IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received June 21, 2019, accepted July 2, 2019, date of publication July 8, 2019, date of current version July 26, 2019.

*Digital Object Identifier 10.1109/ACCESS.2019.2927433*

# Brain Tumor Segmentation Using Multi-Cascaded Convolutional Neural Networks and Conditional Random Field

**KAI HU[1,2], (Member, IEEE), QINGHAI GAN[1], YUAN ZHANG[1], SHUHUA DENG[1], FEN XIAO[1], WEI HUANG[4], CHUNHONG CAO[1], AND XIEPING GAO[1,3], (Member, IEEE)**

[1]Key Laboratory of Intelligent Computing and Information Processing, Ministry of Education, Xiangtan University, Xiangtan 411105, China
[2]Postdoctoral Research Station for Mechanics, Xiangtan University, Xiangtan 411105, China
[3]College of Software and Communication Engineering, Xiangnan University, Chenzhou 423043, China
[4]Department of Radiology, The First Hospital of Changsha, Changsha 410005, China

Corresponding authors: Chunhong Cao (caoch@xtu.edu.cn) and Xieping Gao (xpgao@xtu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61802328 and Grant 61771415, in part by the Natural Science Foundation of Hunan Province in China under Grant 2019JJ50606, in part by the Cernet Innovation Project under Grant NGII20170702, and in part by the Scientific Research Fund of Hunan Provincial Education Department of China under Grant 17A211.

**ABSTRACT** Accurate segmentation of brain tumor is an indispensable component for cancer diagnosis and treatment. In this paper, we propose a novel brain tumor segmentation method based on multi-cascaded convolutional neural network (MCCNN) and fully connected conditional random fields (CRFs). The segmentation process mainly includes the following two steps. First, we design a multi-cascaded network architecture by combining the intermediate results of several connected components to take the local dependencies of labels into account and make use of multi-scale features for the coarse segmentation. Second, we apply CRFs to consider the spatial contextual information and eliminate some spurious outputs for the fine segmentation. In addition, we use image patches obtained from axial, coronal, and sagittal views to respectively train three segmentation models, and then combine them to obtain the final segmentation result. The validity of the proposed method is evaluated on three publicly available databases. The experimental results show that our method achieves competitive performance compared with the state-of-the-art approaches.

**INDEX TERMS** Brain tumor segmentation, convolutional neural network, multi-cascaded convolutional neural network, conditional random field, multi-modality.

## I. INTRODUCTION

Brain tumors are one of the common diseases of the nervous system and have great harm to human health, and even lead to death. Glioma is one of the intracranial tumors with the highest mortality and morbidity [1]. It is usually divided into high-grade glioma (HGG) and low-grade glioma (LGG), and the average life expectancy of patients who have evolved into HGG is about two years. Many imaging techniques have been employed for investigating brain tumors such as magnetic resonance imaging (MRI), computed tomography (CT), positron emission tomography (PET), and single-photon emission computed tomography (SPECT). Among them, MRI has become the main imaging technique for the diagnosis and treatment of glioma, since its advantages of good soft tissue contrast, multi-parameter, imaging in arbitrary direction, non-invasive imaging, etc. In addition, multiple modalities, e.g., T1-weighted (T1), T1-weighted with contrast enhancement (T1c), T2-weighted (T2), and Fluid Attenuated Inversion Recovery (FLAIR) can be obtained through MRI [2]. Different MRI modalities focus on different detailed information of images and describe the characteristics of brain tumors from different aspects. The accurate segmentation of brain tumors is of great significance for medical diagnosis, surgical planning, and treatment planning. In particular, it is crucial to separate tumor tissues, e.g., necrosis, edema, enhancing core, and non-enhancing core from normal brain tissues including the gray matter (GM), the white matter (WM), and

The associate editor coordinating the review of this manuscript and approving it for publication was Wei Liu.

the cerebrospinal fluid (CSF). However, it is an extremely challenging problem to segment them accurately, mainly due to the following reasons. First, the shape, location, appearance, and size of gliomas can vary greatly from patient to patient [2]. Second, gliomas usually invade surrounding tissues, making the boundary blurred [3]. Third, the image distortion and noise caused by different factors such as imaging devices or acquisition protocols further increase the difficulty.

Manual segmentation is a tedious, time-consuming task, and if the person delineating the region of interest (ROI) is not a well-trained technologist, it will usually yields poor segmentation results [4], [5]. Therefore, automatic segmentation or semi-automatic segmentation attracts the attention of researchers. The existing automatic segmentation or semi-automatic segmentation methods can be roughly classified as either generative models based methods or discriminative models based methods [6]. The generative models rely on prior knowledge of the brain anatomy, and their learning process is relatively complex. Prastawa *et al.* [7] proposed a typical generative model of MR brain images, which mainly based on probabilistic image atlas. Other methods that require prior knowledge can be found in [8]–[10]. On the contrary, discriminative models often rely on low level image features. They usually employ a discriminative classifier to transform local features into class probability, which is more suitable for multi-category identification problems. Typical discriminative models include conditional random fields [11]–[13], random forests [14]–[16], support vector machines [17], decision forests [18], etc.

In recent years, the deep learning based automatic segmentation methods have shown great advantages in medical image analysis [19], [20]. CNN is one of the most popular deep learning models, which can extract features that are favorable for classification from the original data. Compared with the traditional machine learning methods, CNN has more powerful capability of feature learning and characterization, and can reduce the complexity of model due to the nature of weight sharing. Based on the advantages of deep learning, it has attracted an increasing attention in brain tumor segmentation [21]–[38]. Typically, Pereira *et al.* [21] proposed to design different CNN architectures with deeper layer for different grades of glioma samples using small $3 \times 3$ kernels. Urban *et al.* [22] proposed a two-path CNN model in which one channel extracts local detail features and the other extracts global context features. Most of the above brain tumor segmentation methods built CNN upon image patches. These patch-based methods classify each image patch into different classes, ignoring the spatial consistency of the entire image. They usually assumed that the label for each voxel is independent and did not take the local dependencies of labels into account. To consider the local dependencies of labels, Havaei *et al.* [23] developed a cascaded architecture (InputCascadeCNN), which can obtain a better segmentation performance. Furthermore, they adopted a two-stage training strategy to solve the imbalance of label distributions.

Recently, the advantages of the multi-scale features from CNNs have been demonstrated in segmentation tasks [24], [39]. Generally, there are two ways to extract multi-scale features [24]. One way is to use the feature maps from different levels of a network to represent multi-scale features [24]. In [40], Hu *et al.* proposed a multi-scale CNNs for the retinal vessel segmentation. Multi-scale images at different stages of the CNNs were fused to obtain the probability maps of the retinal vessels. The other is to pass different scaled versions of the input image using the same network [24]. In [13], the multi-scale features were extracted by Kamnitsas *et al.* using two patches with different sizes. In [23], Havaei *et al.* employed an InputCascadeCNN architecture to capture multi-scale features. While in this paper, we propose a new network architecture from InputCascadeCNN to obtain more discriminative multi-scale features through multi-cascaded networks.

It is undeniable that CNN has achieved remarkable achievement. But for a neural network with fixed architecture and parameters, its learning ability is still limited, and some useful information may be ignored, especially for the 3D information of the MRI data. Some researchers adopt 3D-CNN models [13], [32], [41] or 2.5D models [33] to deal with 3D images. Myronenko *et al.* [32] proposed a 3D semantic segmentation network for brain tumor segmentation based on encoder-decoder architecture. Wang *et al.* [33] proposed a hierarchical segmentation system, which converted the multi-class segmentation into three binary segmentation tasks. They also trained the segmentation models from the axial, sagittal, and coronal views respectively. During the testing stage, they averaged the softmax outputs in these three views to obtain the final results. Although these methods achieve excellent performance, they increase computational complexity and memory consumption. Hence, some mathematical models, e.g., Markov Random Fields (MRFs) or Conditional Random Fields (CRFs), are usually introduced to consider the spatial contextual information [12], [13], [25], [42]. Kamnitsas *et al.* [13] adopted 3D CRFs for post-processing, which improved the segmentation results, but the configuration of the 3D CRFs was a laborious process. Zhao *et al.* [25] formulated CRFs as neural networks and trained them with fully convolutional neural networks (FCNNs), their training process was cumbersome and computationally expensive. As a trade-off, we employ fully connected CRFs as a fine segmentation step to take the appearance and spatial consistency of the segmentation results into account.

In this paper, a novel coarse-to-fine segmentation framework for brain tumor segmentation is proposed by combining MCCNN and fully connected CRFs. Firstly, we propose a multi-cascaded network architecture for coarse segmentation to consider the dependencies between neighbouring pixels and the complementary information at different layers and scales. Then, considering that spatial contextual information of 3D MR images is very important for brain tumor segmentation and it is not fully considered by CNN, and

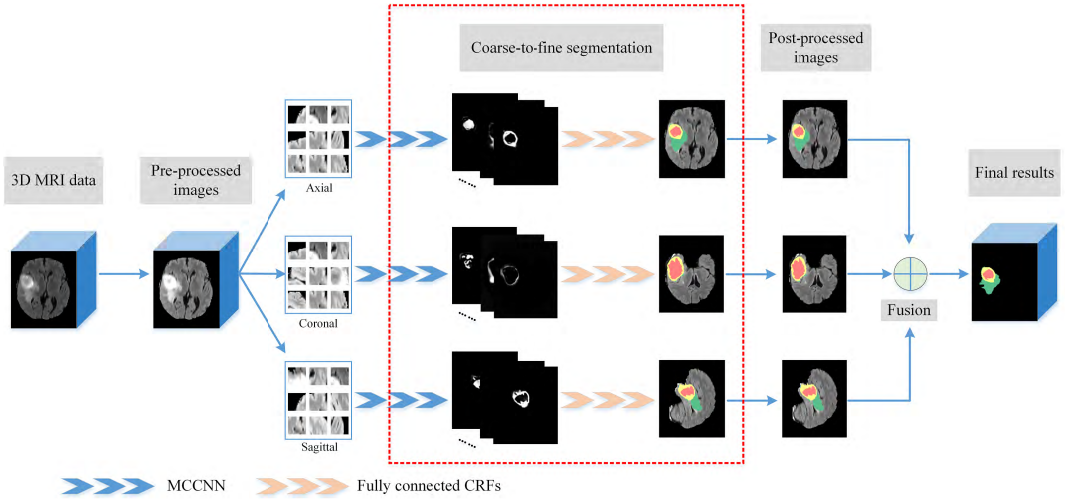



**FIGURE 1.** Flowchart of the proposed method.

we introduce the fully connected CRFs to further refine the probability map obtained by CNN. Finally, to fully consider the spatial information of 3D medical images and amplify the training samples, we train the segmentation models from the axial, sagittal, and coronal views respectively. The segmentation results of these three models are then fused to obtain the final result using a voting based fusion strategy as described in [25]. The evaluation of our method is performed on the imaging data provided by the Brain Tumor Segmentation Challenge (BRATS) 2013, BRATS 2015, and BRATS 2018 databases [6], [43]–[46]. The experimental results show that the proposed method not only yields competitive performance but also is computationally efficient compared with the state-of-the-art approaches. We also investigate the effects of different modalities on the performance of our model. The experimental results show that our model can achieve competitive performance based on three modalities, i.e., FLAIR, T1c and T2, compared with the model based on all four modalities [6], [13], [15], [16], [21]–[23], [26], [30], and can also reduce the cost of computational complexity and data acquisition. Besides, our model can segment the brain images slice by slice in the testing stage, which is much faster than the segmentation method based on the image patches. The contributions of this study can be summarized as follows.

1) We extend InputCascadeCNN to MCCNN architecture to obtain multi-scale features through multi-cascaded networks on brain MR images. The discriminative features are efficiently extracted by passing multiple scaled versions of the input image through three parallel sub-networks. In each sub-network, the high-level and low-level information are utilized jointly by fusing the features of the front-layer. In addition, different cascaded forms are employed to take the dependencies of labels and the neighborhood information of the central pixel into account to achieve a smoother boundary.
2) We propose an efficient coarse-to-fine segmentation framework for brain tumor segmentation. Our framework mainly consists of a coarse segmentation and a fine segmentation. The former uses the proposed MCCNN to obtain a probability map by considering multi-scale features on the multi-modal brain images. The latter employs a fully connected CRFs to refine the coarse segmentation by smoothing tumor edges and eliminating false positives.

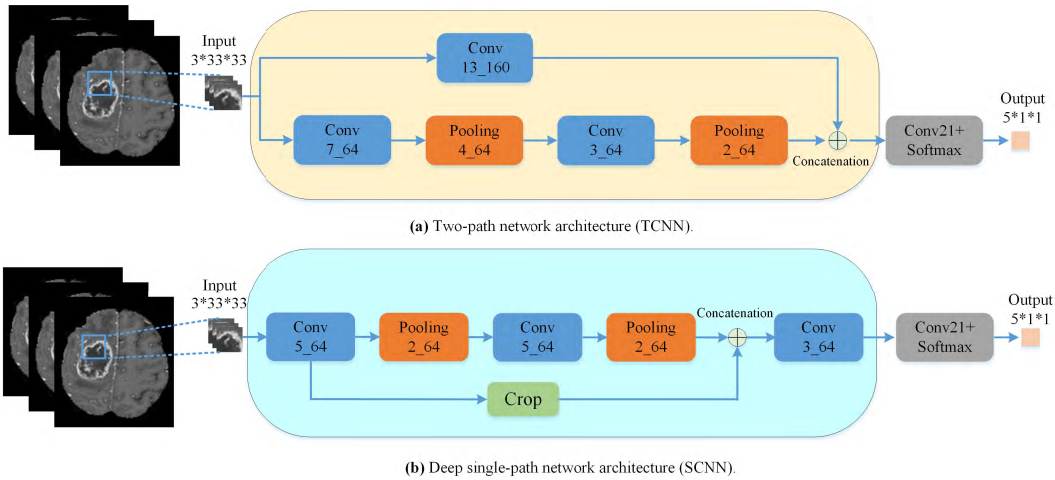
## II. PROPOSED METHOD

As shown in Fig. 1, the proposed method includes following steps, i.e., pre-processing, coarse segmentation based on MCCNN, fine segmentation based on fully connected CRFs, post-processing, and fusion of the segmentation results of three different views. The details of each step are presented as below.

### A. PRE-PROCESSING

Since MRI scans are affected by the bias field distortion [6], which makes the intensity of the same tissues to vary across the image. So we adopt the N4ITK method proposed by Tustison *et al.* [47] to correct the bias field of MRI data. In addition, it will also vary even if the images of the same patient are obtained in the same scanner at different time points, or in the case of pathology [48]. Therefore, in order to reduce the contrast and intensity ranges between patients and acquisitions, we also employ the method proposed by Nyúl *et al.* [48] to perform the intensity normalization on each modality. The pre-processing steps are as follows.

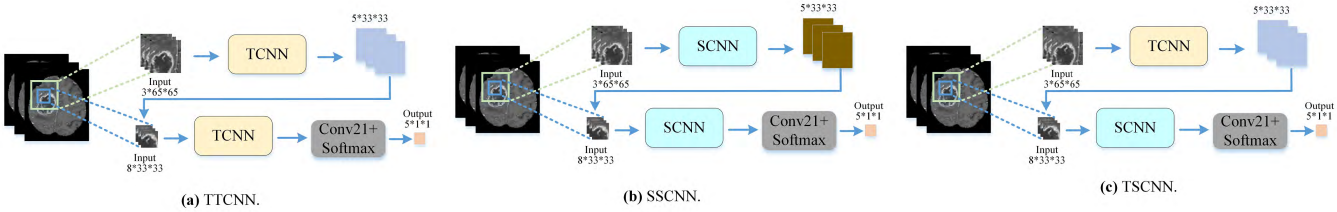
- Step 1. Remove the 1% highest and lowest intensities.
- Step 2. Apply an N4ITK bias correction to T1c modality.
- Step 3. Perform the intensity normalization on each modality.
- Step 4. Normalize the data within each input channel by subtracting the mean of all pixels from each pixel in




**(a)** Two-path network architecture (TCNN).

**(b)** Deep single-path network architecture (SCNN).

**FIGURE 2.** The baseline network architectures.




**(a)** TTCNN. **(b)** SSCNN. **(c)** TSCNN.

**FIGURE 3.** The proposed simple cascaded architectures.

the image of each channel and dividing by the standard deviation.

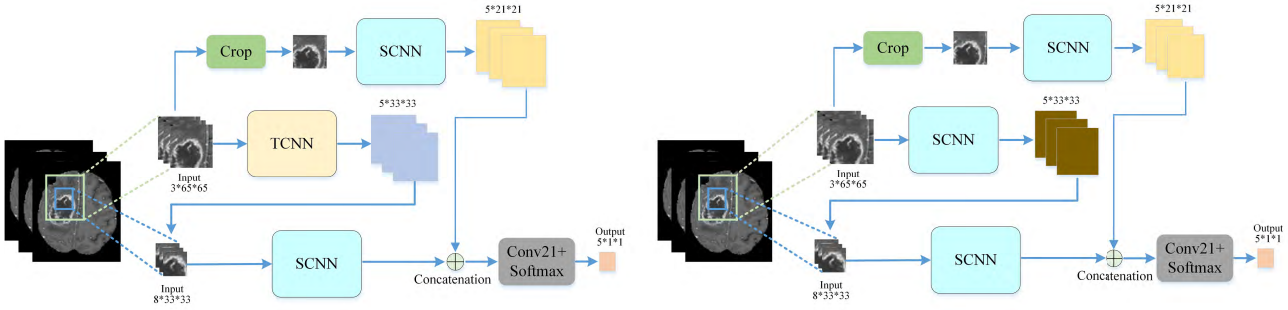
### B. COARSE SEGMENTATION USING OUR MCCNN ARCHITECTURE

CNN is a very powerful method in the field of image segmentation [2]. The basic architecture of CNN includes input layer, hidden layer, and output layer [49]. The value of output can be obtained from the forward propagation of CNN, and the bias and weight can be adjusted by back-propagation. The classic CNN consists of three parts: convolution layers, pooling layers, and fully connected layers [49]. There are usually a multitude of parameters that need to be trained in a deep learning model of CNN. In order to ensure that there are enough training samples and consider the complexity of training, our MCCNN is trained using image patches, which are extracted randomly from slice of different perspectives i.e., axial, coronal, and sagittal views in 3D brain MRI data. Overall, the glioma images are roughly segmented in this process, that is, the entire glioma contour is obtained as completely and accurately as possible. For this purpose, we conduct a large number of experiments to design and select the network. Firstly, we propose two basic CNN architectures without cascade as baseline connected components. Then, we explore the performance of simple cascade architectures and the multi-cascaded architectures. Details of these models are as follows.

The basic network architectures are shown in Fig. 2. The two-path convolutional neural network (TCNN), as shown in Fig. 2(a), is based on Havaei *et al.* [23], and we improve it to obtain more accurate tumor contour. At the same time, we design a deep single-path convolutional neural network (SCNN) with front-layer information as alternative comparison, which is shown in Fig. 2(b).

#### 1) SIMPLE CONCATENATION

As shown in Fig. 3, in this study we propose three simple cascaded network architectures that concatenate the output of the first CNN at the input of the subsequent CNN, respectively named TTCNN, SSCNN, and TSCNN. In the TTCNN architecture, as shown in Fig. 3(a), two basic two-path networks are adopted. The first network processes larger image patches to obtain more comprehensive context information. The output of the first network is concatenated to the input of the second network, which guides the training of the second network. The basic network consists of two channels, which extract useful image features by using different kernels. One channel with larger kernel of size $13 \times 13$ is used to extract global contextual information, while another with smaller kernels of size $7 \times 7$ and $3 \times 3$ are used to extract local details. The global channel takes three planes of size $33 \times 33$ and generates the feature maps of size $21 \times 21$. The local channel also takes patches of size $33 \times 33$ and generates the feature maps of size $21 \times 21$. These feature maps are




**(a)** MCCNN1. **(b)** MCCNN2.

**FIGURE 4.** The proposed MCCNN architectures.

combined using the concatenating ability of the convolution layer.

As shown in Fig. 3(b), the proposed SSCNN architecture employs two deep single-path networks with front-layer information. We herein use smaller $5 \times 5$ and $3 \times 3$ receptive fields to take more details into account. It integrates high-level information with low-level information through fusing the front-layer features to utilize the features of different layers for more accurate segmentation. In addition, larger filters extract more local features, whereas smaller filters focus more on edge detection. Therefore, smaller kernels are used in this structure to take more edge information into account for a more accurate and complete tumor contour.

The proposed TSCNN architecture is a combination of a two-path network and a deep single-path network as shown in Fig. 3(c). Similar to the previous two architectures, the output of the first half of TSCNN is cascaded at the input of the second half. The first network of this architecture uses larger kernels to learn more global features, while the second network has smaller kernels to consider the local details.

#### 2) MCCNN ARCHITECTURES

Fig. 4 shows the proposed MCCNN, which combined multiple networks to form a multi-cascaded network architecture. The InputCascadeCNN architecture proposed by Havaei *et al.* [23] has achieved state-of-the-art segmentation performance in brain glioma segmentation. We extend and improve this architecture by taking two cascaded forms. One form is cascading at the input of the bottom network, and the output of the first network is treated as an additional image channel to guide the training of the second network, which is called simple concatenation. The other is to add a cascaded network before the feature map obtained by simple concatenation being fed to the final classification layer, thereby forming a multi-cascaded CNN architecture. Since the neurons of the soft-max output layer are cascaded directly with the previous output, the network learning is more inclined to assume that the central pixel label should be similar to its surroundings, so that smoother boundaries can be produced. In this work, we investigate two multi-cascaded network architectures, named MCCNN1 and MCCNN2, respectively.

As shown in Fig. 4(a), MCCNN1 uses the TSCNN architecture as a simple concatenation. In this architecture, we adopt a two-path network to extract the features of larger image patches, and employ a deep single-path network to learn the features of smaller image patches. In addition, another single-path network is added to detect tumor edge information, which helps to extract contextual information and more local details.

In MCCNN2, as shown in Fig. 4(b), on the basis of the SSCNN architecture, we crop the larger image patches, which are used as input to another single-path network to identify local features of the image, e.g., the edge information of brain glioma. The intermediate results obtained are connected in the previous layer of the final output. The size of the cropping patches can be adjusted according to the specific requirements. By this manner, it can not only consider the dependence of labels, but also enhance the performance of glioma segmentation task by combining different cascaded forms and fusing the multi-scale information.

### C. FINE SEGMENTATION USING FULLY CONNECTED CRFS

Although the form of cascading has been used to consider the local dependencies of labels, it is not enough for medical images. This is mainly because anatomical structures have complex shapes that are difficult to model [50]. Further, the temporal or spatial correlation of MRI data also plays important role in the segmentation, which needs to be considered in the method. Therefore, it is necessary to further refine the probability map obtained by CNN. In recent years, CRFs have been widely used in many medical image applications, since they show good performance for modeling the complex dependencies in spatial data. For the segmentation of brain tumors, CRFs can be used not only to model the relationship between image pixel features and its label, but also to model the dependencies between neighborhood pixel features and their labels [11], [51]. As mentioned before, Zhao *et al.* [25] applied CRFs to image segmentation by formulating it as neural networks. However, the training process of their method is cumbersome and computationally expensive. By contrast, in this paper we employ the CRFs as a fine segmentation to

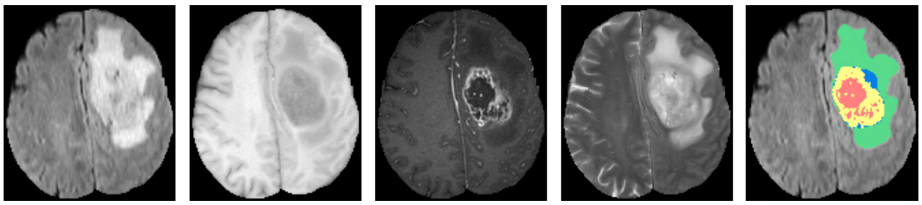


**FIGURE 5.** Examples of the brain MRI data from the BRATS 2013 dataset. From left to right: Flair, T1, T1c, T2, and the ground truth. Each color represents a tumor class: red—necrosis, green—edema, blue—non-enhancing, and yellow—enhancing tumor.

refine the segmentation of MCCNN. It not only improves the segmentation accuracy of MCCNN, but also has lower computational complexity.

For a 2D image $I$, which consists of pixels $\{I_i|i = 1, 2, \ldots, N\}$ and its corresponding label configuration $\{Y_l|l = 1, 2, \ldots, C\}$, the image segmentation problem is transformed into an optimization problem. We train a suitable model by minimizing the energy function of CRFs[1] [51], which can be modeled as

$$E(Y) = \sum_{i}^{N} \Psi_u(y_i) + \sum_{\forall_{i,j,i\neq j}}^{N} \Psi_p(y_i, y_j), \tag{1}$$

where $u, p \in \{1, 2, \ldots, C\}$ are segmentation labels, and $i$, $j \in \{1, 2, \ldots, N\}$ are certain pixels of image $I$. The unary potential, $\Psi_u(y_i) = -\log P(y_i|I)$, is the negative log-likelihood, where $P(y_i|I)$ is the probability obtained by CNN for pixel $i$. While the pairwise potential measures the influence between any pair of pixels [49], which is defined as

$$\Psi_p(y_i, y_j) = \mu(y_i, y_j) \sum_{m=1}^{M} w^{(m)} k^{(m)}(f_i, f_j), \tag{2}$$

where $M = 2$ is the number of Gaussian kernel and $w^{(m)}$ denotes a weight for the $m$-th Gaussian kernel. $\mu(y_i, y_j) = [y_i \neq y_j]$ is the label compatibility function [51]. $k^{(1)}$ represents the appearance kernel, which tries to assign the same class labels to nearby pixels with similar intensity. $k^{(2)}$ represents the smoothness kernel, aiming to remove small isolated regions. They are respectively defined by

$$k^{(1)}(f_i, f_j) = \exp(-\frac{|s_i - s_j|}{2\theta_\alpha^2} - \frac{|e_i - e_j|}{2\theta_\beta^2}), \tag{3}$$

$$k^{(2)}(f_i, f_j) = \exp(-\frac{|s_i - s_j|}{2\theta_\gamma^2}), \tag{4}$$

where $e_i$ and $e_j$ are the intensities of pixel $i$ and $j$, $s_i$ and $s_j$ are the corresponding spatial coordinates. $f_i$ and $f_j$ represent the feature vectors of the pair of pixels, namely the intensity and position information. $\theta_\alpha$, $\theta_\beta$, and $\theta_\gamma$ represent the parameters of the Gaussian kernels, respectively.

[1]https://github.com/lucasb-eyer/pydensecrf

### D. POST-PROCESSING AND FUSION OF THE SEGMENTATION RESULTS

For the post-processing, we employ a simple method based on connected components [23] to remove small spurious areas due to the high intensity value near the edges of the brain image.

To utilize the spatial information of 3D MR brain image, we train three segmentation models using patches extracted from different perspectives, i.e., axial, coronal, and sagittal views. Then, the test images can be segmented slice by slice using these three models, respectively. Finally, we perform post-processing on these three results as described above, and then fuse them using a majority voting strategy.

## III. EXPERIMENTAL SETUP

### A. BRAIN MRI DATA

The proposed method is tested and evaluated on the BRATS 2013,[2] BRATS 2015,[3] and BRATS 2018[4] databases. The BRATS 2013 database contains training dataset of 30 glioma patients, comprising of 20 HGG cases and 10 LGG cases. The testing dataset contains 10 patient subjects with HGG. All the brain images were skull stripped and had the same orientation. For each patient subject, there are four MRI modalities, i.e., T1, T1c, T2, and FLAIR. All modalities were co-registered to the T1c sequence in order to homogenize these data and then these images were resampled to 1 mm isotropic resolution in a standardized axial orientation with a linear interpolator [6]. In addition, the manual annotations have been used as the ground truth in the training dataset. No ground truths are available for the testing dataset. The labels were divided into five classes, namely healthy tissues (label 0), necrosis (label 1), edema (label 2), non-enhancing (label 3), and enhancing tumor (label 4). Examples of the brain MRI data along with the ground truth are shown in Fig. 5.

The BRATS 2015 database contains 274 training MR images, which are composed of 220 HGG cases and 54 LGG cases. Please note that the BRATS 2015 dataset includes imaging data of the BRATS 2013, where the ground truth was manually annotated. The rest produced by voted average of the segmentation results of the top ranked methods in

[2]https://www.virtualskeleton.ch/BRATS/Start2013
[3]https://www.virtualskeleton.ch/BRATS/Start2015
[4]http://www.med.upenn.edu/sbia/brats2018.html

the BRATS 2012 and BRATS 2013. These automatically generated annotations ultimately manually modified by experienced raters.

The BRATS 2018 database contains training dataset of 285 glioma patients, comprising of 210 HGG cases and 75 LGG cases. The validation dataset contains 66 patient subjects of unknown grade. The ground truths of training dataset were provided by the BRATS organizers.

### B. EVALUATION METRICS

According to the protocol in the BRATS databases [6], the tumor structures are divided into three sub-regions for each patient, which are defined as follows:

a. The complete tumor region (including all four sub-tumoral classes, i.e, label 1, 2, 3, and 4).
b. The core tumor region (including complete tumor region but excluded ''edema'' region, i.e., label 1, 3, and 4).
c. The enhancing tumor region (only including ''enhancing'' region, i.e., label 4).

For each tumor region, we use Dice Similarity Coefficient (DSC) as well as Positive Predictive Value (PPV) and Sensitivity for the evaluation. DSC provides a measurement criterion for overlapping areas between manual delineated brain tumor regions and our segmentation results. It is defined as

$$DSC = \frac{2TP}{FP + 2TP + FN}, \tag{5}$$

where $TP$, $FP$, and $FN$ denote the numbers of true positive, false positive, and false negative measurements, respectively. PPV is used to evaluate the number of $TP$ and $FP$, which is defined as

$$PPV = \frac{TP}{TP + FP}. \tag{6}$$

In addition, Sensitivity is a useful measure of the number of $TP$ and $FN$ and it is given by

$$Sensitivity = \frac{TP}{TP + FN}. \tag{7}$$

### C. EXPERIMENTAL SCHEME AND IMPLEMENTATION DETAILS

In order to verify the effectiveness of the proposed method, a series of comparative experiments are conducted on the BRATS 2013 database. Firstly, we perform an experiment to evaluate the segmentation performance of the proposed two MCCNN architectures in comparison with two basic networks described previously and the simple concatenation forms between them. Secondly, to verify the validity of the proposed coarse-to-fine segmentation framework, we compare the segmentation performance of MCCNN with and without used CRFs. In addition, to investigate the effect of each modality on the proposed method, we carry out an experiment to compare the performance of our models based on input data of four imaging modalities (Flair, T1, T1c, and T2) and three imaging modalities (any three of these four imaging modalities). It should be noted that all the experiments above are conducted on the images extracted from axial view. Moreover, in order to consider the information from different views, we evaluate the segmentation results of three different perspectives and demonstrate the effectiveness of fusing them. Finally, we compare the proposed method with some state-of-the-art approaches on the BRATS 2013, BRATS 2015, and BRATS 2018 databases.

In the experiments, the training process consists of two stages. In the first stage, the coarse segmentation is carried out through the proposed MCCNN network, and the probability maps of three perspectives are obtained. In the second stage, refined segmentation results are obtained through the fully connected CRFs. The hyper-parameters of the different network architectures (kernel and pooling size for each layer and the number of feature maps) are found using the validation set, consisting of two HGG cases, which is shown in Fig. 4. Other hyper-parameters of the model are also tuned with the validation set. The final results of our experiments are obtained using ReLU as activation function in convolution layers. The learning rate is set by $\alpha = 0.005$. Momentum is 0.9 and weight decay is $10^{-3}$. Moreover, we use Dropout [52] and L2 regularization to prevent overfitting and use the Stochastic Gradient Descent (SGD) [53] algorithm for training the network. Besides, the parameters of CRFs in our experiments are obtained by grid search. The optimal performance in our experiments is obtained when $\theta_\alpha = 10$, $\theta_\beta = 50$, and $\theta_\gamma = 3$. The linear combination weights are set by $w^{(1)} = 3$ and $w^{(2)} = 3$.

Since the true distribution of each label in the brain is highly unbalanced, in which the majority of voxels are healthy (98%), and non-healthy voxels account for a very small proportion (2%). This makes neural network training algorithms such as SGD perform poorly. Therefore, we adopt the similar two-phase training procedure [23] to train the network. In the first phase, we construct a training data with uniform distribution of all labels. The model iterates 50–75 epochs using the equal distributed instances. In the second phase, unbalanced training data is used. The model iterates through 5 epochs using the instances with true distribution.

All models are trained and tested with keras on a computing server with four Intel I5-4590 CPUs and eight NVIDIA GTX1080 GPUs graphics cards. Each program takes up only one CPU and GPU at a time.

## IV. RESULTS AND DISCUSSION

### A. PERFORMANCE COMPARISON OF THE PROPOSED NETWORK ARCHITECTURES

It is generally accepted that the learning ability of the network will improve by increasing the complexity of the network. In this study, we provide a comparison based on different network architectures, i.e., basic CNN architectures (TCNN and SCNN), simple concatenation architectures (TTCNN, SSCNN and TSCNN), and multi-cascaded architectures (MCCNN1 and MCCNN2). These architectures are

**TABLE 1.** Performance of the basic network models and variations on the BRATS 2013 testing dataset.

| Methods | DSC | | | PPV | | | Sensitivity | | |
|---|---|---|---|---|---|---|---|---|---|
| | Complete | Core | Enhanced | Complete | Core | Enhanced | Complete | Core | Enhanced |
| TCNN | 0.80 | 0.68 | 0.67 | 0.75 | 0.66 | **0.61** | 0.88 | 0.78 | 0.79 |
| SCNN | 0.81 | 0.68 | 0.67 | 0.76 | 0.65 | 0.59 | 0.88 | 0.78 | 0.80 |
| TTCNN | **0.82** | 0.67 | 0.67 | **0.78** | 0.61 | 0.59 | 0.87 | 0.80 | 0.80 |
| SSCNN | **0.82** | **0.70** | **0.68** | 0.76 | **0.67** | 0.60 | 0.89 | 0.79 | 0.82 |
| TSCNN | 0.78 | 0.66 | 0.67 | 0.71 | 0.59 | 0.59 | 0.90 | **0.81** | **0.83** |
| MCCNN1 | 0.78 | 0.69 | **0.68** | 0.69 | 0.65 | 0.60 | **0.91** | 0.80 | 0.82 |
| MCCNN2 | **0.82** | **0.70** | **0.68** | 0.77 | **0.67** | 0.60 | 0.89 | 0.80 | **0.83** |

**TABLE 2.** Evaluation results of the proposed coarse-to-fine segmentation framework on the BRATS 2013 testing dataset.

| Methods | DSC | | | PPV | | | Sensitivity | | |
|---|---|---|---|---|---|---|---|---|---|
| | Complete | Core | Enhanced | Complete | Core | Enhanced | Complete | Core | Enhanced |
| MCCNN | 0.82 | 0.70 | 0.68 | 0.77 | 0.67 | 0.60 | **0.89** | 0.80 | 0.83 |
| MCCNN+CRFs | **0.86** | **0.77** | **0.70** | **0.86** | **0.75** | **0.62** | 0.87 | **0.84** | **0.85** |

illustrated in Figs. 2-4 respectively. It is need to note that the proposed multi-cascaded architectures consist of three sub-networks, because going deeper will not improve the performance.

Table 1 presents the comparison of quantitative results on brain axial view images using different network architectures. All the above results are obtained without any post-processing. As described in Section II-B, at this stage we need to obtain the contour that contains the entire tumor as much as possible. This is the higher the value of DSC and Sensitivity, the better the subsequent refinement results. From Table 1, it shows that the segmentation results of simple concatenation and multi-cascaded architectures have outstanding performance on these two indicators. Especially, SSCNN and MCCNN2 achieve similar performance in DSC, but the latter has higher Sensitivity than the former. Moreover, MCCNN2 achieves the best performance for the overall. From Fig. 4 (b), each sub-network of this architecture takes the front-layer information into account, and uses multiple concatenation forms to seamlessly integrate the dependencies of labels and multi-scale features into our network, thus utilizing the complementary information of features at different layers and scales. Besides, this architecture can process each sub-network in parallel. Experiments demonstrate that this architecture achieves comparable segmentation results while guaranteeing the time efficiency. Therefore, MCCNN2 is selected as the final recommended multi-cascaded network architecture. In the following part, we will use MCCNN to represent this network architecture.

### B. EVALUATION OF THE EFFECTIVENESS OF THE PROPOSED COARSE-TO-FINE SEGMENTATION FRAMEWORK

Table 2 shows the evaluation results of MCCNN with and without used fully connected CRFs on the BRATS 2013 dataset. From Table 2, we can find that the results of MCCNN + CRFs are obviously higher than those of MCCNN except the Sensitivity of the complete tumor is slightly lower. Therefore, we can demonstrate that the coarse segmentation using MCCNN achieves higher Sensitivity, and the fine segmentation using CRFs can effectively refine the segmentation results. Furthermore, Fig. 6 shows the visual effects of the segmentation results. The results in the third row after the refinement using CRFs are significantly better than those in the second row, not only obtains more accurate tumor areas, but also effectively removes the false positive areas.

In summary, in our coarse-to-fine segmentation framework, the tumor profile with higher accuracy is obtained by using the proposed MCCNN due to the advantages of multiple cascaded combinations and multi-scale feature fusion. And we use CRFs to refine the segmentation results by considering the spatial and appearance consistency of segmentation results. Both the quantitative results and visual effects demonstrate the performance of the proposed coarse-to-fine segmentation framework.

### C. EVALUATION OF THE IMPACT OF EACH MODALITY

As can be seen from Fig. 5, different modalities in MRI data emphasize different information. In order to explore the effect of each modality on training, we establish segmentation models respectively using all four imaging modalities, i.e., Flair, T1, T1c, and T2, and the combination of any three imaging modalities. Comparisons are presented in Table 3.

In general, different modalities have different features. Since the T1c sequence strengthens the features of tumor boundary, making the boundary clear and easy to distinguish [2]. The information reflected by this sequence is effective for the identification and classification of core tumor and enhancing tumor regions. Therefore, in the case of removing the T1c sequence, the model cannot learn robust features of the tumor edge, resulting in poor segmentation results.

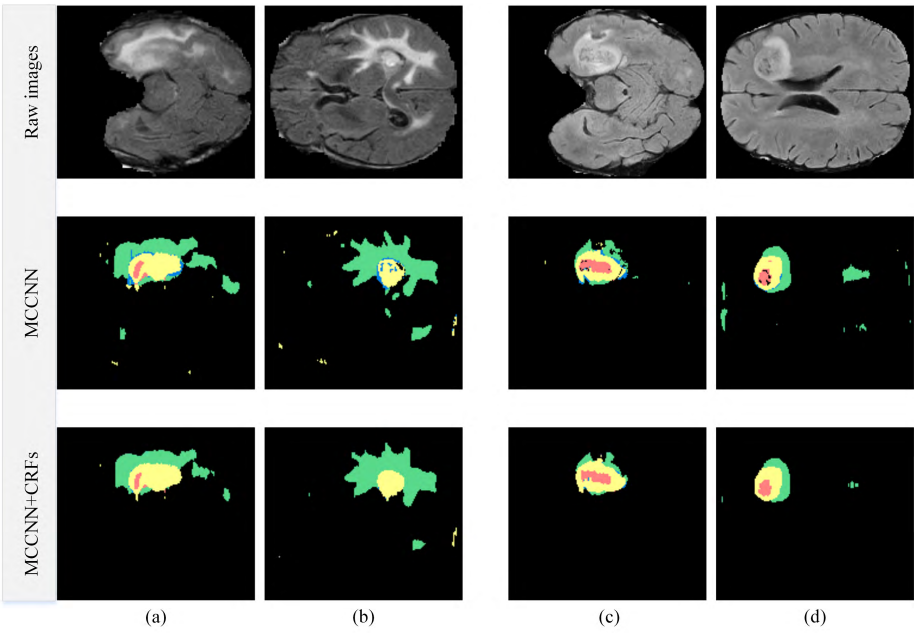



**FIGURE 6.** Visual effects on the BRATS 2013 dataset. (a) and (b) show the segmentation results of the 45-th and 75-th slice of the axial view of case 0301. (c) and (d) show the segmentation results of 50-th and 70-th slice of the axial view of case 0310. Colors in the results represent different tumor classes: red—necrosis, green—edema, blue—non-enhancing, and yellow—enhancing tumor.

**TABLE 3.** Performance comparison of the segmentation models trained using different imaging modalities on the BRATS 2013 testing dataset.

| Modalities | DSC | | | PPV | | | Sensitivity | | |
|---|---|---|---|---|---|---|---|---|---|
| | Complete | Core | Enhanced | Complete | Core | Enhanced | Complete | Core | Enhanced |
| Flair,T1,T1c,T2 | **0.88** | 0.79 | **0.75** | 0.90 | 0.81 | 0.70 | 0.86 | 0.82 | 0.83 |
| Flair,T1,T2 | 0.86 | 0.54 | 0.27 | 0.89 | 0.73 | 0.44 | 0.84 | 0.46 | 0.24 |
| Flair,T1c,T2 | **0.88** | **0.80** | 0.74 | 0.90 | 0.80 | 0.67 | **0.87** | **0.84** | **0.85** |
| Flair,T1,T1c | 0.86 | 0.77 | 0.73 | 0.89 | 0.79 | 0.66 | 0.85 | 0.82 | 0.84 |
| T1,T1c,T2 | 0.82 | 0.79 | 0.74 | **0.91** | **0.86** | **0.72** | 0.76 | 0.76 | 0.79 |

The T1 sequence is helpful for the differentiation of normal tissues, but it weakens the characteristics of the tumor. Therefore, the model based on the T1 sequence is more favorable for the segmentation of normal tissues, that is, the value of PPV is relatively higher, accompanied by lower Sensitivity. The signals of water molecules are inhibited in the Flair sequence, which helps to distinguish between edema regions and CSF [2]. Therefore, the Flair sequence has a significant effect on the segmentation of the complete tumor region as well as different tumor sub-regions. The T2 sequence is mainly used to delineate the edema area, and strengthens the signal of this area, which can provide useful information for the training of our model. It can be seen that the overall segmentation accuracy of our model decreased in the absence of the T2 sequence.

In summary, the Flair, T1c, and T2 modalities are indispensable, and the information they emphasized are very useful and irreplaceable for model training. Although the T1 sequence also plays an important role in normal tissue segmentation, it is not prominent for our segmentation task. In addition, we find that the models achieve similar segmentation performance based on all four modalities or three modalities, i.e., Flair, T1c, and T2, while the latter has fewer training parameters and lower computational complexity. Therefore, we choose the Flair, T1c, and T2 modalities as our final multi-modal MRI data to conduct the experiments.

### *D. EVALUATION OF THE VALIDITY OF FUSING THE SEGMENTATION RESULTS OBTAINED BY THREE PERSPECTIVES*

Considering that more robust information can be extracted from different perspectives, we separately train three segmentation models using patches obtained from three different views, i.e., axial, coronal, and sagittal views. These three models are used to segment the brain images to produce three segmentation results, which are then fused to obtain the final segmentation result. The quantitative results of each stage from three perspectives and their fusion results are shown in Table 4. It shows that fusing the segmentation results usually yields better segmentation performance at any stage. It is worth noting that after the CRFs, the fusion

**TABLE 4.** Evaluations of segmentation results obtained in three different views on the BRATS 2013 testing dataset.

| Methods | | DSC | | | PPV | | | Sensitivity | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Complete | Core | Enhanced | Complete | Core | Enhanced | Complete | Core | Enhanced |
| MCCNN | axial | **0.82** | 0.70 | 0.68 | **0.77** | 0.67 | 0.60 | 0.89 | 0.80 | **0.83** |
| | coronal | 0.78 | 0.68 | 0.66 | 0.69 | 0.64 | 0.60 | 0.91 | 0.80 | 0.79 |
| | sagittal | 0.72 | 0.67 | 0.67 | 0.60 | 0.60 | 0.60 | **0.92** | **0.81** | 0.80 |
| | fusion | 0.80 | **0.71** | **0.70** | 0.72 | **0.69** | **0.63** | 0.91 | 0.79 | 0.81 |
| MCCNN+CRFs | axial | 0.86 | 0.77 | 0.70 | **0.86** | 0.75 | 0.62 | 0.87 | 0.84 | 0.85 |
| | coronal | 0.86 | 0.79 | 0.73 | 0.83 | 0.77 | 0.68 | 0.90 | 0.85 | 0.83 |
| | sagittal | 0.81 | 0.75 | 0.73 | 0.74 | 0.69 | 0.66 | **0.92** | **0.87** | 0.85 |
| | fusion | **0.88** | **0.81** | **0.76** | **0.86** | **0.81** | **0.69** | 0.90 | 0.84 | **0.86** |
| MCCNN+CRFs+post-process | axial | 0.88 | 0.80 | 0.74 | **0.90** | 0.80 | 0.67 | 0.87 | 0.84 | 0.85 |
| | coronal | 0.88 | 0.80 | 0.75 | 0.86 | 0.79 | 0.70 | 0.90 | 0.85 | 0.83 |
| | sagittal | 0.85 | 0.79 | 0.76 | 0.81 | 0.75 | 0.70 | **0.92** | **0.87** | 0.85 |
| | fusion | **0.89** | **0.82** | **0.77** | 0.87 | **0.83** | **0.71** | 0.90 | 0.84 | **0.86** |

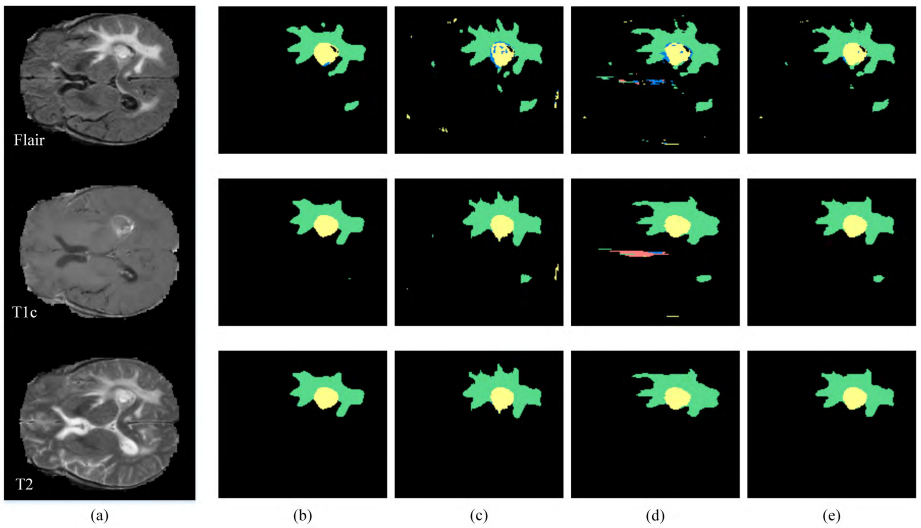



**FIGURE 7.** Visual effects from three views on case 0301 of the BRATS 2013 dataset. (a) shows raw images. (b)-(d) represent the segmentation results from axial, coronal, and sigattal views, respectively. (e) shows the fusion results of three views. For a better visual comparison, these segmentation results are converted into the same view. In (b)-(e), each row from top to bottom shows: segmentation results of MCCNN, segmentation results of MCCNN + CRFs, and segmentation results of MCCNN + CRFs + post-processing. Colors in the results represent different tumor classes: red—necrosis, green—edema, blue—non-enhancing, and yellow—enhancing tumor.

results are greatly improved without the post-processing compared with the negligible improvement after applying the post-processing.

To better illustrate the effectiveness of the fusion segmentation results, we show the visual effects of each segmentation process in Fig. 7. From Table 4 and Fig. 7, we can find that the segmentation results of the axial view are optimal in the coarse segmentation stage. In the fine and final segmentation stages with post-processing, the coronal view is the best. And in any stage, the segmentation results from sagittal view are the worst. Overall, the best performance is achieved by combining these three different perspectives.

### *E. COMPARISON WITH THE STATE-OF-THE-ART METHODS*

#### 1) RESULTS ON THE BRATS 2013 DATABASE

As listed in Table 5, we summarize the evaluation results of the top ranked participants, as well as the results of our method and other six state-of-the-art approaches. In the ranking of the testing dataset, Pereira *et al.*'s method [21] ranked the first, Zhao *et al.*'s method [25] ranked the second, and our method ranked the third. As seen from Table 5, none of approaches can produce optimal segmentation results for all sub-tumoral regions. However, we can find that the proposed method achieves forefront and competitive performance compared to the state-of-the-art approaches. In particular, our method achieves the best values for DSC of the complete tumor and Sensitivity of the enhancing tumor region.

Our method is inspired by Havaei *et al.*'s [23] approach, in which they proposed a cascaded network architecture for brain tumor segmentation. In this paper, we propose a novel multi-cascaded network architecture by utilizing multiple concatenation forms, in which the higher accuracy tumor profile and smoother boundaries can be obtained. We herein provide a one-to-one comparison with the results of the proposed method and those of Havaei *et al.*'s approach [23].

**TABLE 5.** Comparison of the proposed method with other state-of-the-art methods on the BRATS 2013 testing dataset.

| Methods | DSC | | | PPV | | | Sensitivity | | |
|---|---|---|---|---|---|---|---|---|---|
| | Complete | Core | Enhanced | Complete | Core | Enhanced | Complete | Core | Enhanced |
| Meier | 0.82 | 0.73 | 0.69 | 0.76 | 0.78 | 0.71 | **0.92** | 0.72 | 0.73 |
| Reza | 0.83 | 0.72 | 0.72 | 0.82 | 0.81 | 0.70 | 0.86 | 0.69 | 0.76 |
| L. Zhao | 0.84 | 0.70 | 0.65 | 0.80 | 0.67 | 0.65 | 0.89 | 0.79 | 0.70 |
| Cordier | 0.84 | 0.68 | 0.65 | 0.88 | 0.63 | 0.68 | 0.81 | 0.82 | 0.66 |
| Tustison et al. [14] | 0.87 | 0.78 | 0.74 | 0.85 | 0.74 | 0.69 | 0.89 | **0.88** | 0.83 |
| Davy et al. [30] | 0.85 | 0.74 | 0.68 | 0.85 | 0.74 | 0.62 | 0.85 | 0.78 | 0.77 |
| Havaei et al. [23] | 0.88 | 0.79 | 0.73 | 0.89 | 0.79 | 0.68 | 0.87 | 0.79 | 0.80 |
| Pereira et al. [21] | 0.88 | 0.83 | **0.77** | 0.88 | **0.87** | 0.74 | 0.89 | 0.83 | 0.81 |
| Zhao et al. [25] | 0.88 | **0.84** | **0.77** | **0.90** | **0.87** | **0.76** | 0.86 | 0.82 | 0.80 |
| Zhan et al. [26] | 0.88 | 0.81 | 0.74 | - | **0.87** | 0.74 | - | 0.84 | 0.83 |
| Our method (axial) | 0.88 | 0.80 | 0.74 | **0.90** | 0.80 | 0.67 | 0.87 | 0.84 | 0.85 |
| Our method (coronal) | 0.88 | 0.80 | 0.75 | 0.86 | 0.79 | 0.70 | 0.90 | 0.85 | 0.83 |
| Our method (sagittal) | 0.85 | 0.79 | 0.76 | 0.81 | 0.75 | 0.70 | **0.92** | 0.87 | 0.85 |
| Our method (fusion) | **0.89** | **0.82** | **0.77** | **0.87** | **0.83** | **0.71** | **0.90** | **0.84** | **0.86** |

**TABLE 6.** Comparison of the proposed method with other state-of-the-art approaches on the BRATS 2015 dataset.

| Methods | DSC | | | PPV | | | Sensitivity | | |
|---|---|---|---|---|---|---|---|---|---|
| | Complete | Core | Enhanced | Complete | Core | Enhanced | Complete | Core | Enhanced |
| Piedra et al. [31] | 0.74 | 0.54 | 0.54 | - | - | - | - | - | - |
| Isensee et al. [28] | 0.85 | 0.74 | 0.64 | - | - | - | - | - | - |
| Chen et al. [29] | 0.85 | 0.70 | 0.63 | - | - | - | - | - | - |
| Kamnitsas et al. [13] | 0.847 | 0.670 | 0.629 | 0.850 | **0.848** | 0.634 | 0.876 | 0.607 | 0.662 |
| Zhao et al. [25] | 0.84 | 0.73 | 0.62 | **0.89** | 0.76 | 0.63 | 0.82 | 0.76 | 0.67 |
| Our method (axial) | 0.86 | 0.75 | 0.74 | 0.85 | 0.76 | **0.79** | **0.89** | **0.79** | 0.73 |
| Our method (coronal) | 0.85 | 0.75 | 0.74 | 0.85 | 0.79 | 0.74 | 0.88 | 0.75 | 0.79 |
| Our method (sagittal) | 0.84 | 0.74 | 0.72 | 0.84 | 0.73 | 0.69 | 0.86 | 0.77 | 0.79 |
| Our method (fusion) | **0.87** | **0.76** | **0.75** | **0.88** | **0.83** | **0.75** | **0.87** | **0.74** | **0.80** |

From Table 5, we can observe that our results are significantly better than their method except for a slightly lower PPV value of the complete tumor region.

In addition, we have noticed that Zhao *et al.* [25] proposed a brain tumor segmentation method of combining FCNNs and CRFs. Different from them, in this study we employ CRFs as a fine segmentation process, which can achieve competitive performance while saving time overhead. As reported, they have a training duration of about 12 days, whereas ours is about 3 days on the shared GPU server. From Table 5, our results are slightly lower than Zhao *et al.*'s in some indicators because we only use a simple post-processing procedure and they carry out a series of complex post-processing operations. Besides, although their post-processing method is high-performance, it is also highly hand-crafted and may not easily be transferred to other related segmentation tasks. Furthermore, in their study they published segmentation results without post-processing on the BRATS 2013 testing dataset, for complete, core, and enhancing tumor regions, their results are 0.85, 0.83, 0.74 for DSC, 0.85, 0.84, 0.68 for PPV, and 0.88, 0.82, 0.82 for Sensitivity, respectively. While our results are 0.88, 0.81, 0.76 for DSC, 0.86, 0.81, 0.69 for PPV, and 0.90, 0.84, 0.86 for Sensitivity, which are better than Zhao *et al.*'s.

### 2) RESULTS ON THE BRATS 2015 DATABASE

To illustrate the robustness of our method, we also use the proposed method to segment the 3D MR images provided by the BRATS 2015. Similar to Chen *et al.* [29] and Zhan *et al.* [26] did in their works, we randomly selected 20% MRI scans as testing sets, and the remaining scans are used as training sets. As shown in Table 6, we herein present a comparison of the proposed method with other approaches in terms of DSC, PPV, and Sensitivity. It can be seen that the proposed method achieves better results on the overall performance, especially the DSC scores exceed all the methods in Table 6. In addition, the boxplots of the results are shown in Fig. 8. Obviously, our method performs well on all images in the testing sets except for some outliers.

### 3) RESULTS ON THE BRATS 2018 DATABASE

To further evaluate the effectiveness of the proposed method, we conduct an experiment on the BRATS 2018 dataset. The data provided by BRATS 2018 differs from the data provided during the previous BRATS challenges. Specifically, the tumor sub-regions consist of the enhancing tumor (ET), the whole tumor (WT), and the tumor core (TC). The labels in this dataset include label 1 for necrosis and non-enhancing, label 2 for edema, label 4 for enhancing tumor. The segmentation performance of our algorithm on the validation dataset can be assessed by DSC, Hausdorff Distance (HD), Specificity, and Sensitivity, which are calculated through the online evaluation system. Table 7 shows the quantitative performance of the proposed method and other six state-of-the-art approaches on 66 validation unseen cases. Particularly, the method proposed by Myronenko [32] ranked first at the BRATS 2018 challenge. Their approach achieved excellent segmentation performance by integrating 10 models and adopting test-time augmentation (TTA) strategy, but their

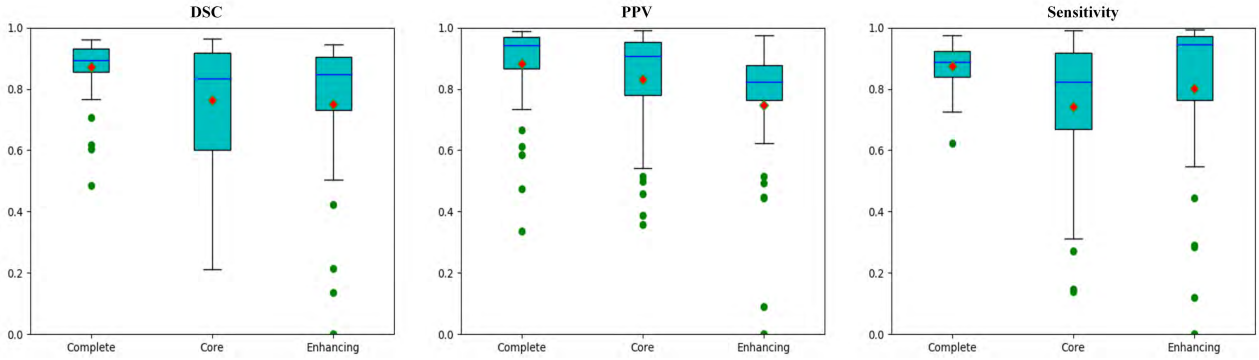



**FIGURE 8.** Boxplots of the results on the BRATS 2015 dataset using the proposed method. Red indicates the mean, blue indicates the median, and green indicates the outliers.

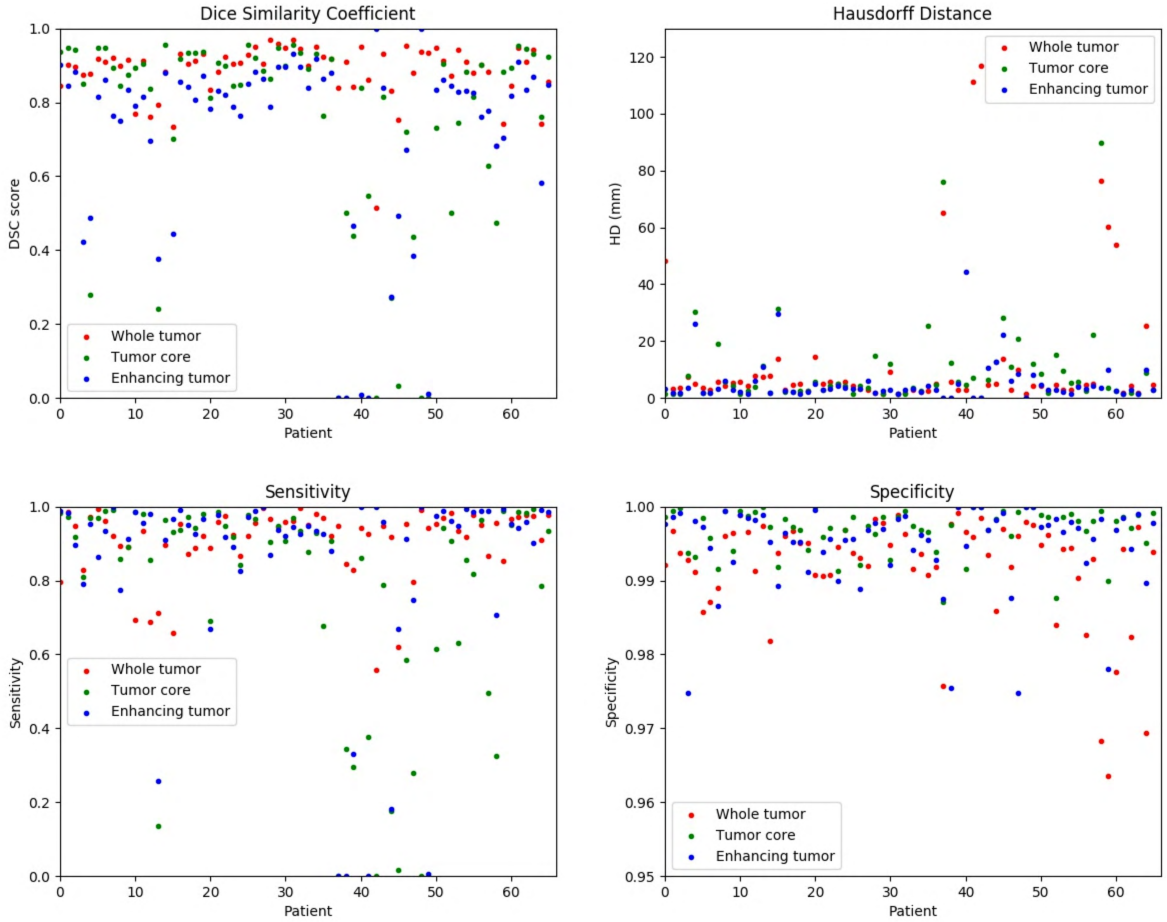



**FIGURE 9.** Scatter diagram of the results on the BRATS 2018 dataset using the proposed method.

models were quite complex and required high computing resources.

Relatively speaking, our model is not so complicated, and promising performance can be obtained on the BRATS 2013, 2015, and 2018 databases. In addition, Fig. 9 gives a scatter diagram of the validation set. It shows that our algorithm performs well for most brain images. Since the data of BRATS 2018 differ significantly and the class distribution is highly imbalanced, thus there are a few outliers that cause the average score to decrease.

More importantly, our method is mainly aimed at improving the method of Havaei *et al.* [23] and the improvements are reflected in the network architecture and the coarse-to-fine segmentation framework. The experimental results have

**TABLE 7.** Comparison of the proposed method with other state-of-the-art approaches on the BRATS 2018 dataset.

| Methods | DSC | | | Sensitivity | | | Specificity | | | HD | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | ET | WT | TC | ET | WT | TC | ET | WT | TC | ET | WT | TC |
| Zhou et al. [34] | 0.7922 | 0.9074 | 0.8358 | 0.8135 | **0.9080** | **0.8101** | **0.9982** | **0.9952** | **0.9982** | 2.80 | 4.48 | 7.07 |
| Wang et al. [35] | 0.7707 | 0.8956 | 0.7304 | - | - | - | - | - | - | 4.44 | 4.92 | 11.13 |
| | 0.7543 | 0.8731 | 0.7832 | - | - | - | - | - | - | 4.53 | 5.90 | 8.03 |
| Myronenko et al. [32] | **0.8233** | **0.9100** | **0.8668** | - | - | - | - | - | - | **3.9257** | **4.5160** | **6.8545** |
| Chandra et al. [36] | 0.7405 | 0.8718 | 0.7946 | 0.7946 | 0.8299 | 0.7887 | 0.9970 | 0.9964 | 0.9973 | 5.575 | 5.0379 | 9.5883 |
| Marcinkiewicz et al. [37] | 0.7229 | 0.8623 | 0.7675 | 0.7773 | 0.8553 | 0.7564 | 0.9976 | 0.9931 | 0.9967 | - | - | - |
| Zhang et al. [38] | 0.7486 | 0.8895 | 0.7395 | 0.7325 | 0.8803 | 0.7209 | 0.9832 | 0.9806 | 0.9833 | 21.9397 | 7.2743 | 27.3519 |
| Our method | **0.7178** | **0.8824** | **0.7481** | **0.8684** | **0.9074** | **0.7621** | **0.9947** | **0.9918** | **0.9969** | **5.6864** | **12.6069** | **9.6223** |

**TABLE 8.** Comparison of computational complexity of different methods for one subject patient.

| Methods | Meier | Reza | Tustison et al. [14] | Havaei et al. [23] | Pereira et al. [21] | Zhao et al. [25] | Chen et al. [29] | Our method |
|---|---|---|---|---|---|---|---|---|
| Time | 6 min | 90 min | 100 min | 3 min | 8 min | 2-4 min | 186.93 s | **1.5-3 min** |

shown the superior of the proposed method when compared with Havaei *et al.*'s method.

### F. SEGMENTATION TIME

Thanks to the convolutional nature of the models and the use of an efficient GPU implementation, the proposed segmentation system can segment the brain image slice by slice in the testing stage. In general, on the shared GPU server, our method took 1.5-3 minutes to segment an entire brain for the three views. A comparison of the testing time for the different segmentation methods is shown in Table 8. The results demonstrate that our method not only has stable segmentation results, but also has relatively low computational complexity.

## V. CONCLUSION

In this paper, we firstly proposed a novel MCCNN architecture to extract more discriminative multi-scale features for brain tumor segmentation. Especially, we considered different network frameworks and studied their impact on the segmentation performance and found that high performance can be achieved by stacking three CNNs to model the dependencies of labels and exploiting the nature of fully convolution. Secondly, we presented a coarse-to-fine segmentation framework by combining MCCNN and fully connected CRFs. In this framework, we first performed a coarse segmentation using MCCNN in order to obtain the tumor contour. Then, we used CRFs to refine the segmentation results by minimizing an energy function. In particular, we train three segmentation models using 2D patches obtained from different perspectives (axial, coronal, and sagittal views), respectively, and then to obtain the final segmentation results using a voting based fusion strategy. Finally, the experimental results have demonstrated the performance of the proposed method compared with some state-of-the-art approaches. Our experimental results also show that different modalities have different effects on the performance. In addition, we found that the Flair and T1c sequences play an essential role in the training of a segmentation model, and the segmentation model based on Flair, T1c, and T2 modalities obtained similar segmentation results and lower computation time compared to the segmentation model based on all four modalities.

Overall, our method provided promising performance on the BRATS 2013, 2015, and 2018 datasets compared to the state-of-the-art approaches. It took about 1.5-3 minutes to segment an entire brain, which is faster than most patch-based methods. However, our model has a decrease in performance when the data differing significantly. Maybe 2D CNN still cannot fully utilize 3D information of MRI data [13]. Our experimental results have shown that the fusion of information from different views can improve the performance of tumor segmentation, but not each single view can achieve good segmentation results. Therefore, our future research is to consider the interaction information of images from more perspectives and effectively integrate them into a 3D CNN for brain tumor segmentation.

## ACKNOWLEDGMENT

The authors would like to thank Dr. Mohammad Havaei from Université de Sherbrooke for some helpful discussions about the experiment. The authors would also like to thank the anonymous reviewers for their insightful comments, which have greatly helped to improve the quality of this paper.

## REFERENCES

[1] S. Bauer, R. Wiest, L.-P. Nolte, and M. Reyes, "A survey of MRI-based medical image analysis for brain tumor studies," *Phys. Med. Biol.*, vol. 58, no. 13, p. R97, 2013.

[2] A. Işın, C. Direkoğlu, and M. Şah, "Review of MRI-based brain tumor image segmentation using deep learning methods," *Procedia Comput. Sci.*, vol. 102, pp. 317–324, 2016.

[3] M. Goetz, C. Weber, F. Binczyk, J. Polanska, R. Tarnawski, B. Bobek-Billewicz, U. Koethe, J. Kleesiek, B. Stieltjes, and K. H. Maier-Hein, "DALSA: Domain adaptation for supervised learning from sparsely annotated MR images," *IEEE Trans. Med. Imag.*, vol. 35, no. 1, pp. 184–196, Jan. 2016.

[4] S. D. Olabarriaga and A. W. Smeulders, "Interaction in the segmentation of medical images: A survey," *Med. Image Anal.*, vol. 5, no. 2, pp. 127–142, 2001.

[5] J. Yao, "Image processing in tumor imaging," in *New Techniques in Oncologic Imaging*. London, U.K.: Taylor & Francis, 2006, pp. 79–102.

[6] B. H. Menze *et al.*, "The multimodal brain tumor image segmentation benchmark (BRATS)," *IEEE Trans. Med. Imag.*, vol. 34, no. 10, pp. 1993–2024, Oct. 2015.

[7] M. Prastawa, E. Bullitt, S. Ho, and G. Gerig, "A brain tumor segmentation framework based on outlier detection," *Med. Image Anal.*, vol. 8, no. 3, pp. 275–283, 2004.
[8] M. B. Cuadra, C. Pollo, A. Bardera, O. Cuisenaire, J. G. Villemure, and J. P. Thiran, "Atlas-based segmentation of pathological MR brain images using a model of lesion growth," *IEEE Trans. Med. Imag.*, vol. 23, no. 10, pp. 1301–1314, Oct. 2004.
[9] B. H. Menze, K. van Leemput, D. Lashkari, M.-A. Weber, N. Ayache, and P. Golland, "A generative model for brain tumor segmentation in multi-modal images," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Beijing, China: Springer, 2010, pp. 151–159.
[10] A. Gooya, K. M. Pohl, M. Bilello, L. Cirillo, G. Biros, E. R. Melhem, and C. Davatzikos, "GLISTR: Glioma image segmentation and registration," *IEEE Trans. Med. Imag.*, vol. 31, no. 10, pp. 1941–1954, Oct. 2012.
[11] C.-H. Lee, M. Schmidt, A. Murtha, A. Bistritz, J. Sander, and R. Greiner, "Segmenting brain tumors with conditional random fields and support vector machines," in *Proc. Int. Workshop Comput. Vis. Biomed. Image Appl.* Beijing, China: Springer, 2005, pp. 469–478.
[12] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Semantic image segmentation with deep convolutional nets and fully connected CRFs," Dec. 2014, *arXiv:1412.7062*. [Online]. Available: https://arxiv.org/abs/1412.7062
[13] K. Kamnitsas, C. Ledig, V. F. J. Newcombe, J. P. Simpson, A. D. Kane, D. K. Menon, D. Rueckert, and B. Glocker, "Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation," *Med. Image Anal.*, vol. 36, pp. 61–78, Feb. 2017.
[14] N. J. Tustison, K. L. Shrinidhi, M. Wintermark, C. R. Durst, B. M. Kandel, J. C. Gee, M. C. Grossman, and B. B. Avants, "Optimal symmetric multimodal templates and concatenated random forests for supervised brain tumor segmentation (simplified) with *ANTsR*," *Neuroinformatics*, vol. 13, no. 2, pp. 209–225, 2015.
[15] R. Meier, S. Bauer, J. Slotboom, R. Wiest, and M. Reyes, "Appearance-and context-sensitive features for brain tumor segmentation," in *Proc. MICCAI BRATS Challenge*, 2014, pp. 20–26.
[16] S. Reza and K. M. Iftekharuddin, "Improved brain tumor tissue segmentation using texture features," in *Proc. MICCAI BraTS (Brain Tumor Segmentation Challenge)*, 2014, pp. 27–30.
[17] S. Bauer, L.-P. Nolte, and M. Reyes, "Fully automatic segmentation of brain tumor images using support vector machine classification in combination with hierarchical conditional random field regularization," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Toronto, ON, Canada: Springer, 2011, pp. 354–361.
[18] D. Zikic, B. Glocker, E. Konukoglu, A. Criminisi, C. Demiralp, J. Shotton, O. M. Thomas, T. Das, R. Jena, and S. J. Price, "Decision forests for tissue-specific segmentation of high-grade gliomas in multi-channel MR," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Nice, France: Springer, 2012, pp. 369–376.
[19] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. W. M. van der Laak, B. Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Med. Image Anal.*, vol. 42, pp. 60–88, Dec. 2017.
[20] J. Ker, L. Wang, J. Rao, and T. Lim, "Deep learning applications in medical image analysis," *IEEE Access*, vol. 6, pp. 9375–9389, 2018.
[21] S. Pereira, A. Pinto, V. Alves, and C. A. Silva, "Brain tumor segmentation using convolutional neural networks in MRI images," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1240–1251, May 2016.
[22] G. Urban, M. Bendszus, F. Hamprecht, and J. Kleesiek, "Multi-modal brain tumor segmentation using deep convolutional neural networks," in *Proc. MICCAI BraTS (Brain Tumor Segmentation) Challenge, Winning Contribution*, 2014, pp. 31–35.
[23] M. Havaei, A. Davy, D. Warde-Farley, A. Biard, A. Courville, Y. Bengio, C. Pal, P.-M. Jodoin, and H. Larochelle, "Brain tumor segmentation with deep neural networks," *Med. Image Anal.*, vol. 35, pp. 18–31, Jan. 2017.
[24] G. Wang, M. A. Zuluaga, W. Li, R. Pratt, P. A. Patel, M. Aertsen, T. Doel, A. L. David, J. Deprest, S. Ourselin, and T. Vercauteren, "DeepIGeoS: A deep interactive geodesic framework for medical image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 7, pp. 1559–1572, Jul. 2019.
[25] X. Zhao, Y. Wu, G. Song, Z. Li, Y. Zhang, and Y. Fan, "A deep learning model integrating FCNNs and CRFs for brain tumor segmentation," *Med. Image Anal.*, vol. 43, pp. 98–111, Jan. 2017.
[26] T. Zhan, F. Shen, X. Hong, X. Wang, Y. Chen; Z. Lu, and G. Yang, "A glioma segmentation method using cotraining and superpixel-based spatial and clinical constraints," *IEEE Access*, vol. 6, pp. 57113–57122, 2018.
[27] J. Wei and Y. Xia, "Multi-scale networks for segmentation of brain magnetic resonance images," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Cham, Switzerland: Springer, 2017, pp. 312–320.
[28] F. Isensee, P. Kickingereder, D. Bonekamp, M. Bendszus, W. Wick, H.-P. Schlemmer, and K. Maier-Hein, "Brain tumor segmentation using large receptive field deep convolutional neural networks," in *Bildverarbeitung für die Medizin 2017*. Berlin, Germany: Springer-Vieweg, 2017, pp. 86–91.
[29] S. Chen, C. Ding, and M. Liu, "Dual-force convolutional neural networks for accurate brain tumor segmentation," *Pattern Recognit.*, vol. 88, pp. 90–100, Apr. 2019.
[30] A. Davy, M. Havaei, D. Warde-Farley, A. Biard, A. Courville, Y. Bengio, C. Pal, P. M. Jodoin, and H. Larochelle, "Brain tumor segmentation with deep neural networks," in *Proc. MICCAI BraTS (Brain Tumor Segmentation) Challenge, Winning Contribution*, Boston, MA, USA, 2014, pp. 1–5.
[31] E. A. R. Piedra, B. M. Ellingson, R. K. Taira, S. El-Saden, A. A. Bui, and W. Hsu, "Brain tumor segmentation by variability characterization of tumor boundaries," in *Proc. Int. Workshop Brainlesion, Glioma, Multiple Sclerosis, Stroke Traumatic Brain Injuries*. Athens, Greece: Springer, 2016, pp. 206–216.
[32] A. Myronenko, "3D MRI brain tumor segmentation using autoencoder regularization," in *Proc. Int. MICCAI Brainlesion Workshop*. Granada, Spain: Springer, 2018, pp. 311–320.
[33] G. Wang, W. Li, S. Ourselin, and T. Vercauteren, "Automatic brain tumor segmentation using cascaded anisotropic convolutional neural networks," in *Proc. Int. MICCAI Brainlesion Workshop*. Quebec City, QC, Canada: Springer, 2017, pp. 178–190.
[34] C. Zhou, S. Chen, C. Ding, and D. Tao, "Learning contextual and attentive information for brain tumor segmentation," in *Proc. Int. MICCAI Brainlesion Workshop*. Granada, Spain: Springer, 2018, pp. 497–507.
[35] G. Wang, W. Li, S. Ourselin, and T. Vercauteren, "Automatic brain tumor segmentation using convolutional neural networks with test-time augmentation," in *Proc. Int. MICCAI Brainlesion Workshop*. Granada, Spain: Springer, 2018, pp. 61–72.
[36] S. Chandra, M. Vakalopoulou, L. Fidon, E. Battistella, T. Estienne, R. Sun, C. Robert, E. Deutsch, and N. Paragios, "Context aware 3D CNNs for brain tumor segmentation," in *Proc. Int. MICCAI Brainlesion Workshop*. Granada, Spain: Springer, 2018, pp. 299–310.
[37] M. Marcinkiewicz, J. Nalepa, P. R. Lorenzo, W. Dudzik, and G. Mrukwa, "Automatic brain tumor segmentation using a two-stage multi-modal FCNN," in *Proc. Int. MICCAI Brainlesion Workshop*. Granada, Spain: Springer, 2018, pp. 314–321.
[38] X. Zhang, W. Jian, and K. Cheng, "3D dense U-nets for brain tumor segmentation," in *Proc. Int. MICCAI Brainlesion Workshop*. Granada, Spain: Springer, 2018, pp. 562–570.
[39] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 3431–3440.
[40] K. Hu, Z. Zhang, X. Niu, Y. Zhang, C. Cao, F. Xiao, and X. Gao, "Retinal vessel segmentation of color fundus images using multiscale convolutional neural network with an improved cross-entropy loss function," *Neurocomputing*, vol. 309, pp. 179–191, Oct. 2018.
[41] D. Yi, M. Zhou, Z. Chen, and O. Gevaert, "3-D convolutional neural networks for glioblastoma segmentation," Nov. 2016, *arXiv:1611.04534*. [Online]. Available: https://arxiv.org/abs/1611.04534
[42] S. Zheng, S. Jayasumana, B. Romera-Paredes, V. Vineet, Z. Su, D. Du, C. Huang, and P. H. S. Torr, "Conditional random fields as recurrent neural networks," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2015, pp. 1529–1537.
[43] M. Kistler, S. Bonaretti, M. Pfahrer, R. Niklaus, and P. Büchler, "The virtual skeleton database: An open access repository for biomedical research and collaboration," *J. Med. Internet Res.*, vol. 15, no. 11, p. e245, 2013.
[44] S. Bakas, H. Akbari, A. Sotiras, M. Bilello, M. Rozycki, J. S. Kirby, J. B. Freymann, K. Farahani, and C. Davatzikos, "Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features," *Sci. Data*, vol. 4, p. 170117, Sep. 2017.
[45] S. Bakas, H. Akbari, A. Sotiras, M. Bilello, M. Rozycki, J. Kirby, J. Freymann, K. Farahani, and C. Davatzikos, "Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-GBM collection," *Cancer Imag. Arch.*, 2017. doi: 10.7937/K9/TCIA.2017.KLXWJJ1Q.

[46] S. Bakas, H. Akbari, A. Sotiras, M. Bilello, M. Rozycki, J. Kirby, J. Freymann, K. Farahani, and C. Davatzikos, "Segmentation labels and radiomic features for the pre-operative scans of the TCGA-LGG collection," *Cancer Imag. Arch.*, 2017. doi: 10.7937/K9/TCIA.2017.GJQ7R0EF.
[47] N. J. Tustison, B. B. Avants, P. A. Cook; Y. Zheng, A. Egan, P. A. Yushkevich, and J. C. Gee, "N4ITK: Improved N3 bias correction," *IEEE Trans. Med. Imag.*, vol. 29, no. 6, pp. 1310–1320, Jun. 2010.
[48] L. G. Nyúl, J. K. Udupa, and X. Zhang, "New variants of a method of mri scale standardization," *IEEE Trans. Med. Imag.*, vol. 19, no. 2, pp. 143–150, Feb. 2000.
[49] L. Bottou, Y. Bengio, and Y. Le Cun, "Global training of document processing systems using graph transformer networks," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 1997, pp. 489–494.
[50] J. D. Lafferty, A. McCallum, and F. C. N. Pereira, "Conditional random fields: Probabilistic models for segmenting and labeling sequence data," in *Proc. 18th Int. Conf. Mach. Learn.*, San Francisco, CA, USA, 2001, pp. 282–289.
[51] P. Krähenbühl and V. Koltun, "Efficient inference in fully connected crfs with Gaussian edge potentials," in *Proc. Adv. Neural Inf. Process. Syst.*, 2011, pp. 109–117.
[52] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, 2014.
[53] Y. Bengio, "Practical recommendations for gradient-based training of deep architectures," in *Neural Networks: Tricks of the Trade*. Berlin, Germany: Springer, 2012, pp. 437–478.

**KAI HU** (M'16) received the B.S. degree in computer science and the Ph.D. degree in computational mathematics from Xiangtan University, Hunan, China, in 2007 and 2013, respectively. He was a Visiting Scholar with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore, from 2016 to 2017. He is currently an Associate Professor with the College of Information Engineering, Xiangtan University. His current research interests include machine learning, pattern recognition, and medical image processing.

**QINGHAI GAN** received the B.S. degree in information and computational science from the Hunan University of Technology, Hunan, China, in 2016. She is currently pursuing the M.S. degree in computer technology with Xiangtan University, China. Her current research interests include deep learning and medical image processing.

**YUAN ZHANG** received the B.S. degree in biomedical engineering from Zhengzhou University, Henan, China, in 2009, and the M.S. degree in information and communication engineering from Xiangtan University, Hunan, China, in 2012, where she is currently pursuing the Ph.D. degree in computational mathematics. Her research interests include wavelet analysis, machine learning, and biomedical signal processing.

**SHUHUA DENG** received the B.S. degree in computer science and the Ph.D. degree in computational mathematics from Xiangtan University, Hunan, China, in 2013 and 2018, respectively. His current research interests include software-defined networks, network security, and machine learning.

**FEN XIAO** received the B.S. degree in computer science and the Ph.D. degree in computational mathematics from Xiangtan University, Hunan, China, in 2002 and 2008, respectively. She was a Visiting Scholar with the Pacific Northwest National Laboratory, USA, from 2015 to 2016. She is currently a Professor with the College of Information Engineering, Xiangtan University. Her research interests include wavelet analysis theory, neural networks, and image processing.

**WEI HUANG** received the B.S. degree in medical imaging from the University of South China, Hunan, China, in 2005, and the M.S. degree in clinical medicine from Central South University, Hunan, China, in 2014. He is currently an Associate Professor with the Radiology Department, The First Hospital of Changsha. His research interests include medical imaging, medical image processing, and disease diagnosis.

**CHUNHONG CAO** received the B.S. degree in computational mathematics and applied software from Central South University, Hunan, China, in 1999, and the M.S. degree in computer software and theory and the Ph.D. degree in computational mathematics from Xiangtan University, Hunan, China, in 2005 and 2017, respectively. She is currently an Associate Professor with the College of Information Engineering, Xiangtan University. Her current research interests include compressed sensing, pattern recognition, and image processing.

**XIEPING GAO** (M'16) was born in 1965. He received the B.S. and M.S. degrees from Xiangtan University, China, in 1985 and 1988, respectively, and the Ph.D. degree from Hunan University, China, in 2003. He was a Visiting Scholar with the National Key Laboratory of Intelligent Technology and Systems, Tsinghua University, China, from 1995 to 1996, and also with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore, from 2002 to 2003. He is currently a Professor with the College of Software and Communication Engineering, Xiangnan University, China. He has authored and coauthored more than 80 journal papers, conference papers, and book chapters. His current research interests include wavelets analysis, neural networks, image processing, computer networks, mobile communication, and bioinformatics. He has been a member of the technical committees of several scientific conferences. He is a Regular Reviewer for several journals.