EFFICIENT SYMMETRY-DRIVEN FULLY CONVOLUTIONAL NETWORK FOR MULTIMODAL BRAIN TUMOR SEGMENTATION

Haocheng Shen[⋆] *Jianguo Zhang*[⋆] *Weishi Zheng*[†]

[⋆] Computing, School of Science and Engineering, University of Dundee, UK
[†] School of Data and Computer Science, Sun Yat-sen University, China

ABSTRACT

In this paper, we present a novel and efficient method for brain tumor (and sub regions) segmentation in multimodal MR images based on a fully convolutional network (FCN) that enables end-to-end training and fast inference. Our structure consists of a downsampling path and three upsampling paths, which extract multi-level contextual information by concatenating hierarchical feature representation from each upsampling path. Meanwhile, we introduce a symmetry-driven FCN by the proposal of using symmetry difference images. The model was evaluated on Brain Tumor Image Segmentation Benchmark (BRATS) 2013 challenge dataset and achieved the state-of-the-art results while the computational cost is less than competitors.

Index Terms— FCN, brain tumor segmentation

1. INTRODUCTION

Gliomas are the most frequent primary brain tumors in adults [1] and can be classified as high-grade (HG) or low-grade (LG) based on the aggressive form of the disease. Multimodal MRI is usually utilized to enhance the contrast of tumor and its structures. Fig. 1 shows a typical HG gliomas tumor with experts' delineation of tumor and its sub-regions. Normally there are four structures in the tumor: edema (green), necrosis (red), non-enhancing (blue) and enhancing (yellow). The last three also make up a super-structure called *tumor core*.

Clinically, precise segmentations of tumors are crucial for making treatment plans, guiding surgery and follow-up of individual patients. Unreliable segmentation may mislead surgery, which could cause irreversible impact, e.g., the loss of brain functions such as speaking or reading. However, it is tedious and time-consuming to segment brain tumor manually, especially in 3D MR images. Automatic and reliable segmentation of gliomas brain tumor is an active topic for decades with challenges on the diversity and variation of tumor size, shape, and location and appearance.

A common approach is to pose this problem as classifying voxels into different tissues using hand-crafted features, followed by conditional random fields (CRF) models, incorporating smoothness terms of the classification results and maximizing label agreement between pixels in the neighborhood [1, 2, 3]. Nowadays, deep convolutional neural network (CNN) achieved several substantial breakthrough in several image recognition benchmarks, such as image classification [4, 5], semantic segmentation [6, 7] and object detection [8] and is getting popular in applications of medical imaging. Basically, CNN automatically learns high-level discriminative feature representations through a supervised manner. When CNN were applied to MRI brain tumor segmentation they achieved the state-of-the-art results [9, 10, 11]. Specifically, [10] trained a traditional 2D CNN on 2D image patches. During testing, 2D patches were extracted from the new images by a sliding window and assigned labels for the central pixel. [9] used 2D CNN on larger 2D patches in a cascade way to capture both local and global contextual information. [11] learned 3D CNN on 3D patches and considered global contextual feature via downsamping the original 3D patches. All these methods are *patch-level* based. Fully convolutional networks (FCN) are recently studied by [6, 7] and achieved promising results for natural image segmentation. It replaces the fully connected layers in the traditional CNN with all convolutional kernels and includes upsampling or deconvolutional layers to transform back to original spatial size. Thus, it can take input of arbitrary size and enables image end-to-end training as well as fast inference. Although FCN has been recently applied to medical image segmentation tasks [12, 13, 14], but not for brain tumor.

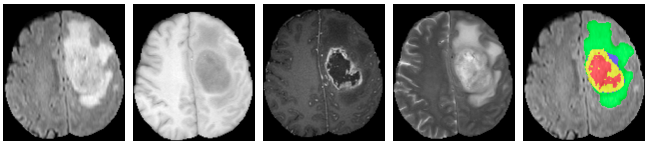


Fig. 1: An HG brain tumor example with 4 MRI modalities (Flair, T1, T1c, T2) and experts' delineation of tumor structure: edema (green), necrosis (red), non-enhancing (blue) and enhancing tumor (yellow). Best viewed in colour.

In this paper, we propose an automatic method for brain tumor segmentation based on FCN. The main contributions of our paper are: 1) to our best knowledge, we are the first to apply FCN for multimodal brain tumor (and sub-structure)

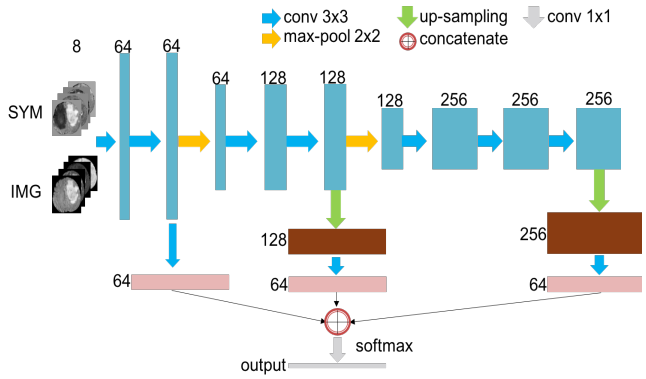


Fig. 2: CNN structure. The original images and the corresponding symmetry maps are concatenated as the input of the network. Best viewed in color.

segmentation; 2) our proposed FCN structure is simple and efficent, with only one loss layer coupling features at different levels; 3) we introduce brain symmetry inputs to FCN to further improve the segmentation performance; 4) our model is ranked top on BRATS 2013 testing set, and more efficient than the other competitors.

2. METHODOLOGY

In this section, we present an efficient fully convolutional network (FCN) for brain tumor segmentation. The proposed network is a variant which combines multiscale information from different stages and also takes full advantage of convolutional kernels for efficient and effective image segmentation.

2.1. Our FCN structure

The architecture of the proposed method is illustrated in Fig. 2. It contains two modules, i.e., one downsampling path with convolutional and maxpooling layers and three upsampling paths with upsampling and convolutional layers. The downsampling path aims at enlarging receptive fields to encode high level abstract and contextual information to detect tumors, while the upsampling paths reconstruct the fine details such as tumor boundaries. We designed the upsampling paths in a hierarchical manner to take full advantage of including multiple scale feature maps from downsampling path.

The downsampling path is similar to VGG-16 network [5], but instead of using total 5 convolutional blocks (one convolutional blocks contains two or three convolutional layers with 3×3 kernels and 1 maxpooling layer with 2×2 strides), we only use the first 3 convolutional blocks. There are three main reasons for only adopting the first 3 convolutional blocks: 1) unlike natural images which mostly contains rich high-level semantic features, medical images are mostly based on low-level texture features. Thus going *deeper* may not be helpful for medical images as limited high-level feature information could be learned from medical images; 2) in medical images, the sizes of lesions (e.g., brain tumors) are normally small compared to the entire image. Going too *deeper* may cause some tiny lesions *vanished* in the later convolutional blocks as each block shrinks image size by maxpooling operation using a factor of stride (usually stride is set to 2). Therefore, the filters learned in the later blocks has less capability of detecting lesions; 3) adding more convolutional blocks will introduce *more* parameters (e.g., adding the forth block will introduce 6 millions more parameters), which could lead to overfit the medical datasets of small scale. Our experiments confirmed that adding more convolutional blocks is not helpful in terms of brain tumor segmentation accuracy (described in Section 3.2).

For upsampling paths, we simply upsample the feature maps from the last convolutional layer of each convolutional block (before maxpooling layer) to the original spatial size. Then another three convolutional layers are applied to encode multi-scale feature representations. The resulting feature maps from three upsampling paths are concatenated before the final classification layer. Note that we did not use backwards strided convolutional (a.k.a deconvolutional) layer to perform upsampling as it will also introduce more parameters in the network, which may potentially lead to over-fitting.

Our structure shares some similarity with the one used in [14], but instead of injecting 3 auxiliary classifiers for each upsampling path for regularization, we extract multi-level contextual information by concatenating hierarchical feature representation from each upsampling path before the classification layer. We formulate the training of whole network as a per-pixel classification problem with respect to the ground-truth segmentation masks and choose categorical cross entropy as the loss function.

After each convolutional layer, we use Relu [4] as activation function to ensure non-linear mapping and batch normalization [15] to reduce the internal-covariate-shift. We observe batch normalization is crucial to optimize out network in experiments: it can accelerate training process by allowing a larger learning rate and avoid optimization ending with poor local minimals. In our architecture, we use 2D slices split from 3D MR volumes from axial view as the input of the proposed network. This is for two reasons: 1) it can significantly increase the number of training samples as each 3D volumes contains about 100$\sim$150 slices; 2) 2D axial slices might have enough discriminative information to differentiate tumor tissues as the experts' annotations in BRATS dataset were drawn in 2D axial slices rather than in a 3D version [1].

2.2. Symmetry-Driven FCN

It was noted that symmetry in axial view is an important cue for brain tumor segmentation as tumors usually break symmetric appearance of a health brain (see Fig. 1). Although,

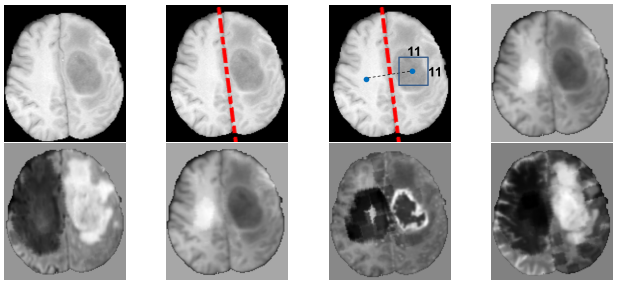


Fig. 3: top row: each step of computing the symmetry maps (from left to right: Flair, symmetry axis, local search, and symmetry map); bottom row: the final symmetry maps for each MR modality.

brain symmetry or asymmetry information has been used in shallow methods [2, 16, 17], however, they are not explored in deep method such as CNN.

In this paper, we encode brain symmetry information to the CNN framework by adding extra symmetry maps. Our symmetry maps are computed as follows: 1) we first locate the symmetric axis in T1 modality axial slices through the approach presented in [18]; 2) given the symmetric axis (the red dash line in Fig. 3), we found the corresponding matching pixel pairs and calculated their intensity differences. In order to reduce the effects of the errors of symmetric axis and image noises, each image was smoothed beforehand using a Gaussian filter with 5×5 kernel. The most matched pixel was searched in a 11×11 local window (the blue square in Fig. 3) centered on the mirrored the location w.r.t the symmetry axis. The resulting intensity differences are then converted into range [0, 1] by a sigmoid function.

The results of each step when computing the symmetry maps are visualized in the top row of Fig. 3, while the final symmetry maps for each MR modality are shown in the bottom row. We combine them with the four original images as the inputs of our CNN framework as show in Fig. 2.

3. EVALUATION

We evaluated our model on BRATS 2013 clinical dataset. The dataset contains 20 HG patients with pixel-level annotations for the training (10 LG patients were not used for the HG segmentation task) and 10 HGs for the testing. For each patient, there exists 4 modalities, namely T1, T1-contrast (T1c), T2 and Flair, which are skull-stripped and co-registered.

Quantitative evaluation on the testing set is through the online VSD evaluation system [1] for three sub-tasks: 1) the *complete* tumor region (including all four tumor structures); 2) the *core* tumor region (including all tumor structure except "edema"); 3) the *enhancing* tumor region (including only the "enhancing tumor" structure). For each tumor region, *Dice*, *Sensitivity* and *Positive Predictive Value* are computed.

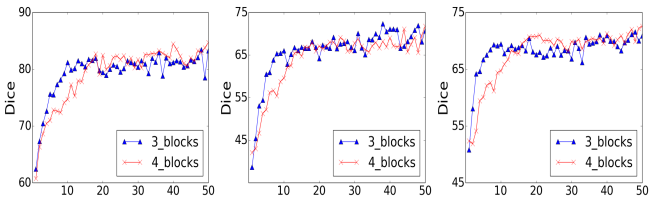
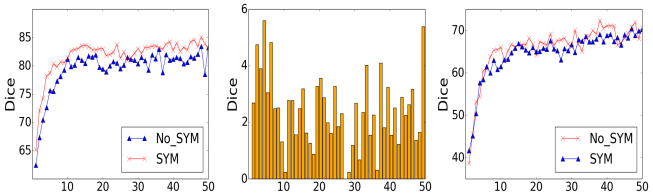


Fig. 4: performance curves of 3 blocks vs 4 blocks. From left to right: *complete* , *core* and *enhancing*. The vertical axis is Dice while horizontal axis is the number of epochs.



(a) (b) (c)

Fig. 5: performance curves of with and w/o symmetry maps. (a) *complete tumor*; (b) the residual of (a); (c) *tumor core*.

3.1. Implementation

Each brain MR image is standardized with zero mean and unit standard deviation. We augmented the data set by scaling, rotating, flipping each image; thus resulting a new dataset that is 3 times larger than the original one.

Our model was implemented with Keras library and Theano backend. All images were cropped to have the same size $144 \times 192 \times 128$ as the input into the network, which was trained with standard back-propagation using Adam optimizer. We set the learning rate as 0.001 and never changed it during the training. The downsampling path was initialized by VGG-16 weights [5] while the upsampling paths were initialized randomly using the method proposed by [19]. The training time on the augmented dataset is about five hours using a standard PC with a NVIDIA Titan Pascal GPU.

3.2. Cross Validation

We perform a 5-fold cross validation for 20 HG training images, and conducted two experiments to test the effects of 1) going 'deeper' and 2) symmetry maps. The augmented dataset was not used in 5-fold CV to save computational cost.

Firstly, we compare the performance with 3 or 4 convolutional blocks to see whether going 'deeper' of the model is helpful for our tasks. We plot the Dice scores for three sub-tasks at different training epoch (up to 50). The curves are shown Fig 4. It can be observed that there is no obvious improvement using 4 convolutional blocks over 3 for all subtasks in terms of Dice score. On the other hand, going 'deeper' added more parameters and thus resulted in more training time and slower convergence.

We then evaluate the effect of symmetry maps. Fig.5 shows the validation Dice scores w.r.t epochs of FCN and symmetry-driven FCN. It could be observed that FCN ben-

Table 1: Comparison with the state-of-the-arts on the testing set (ranked by VSD evaluation system [1])

Method	Dice			Positive Predictive Value			Sensitivity		
	Complete	Core	Enhancing	Complete	Core	Enhancing	Complete	Core	Enhancing
Pereira [10]	0.88	0.83	0.77	0.88	0.87	0.74	0.89	0.83	0.81
Proposed	0.87	0.82	0.75	0.85	0.87	0.72	0.89	0.79	0.80
Kwon [20]	0.88	0.83	0.72	0.92	0.90	0.74	0.84	0.78	0.72
Havaei [9]	0.88	0.79	0.73	0.89	0.79	0.68	0.87	0.79	0.80
Tustison [2]	0.87	0.78	0.74	0.85	0.74	0.69	0.89	0.88	0.83
Meier [1]	0.82	0.73	0.69	0.76	0.78	0.71	0.92	0.72	0.73
Reza [1]	0.83	0.72	0.72	0.82	0.81	0.70	0.86	0.69	0.76

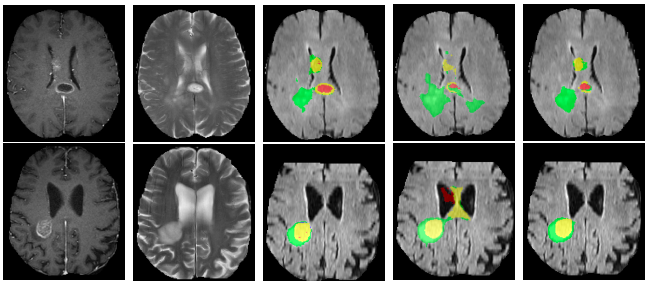


Fig. 6: Examples of segmentation results. From left to right: T1c, T2, Flair with ground truth, results without symmetry maps and results with symmetry maps. Best viewed in colour.

efits from symmetry maps, especially for the *Complete* tumor task (Fig.5(a)). For this task, we further calculate the residual between FCN and symmetry-driven FCN along training epoch (Fig.5(b)). Using symmetry maps improves the performance over most training epochs and gives an average of 3% improvement for Dice score. The *Core* tumor segmentation also benefit from adding symmetry maps, though not as big as *Complete* tumor (Fig.5(c)). Alternatively, the performances of FCN and symmetry-driven FCN at their respected best performing epoch are 0.83 vs 0.85 for the *Complete* tumor task (0.70 vs 0.72 for the *Core* tumor task), resulting an 2% improvement. However, there is no obvious improvement observed for *Enhancing* tumor task. Examples of results with and without symmetry maps are shown in Fig 6. It shows that symmetry-driven FCN greatly removes the false positives, which confirms its efficacy.

3.3. Comparison with Best Performers on Testing Set

We compare the proposed method with state-of-the-arts on BRATS13 testing set. As it only contains HG images, we only use the 20 HG images for training. The proposed method is among the top-ranking in the state-of-the-arts (see Table 1).

Specifically, Tustison, Meier and Reza were the best performers of BRATS13 challenge [1]. Our method outperforms them all (by a big margin over Meier and Reza, e.g., 0.82 vs 0.72 in terms of Dice for *Core* tumor segmentation). Particularly, Tustison[2], the winner of BRATS13 challenge, is less efficient than ours as it needs an auxiliary health brain dataset for registration to calculate the features, while we only use the data provided by the challenge. Our model is *fully* automatic and overall ranked higher than a *semi-automatic* method [20].

For CNN methods, our results are competitive with [10] and better than [9]. Compared to the cascade structure [9], our network structure is simpler, showing the effectiveness of FCN framework and combining multi-scale features. Note that although [10] performs best on this dataset, they evaluated different experimental settings on the testing set, which might lead to overfit and produce optimistically-biased results. A fair comparison with 3D CNN [11] is not available as they did not evaluate on this dataset.

One advantage of our model over the others is the computational efficiency for a new test image. [20] reported an average running time of 85 minutes for each 3D volume on CPU, which is a bottleneck for daily clinical use. The two CNN approaches, [10] reported an average running time of 8 minutes while 3 minutes was reported by [9], both using a modern GPU. For an indicative comparison, our method took about 2 minutes for each 3D volume in its current implementation. Note that 95% of the time was used to compute the symmetry inputs on CPU. Thus it expects the time could be much less if the computation of symmetry maps is parallelized on GPU.

4. CONCLUSION

We propose an automatic brain tumor segmentation method based on fully convolutional neural network. Our method contains three convolutional blocks and encodes multi-scale features from different layers in one loss function. Going *deeper* did not make a big difference. We also present a symmetry-driven FCN, which further improves segmentation performance, especially for the *Complete* tumor region. Our method achieved state-of-the-art results, and is more efficient than others. In the future, we will evaluate our model on a larger dataset.

Acknowledgments

This work was supported partially by the National Natural Science Foundation of China (No. 61628212).

5. REFERENCES

[1] Bjoern H Menze et al., "The multimodal brain tumor image segmentation benchmark (BRATS)," *IEEE Transactions on Medical Imaging*, vol. 34, no. 10, pp. 1993–2024, 2015.

[2] Nicholas J Tustison, KL Shrinidhi, Max Wintermark, et al., "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.

[3] Nagesh Subbanna, Doina Precup, and Tal Arbel, "Iterative multilevel MRF leveraging context and voxel information for brain tumour segmentation in MRI," in *CVPR*, 2014, pp. 400–405.

[4] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton, "Imagenet classification with deep convolutional neural networks," in *NIPS*, 2012, pp. 1097–1105.

[5] Karen Simonyan and Andrew Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[6] Jonathan Long, Evan Shelhamer, and Trevor Darrell, "Fully convolutional networks for semantic segmentation," in *CVPR*, 2015, pp. 3431–3440.

[7] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille, "Semantic image segmentation with deep convolutional nets and fully connected CRFs," *arXiv preprint arXiv:1412.7062*, 2014.

[8] Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *CVPR*, 2014, pp. 580–587.

[9] Mohammad Havaei, Axel Davy, David Warde-Farley, Antoine Biard, Aaron Courville, Yoshua Bengio, Chris Pal, Pierre-Marc Jodoin, and Hugo Larochelle, "Brain tumor segmentation with deep neural networks," *Medical Image Analysis*, vol. 35, pp. 18–31, 2017.

[10] Sérgio Pereira, Adriano Pinto, Victor Alves, and Carlos A Silva, "Brain tumor segmentation using convolutional neural networks in MRI images," *IEEE transactions on Medical Imaging*, vol. 35, no. 5, pp. 1240–1251, 2016.

[11] Konstantinos Kamnitsas, Christian Ledig, Virginia FJ Newcombe, Joanna P Simpson, Andrew D Kane, David K Menon, Daniel Rueckert, and Ben Glocker, "Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation," *Medical Image Analysis*, vol. 36, pp. 61–78, 2017.

[12] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, "U-net: Convolutional networks for biomedical image segmentation," in *MICCAI*, 2015, pp. 234–241.

[13] Hao Chen, Xiaojuan Qi, Lequan Yu, et al., "Dcan: Deep contour-aware networks for accurate gland segmentation," in *CVPR*, 2016, pp. 2487–2496.

[14] Hao Chen, Xiao Juan Qi, Jie Zhi Cheng, and Pheng Ann Heng, "Deep contextual networks for neuronal structure segmentation," in *AAAI*, 2016.

[15] Sergey Ioffe and Christian Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *arXiv preprint arXiv:1502.03167*, 2015.

[16] Anthony Bianchi, James V Miller, Ek Tsoon Tan, and Albert Montillo, "Brain tumor segmentation with symmetric texture and symmetric intensity-based decision forests," in *ISBI*, 2013, pp. 748–751.

[17] Yu Sun, Bir Bhanu, and Shiv Bhanu, "Symmetry-integrated injury detection for brain MRI," in *ICIP*, 2009, pp. 661–664.

[18] Gareth Loy and Jan-Olof Eklundh, "Detecting symmetry and symmetric constellations of features," in *ECCV*, 2006, pp. 508–521.

[19] Kaiming He, Xiangyu Zhang, Shaoqing Ren, et al., "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," in *ICCV*, 2015, pp. 1026–1034.

[20] Dongjin Kwon, Russell T Shinohara, Hamed Akbari, and Christos Davatzikos, "Combining generative models for multifocal glioma segmentation and registration," in *MICCAI*, 2014, pp. 763–770.