Automatic Brain Tumor Segmentation using Cascaded Anisotropic Convolutional Neural Networks

Guotai Wang, Wenqi Li, Sébastien Ourselin, and Tom Vercauteren

Translational Imaging Group, CMIC, University College London, UK
Wellcome/EPSRC Centre for Interventional and Surgical Sciences, UCL, London, UK
`guotai.wang.14@ucl.ac.uk`

Abstract. A cascade of fully convolutional neural networks is proposed to segment multi-modality MR images with brain tumor into background and three subregions: enhanced tumor core, whole tumor and tumor core. The cascade is designed to decompose the multi-class segmentation into a sequence of three binary segmentations according to the subregion hierarchy. Segmentation of the first (second) step is used as a crisp binary mask for the second (third) step. Each network consists of multiple layers of anisotropic and dilated convolution filters that were obtained by training each network end-to-end. Residual connections and multi-scale predictions were employed in these networks to boost the segmentation performance. Experiments with BraTS 2017 validation set shows the proposed method achieved average Dice scores of 0.7859, 0.9050, 0.8378 for enhanced tumor core, whole tumor and tumor core respectively[1].

Keywords: Brain tumor, convolutional neural network, segmentation

1 Introduction

Gliomas are the most common brain tumors that arise from glial cells. They can be categorized into two basic grades: low-grade gliomas (LGG) that tend to exhibit benign tendencies and indicate a better prognosis for the patient, and high-grade gliomas (HGG) that are malignant and more aggressive. With the development of medical imaging, brain tumors can be imaged by various Magnetic Resonance (MR) modalities, such as T1, T1-contrast, T2 and Fluid Attenuation Inversion Recovery (FLAIR). Different modalities can provide complementary information to analyze different sub-regions of gliomas, such as tumor cores and edema regions.

Automatic segmentation of brain tumors and substructures is promising to provide accurate and reproducible measurements of the tumors. It has great potential for better diagnosis, surgical planning and treatment assessment for brain

---

[1] The BraTS challenge organizers have not revealed the performance on the testing dataset yet, but they have ranked our method as one of the top-performing ones.
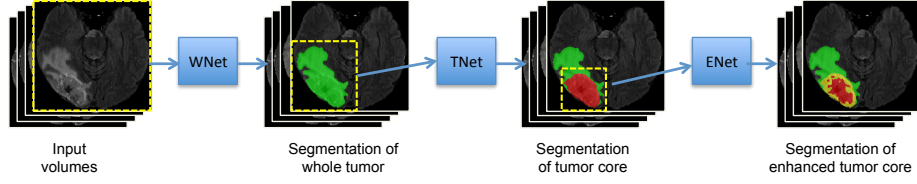
Fig. 1. The proposed triple cascaded framework for brain tumor segmentation. Three networks are proposed to hierarchically segment whole tumor (WNet), tumor core (TNet) and enhanced tumor core (ENet) sequentially.

tumors [15,2]. However, this segmentation task is challenging because 1) the size, shape, and localization of brain tumors have considerable variations among patient; 2) the boundaries between adjacent structures are often ambiguous due to the smooth intensity gradients, partial volume effects and bias field artifacts.

Discriminative methods based on deep neural networks have achieved state-of-the-art performance for multi-modality brain tumor segmentation tasks. Several key ideas to improve the performance of segmentation networks have been explored in the literature. These include efficient end-to-end training using a fully convolutional approach [1,8], incorporating large visual contexts by employing a mixture of convolution and downsampling operations [12,9], maintaining high resolution multi-scale features with dilated convolution and residual connection [17,14,5], and handling training data imbalance issue by designing new loss functions [7,16] and sampling strategies [16].

Inspired by the previous work of cascaded neural networks for liver segmentation [6], we propose a cascade of CNNs for brain tumor subregion segmentation. We take advantage of dilated convolution, residual connection and multi-scale prediction to boost performance of the networks. In addition, we use anisotropic convolution to deal with 3D images as a trade-off between memory consumption and model complexity.

2   Methods

2.1   Triple Cascaded Framework

The proposed cascaded framework is shown in Fig. 1. We use three networks to hierarchically and sequentially segment substructures of brain tumors. The first network (WNet) segments the whole tumor from multi-modality 3D volumes of the same patient. The second network (TNet) segments the tumor core from the whole tumor region given by WNet, and the third network (ENet) segments the enhanced tumor core from the tumor core region given by TNet. Segmentation of the first (second) network is used as a crisp binary mask for the second (third) network. These networks deal with binary segmentations and have different receptive fields. For WNet, the receptive field is the whole image region. The receptive field of TNet and ENet is the whole tumor region and tumor core region
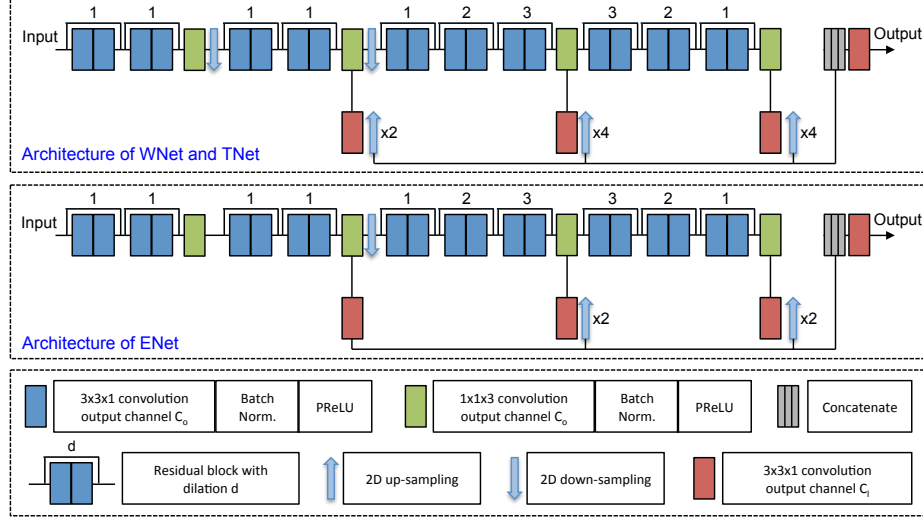
Fig. 2. Our anisotropic convolutional networks with dilated convolution, residual connection and multi-scale fusion. ENet uses only one downsampling layer considering its smaller input size.

respectively. There are several benefits of using such a cascaded segmentation framework. First, compared with training a single network for all substructures which requires complex network architectures, using three binary segmentation networks allows for a simpler network for each task. Therefore they are easier to train and can reduce over-fitting. Second, this helps reduce false positives since TNet only works on the region extracted by WNet and ENet only works on the region extracted by TNet. Third, this hierarchical pipeline follows the anatomic structure of tumors. It restricts the tumor core to be inside the whole tumor region and enhanced tumor core to be inside the tumor core region.

2.2   Anisotropic Convolutional Neural Networks

For 3D neural networks, the balance between memory consumption and feature representation ability should be considered. Many 2D networks take a whole 2D slice as input and can capture features in a large receptive field. However, taking a whole 3D volume as input consumes a lot of memory and therefore limits the resolution and number of features in the network, leading to a low representation ability. As a trade-off, we propose anisotropic networks that take a stack of slices as input with a large receptive field in 2D and a smaller receptive field along the direction orthogonal to the 2D slices. The architectures of our proposed MNet, TNet and ENet are shown in Fig. 2. All the networks are fully convolutional and use 10 residual connection blocks with anisotropic convolution, dilated convolution, and multi-scale prediction.

Anisotropic and Dilated Convolution. To deal with anisotropic receptive fields, we decompose a 3D kernel of size $3{\times}3{\times}3$ into an intra-slice kernel with size $3{\times}3{\times}1$ and an inter-slice kernel with size $1{\times}1{\times}3$. Convolutional layers with either of these kernels have C_o output channels and each is followed by a batch normalization layer and an activation layer, as illustrated by blue and green blocks in Fig. 2. We use Parametric Rectified Linear Unit (PReLU) [10] in the activation layers. WNet and TNet use 20 intra-slice convolutional layers and four inter-slice convolutional layers with two 2D downsampling layers. ENets use the same set of convolutional layers as WNet but only one downsampling layer considering its smaller input size. We only employ up to two layers of downsampling in order to avoid large image resolution reduction and loss of segmentation details. After the downsampling layers, we use dilated convolution for intra-slice kernels to enlarge the receptive field within a slice. The dilation parameter is set to 1 to 3 as shown in Fig. 2. Since the anisotropic convolution has small receptive field along the out-plane direction, to take advantage of 3D features, we fuse segmentation from three different orthogonal views. Each of these three networks was trained along axial, sagittal and coronal view respectively. During testing time, predictions in these three views are averaged to get the final segmentation.

Residual Connection. For effective training of deep CNNs, residual connections [11] were introduced to create identity mapping connections to bypass the parameterized layers in a network. Our MNet, TNet and ENet have 10 residual blocks. Each of the block contains two intra-slice convolutional layers, and the input of a residual block is directly added to the output, encouraging the block to learn residual functions with reference to the input. This can make information propagation smooth and speed the convergence of training [11,14].

Multi-scale Prediction. In deep convolutional neural networks, sequential convolutional layers increase the receptive field and they capture features at different scales. Shallow layers learn to represent local and simple features while deep layers learn to represent global and abstract features. To combine both local and global features, we use three $1{\times}3{\times}3$ convolutional layers at different scales of the networks to get intermediate predictions and upsample them to the resolution of the input. A concatenation of these predictions are fed into an additional $1{\times}3{\times}3$ convolutional layer to obtain the final score map. These layers are illustrated by red blocks in Fig. 2. The outputs of these layers have C_l channels where C_l is the number of classes for segmentation in each network.

3   Experiments and Preliminary Results

Data and Implementation Details. We used the BraTS 2017[2] [15,2,4,3] training and validation set for experiments. The training set contains images

---

[2] http://www.med.upenn.edu/sbia/brats2017.html

from 285 patients (210 HGG and 75 LGG). The BraTS 2017 validation set contains images from 46 patients with brain tumores of unknown grade. Each patient was scanned with four modalities: T1, T1c, T2 and FLAIR. We uploaded the segmentation results to the BraTS 2017 server which evaluated the segmentation and provided quantitative measurements in terms of Dice score, sensitivity, specificity and Hausdorff distance of enhanced tumor core, whole tumor, and tumor core respectively.

Our networks were implemented in Tensorflow[3] using NiftyNet[4]. We used Adaptive Moment Estimation (Adam) [13] for training, with initial learning rate 10^{-3}, weight decay 10^{-7}, batch size 5. Training was implemented on a an NVIDIA TITAN X GPU. We set C_o to 48 and C_l to 2 for MNet, TNet and ENet.

Table 1. Dice and Hausdorff measurements of our method (UCL-TIG) compared with top performance achieved by other teams. The results were provided by the BraTS 2017 validation leaderboard up to August 31, 2017. EN, WT, TC denote enhanced tumor core, whole tumor and tumor core respectively.

	Dice			Hausdorff		
	ET	WT	TC	ET	WT	TC
UCL-TIG*	0.7859	0.9050	0.8378	3.2821	3.8901	6.4790
biomedia1	0.7570	0.9016	0.8202	4.2225	4.5576	6.1055
MIC-DKFZ	0.7320	0.8964	0.7971	4.5470	6.9741	9.4767
pvg	0.7353	0.8885	0.7711	6.3246	4.3354	8.6320
Zhouch	0.7605	0.9034	0.8246	3.7199	4.8768	6.7466

Table 2. Sensitivity and specificity measurements of our method (UCL-TIG) compared with top performance achieved by other teams. The results were provided by the BraTS 2017 validation leaderboard up to August 31, 2017. EN, WT, TC denote enhanced tumor core, whole tumor and tumor core respectively.

	Sensitivity			Specificity		
	ET	WT	TC	ET	WT	TC
UCL-TIG*	0.7748	0.9118	0.8412	0.9985	0.9942	0.9973
biomedia1	0.7895	0.9088	0.7829	0.9982	0.9946	0.9986
MIC-DKFZ	0.7900	0.8965	0.7807	0.9984	0.9956	0.9988
pvg	0.7676	0.8941	0.7558	0.9980	0.9950	0.9980
Zhouch	0.8004	0.9058	0.8198	0.9980	0.9952	0.9970

---

[3] https://www.tensorflow.org/
[4] http://niftynet.io/

Segmentation Results. Quantitative evaluation are shown on the BraTS 2017 leaderboard[5,6]. Table 1 presents Dice and Hausdorff measurements according to the leaderboard. It shows that our method achieves competitive results in terms of dice scores averaged over patients. Table 1 also shows our method achieves low Hausdorff distances for different tumor subregions. Table 2 presents sensitivity and specificity measurements according to the leaderboard.

4   Conclusion

We developed a cascaded system to segment glioma subregions from multi-modality brain MR images. Results on BraTS 2017 online validation set predicted average Dice scores of 0.7859, 0.9050, 0.8378 for enhanced tumor core, whole tumor and tumor core respectively.

References

1. B, M.H., Guizard, N., Chapados, N.: HeMIS : Hetero-Modal Image Segmentation. In: MICCAI. vol. 1, pp. 469–477 (2016)
2. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J., Freymann, J., Farahani, K., Davatzikos, C.: Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features. Nature Scientific Data (2017)
3. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J., Freymann, J., Farahani, K., Davatzikos, C.: Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-LGG collection. The Cancer Imaging Archive (2017)
4. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J., Freymann, J., Farahani, K., Davatzikos, C.: Segmentation Labels for the Pre-operative Scans of the TCGA-GBM collection. The Cancer Imaging Archive (2017)
5. Chen, H., Dou, Q., Yu, L., Heng, P.A.: VoxResNet: Deep Voxelwise Resid-ual Networks for Volumetric Brain Segmentation. NeuroImage pp. 1–9 (2016), http://arxiv.org/abs/1608.05895

---

[5] https://www.cbica.upenn.edu/BraTS17/lboardValidation.html
[6] Results retrieved on August 31, 2017.

6. Christ, P.F., Elshaer, M.E.A., Ettlinger, F., Tatavarty, S., Bickel, M., Bilic, P., Rempfler, M., Armbruster, M., Hofmann, F., Anastasi, M.D., Sommer, W.H., Ahmadi, S.a., Menze, B.H.: Automatic Liver and Lesion Segmentation in CT Using Cascaded Fully Convolutional Neural Networks and 3D Conditional Random Fields. In: MICCAI. vol. 1, pp. 415–423 (2016)
7. Fidon, L., Li, W., Garcia-peraza herrera, L.C.: Generalised Wasserstein Dice Score for Imbalanced Multi-class Segmentation using Holistic Convolutional Networks. arxiv pp. 1–11 (2017), `https://arxiv.org/abs/1707.00478`
8. Fidon, L., Li, W., Garcia-peraza herrera, L.C., Ekanayake, J., Kitchen, N., Ourselin, S., Vercauteren, T.: Scalable multimodel convolutional networks for brain tumour segmentation. In: MICCAI (2017)
9. 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. Medical Image Analysis 35, 18–31 (2016)
10. He, K., Zhang, X., Ren, S., Sun, J.: Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification. In: ICCV (2015), `http://arxiv.org/abs/1502.01852`
11. He, K., Zhang, X., Ren, S., Sun, J.: Deep Residual Learning for Image Recognition. In: CVPR (2016)
12. Kamnitsas, K., Ledig, C., Newcombe, V.F.J., Simpson, J.P., Kane, A.D., Menon, D.K., Rueckert, D., Glocker, B.: Efficient Multi-Scale 3D CNN with Fully Connected CRF for Accurate Brain Lesion Segmentation. Medical Image Analysis 36, 61–78 (2017)
13. Kingma, D.P., Ba, J.L.: Adam: a Method for Stochastic Optimization. International Conference on Learning Representations 2015 pp. 1–15 (2015)
14. Li, W., Wang, G., Fidon, L., Ourselin, S., Cardoso, M.J., Vercauteren, T.: On the Compactness , Efficiency , and Representation of 3D Convolutional Networks : Brain Parcellation as a Pretext Task. In: IPMI (2017)
15. Menze, B.H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., Lanczi, L., Gerstner, E., Weber, M.A., Arbel, T., Avants, B.B., Ayache, N., Buendia, P., Collins, D.L., Cordier, N., Corso, J.J., Criminisi, A., Das, T., Delingette, H., Demiralp, Ç., Durst, C.R., Dojat, M., Doyle, S., Festa, J., Forbes, F., Geremia, E., Glocker, B., Golland, P., Guo, X., Hamamci, A., Iftekharuddin, K.M., Jena, R., John, N.M., Konukoglu, E., Lashkari, D., Mariz, J.A., Meier, R., Pereira, S., Precup, D., Price, S.J., Raviv, T.R., Reza, S.M., Ryan, M., Sarikaya, D., Schwartz, L., Shin, H.C., Shotton, J., Silva, C.A., Sousa, N., Subbanna, N.K., Szekely, G., Taylor, T.J., Thomas, O.M., Tustison, N.J., Unal, G., Vasseur, F., Wintermark, M., Ye, D.H., Zhao, L., Zhao, B., Zikic, D., Prastawa, M., Reyes, M., Van Leemput, K.: The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). TMI 34(10), 1993–2024 (2015)
16. Sudre, C.H., Li, W., Vercauteren, T., Ourselin, S., Cardoso, M.J.: Generalised Dice overlap as a deep learning loss function for highly unbalanced segmentations. arXiv pp. 1–8 (2017), `http://arxiv.org/abs/1707.03237`
17. Wang, G., Zuluaga, M.A., Li, W., Pratt, R., Patel, P.A., Aertsen, M., Doel, T., Klusmann, M., David, A.L., Deprest, J., Vercauteren, T., Ourselin, S.: DeepIGeoS: A Deep Interactive Geodesic Framework for Medical Image Segmentation. arXiv (2017), `https://arxiv.org/abs/1707.00652`