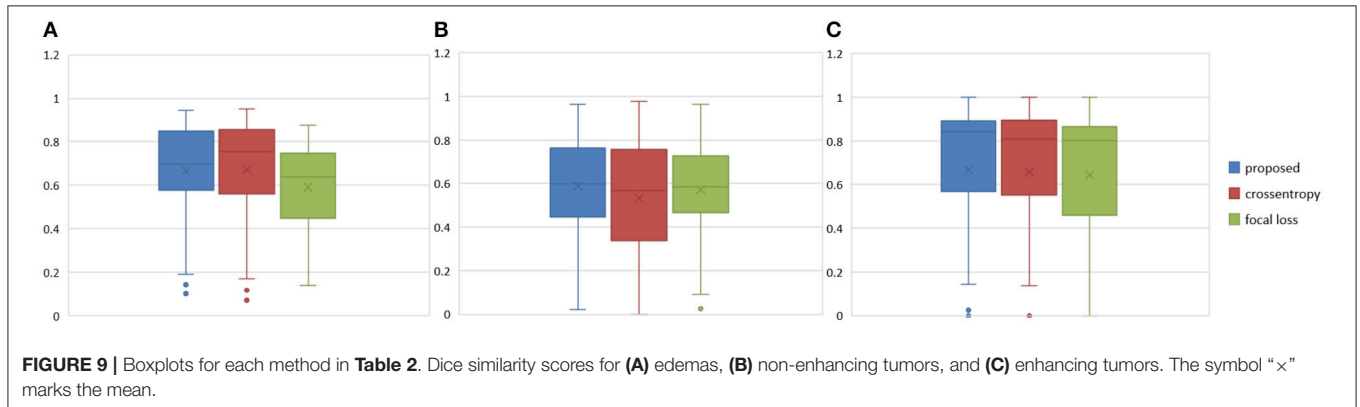
Table 3 indicates that the Dice similarity scores of our proposed method are 0.7043, 0.5889, and 0.7206 for edemas, non-enhancing tumors, and enhancing tumors, respectively, which are higher than those of all comparison methods for every class. Moreover, the results are 3.9 and 5.2% higher for edemas and enhancing tumors than when the three classes are trained together, and the results for the non-enhancing tumors do not worsen. These results prove that the cascade training strategy can improve the accuracy for brain tumor segmentation. **Figure 12** shows the ground truth and prediction results for different state-of-the-art methods.

4. DISCUSSION

4.1. Competitive Segmentation Results

U-Net increases the number of up-sampling and skip connections compared with FCN, which can supplement more location information for semantic information. The U-Net architecture has received increasing attention recently and has been shown that it is a stable algorithm for many segmentation tasks. Despite its great success, however, U-Net still has limitations for some specialized tasks.

We found that stacking residual blocks instead of simple convolution layers can improve the brain tumor segmentation performance. This is because residual blocks can fuse receptive fields of different sizes and ease the training of the networks. Attention mechanisms have shown their utility for many computer vision tasks. SE blocks work as an attention mechanism that can explore the relationship between channels to suppress useless information and enhance useful information by fusing global information. They can help a network notice essential features and make correct decisions. Nesting the SE blocks into our base structure causes the corresponding Dice similarity
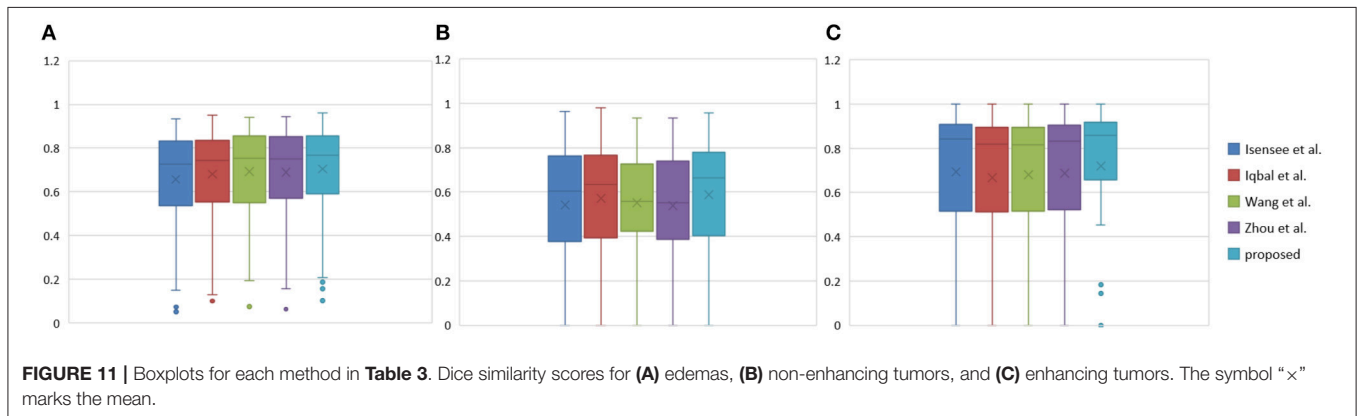

FIGURE 9 | Boxplots for each method in **Table 2**. Dice similarity scores for **(A)** edemas, **(B)** non-enhancing tumors, and **(C)** enhancing tumors. The symbol "×" marks the mean.

FIGURE 10 | Brain tumor segmentation results predicted by NDNs with different loss functions. The rows represent three samples from different patients, and the columns represent results predicted by NDNs with different losses. Organizers provided ground truth images. **(A-C)** list three samples for different patients.

FIGURE 11 | Boxplots for each method in **Table 3**. Dice similarity scores for **(A)** edemas, **(B)** non-enhancing tumors, and **(C)** enhancing tumors. The symbol "×" marks the mean.
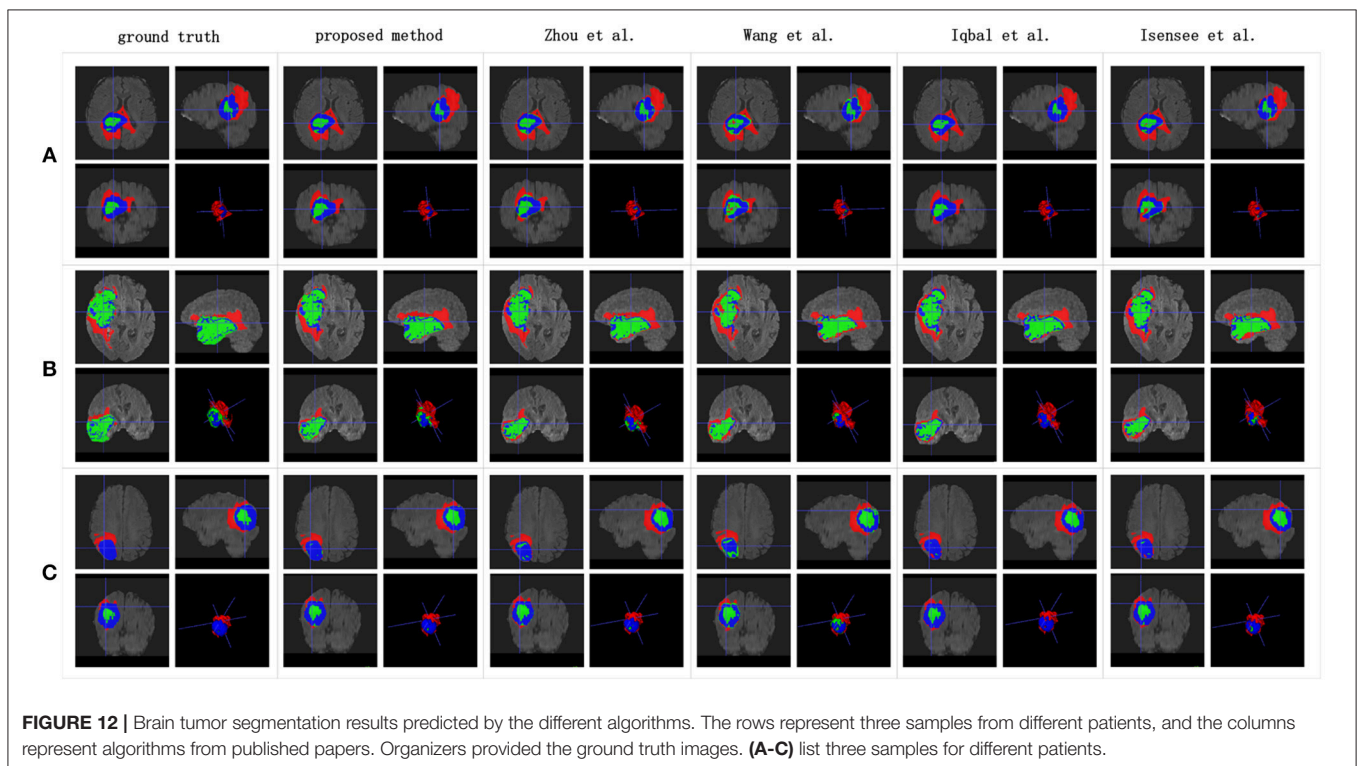

FIGURE 12 | Brain tumor segmentation results predicted by the different algorithms. The rows represent three samples from different patients, and the columns represent algorithms from published papers. Organizers provided the ground truth images. **(A-C)** list three samples for different patients.

scores of the edemas, non-enhancing tumors, and enhancing tumors to reach 0.6725, 0.536, and 0.638, respectively. To solve the problem of insufficient receptive fields and to simultaneously avoid the gridding issue, we add RnD blocks to the network. By learning from the HDC framework, RnD blocks can enlarge receptive fields by using dilated convolutions with different dilation rates. Based on this, our method obtains results of 0.6652, 0.5880, and 0.6682, respectively.

An extreme imbalance between categories affects the segmentation results, especially for edemas, and needs to be addressed. Non-enhancing tumors usually have smaller regions than the other two classes, as shown in **Figure 14**, which will have a negative effect on the segmentation results. In order to alleviate the class imbalance, two measures are taken. First, different losses

are employed by NDNs to determine the best performance, and a Dice loss function is eventually selected. Moreover, we borrowed the cascade training strategy adopted by many state-of-the-art methods for brain tumor segmentation. Cascade training can balance the quantitative differences among different classes to some extent. The final results obtained by our proposed method were 0.7043 for edema, 0.5889 for non-enhancing tumors, and 0.7206 for enhancing tumors. The experimental results are shown in **Figure 13**. These reasonable results are attributed to both the network structure and training strategies.

4.2. Limitations

This study is limited by the class imbalance problem, even though some measures have been taken to alleviate it. Some small

FIGURE 13 | Histogram for each method. Dice similarity scores for **(A)** edemas, **(B)** non-enhancing tumors, and **(C)** enhancing tumors.

FIGURE 14 | Limitations caused by the class imbalance problem. **(A,B)** present two different samples.
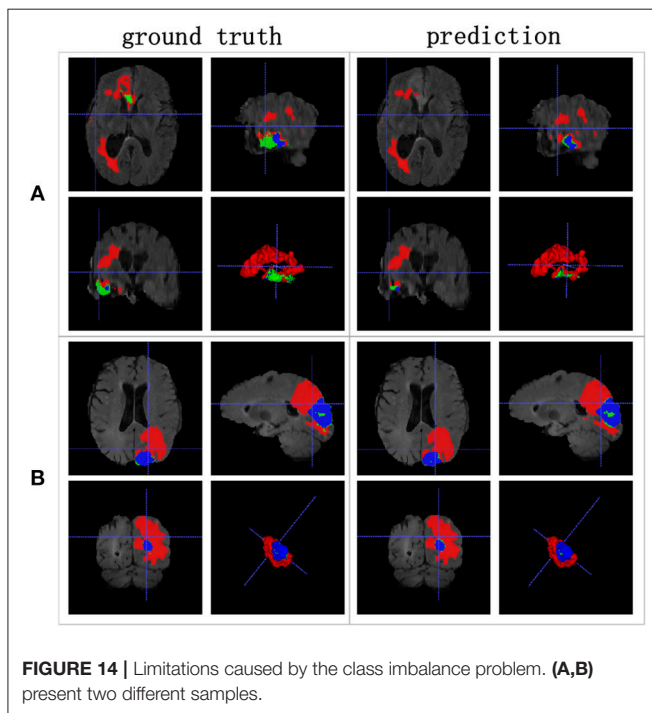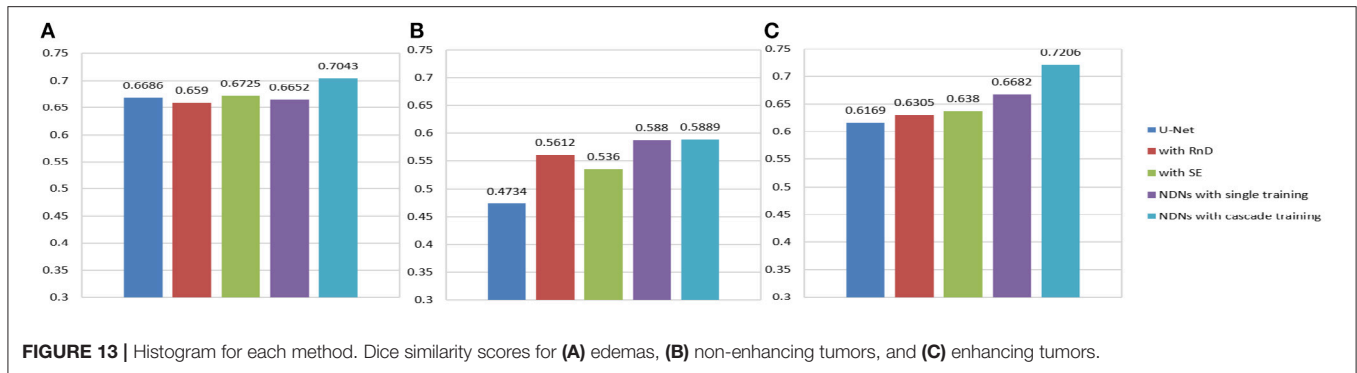
regions in brain tumors like non-enhancing tumors could not be predicted very well. For example, in the two samples in **Figure 14**, only 8.5% of the entire tumor is non-enhancing in sample A and 2.25% in sample B. This huge category imbalance lead to inaccurate segmentation results of 0.279 and 0.402 for non-enhancing tumors in samples A and B, respectively. The class imbalance problem remains a challenge that should be addressed in the future.

5. CONCLUSION

Clinical applications of computer-aided systems have gained a great deal of research attention. Supremely accurate brain tumor segmentation is a tedious but vital task for clinicians because of various sizes and shapes of tumors. Quantitative analysis of brain tumors is critical to relieve pressure on doctors

and obtain more accurate segmentation results. We developed a new deep learning framework based on U-Net, NDNs, for segmenting brain tumors. Our results showed that NDNs can extract discriminative features of edemas, non-enhancing tumors, and enhancing tumors by obtaining large receptive fields and integrating channel information. Compared with other state-of-the-art methods, NDNs obtained higher Dice similarity scores. The proposed method makes it possible to generate accurate segmentation result for brain tumors without manual interference and provides considerable insight on the application of computer-aided systems to clinical tasks.

DATA AVAILABILITY

The datasets for this study can be found at Medical Segmentation Decathlon, and the low-grade data can be downloaded from MICCAI BraTS 2018.

AUTHOR'S NOTE

AUTHOR CONTRIBUTIONS

FUNDING

REFERENCES

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., et al. (2016). "Tensorflow: a system for large-scale machine learning," in *OSDI'16 Proceedings of the 12th USENIX Conference on Operating Systems Design and Implementation*, Vol. 16 (Savannah, GA), 265–283.

Bauer, S., Wiest, R., Nolte, L.-P., and Reyes, M. (2013). A survey of mri-based medical image analysis for brain tumor studies. *Phys. Med. Biol.* 58:R97. doi: 10.1088/0031-9155/58/13/R97

Bogdańska, M. U., Bodnar, M., Piotrowska, M. J., Murek, M., Schucht, P., Beck, J., et al. (2017). A mathematical model describes the malignant transformation of low grade gliomas: prognostic implications. *PLoS ONE* 12:e0179999. doi: 10.1371/journal.pone.0179999

Castillo, L. S., Daza, L. A., Rivera, L. C., and Arbeláez, P. (2017). "Volumetric multimodality neural network for brain tumor segmentation," in *13th International Conference on Medical Information Processing and Analysis*, Vol. 10572 (International Society for Optics and Photonics), 105720E.

Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K., and Yuille, A. L. (2014). Semantic image segmentation with deep convolutional nets and fully connected crfs. *arXiv preprint arXiv:1412.7062*.

Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K., and Yuille, A. L. (2018). Deeplab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE Trans. Pattern Anal. Mach. Intel.* 40, 834–848. doi: 10.1109/TPAMI.2017.2699184

Chen, L.-C., Papandreou, G., Schroff, F., and Adam, H. (2017). Rethinking atrous convolution for semantic image segmentation. *arXiv preprint arXiv:1706.05587*.

Cho, H.-H., and Park, H. (2017). "Classification of low-grade and high-grade glioma using multi-modal image radiomics features," in *Engineering in Medicine and Biology Society (EMBC), 2017 39th Annual International Conference of the IEEE* (Jeju Island: IEEE), 3081–3084.

Chollet, F., Rahman, F., Lee, T., de Marmiesse, G., Zabluda, O., and Pumperla, M. (2015). *Keras*. https://github.com/fchollet/keras (accessed February 02, 2019).

Dong, H., Yang, G., Liu, F., Mo, Y., and Guo, Y. (2017). "Automatic brain tumor detection and segmentation using u-net based fully convolutional networks," in *Annual Conference on Medical Image Understanding and Analysis* (Edinburgh: Springer), 506–517.

Gerlee, P., and Nelander, S. (2012). The impact of phenotypic switching on glioblastoma growth and invasion. *PLoS Comp. Biol.* 8:e1002556. doi: 10.1371/journal.pcbi.1002556

He, K., Zhang, X., Ren, S., and Sun, J. (2016). "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Las Vegas, NV), 770–778.

Hinton, G. E., Srivastava, N., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. R. (2012). Improving neural networks by preventing co-adaptation of feature detectors. *arXiv preprint arXiv:1207.0580*.

Hu, J., Shen, L., and Sun, G. (2017). Squeeze-and-excitation networks. *arXiv preprint arXiv:1709.01507*.

Huang, G., Liu, Z., Van Der Maaten, L., and Weinberger, K. Q. (2017). "Densely connected convolutional networks," in *Computer Vision and Pattern Recognition*, Vol. 1 (Honolulu), 3.

Ioffe, S., and Szegedy, C. (2015). Batch normalization: accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167*.

Iqbal, S., Ghani, M. U., Saba, T., and Rehman, A. (2018). Brain tumor segmentation in multi-spectral mri using convolutional neural networks (cnn). *Microsc. Res. Tech.* 81, 419–427. doi: 10.1002/jemt.22994

Isensee, F., Kickingereder, P., Wick, W., Bendszus, M., and Maier-Hein, K. H. (2017). "Brain tumor segmentation and radiomics survival prediction: Contribution to the brats 2017 challenge," in *International MICCAI Brainlesion Workshop*, (Quebec City, QC: Springer), 287–297.

Kamnitsas, K., Ferrante, E., Parisot, S., Ledig, C., Nori, A. V., Criminisi, A., et al. (2016a). "Deepmedic for brain tumor segmentation," in *International Workshop on Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries* (Athens: Springer), 138–149.

Kamnitsas, K., Ledig, C., Newcombe, V. F., Simpson, J. P., Kane, A. D., Menon, D. K., et al. (2016b). Efficient multi-scale 3d cnn with fully connected crf for accurate brain lesion segmentation. *arXiv [Preprint]. arXiv:1603.05959*.

Kayalibay, B., Jensen, G., and van der Smagt, P. (2017). Cnn-based segmentation of medical imaging data. *arXiv [Preprint]. arXiv:1701.03056*.

Kingma, D. P. and Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv [Preprint]. arXiv:1412.6980*.

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems*, (Lake Tahoe, NV), 1097–1105.

Le, H. T. and Pham, H. T.-T. (2018). Brain tumour segmentation using u-net based fully convolutional networks and extremely randomized trees. *Vietnam J. Sci. Tech. Eng.* 60, 19–25. doi: 10.31276/VJSTE.60(3).19

LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proc. IEEE* 86, 2278–2324.

Li, X., Chen, H., Qi, X., Dou, Q., Fu, C.-W., and Heng, P. A. (2017). H-denseunet: Hybrid densely connected unet for liver and liver tumor segmentation from ct volumes. *arXiv [Preprint.] arXiv:1709.07330*. doi: 10.1109/TMI.2018.2845918

Lin, T. Y., Goyal, P., Girshick, R., He, K., and Dollar, P. (2017). Focal loss for dense object detection. *IEEE Trans. Pattern Anal. Mach. Intel.* 99, 2999–3007. doi: 10.1109/TPAMI.2018.2858826

Lin, W., Tong, T., Gao, Q., Guo, D., Du, X., Yang, Y., et al. (2018). Convolutional neural networks-based mri image analysis for the alzheimers disease prediction from mild cognitive impairment. *Front. Neurosci.* 12:777. doi: 10.3389/fnins.2018.00777

Long, J., Shelhamer, E., and Darrell, T. (2015). "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Boston, MA), 3431–3440.

Mazzara, G. P., Velthuizen, R. P., Pearlman, J. L., Greenberg, H., and Wagner, H. N. (2004). Brain tumor target volume determination for radiation treatment planning through automated mri segmentation. *Int. J. Radiat. Oncol. Biol. Phys.* 59, 300–312. doi: 10.1016/j.ijrobp.2004.1.026

Ren, S., He, K., Girshick, R., and Sun, J. (2015). "Faster r-cnn: towards real-time object detection with region proposal networks," in *Advances in Neural Information Processing Systems* (Montréal, QC), 91–99.

Ronneberger, O., Fischer, P., and Brox, T. (2015). "U-net: convolutional networks for biomedical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, (Springer), 234–241.

Simonyan, K., and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv [Preprint]. arXiv:1409.1556*.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., et al. (2015). "Going deeper with convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Boston, MA), 1–9.

Wang, G., Li, W., Ourselin, S., and Vercauteren, T. (2017). "Automatic brain tumor segmentation using cascaded anisotropic convolutional neural networks," in *International MICCAI Brainlesion Workshop* (Quebec City, QC), 178–190.

Wang, P., Chen, P., Yuan, Y., Liu, D., Huang, Z., Hou, X., et al. (2018a). "Understanding convolution for semantic segmentation," in *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)* (Lake Tahoe, NV/CA: IEEE), 1451–1460.

Wang, X., Wang, D., Yao, Z., Xin, B., Wang, B., Lan, C., et al. (2018b). Machine learning models for multiparametric glioma grading with quantitative result interpretations. *Front. Neurosci.* 12:1046. doi: 10.3389/fnins.2018.01046

Wen, P. Y., Macdonald, D. R., Reardon, D. A., Cloughesy, T. F., Sorensen, A. G., Galanis, E., et al. (2010). Updated response assessment criteria for high-grade gliomas: response assessment in neuro-oncology working group. *J. Clin. Oncol.* 28, 1963–1972. doi: 10.1200/JCO.2009.26.3541

Yang, Y., Yan, L.-F., Zhang, X., Han, Y., Nan, H.-Y., Hu, Y.-C., et al. (2018). Glioma grading on conventional mr images: a deep learning study with transfer learning. *Front. Neurosci.* 12:804. doi: 10.3389/fnins.2018.00804

Yu, F., and Koltun, V. (2015). Multi-scale context aggregation by dilated convolutions. *arXiv [Preprint]. arXiv:1511.07122.*

Zeiler, M. D., and Fergus, R. (2014). "Visualizing and understanding convolutional networks. in *European Conference on Computer Vision*, (Springer), 818–833.

Zhao, X., Wu, Y., Song, G., Li, Z., Zhang, Y., and Fan, Y. (2018). A deep learning model integrating FCNNs and CRFs for brain tumor segmentation. *Med. Image Anal.* 43, 98–111. doi: 10.1016/j.media.2017.10.002

Zhou, C., Ding, C., Lu, Z., Wang, X., and Tao, D. (2018). "One-pass multi-task convolutional neural networks for efficient brain tumor segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, (Springer), 637–645.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.