3D Large Kernel Anisotropic Network for Brain Tumor Segmentation

Dongnan Liu[1]([⊠]), Donghao Zhang[1], Yang Song[1], Fan Zhang[2],
Lauren J. O'Donnell[2], and Weidong Cai[1]

[1] School of Information Technologies, University of Sydney, NSW 2006, Australia
`dliu5812@uni.sydney.edu.au`
[2] Brigham and Women's Hospital, Harvard Medical School, Boston, USA

Abstract. Brain tumor segmentation in magnetic resonance images is a key step for brain cancer diagnosis and clinical treatment. Recently, deep convolutional neural network (DNN) based models have become a popular and effective choice due to their learning capability with a large amount of parameters. However, in traditional 3D DNN models, the valid receptive fields are not large enough for global details from the objective and the large amount of parameters are easy to cause high computational cost and model overfitting. In order to address these problems, we propose a 3D large kernel anisotropic network. In our model, the large kernels in the decoders ensure the valid receptive field is large enough and the anisotropic convolutional blocks in the encoders simulate the traditional isotropic ones with fewer parameters. Our proposed model is evaluated on datasets from the MICCAI BRATS 17 challenge and outperforms several popular 3D DNN architectures.

Keywords: brain tumor segmentation, magnetic resonance image, 3D deep neural network

1 Introduction

Brain tumor is one of the leading causes of cancer deaths, which is difficult to cure and remains high mortality [23]. Among all varieties of brain tumors diagnosed in adults, gliomas account for about 70% [26]. Thus the detection and segmentation of gliomas is a necessary task for clinical diagnosis of brain cancer. Gliomas are caused by glial cells [8] and can be divided into high and low grade categories. High grade gliomas (HGG) turn out to be more aggressive with a survival time of no more than two years, while the low grade gliomas (LGG) grow slowly and leave a longer life expectancy of several years [19].

Although surgical treatment is the most effective way to directly remove the tumors, radiological treatment is also necessary to slow the growth of the tumors which can not be removed. Nowadays, magnetic resonance imaging (MRI) is one of the most common tools for radiological diagnosis since it is able to image the brain structure with detailed information. Various imaging modalities are used to describe the information for different subregions of the tumor, as shown in

Fig 1. For example, T2- and Flair-weighted brain images highlight the tissues with water content, which represent the whole tumor with edema. T1-weighted images highlight the tumor core which contains no water, and contrast enhance T1-weighted (T1ce) images represent the enhanced parts with hyper-intensity in the tumor core [16].
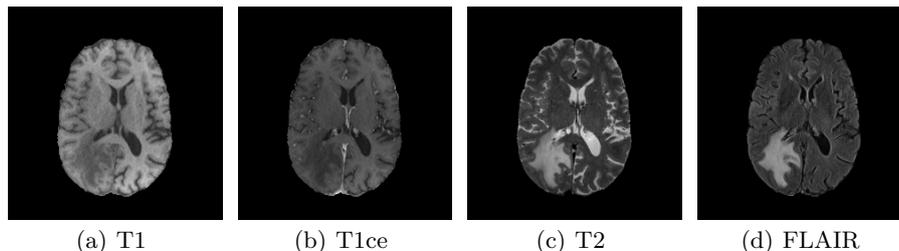

(a) T1   (b) T1ce   (c) T2   (d) FLAIR

Fig. 1. An example of one slice of a brain MR image in different modalities: T1, T1ce, T2, and FLAIR.

However, brain tumor segmentation is challenging due to the following factors: 1) Brain tumors can appear anywhere in the whole brain with different shapes and sizes across patients. Thus the methods based on processing the shape or size features become less effective. 2) The gradual transition from tumor to edema makes the boundaries between the edema and the tumor core ambiguous and hard to segment. 3) Although T1ce modality is able to highlight the enhanced tumor core, some other parts such as blood vessels and cortical cerebrospinal fluid (CSF) are highlighted as well. In this way, a thresholding method can not be directly applied to segment the enhanced tumor [21]. 4) Due to the pre-processing methods and poor contrast of tumors, the boundaries between tumors and healthy tissues are fuzzy.

In this work, we propose an end-to-end 3D convolutional neural network (CNN) model for the brain tumor segmentation task. The contributions of our work are three-folds. First, we apply anisotropic convolutional blocks to simulate the isotropic blocks with less parameters for memory efficiency. Second, inspired by 2D global convolutional network (GCN) [20], we propose a 3D large convolutional block which is capable of enlarging the receptive field for the feature maps. Third, our proposed model is evaluated on part of the MICCAI brain tumor segmentation (BRATS) challenge 2017 dataset [16] and outperforms some state-of-the-art 3D CNN models.

2   Related Work

Expert annotation for brain tumor segmentation is time-consuming and labor-intensive, thus automatic segmentation algorithms are urgently needed for high

efficiency. Brain tumor segmentation approaches can be categorized into generative and discriminative methods. Generative methods apply specific prior knowledge about appearance of the brain tissues, including tumors and healthy parts. Such approaches require little training and largely rely on encoding the prior probability distribution of the spatial relationship between tissues [4,17]. In [22], Prastawa *et al* propose a typical generative method which aligns the brain tissues to the ICBM brain digital atlas. Then the tumor is detected by comparing its posterior probabilities with that of the healthy tissues. Although generative methods process the MRI image with high efficiency, the accurate probability distribution is hard to encode and model. Discriminative methods directly learn the characteristic differences of the appearance between tumor and healthy tissues. In these methods, first the dense voxel features are extracted from the original images [24,14]. Then the features are fed into a classifier such as decision forests [30], Markov Random Field [24], or clique-based graphic model [14].

Recently, deep neural network (DNN) based architectures have shown competitive performances on the segmentation tasks for biomedical images [28,13,27]. Discriminative methods based on DNN achieve competitive performance compared with other state-of-the-art methods as DNN is able to learn the feature information in accurate details with numerous parameters. In [6], Havaei *et al* proposes a dual path CNN architecture for analyzing the local and global feature details of the brain tumors. However, the model only applies 2D convolutional blocks, which means that it fails to process the information between different slices. DeepMedic [9] is a dual path architecture which contains 3D convolutional blocks and residual connections. HighRes3DNet [12] is a residual connected DNN model with dilated convolutional blocks which is capable of getting high spatial resolution features. However, in these single connection models, the feature information from low resolution level could easily be lost after passing through a number of convolutional and pooling layers. In DNN, the information from each resolution level represents different features of the original image. Losing the low level feature information results in a lack of detailed information such as curves and edges in the final segmentation. In order to solve this problem, skip connections between the encoders and decoders are applied to 3D DNN such as 3D U-Net [3] and V-Net [18]. In V-Net [18], residual connections are applied on the encoders and decoders to prevent the gradient vanishing during the training process. Even though these methods achieve competitive performance, the complexity of the model tends to be high due to the huge amount of parameters from 3D convolutional kernels, which makes the model difficult to train and easy to overfit. Wang *et al* [25] decomposes a traditional $3 \times 3 \times 3$ block to one $3 \times 1 \times 1$ block and one $1 \times 3 \times 3$ block in the proposed architecture to process the inter- and intra- slices information respectively for a higher efficiency and lower computational cost. Moreover, 3D anisotropic convolutional blocks also show their effectiveness on the segmentation tasks for other 3D biomedical datasets, such as membrane segmentation for electron microscopy images [11], and liver tumor segmentation for CT scans [15].

3   Methods

In this section, we introduce our proposed 3D large kernel anisotropic network. Our model is composed of residual connection encoders with anisotropic convolutional blocks and decoders with 3D large kernel blocks to enlarge the actual receptive field, as shown in Fig. 2.
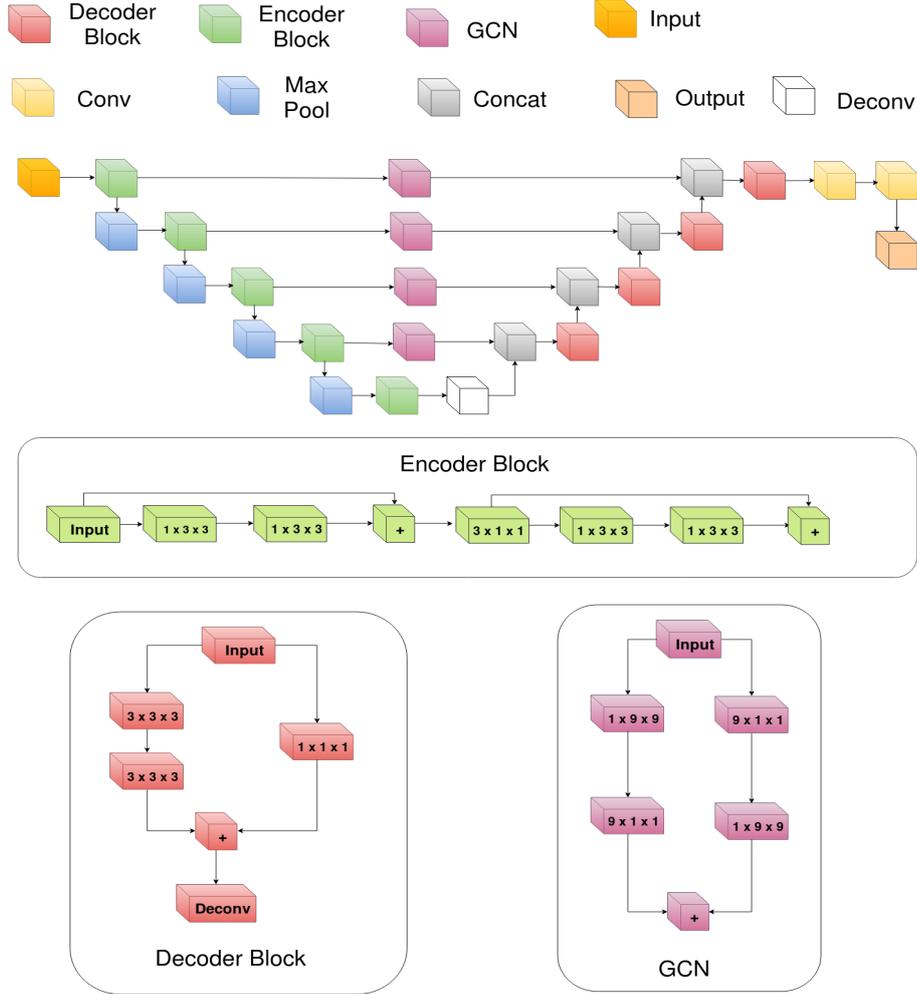

Fig. 2. The network architecture of our proposed network. The sizes of the last 2 Conv blocks are $3 \times 3 \times 3$ and $1 \times 1 \times 1$ respectively. The size and the stride for the Deconv at the highest resolution are $2 \times 2 \times 2$ and $(2, 2, 2)$. We omit the batch nomalization and PReLU layers before each convolutional layers in this graph for brevity.

As the pre-activation residual connections proposed in [7] are proved to outperform the traditional ones, we apply them on encoders. In the first residual connection unit of the first encoder, the output feature channel of the convolutional path is 16. The input image is padded from $H \times W \times D \times 4$ to $H \times W \times D \times 16$. In each encoder, the feature channels are doubled after passing the $3 \times 1 \times 1$ kernel for the next residual unit. For each encoder at different resolution levels, the number of output feature maps are 32, 64, 128, 256 and 512 respectively. After each encoders except for the one at the highest resolution level, the output feature maps pass through two ways: downsampling by max pooling layers with a size of $2 \times 2 \times 2$ and a stride of $(2, 2, 2)$ for the encoder at a higher resolution level, and feeding into a 3D large kernel convolutional block for a large actual receptive field.

The number of output channels for each 3D large kernel is the same as the number of feature maps from the corresponding encoder at each resolution level. In order to have a large receptive field for the model, we set all the sizes of 3D large kernels as 9. The output of each 3D large kernel block is concatenated with the feature map upscaled from the decoder at a lower resolution level. Then the result passes through another residual connected decoder block. In each decoder, the number of feature maps reduces to $1/3$ after the $3 \times 3 \times 3$ block path and $1 \times 1 \times 1$ path respectively.

Then the results of the two paths are summed together and upscaled by a deconvolutional block with a kernel size of $2 \times 2 \times 2$ and a stride of $(2, 2, 2)$. Compared with the upsampling methods applied in GCN[20] and LinkNet[2], the deconvolutional block is able to learn about the features when upscaling, which makes the model consider more information for the final segmentation. The number of the output feature maps from the decoders are 256, 128, 64, and 32 respectively. After passing the last convolutional block, the output channel number becomes the same as the input.

3.1   3D Large Kernel Convolutional Block

The work in [29] shows that in the deep CNN model, the empirical size of the receptive field is typically smaller than the theoretical one. Thus in the traditional segmentation models, the actual receptive fields are always smaller than expected, which means the models are only capable of learning limited feature information from the input image. It is thus harmful for the segmentation result due to the lack of global details. To this end, we design a 3D large kernel convolutional block with a large kernel size to enlarge the actual receptive field for the model. However, if we directly apply a large 3D kernel with a size of $K \times K \times K$, it will greatly increase the number of model parameters which is a large computational burden. In order to reduce the computational cost, we apply two convolutional kernels with sizes of $1 \times K \times K$ and $K \times 1 \times 1$ combined in different orders to simulate a convolutional kernel with a size of $K \times K \times K$, as illustrated in Fig. 2.

3.2   Anisotropic Convolutional Block

Due to the large amount of parameters from the 3D DNN model, the isotropic convolutional blocks with a size of $N \times N \times N$ make the computational cost and memory consumption high. Thus we decompose a $N \times N \times N$ kernel into one $1 \times N \times N$ kernel which fuses the features within each slice and one $N \times N \times 1$ kernel which fuses the features between slices. Inspired by the architecture proposed in [25] and [15], we employ the anisotropic convolutional blocks in the encoders. We first apply two $1 \times 3 \times 3$ kernels with residual connection for learning the intra-slice information in sagittal and coronal directions. Then we make the result pass through a $3 \times 1 \times 1$ kernel for inter-slice information processing along the axial direction. In this way, the anisotropic convolutional blocks are able to learn the information of the dataset in all dimensions of the isotropic ones with a higher memory efficiency.

4   Experiments and Results

4.1   Data Description and Implementation Settings

The dataset we used to evaluate our model is from 2017 MICCAI BRATS Challenge [16], which contains 210 HGG cases and 75 LGG cases. For each case, there are four MR images in different modalities: T1, T2, T1ce and FLAIR , with image size $155 \times 240 \times 240$. The organizers have already finished the pre-processing for the dataset, which includes skull-strip and co-registering for the four sequences. All the ground truths are labeled by experts and the results are evaluated by the official evaluation server named CBICAs Image Processing Portal. For each image, the segmentation ground truth contains three classes besides the background, which is shown in Fig. 4.1: the green part is edema (label 2), the yellow part is enhanced tumor core (label 4), and the red part is tumor core (label 1). In this experiment, we only focus on the segmentation for HGG brain tumor. We select 87 cases from HGG set for training and 40 from the rest of HGG set for testing.
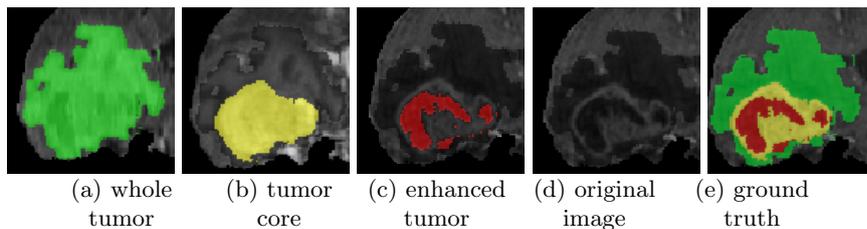

| (a) whole   (b) tumor   (c) enhanced   (d) original   (e) ground |
| tumor   core   tumor   image   truth |

Fig. 3. An example of the ground truth label for one HGG brain tumor case.

We use ADAM [10] as the optimiser with $\beta_1 = 0.9$, $\beta_2 = 0.99$, and $\epsilon = 10^{-8}$. The learning rate for training our model is set as 0.001 and the loss function is

Dice loss. For each model in the experiment, the training iteration is 8700. All of our networks are trained on 2 NVIDIA 1080Ti GPUs implemented in Tensorflow [1] with NiftyNet [5].

4.2 Evaluation Metrics

In order to evaluate the performances of the models, we apply the Dice score and Hausdorff distance mentioned in [16] for the segmentation results of the three tumor regions. For each part of the tumor, we obtain the binary map of the predicted result P and ground truth T all in range $[0, 1]$ and calculate the Dice score:

$$Dice(P, T) = \frac{|P_1 \wedge T_1|}{(|P_1| + |T_1|)/2} \tag{1}$$

where $\wedge$ is the logical AND operation, $|.|$ represents the size of the set and P_1, T_1 mean the numbers of voxels in set P and T having a value of 1 (tumor voxels) respectively. While the Dice score measures the segmentation overlap results, Hausdorff distance is applied to calculate the distance from all the points on the surface of predicted set to those of the ground truth set:

$$Haus(P, T) = max\{ \sup_{p \in \partial P_1} \inf_{t \in \partial T_1} d(p, t), \sup_{t \in \partial T_1} \inf_{p \in \partial P_1} d(t, p)\} \tag{2}$$

where $d(p, t)$ represents the shortest least-squares distance from point p to t. ∂P_1 and ∂T_1 mean the surface of set P_1 and T_1, respectively.

4.3 Results and Comparison

Table 1 and Fig. 4 show the comparison between the result of our proposed method and the state-of-the-art. It can be seen that our proposed method outperforms others in most of the metrics.

Table 1. The experiment results from different models. All the results shown are the average results on the 40 testing images. The unit of Hausdorff distance is mm.

Metrics	3D U-Net [3]	V-Net [18]	HighRes3DNet [12]	Proposed
Dice-ET	0.7731 ± 0.1673	0.7416 ± 0.1573	0.6593 ± 0.2174	$\mathbf{0.7930 \pm 0.1351}$
Dice-WT	0.8417 ± 0.1188	0.8257 ± 0.1281	0.8032 ± 0.1619	$\mathbf{0.8644 \pm 0.0909}$
Dice-TC	0.8022 ± 0.1764	0.7546 ± 0.1876	0.7205 ± 0.2440	$\mathbf{0.8189 \pm 0.1465}$
Hausdorff95-ET	$\mathbf{10.0988 \pm 20.1343}$	10.2352 ± 18.5055	21.0615 ± 29.4237	10.4811 ± 21.9127
Hausdorff95-WT	15.3867 ± 21.7618	19.6240 ± 24.1664	31.9161 ± 25.6838	$\mathbf{15.0470 \pm 21.8504}$
Hausdorff95-TC	13.687 ± 21.4019	13.4893 ± 18.9466	30.0803 ± 32.3823	$\mathbf{13.1196 \pm 21.9502}$

As shown in Table 1, the standard deviations of the Hausdorff distances are large. It is because there remain large differences of the brain tumor between cases, which makes the difficulty of the brain tumor segmentation for each case
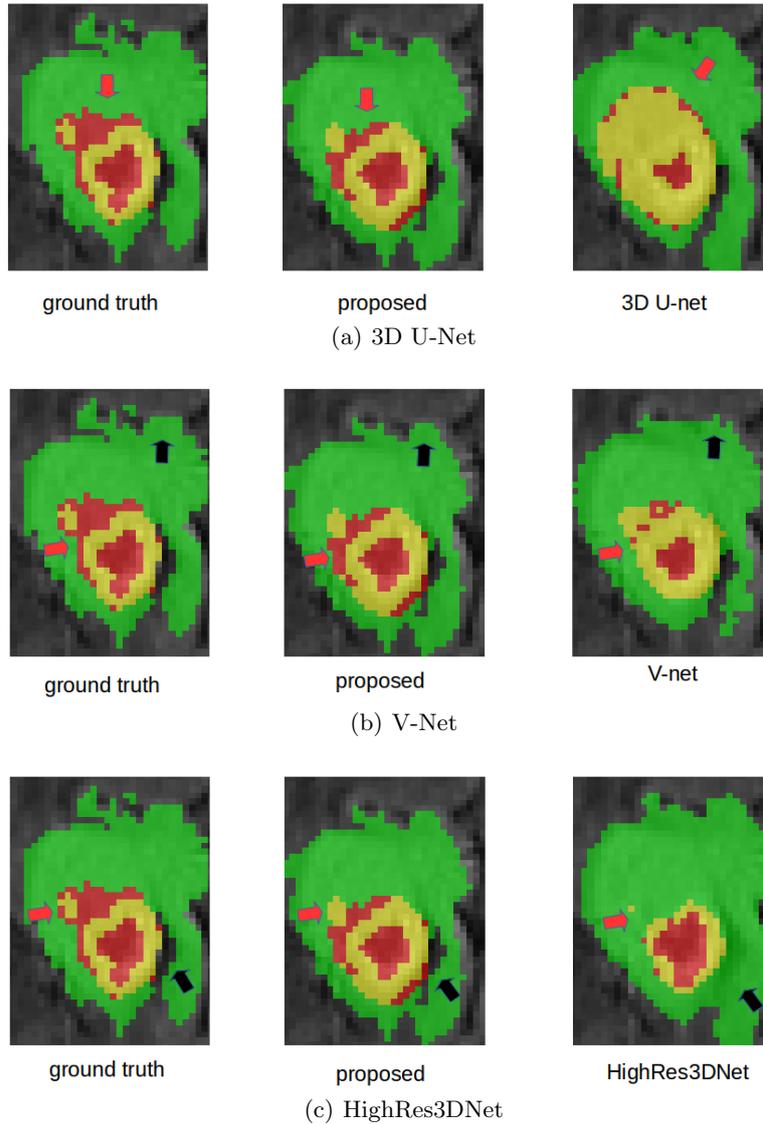
(a) 3D U-Net

(b) V-Net

(c) HighRes3DNet

Fig. 4. Visual comparison between our proposed method and the state-of-the-art.

vary. In HighRes3DNet [12], although the dialated convolutional blocks produce a large receptive field, the detailed information from low resolution levels would be lost during training due to the lack of skip connections. From the result shown in Fig. 4(c), HighRes3DNet fails to segment part of the details of the tumor core and enhanced tumor, as pointed by the red arrow. As for the whole tumor, part of the background is still misclassified as edema, as pointed by the black arrow.

3D U-Net [3] solves the problem with skip connections and provides a better segmentation result. However, the small actual field produced by the limited size of convolutional blocks makes the model fail to learn the global details of the object. From the result in Fig. 4(a), the ratio of the tumor core to whole tumor is larger than that of the ground truth. Even though the segmentation result of tumor core is correct in a local view, the high ratio from the global view makes the result less accurate. From the result of V-Net [18], although the skip connection keeps the low level details, the large amount of parameters from the isotropic convolutional blocks causes high memory consumption and makes the training hard to converge. In Fig. 4(b), it can be seen that some details at the boundaries of the whole tumor and parts of the enhanced tumor are misclassified.

5   Conclusion

In this work, we propose a 3D large kernel anisotropic network for brain tumor segmentation in MR images. The 3D large kernels after the encoders produce a large enough receptive field for the model to capture global information. Additionally, the simulation of 3D large kernel blocks with several 2D convolutional kernels reduces the computational cost. The anisotropic convolutional blocks in the encoders have the same effect as the traditional isotropic ones, and are proposed for memory efficiency and low risk of overfitting. Evaluated on part of the BRATS 17 dataset, our model is proved to be effective by outperforming some popular 3D DNN architectures.

Acknowledgements

References

1. Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., et al.: Tensorflow: A system for large-scale machine learning. In: OSDI. vol. 16, pp. 265–283 (2016)
2. Chaurasia, A., Culurciello, E.: Linknet: Exploiting encoder representations for efficient semantic segmentation. arXiv preprint arXiv:1707.03718 (2017)
3. Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O.: 3D U-Net: learning dense volumetric segmentation from sparse annotation. In: Ourselin S., Joskowicz L., Sabuncu M., Unal G., Wells W. (eds) MICCAI 2016. vol. 9901, pp. 424–432. Springer (2016) doi:10.1007/978-3-319-46723-8_49
4. Doyle, S., Vasseur, F., Dojat, M., Forbes, F.: Fully automatic brain tumor segmentation from multiple MR sequences using hidden Markov fields and variational EM. Procs. NCI-MICCAI BraTS pp. 18–22 (2013)

5. Gibson, E., Li, W., Sudre, C., Fidon, L., Shakir, D., Wang, G., Eaton-Rosen, Z., Gray, R., Doel, T., Hu, Y., et al.: NiftyNet: a deep-learning platform for medical imaging. arXiv preprint arXiv:1709.03485 (2017)

6. 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 (2017)

7. He, K., Zhang, X., Ren, S., Sun, J.: Identity mappings in deep residual networks. In: Leibe B., Matas J., Sebe N., Welling M. (eds) ECCV 2016. vol. 9908, pp. 630–645. Springer (2016) doi:10.1007/978-3-319-46493-0_38

8. Holland, E.C.: Progenitor cells and glioma formation. Current opinion in neurology 14(6), 683–688 (2001)

9. Kamnitsas, K., Ferrante, E., Parisot, S., Ledig, C., Nori, A.V., Criminisi, A., Rueckert, D., Glocker, B.: Deepmedic for brain tumor segmentation. In: Crimi A., Menze B., Maier O., Reyes M., Winzeck S., Handels H. (eds) International Workshop on Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries 2016. vol. 10154, pp. 138–149. Springer (2016) doi:10.1007/978-3-319-55524-9_14

10. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. ICLR (2015)

11. Lee, K., Zung, J., Li, P., Jain, V., Seung, H.S.: Superhuman accuracy on the SNEMI3D connectomics challenge. NIPS (2017)

12. 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: Niethammer M. et al. (eds) IPMI 2017. vol. 10265, pp. 348–360. Springer (2017) doi:10.1007/978-3-319-59050-9_28

13. Liu, D., Zhang, D., Song, Y., Zhang, C., Huang, H., Chen, M., Cai, W.: Large kernel refine fusion net for neuron membrane segmentation. In: CVPR Workshops. pp. 2212–2220 (2018)

14. Liu, S., Song, Y., Zhang, F., Feng, D., Fulham, M., Cai, W.: Clique identification and propagation for multimodal brain tumor image segmentation. In: Ascoli G., Hawrylycz M., Ali H., Khazanchi D., Shi Y. (eds) BIH 2016. vol. 9919, pp. 285–294. Springer (2016) doi:10.1007/978-3-319-47103-7_28

15. Liu, S., Xu, D., Zhou, S.K., Mertelmeier, T., Wicklein, J., Jerebko, A., Grbic, S., Pauly, O., Cai, W., Comaniciu, D.: 3D anisotropic hybrid network: Transferring convolutional features from 2D images to 3D anisotropic volumes. arXiv preprint arXiv:1711.08580 (2017)

16. 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 transactions on medical imaging 34(10), 1993–2024 (2015)

17. Menze, B.H., Van Leemput, K., Lashkari, D., Weber, M.A., Ayache, N., Golland, P.: A generative model for brain tumor segmentation in multi-modal images. In: Jiang T., Navab N., Pluim J.P.W., Viergever M.A. (eds) MICCAI 2010. vol. 6362, pp. 151–159. Springer (2010) doi:10.1007/978-3-642-15745-5_19

18. Milletari, F., Navab, N., Ahmadi, S.A.: V-Net: Fully convolutional neural networks for volumetric medical image segmentation. In: 3D Vision (3DV), 2016 Fourth International Conference on. pp. 565–571. IEEE (2016)

19. Ohgaki, H., Kleihues, P.: Population-based studies on incidence, survival rates, and genetic alterations in astrocytic and oligodendroglial gliomas. Journal of Neuropathology & Experimental Neurology 64(6), 479–489 (2005)

20. Peng, C., Zhang, X., Yu, G., Luo, G., Sun, J.: Large kernel matters–improve semantic segmentation by global convolutional network. CVPR pp. 4353–4361 (2017)

21. Prastawa, M., Bullitt, E., Gerig, G.: Synthetic ground truth for validation of brain tumor MRI segmentation. In: Duncan J.S., Gerig G. (eds) MICCAI 2005. vol. 3749, pp. 26–33. Springer (2005) doi:10.1007/11566465_4

22. Prastawa, M., Bullitt, E., Ho, S., Gerig, G.: A brain tumor segmentation framework based on outlier detection. Medical Image Analysis 8(3), 275–283 (2004)

23. Singh, S.K., Clarke, I.D., Terasaki, M., Bonn, V.E., Hawkins, C., Squire, J., Dirks, P.B.: Identification of a cancer stem cell in human brain tumors. Cancer research 63(18), 5821–5828 (2003)

24. Subbanna, N.K., Precup, D., Collins, D.L., Arbel, T.: Hierarchical probabilistic Gabor and MRF segmentation of brain tumours in MRI volumes. In: Mori K., Sakuma I., Sato Y., Barillot C., Navab N. (eds) MICCAI 2013. vol. 9351, pp. 751–758. Springer (2013) doi:10.1007/978-3-642-40811-3_94

25. Wang, G., Li, W., Ourselin, S., Vercauteren, T.: Automatic brain tumor segmentation using cascaded anisotropic convolutional neural networks. arXiv preprint arXiv:1709.00382 (2017)

26. Wen, P.Y., Kesari, S.: Malignant gliomas in adults. New England Journal of Medicine 359(5), 492–507 (2008)

27. Zhang, D., Song, Y., Liu, D., Jia, H., Liu, S., Xia, Y., Huang, H., Cai, W.: Panoptic segmentation with an end-to-end cell r-cnn for pathology image analysis. MICCAI 2018 Springer (2018)

28. Zhang, D., Song, Y., Liu, S., Feng, D., Wang, Y., Cai, W.: Nuclei instance segmentation with dual contour-enhanced adversarial network. In: ISBI. pp. 409–412. IEEE (2018)

29. Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., Torralba, A.: Object detectors emerge in deep scene CNNs. ICLR (2015)

30. Zikic, D., Glocker, B., Konukoglu, E., Criminisi, A., Demiralp, C., Shotton, J., Thomas, O.M., Das, T., Jena, R., Price, S.J.: Decision forests for tissue-specific segmentation of high-grade gliomas in multi-channel MR. In: Ayache N., Delingette H., Golland P., Mori K. (eds) MICCAI 2012. vol. 7512, pp. 369–376. Springer (2012) doi:10.1007/978-3-642-33454-2_46