Brain Tumor Segmentation Using a Fully Convolutional Neural Network with Conditional Random Fields

Xiaomei Zhao[1(✉)], Yihong Wu[1(✉)], Guidong Song[2], Zhenye Li[3], Yong Fan[4], and Yazhuo Zhang[2,3,5,6]

[1] National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing, China
zhaoxiaomei14@mails.ucas.ac.cn, yhwu@nlpr.ia.ac.cn
[2] Beijing Neurosurgical Institute, Capital Medical University, Beijing, China
[3] Department of Neurosurgery, Beijing Tiantan Hospital, Capital Medical University, Beijing, China
[4] Department of Radiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, USA
[5] Beijing Institute for Brain Disorders Brain Tumor Center, Beijing, China
[6] China National Clinical Research Center for Neurological Diseases, Beijing, China

Abstract. Deep learning techniques have been widely adopted for learning task-adaptive features in image segmentation applications, such as brain tumor segmentation. However, most of existing brain tumor segmentation methods based on deep learning are not able to ensure appearance and spatial consistency of segmentation results. In this study we propose a novel brain tumor segmentation method by integrating a Fully Convolutional Neural Network (FCNN) and Conditional Random Fields (CRF), rather than adopting CRF as a post-processing step of the FCNN. We trained our network in three stages based on image patches and slices respectively. We evaluated our method on BRATS 2013 dataset, obtaining the second position on its Challenge dataset and first position on its Leaderboard dataset. Compared with other top ranking methods, our method could achieve competitive performance with only three imaging modalities (Flair, T1c, T2), rather than four (Flair, T1, T1c, T2), which could reduce the cost of data acquisition and storage. Besides, our method could segment brain images slice-by-slice, much faster than the methods patch-by-patch. We also took part in BRATS 2016 and got satisfactory results. As the testing cases in BRATS 2016 are more challenging, we added a manual intervention post-processing system during our participation.

Keywords: Brain tumor segmentation · Magnetic resonance image · Fully Convolutional Neural Network · Conditional Random Fields · Recurrent Neural Network

A. Crimi et al. (Eds.): BrainLes 2016, LNCS 10154, pp. 75–87, 2016.
DOI: 10.1007/978-3-319-55524-9_8

1 Introduction

Accurate automatic or semi-automatic brain tumor segmentation is very helpful in clinical, however, it remains a challenging task up to now [1]. Gliomas are the most frequency primary brain tumors in adults [2]. Therefore, the majority of brain tumor segmentation methods focus on gliomas. So do we in this paper. Accurate segmentation of gliomas is very difficult for the following reasons: (1) in MR images, gliomas may have the same appearance with gliosis, stroke and so on [3]; (2) gliomas have a variety of shape, appearance, and size, and may appear in any position in the brain; (3) gliomas invade the surrounding tissue rather than displacing it, causing fuzzy boundaries [3]; (4) there exists intensity inhomogeneity in MR images.

The existing brain tumor segmentation methods can be roughly divided into two groups: generative models and discriminative models. Generative models usually acquire prior information through probabilistic atlas image registration [4,5]. However, the image registration is unreliable when the brain is deformed due to large tumors. Discriminative models typically segment brain tumors by classifying voxels based on image features [6,7]. Their segmentation performance is hinged on the image features and classification models. Since deep learning techniques are capable of learning high level and task-adaptive features from training data, they have been adopted in brain tumor segmentation studies [8–14]. However, most of the existing brain tumor segmentation methods based on deep learning do not yield segmentation results with appearance and spatial consistency [15]. To overcome such a limitation, we propose a novel deep network by integrating a fully convolutional neural network (FCNN) and a CRF to segment brain tumors. Our model is trained in three steps and is able to segment brain images slice-by-slice, which is much faster than the segmentation method patch-by-patch [14]. Moreover, our method requires only three MR imaging modalities (Flair, T1c, T2), rather than four modalities (Flair, T1, T1c, T2) [1,6–14], which could help reduce the cost of data acquisition and storage.

2 The Proposed Method

The proposed brain tumor segmentation method consists of three main steps: pre-processing, segmentation using the proposed deep network model, and post-processing. In the following, we will introduce each step in detail respectively.

2.1 Pre-processing

As magnetic resonance imaging devices are not perfect and each imaging object is specific, the intensity ranges and bias fields of different MR images are different. Therefore, the absolute intensity values in different MR images or even in the same MR image do not have fixed tissue meanings. It is necessary to pre-process MR images in an appropriate way.

In this paper, we firstly use N4ITK [16] to correct the bias field of each MR image. Then, we normalize the intensity by subtracting the gray-value of the highest frequency and dividing the revised deviation. We denote the revised deviation by $\tilde{\sigma}$ and the MR image ready to be normalized by V, which is composed by a set of voxels $\{v_1, v_2, v_3, \ldots, v_N\}$. The intensity value of each voxel v_k is denoted as I_k. Then, the revised deviation $\tilde{\sigma}$ can be calculated by $\tilde{\sigma} = \sqrt{\sum_{k=1}^{N}(I_k - \hat{I})^2/N}$, where $\hat{I}$ denotes the gray-value of the highest frequency. Besides, in order to process the MR images as common images, we also change their intensity range to 0–255 linearly.

We take T2 for an example to show the effect of our normalization method. Figure 1 shows 30 T2 MR images' intensity histograms before and after normalization. The 30 T2 MR images come from BRATS 2013 training dataset. It can be seen from Fig. 1 that our normalization method can try to make different MR images have similar intensity distributions, while guarantee their histogram shapes unchanged. In most cases, the gray value of the highest frequency is close to the intensity of white matter. Therefore, transforming the gray value of the highest frequency to the same level is equivalent to transforming the intensity of white matter to the same level. Then, after normalizing the revised deviation, the similar intensities in different MR images can roughly have the similar tissue meaning.

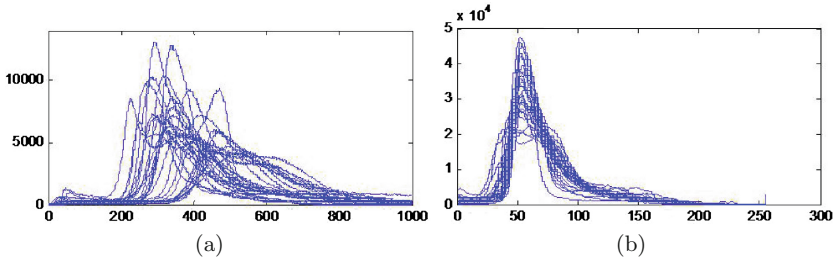

Fig. 1. Comparison of 30 T2 intensity histograms before and after intensity normalization. (a). Before normalization (after N4ITK); (b). After normalization

2.2 Brain Tumor Segmentation Model

Our brain tumor segmentation model consists of two parts, a Fully Convolutional Neural Network (FCNN) and Conditional Random Field (CRF), as shown in Fig. 2. The proposed model was trained by three steps, using image patches and slices respectively. In the testing phase, it can segment brain images slice by slice. Next, we will introduce each part of the proposed segmentation model in detail.

FCNN. FCNN contains the majority of parameters in our whole segmentation model. It was trained based on image patches, which were extracted from slices

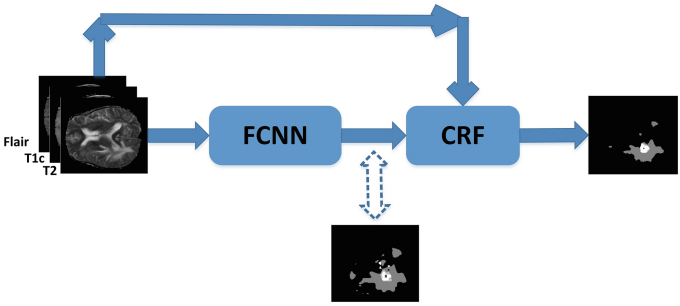

Fig. 2. The structure of our brain tumor segmentation model

of the axial view. Training FCNN by patches can avoid the problem of lacking training samples, as thousands of patches can be extracted from one image. It can also help to avoid the training sample imbalance problem, because the number and position of training samples for each class can be easily controlled by using different patch sampling schemes. In our experiment, we sampled training patches randomly from each training subject and kept the number of training samples for each class equal (5 classes in total, including normal tissue, necrosis, edema, non-enhancing core, and enhancing core). As we didn't reject patches sampled in the same place, there existed duplicated training samples. Figure 3 shows the structure of the proposed FCNN. Similar to the cascaded architecture proposed in [12], the inputs of our FCNN network also have two different sizes. Passing through a series of convolutional and pooling layers, the large inputs turn into feature maps with the same size of small inputs. These feature maps and small inputs are sent into the following network together. In this way, when we predict the center pixel's label, the local information and the context information in larger scale can be taken into consideration at the same time. Compared with the cascaded architecture proposed in [12], the two branches in our FCNN was trained simultaneously, while the two branches in the cascaded architecture in [12] was trained in different steps. Besides, our FCNN network has more convolutional layers.

FCNN is a fully convolutional neural network and the stride of each layer is set to 1. Therefore, even though it was trained by patches, it can segment brain images slice by slice.

CRF. Let's briefly review conditional random field first. Consider an image I composed by a set of pixels $\{I_1, I_2, \ldots, I_M\}$, where M denotes the number of pixels in this image. Each pixel I_i has a label x_i, $x_i \in L = \{l_1, l_2, \ldots, l_k\}$. L is a set of labels, showing the range of value for x_i. The energy function of CRF is written as:

$$E(x) = \sum_i \Phi(x_i) + \sum_{i,j \in N_i} \Psi(x_i, x_j), i \in \{1, 2, 3, \ldots, M\}, j \in N_i \qquad (1)$$

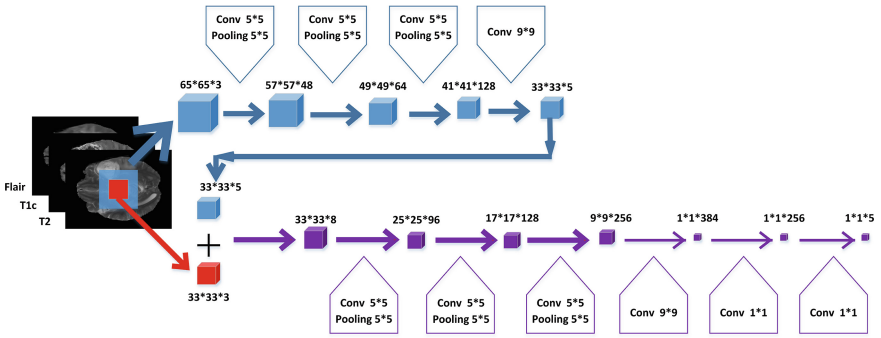

Fig. 3. The structure of our FCNN network

$\Phi(x_i)$ is the unary term, representing the cost of assigning label x_i to the pixel I_i. $\Psi(x_i, x_j)$ is the pairwise term, representing the cost of assigning label x_i and x_j to I_i and I_j respectively. N_i represents the neighborhood of pixel I_i. Using CRF to segment an image is to find a set of x_i to make the energy function have minimum value. In order to improve segmentation accuracy and get a global optimized result, fully connected CRF can be used, which is computing pairwise potentials on all pairs of pixels in the image [17]. The energy function of fully connected CRF is as follows:

$$E(x) = \sum_i \Phi(x_i) + \sum_{i<j} \Psi(x_i, x_j), i, j \in \{1, 2, 3, \ldots, M\} \tag{2}$$

Mean field approximation can be used to solve the optimize problem of minimizing the energy function (2) [17]. Shuai Zheng et al. [15] proposed a neural network formulated fully connected CRF, called CRF-RNN. CRF-RNN performs a mean field iteration by a stack of CNN layers, then the whole iteration steps in the mean field approximation can be formulated as Recurrent Neural Network, making it possible to integrate a CNN and CRF network as one deep network and train it with the usual back-propagation algorithm. CRF-RNN can be implemented on a GPU and has very high computational efficiency. Our CRF model refers to CRF-RNN [15], where the negative of the unary term $-\Phi(x_i)$ is directly provided by the previous segmentation network and the pairwise term $\Psi(x_i, x_j)$ is calculated by the following function:

$$\Psi(x_i, x_j) = \mu(x_i, x_j)[\omega^{(1)} exp(-\frac{|p_i - p_j|^2}{2\theta_\alpha^2} - \frac{|c_i - c_j|^2}{2\theta_\beta^2}) + \omega^{(2)} exp(-\frac{|p_i - p_j|^2}{2\theta_\gamma^2})] \tag{3}$$

In (3), μ is a label compatibility function, representing the penalty for different pixels that are assigned different labels; c_i and c_j denote the color vectors of pixels I_i and I_j; p_i and p_j denote the positions of pixels I_i and I_j; $\omega^{(k)}, k = 1, 2$ is the weight of each Gaussian kernel; θ_α, θ_β, and θ_γ are parameters that control the effect of different features (position and color). For further details about

CRF-RNN and fully connected CRF, please refer to Shuai Zheng et al. [15] and Philipp Krähenbühl et al. [17]. From (3) we can see that, when two adjacent pixels have similar color, the penalty of assigning different labels to them is large. Therefore after CRF, pixels having similar colors and positions are very likely to be assigned same label, ensuring the appearance and spatial consistency of segmentation results.

The Combination of FCNN and CRF-RNN. The proposed brain tumor segmentation network consists of FCNN and CRF-RNN. FCNN provides the preliminary probability of assigning each label to each voxel. These preliminary prediction results are considered as the negative of the unary term of CRF-RNN. CRF-RNN can globally optimize the segmentation results according to each voxel's intensity and position information shown in the pre-processed MR images. Then, the segmentation results' appearance and spatial consistency can be ensured.

In the training phase, we trained the proposed integrated network of FCNN and CRF-RNN in three steps. Firstly we used image patches to train FCNN. Then, we used image slices of the axial view to train the following CRF-RNN with parameters of the FCNN fixed. Finally, we used the image slices to fine-tune the whole network. In the testing phase, we segment brain images slice by slice. All the slices are extracted from the axial view.

2.3 Post-processing

We post-process the segmentation results by removing small 3D-connected regions and correcting some pixels' labels by a simple thresholding method. We validated the values of these thresholds by a small subset of training dataset.

3 Experiment

BRATS is a brain tumor image segmentation challenge. It is organized in conjunction with the International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI). Most of the start of art brain tumor segmentation methods have been evaluated on this benchmark. Since BRATS 2014 dataset is not available and the ranking results of BRATS 2015 testing dataset are not shown on BRATS website, we mainly evaluated our segmentation method on BRATS 2013 dataset. Also, we took part in BRATS 2016 and got satisfactory results. In BRATS 2013, the training dataset contains 20 HGG and 10 LGG. The testing dataset contains two parts. One is Challenge, containing 10 HGG. The other one is Leaderboard, containing 21 HGG and 4 LGG. In BRATS 2016, the training dataset contains 220 HGG and 54 LGG. The testing dataset contains 191 cases, including both grades. Our experiments were performed on our laboratory's server. The GPU of the server is Tesla K80, and the CPU is Intel E5-2620. As the server is public for everyone in our laboratory, we shared

one GPU with other colleagues most time. We used Caffe [18] to implement our neural network.

BRATS provides four MR imaging modalities for each subject, including Flair, T1, T1c, and T2. However, we just used three of them (Flair, T1c, T2), and still achieved competitive performance. We also trained a segmentation model using all the four modalities and tested its segmentation performance. However, there is no obvious performance difference between the segmentation model trained by four modalities (Flair, T1, T1c, T2) and the segmentation model trained by three modalities (Flair, T1c, T2). We speculate that most of effective information shown in T1 is contained in T1c.

3.1 Evaluation on BRATS 2013 Dataset

On BRATS 2013 evaluation website, three metrics of Dice, Positive Predictive Value (PPV), and Sensitivity are used to evaluate a method. Each of the metrics is calculated on three kinds of tumor regions. They are complete, core, and enhancing. The complete tumor region includes necrosis, edema, non-enhancing core, and enhancing core. The core region includes necrosis, non-enhancing core, and enhancing core. The enhancing region only includes the enhancing core. Equations for calculating the three metrics are as follows:

$$Dice(P_*, T_*) = \frac{|P_* \cap T_*|}{(|P_*| + |T_*|)/2} \tag{4}$$

$$PPV(P_*, T_*) = \frac{|P_* \cap T_*|}{|P_*|} \tag{5}$$

$$Sensitivity(P_*, T_*) = \frac{|P_* \cap T_*|}{|T_*|} \tag{6}$$

where $*$ indicates complete, core, or enhancing region. T_* denotes the true region of $*$. P_* denotes the segmented $*$ region. $|P_* \cap T_*|$ denotes the overlap area between P_* and T_*. $|P_*|$ and $|T_*|$ denote the areas of P_* and T_* respectively. BRATS doesn't provide the ground truth for testing subjects. Therefore, all the metrics of testing dataset can only be calculated by BRATS evaluation website[1].

Table 1 shows the Dice scores of FCNN, FCNN+CRF and FCNN+CRF+ post-processing on BRATS 2013 Challenge dataset and Leaderboard dataset. It can be seen from Table 1 that CRF can obviously improve the segmentation accuracy, and post-processing can improve the segmentation accuracy further. FCNN+CRF+post-processing performs best on all three regions of both dataset.

Figure 4 shows some segmentation results on BRATS 2013 Challenge dataset. Figures in each row, from top to bottom, represent: Flair, T1c, T2, segmentation results of FCNN, segmentation results of FCNN+CRF, and segmentation results of FCNN+CRF+post-processing. Compared with segmentation results in Row 4 (FCNN), segmentation results in Row 5 (FCNN+CRF) are smoother and have

[1] https://www.virtualskeleton.ch/BRATS/Start2013.

Table 1. The Dice scores of FCNN, FCNN+CRF, and FCNN+CRF+post-processing on BRATS 2013 Challenge and Leaderboard dataset

Methods	Dice					
	Challenge			Leaderboard		
	Comp.	Core	Enh.	Comp.	Core	Enh.
FCNN	0.74	0.72	0.67	0.70	0.61	0.54
FCNN+CRF	0.85	0.80	0.70	0.83	0.66	0.57
FCNN+CRF+post-processing	**0.87**	**0.83**	**0.76**	**0.86**	**0.73**	**0.62**

more accurate boundaries, representing the effectiveness of CRF. Compared the segmentation results in Row 5 (FCNN+CRF) and Row 6 (FCNN+CRF+post-processing), we can see that, after post-processing, the number of false positives reduces further.

Comparison results with other methods are summarized in Tables 2 and 3. Nick Tustison, Raphael Meier, Syed Reza, and Liang Zhao methods' evaluation results are acquired from BRATS 2013 website[2]. Nick Tustison, Raphael Meier, Syed Reza methods obtained the top 3 positions respectively on Challenge dataset in 2013. Nick Tustison, Liang Zhao, Raphael Meier methods obtained the top 3 positions respectively on Leaderboard dataset in 2013. Sérgio Pereira method [14] ranks first on Challenge dataset and second on Leaderboard dataset right now, while our method ranks second on Challenge dataset and first on Leaderboard dataset right now. In general, the proposed method takes 2–4 min to segment one subject's imaging data, much faster than Sérgio Pereira method (average running time of 8 min). From Tables 2 and 3, we can see that our segmentation method is still competitive with Sérgio Pereira method and much better than the other methods shown in these two tables.

Table 2. Comparison with other methods on BRATS 2013 Challenge dataset

Methods	Dice			PPV			Sensitivity		
	Comp.	Core	Enh.	Comp.	Core	Enh.	Comp.	Core	Enh.
Nick Tustison et al.	0.87	0.78	0.74	0.85	0.74	0.69	0.89	**0.88**	**0.83**
Raphael Meier et al.	0.82	0.73	0.69	0.76	0.78	0.71	**0.92**	0.72	0.73
Syed Reza et al.	0.83	0.72	0.72	0.82	0.81	0.70	0.86	0.69	0.76
Mohammad Havaei et al. [12]	**0.88**	0.79	0.73	0.89	0.79	0.68	0.87	0.79	0.80
Sérgio Pereira et al. [14]	**0.88**	**0.83**	**0.77**	0.88	**0.87**	0.74	0.89	0.83	0.81
Our method	0.87	**0.83**	0.76	**0.92**	**0.87**	**0.77**	0.83	0.81	0.77

[2] http://martinos.org/qtim/miccai2013/results.html.

Table 3. Comparison with other methods on BRATS 2013 Leaderboard dataset

Methods	Dice			PPV			Sensitivity		
	Comp.	Core	Enh.	Comp.	Core	Enh.	Comp.	Core	Enh.
Nick Tustison et al.	0.79	0.65	0.53	0.83	0.70	0.51	0.81	0.73	0.66
Liang Zhao et al.	0.79	0.59	0.47	0.77	0.55	0.50	0.85	0.77	0.53
Raphael Meier et al.	0.72	0.60	0.53	0.65	0.62	0.48	**0.88**	0.69	0.64
Mohammad Havaei et al. [12]	0.84	0.71	0.57	0.88	0.79	0.54	0.84	0.72	**0.68**
Sérgio Pereira et al. [14]	0.84	0.72	**0.62**	0.85	**0.82**	0.60	0.86	0.76	**0.68**
Our method	**0.86**	**0.73**	**0.62**	**0.89**	0.76	**0.64**	0.84	**0.78**	**0.68**

3.2 Participation on BRATS 2016 Challenge

The testing cases in BRATS 2016 are much more challenging. Therefore, we added a manual intervention post-processing system during our participation in BRATS 2016. The manual intervention post-processing system is designed to remove some segmented tumor regions which are obvious false positives. There are two kinds of regions to remove:

① Manually determined rectangular regions.

② Regions which intensities in Flair, T1c, T2 are below three specific thresholds respectively at the same time. The threshold of Flair equals to 0.8×the mean intensity of the segmented tumor region in Flair. The threshold of T1c is a constant. The threshold of T2 equals to 0.9×the mean intensity of the segmented tumor region in T2.

In the manual intervention post-processing system, users just need to decide whether to remove those regions described in ① ② and determine the rectangular regions' sizes and locations described in ①. The manual intervention post-processing system only takes a few minutes on each subject. The regions to remove are 3D. We show an example in Fig. 5 in 2D.

There are 191 cases in BRATS 2016 testing dataset with unknown grades. During our participation on BRATS 2016, we firstly segmented the 191 cases by our proposed integrated network of FCNN and CRF-RNN on our laboratory's sever. We just used one Tesla K80 GPU and one E5-2620 CPU on the sever. It took 2–4 min to segment one case. All 191 cases were segmented in about 11.5 h. We then used the manual intervention post-processing system to post-process segmentation results with one personal computer, in which there is one Q9550 CPU and no GPU. Not every case needed to be manually post-processed. On average, the manual intervention post-processing system only took a few minutes on each case. We successfully segmented the 191 cases in BRATS 2016 testing dataset in 48 h.

There are 19 groups that took part in BRATS 2016. Our method ranked first on the multi-temporal evaluation and ranked in the top 5 on most of items in tumor segmentation. The ranking details of our method are shown in Table 4. The formulation used to calculate Dice is (3), as shown in Sect. 3.1. And the

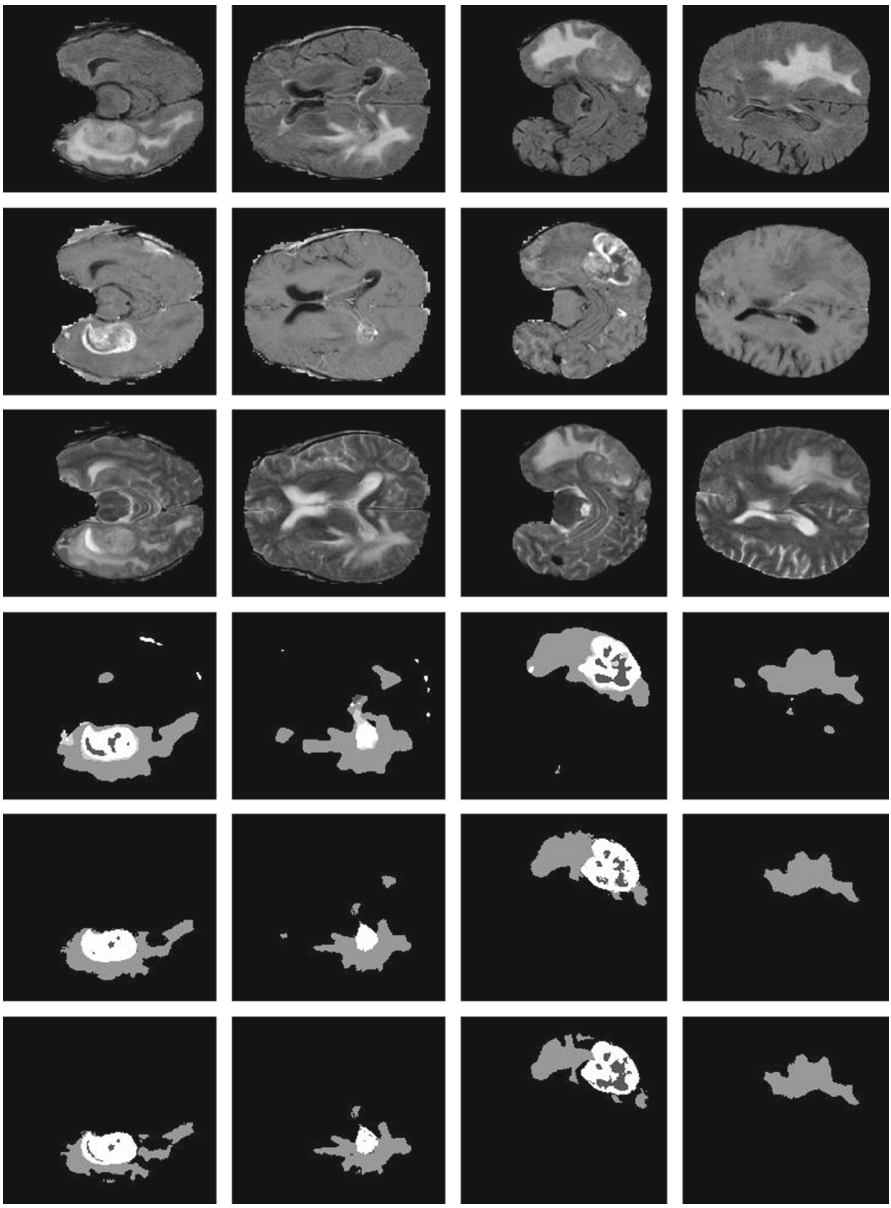

Fig. 4. Some segmentation results on BRATS 2013 Challenge dataset. The first and second columns show the segmentation results of the 50th and 80th slice of the axial view of Subject 0301. The third and fourth columns show the segmentation results of the 40th and 70th slice of the axial view of Subject 0308. Figures in each row, from top to bottom, represent: Flair, T1c, T2, segmentation results of FCNN, segmentation results of FCNN+CRF, and segmentation results of FCNN+CRF+post-processing. Each gray level in segmentation results represents a tumor class, from low to high: necrosis, edema, non-enhancing core, and enhancing core.

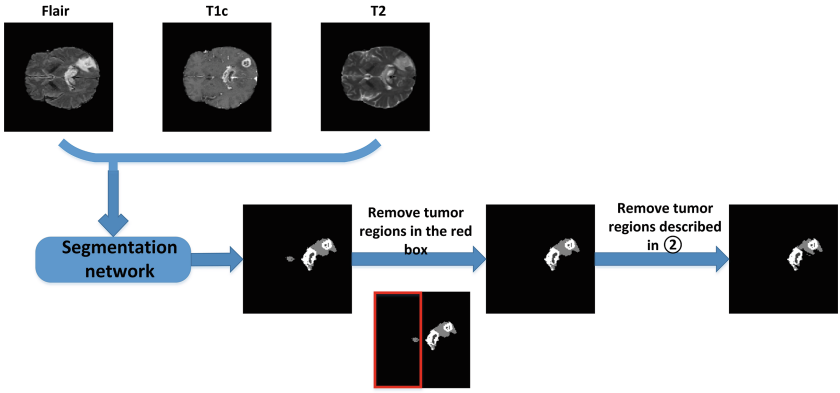

Fig. 5. A manual intervention post-processing example

Table 4. The ranking details of our method on different items on BRATS 2016 (including tie)

Items	Tumor segmentation						Multi-temporal evaluation
	Dice			Hausdorff			
	Comp.	Core	Enh.	Comp.	Core	Enh.	
Ranking	4	3	1	7	6	2	1

formulation used to calculate Hausdorff distance is as follow:

$$Haus(P_*, T_*) = max\{\sup_{p \in \partial P_*} \inf_{t \in \partial T_*} d(p,t), \sup_{t \in \partial T_*} \inf_{p \in \partial P_*} d(t,p)\} \tag{7}$$

The meanings of P_* and T_* have been shown in Sect. 3.1. ∂P_* denotes the surface of P_*, and ∂T_* denotes the surface of T_*. p and t denote points on ∂P_* and ∂T_* respectively. $d(p,t)$ calculates the least-square distance between points p and t. inf denotes the operation of returning the minimum value. sup and max denote the operation of returning the maximum value. Multi-temporal evaluation is designed to evaluate whether the volumetric segmentations provided by the participants are accurate enough to detect the changes indicated by the neuroradiologists[3].

4 Conclusion

Accurate automatic or semi-automatic brain tumor segmentation methods have broad application prospect. In this paper, we propose a novel brain tumor segmentation method by using an integrated model of Fully Convolutional Neural Network (FCNN) and Conditional Random Fields (CRF). This integrated model

[3] http://braintumorsegmentation.org/.

is designed to solve the problem in most existing deep learning brain tumor segmentation methods, by which the appearance and spatial consistency are hard to be ensured. In the CRF part, we use CRF-RNN, which formulates CRF as Recurrent Neural Network, making it possible to integrate FCNN and CRF as one deep network, rather than using CRF as a post-processing step of FCNN. Our integrated network was trained in three steps, using image patches and slices respectively. In the first step, image patches were used to train FCNN. These patches were randomly sampled from training dataset, but we controlled the number of patches for each class equal, in order to avoid the data imbalance problem. Patch-based training strategy could also avoid the problem of lacking training samples, because thousands of patches could be extracted from one subject's MR images. In the second step, slices from the axial view were used to train the following CRF-RNN, with parameters of the FCNN fixed. In the third step, slices from the axial view were used to fine-tune the whole network.

We applied a simple pre-processing strategy and a simple post-processing strategy. We pre-processed each MR image by N4ITK and intensity normalization, which normalized each MR image's intensity mainly by subtracting the gray-value of the highest frequency and dividing the revised deviation. We post-processed the segmentation results by removing small 3D-connected regions and correcting some pixels' labels by a simple thresholding method. The experimental results show that these strategies are effective.

We evaluated our method on BRATS 2013 dataset, obtaining the second position on its Challenge dataset and the first position on its Leaderboard dataset. Compared with other top ranking methods, our method could achieve competitive performance with only 3 imaging modalities (Flair, T1c, T2), rather than 4 (Flair, T1, T1c, T2). We also took part in BRATS 2016 and our method ranked first on the multi-temporal evaluation and ranked in the top 5 on most of items in tumor segmentation.

Acknowledgements. This work was supported by the National High Technology Research and Development Program of China (2015AA020504) and the National Natural Science Foundation of China under Grant No. 61572499, 61421004.

References

1. Menze, B.H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., et al.: The multimodal brain tumor image segmentation benchmark (BRATS). IEEE Trans. Med. Imaging **34**, 1993–2024 (2015)
2. Bauer, S., Wiest, R., Nolte, L.-P., Reyes, M.: A survey of MRI-based medical image analysis for brain tumor studies. Phys. Med. Biol. **58**, 97–129 (2013)
3. Goetz, M., Weber, C., Binczyk, F., Polanska, J., Tarnawski, R., Bobek-Billewicz, B., Koethe, U., Kleesiek, J., Stieltjes, B., Maier-Hein, K.H.: DALSA: domain adaptation for supervised learning from sparsely annotated MR images. IEEE Trans. Med. Imaging **35**, 184–196 (2016)
4. Prastawa, M., Bullitt, E., Ho, S., Gerig, G.: A brain tumor segmentation framework based on outlier detection. Med. Image Anal. **8**, 275–283 (2004)

5. Cobzas, D., Birkbeck, N., Schmidt, M., Jagersand, M., Murtha, A.: 3D variational brain tumor segmentation using a high dimensional feature set. In: IEEE 11th International Conference on Computer Vision, pp. 1–8 (2007)
6. Goetz, M., Weber, C., Bloecher, J., Stieltjes, B., Meinzer, H.-P., Maier-Hein, K.: Extremely randomized trees based brain tumor segmentation. In: Proceedings MICCAI BraTS (Brain Tumor Segmentation Challenge), pp. 6–11 (2014)
7. Kleesiek, J., Biller, A., Urban, G., Kothe, U., Bendszus, M., Hamprecht, F.: Ilastik for multi-modal brain tumor segmentation. In: Proceedings MICCAI BraTS (Brain Tumor Segmentation Challenge), pp. 12–17 (2014)
8. Davy, A., Havaei, M., Warde-farley, D., Biard, A., Tran, L., Jodoin, P.-M., Courville, A., Larochelle, H., Pal, C., Bengio, Y.: Brain tumor segmentation with deep neural networks. In: Proceedings MICCAI BraTS (Brain Tumor Segmentation Challenge), pp. 1–5 (2014)
9. Urban, G., Bendszus, M., Hamprecht, F., Kleesiek, J.: Multi-modal brain tumor segmentatioin using deep convolutional neural networks. In: Proceedings MICCAI BraTS (Brain Tumor Segmentation Challenge), pp. 31–35 (2014)
10. Zikic, D., Ioannou, Y., Brown, M., Criminisi, A.: Segmentation of brain tumor tissues with convolutional neural networks. In: Proceedings MICCAI BraTS (Brain Tumor Segmentation Challenge), pp. 36–39 (2014)
11. Dvorak, P., Menze, B.H.: Structured prediction with convolutional neural networks for multimodal brain tumor segmentation. In: Proceedings MICCAI BraTS (Brain Tumor Segmentation Challenge), pp. 13–24 (2015)
12. Havaei, M., Davy, A., Warde-Farley, D., Biard, A., Courville, A., Bengio, Y., Pal, C., Jodoin, P.-M., Larochelle, H.: Brain tumor segmentation with deep neural networks. Med. Image Anal. **35**, 18–31 (2017)
13. Vaidhya, K., Thirunavukkarasu, S., Alex, V., Krishnamurthi, G.: Multi-modal brain tumor segmentation using stacked denoising autoencoders. In: Proceedings MICCAI BraTS (Brain Tumor Segmentation Challenge), pp. 60–64 (2015)
14. Pereira, S., Pinto, A., Alves, V., Silva, C.A.: Brain tumor segmentation using convolutional neural networks in MRI images. IEEE Trans. Med. Imaging **35**, 1240–1251 (2016)
15. Zheng, S., Jayasumana, S., Romera-Paredes, B., Vineet, V., Su, Z., Du, D., Huang, C., Torr, P.H.: Conditional random fields as recurrent neural networks. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1529–1537 (2015)
16. Tustison, N.J., Avants, B.B., Cook, P.A., Zheng, Y., Egan, A., Yushkevich, P.A., Gee, J.C.: N4ITK: improved N3 bias correction. IEEE Trans. Med. Imaging **29**, 1310–1320 (2010)
17. Krähenbühl, P., Koltun, V.: Efficient inference in fully connected CRFs with Gaussian edge potentials. NIPS (2011)
18. Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S., Darrell, T.: Caffe: convolutional architecture for fast feature embedding. In: Proceedings of the 22nd ACM International Conference on Multimedia, pp. 675–678 (2014)