



# Learning Contextual and Attentive Information for Brain Tumor Segmentation

Chenhong Zhou<sup>1</sup>, Shengcong Chen<sup>1</sup>, Changxing Ding<sup>1</sup>(✉), and Dacheng Tao<sup>2</sup>

<sup>1</sup> School of Electronic and Information Engineering,  
South China University of Technology, Guangzhou, China  
chxding@scut.edu.cn

<sup>2</sup> UBTECH Sydney AI Centre, SIT, FEIT, University of Sydney, Sydney, Australia

**Abstract.** Thanks to the powerful representation learning ability, convolutional neural network has been an effective tool for the brain tumor segmentation task. In this work, we design multiple deep architectures of varied structures to learning contextual and attentive information, then ensemble the predictions of these models to obtain more robust segmentation results. In this way, the risk of overfitting in segmentation is reduced. Experimental results on validation dataset of BraTS 2018 challenge demonstrate that the proposed method can achieve good performance with average Dice scores of 0.8136, 0.9095 and 0.8651 for enhancing tumor, whole tumor and tumor core, respectively. The corresponding scores for BraTS 2018 testing set are 0.7775, 0.8842 and 0.7960, respectively, winning the third position in the BraTS 2018 competition among 64 participating teams.

## 1 Introduction

Brain tumor is one of the most fatal cancers, which consists of uncontrolled, unnatural growth and division of the cells in the brain tissue [1]. The most frequent types of brain tumors in adults are gliomas that arise from glial cells and infiltrating the surrounding tissues [2]. According to the malignant degree of gliomas and their origin, these neoplasms can be categorized into Low Grade Gliomas (LGG) and High Grade Gliomas (HGG) [2, 3]. The former is slower-growing and comes with a life expectancy of several years, while the latter is more aggressive and infiltrative, having a shorter survival period and requiring immediate treatment [2]. Therefore, segmenting brain tumor timely and automatically would be of critical importance for assisting the doctors to improve diagnosis, perform surgery and make treatment planning.

In recent years, convolutional neural networks (CNNs) have been widely applied to automatic brain tumor segmentation tasks. Pereira et al. [15] and Havaei et al. [13] respectively trained a CNN to predict the label of the central

---

C. Zhou and S. Chen—Equal contribution.

voxel only within a patch, which causes that they suffer from high computational cost and time consumption during inference. To reduce the computational burden, Kamnitsas et al. [5] propose an efficient model named DeepMedic that can predict the labels of voxels within a patch simultaneously, in order to achieve dense predictions. Recently, fully convolutional networks (FCNs) have achieved promising results. Shen et al. [6] and Zhao et al. [11] allow end-to-end dense training and testing for brain tumor segmentation at the *slice* level to improve computational efficiency. With a large variety of CNN architectures proposed, the performance of automatic brain tumor segmentation from Magnetic Resonance Imaging (MRI) images has been improved greatly.

In this work, we construct multiple different CNN architectures and approaches to ensemble their prediction results, in order to produce stable and robust segmentation performance. We evaluate our approaches on the validation set of 2018 Brain Tumor Segmentation (BraTS) challenge, where we obtain the good performance with average Dice scores of 0.8136, 0.9095 and 0.8651 for enhancing tumor, whole tumor and tumor core, respectively. Correspondingly, we achieve promising scores for BraTS 2018 testing set are 0.7775, 0.8842 and 0.7960, respectively.

## 2 Data

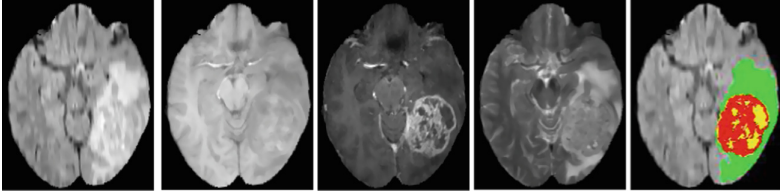
We use the dataset of 2018 Brain Tumor Segmentation challenge [2, 4, 7, 8, 21] for experiments, which consists of the training set, validation set and testing set. The training set contains 210 HGG and 75 LGG cases whose corresponding manual segmentations are provided. As shown in Fig. 1, the provided manual segmentations include four labels: 1 for necrotic (NCR) and the non-enhancing (NET) tumor, 2 for edema (ED), 4 for enhancing tumor (ET), and 0 for everything else, i.e. normal tissue and background (black padding). The validation set and testing set contain 66 cases and 191 cases with unknow grade and hidden segmentations, respectively. Each case has four MRI sequences that are named T1, T1 contrast enhanced (T1ce), T2 and FLAIR, respectively. These datasets are provided after their pre-processing, i.e. co-registered to the same anatomical template, interpolated to the same resolution ( $1\text{ mm}^3$ ) and skull-stripped, where dimensions of each MRI sequence are  $240 \times 240 \times 155$ . Besides, the official evaluation is calculated by merging the predicted labels into three regions: whole tumor (1,2,4), tumor core (1,4) and enhancing tumor (4). The valuation for validation set is conducted via an online system<sup>1</sup>.

## 3 Methods

### 3.1 Basic Networks

As is well known, brain tumor segmentation from MRI images is a very tough and challenging task due to the severe class imbalance problem. Following [14],

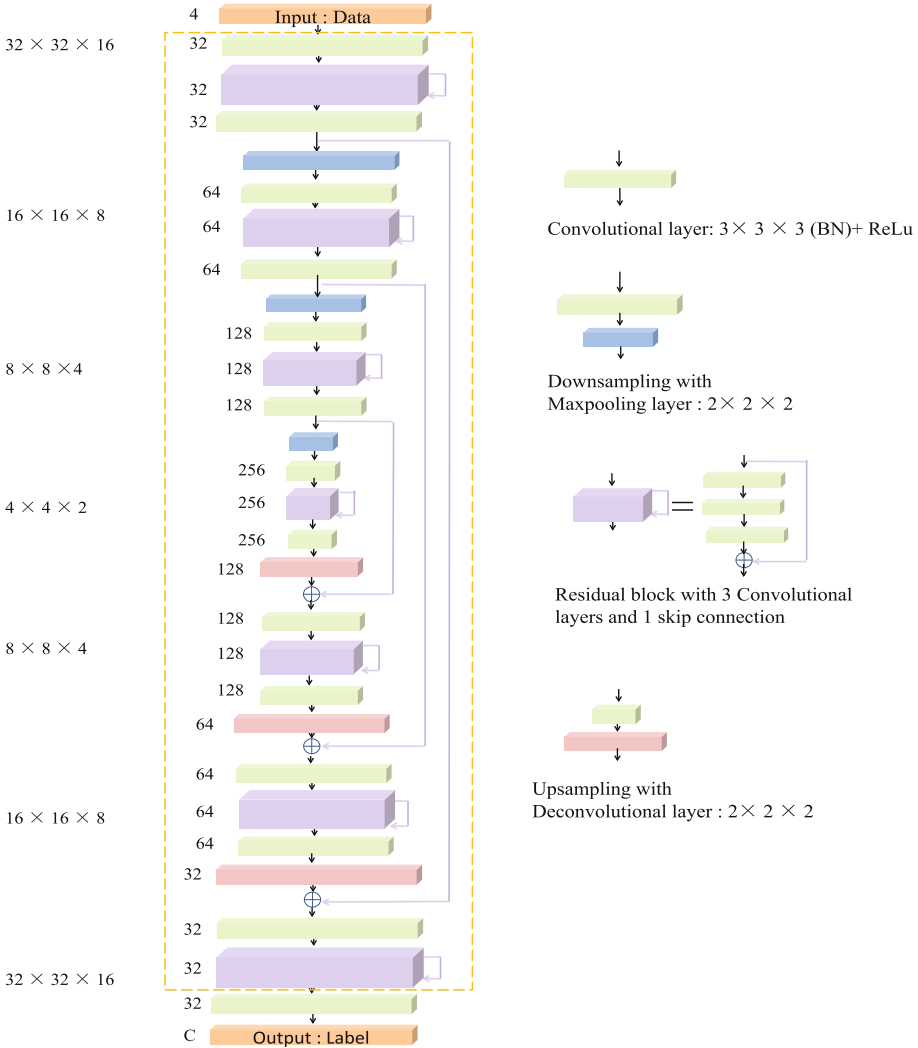
<sup>1</sup> <https://ipp.cbica.upenn.edu/>.



**Fig. 1.** Example of images from the BRATS 2018 dataset. From left to right: Flair, T1, T1ce, T2 and manual annotation overlaid on the Flair image: edema (green), necrosis and non-enhancing (yellow), and enhancing (red). (Color figure online)

we decompose the multi-class brain tumor segmentation into three different but related sub-tasks to deal with the class imbalance problem. (1) *Coarse segmentation to detect whole tumor*. In this sub-task, the region of whole tumor is located. To reduce overfitting, we define the first task being the five-class segmentation problem. (2) *Refined segmentation for whole tumor and its intra-tumoral classes*. The above obtained coarse tumor mask is dilated by 5 voxels as the ROI for the second task. In this sub-task, the precise classes for all voxels within the dilated region are predicted. (3) *Precise segmentation for enhancing tumor*. We specially design the third sub-task to segment the enhancing tumor, due to its high difficulty of segmentation.

**Model Cascade.** In view of the above three sub-tasks, it is probably easy to train a CNN individually for each sub-task, which is the currently popular Model Cascade (MC) strategy. We use a 3D variant of the FusionNet [10], as illustrated in Fig. 2. The network architecture consists of an encoding path (upper half of the network) to extract complex semantic features and a symmetric decoding path (lower half of the network) to recover the same resolution as the input to achieve voxel-to-voxel predictions. The network is constructed by four types of basic building blocks, as shown in Fig. 2. In addition, the network has not only the short shortcuts in residual blocks, but also three long skip connections to merge the feature maps from the same level in the encoding path during decoding by using a voxel-wise addition. We employ the identical network architecture for each sub-task, except for the final convolutional classification layer. The number of channels of last classification layer is equal to 5, 5 and 2 for the first, second and third sub-tasks, respectively. Besides, size of input patches for the network is  $32 \times 32 \times 16 \times 4$ , where the number 4 indicates the four MRI modalities. During inference, we adopt overlap-tile strategy in [9]. Thus, we abandon the prediction results of border region and only retain the predictions in the center region ( $20 \times 20 \times 5$ ). This trick is also used in the following models. Different from [20] that is a typical example of model cascade strategy, we dilate the coarse tumor mask to prevent tumor omitting in the second sub-task and adopt the same 3D basic network architecture for each sub-task instead of sophisticated operations that design different networks for different sub-tasks.

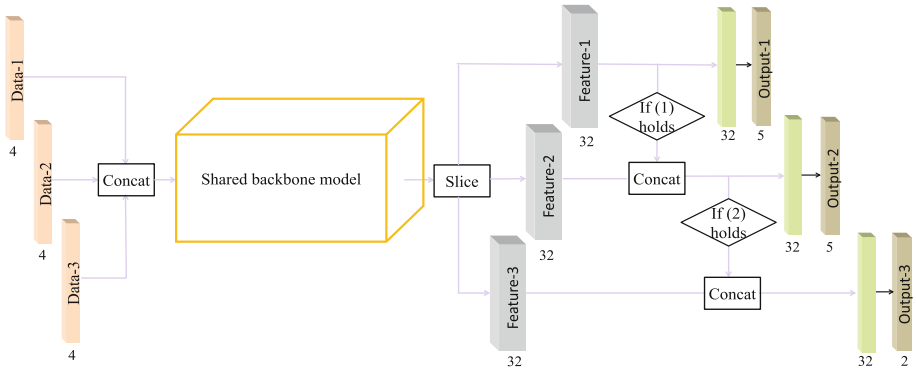


**Fig. 2.** Network architecture used in each sub-task. The building blocks are represented by colored cubes with numbers nearby being the number of feature maps.  $C$  equals to 5, 5, and 2 for the first, second, and third task, respectively. (Best viewed in color) This figure is reproduced from [14].

**One-Pass Multi-task Network.** The above proposed model cascade approach has obtained promising segmentation performance. To a certain extent, it alleviates the problem of class imbalance. However, model cascade approach needs to train a series of deep models individually for the three different sub-tasks, which leads to large memory cost and system complexity during training and testing. In addition, we have observed that the networks used for three sub-tasks are

almost the same except for the training data and the classification layer. It is obvious that the three sub-tasks are relative to each other.

Therefore, we employ the one-pass multi-task network (OM-Net) proposed in [14], which is a multi-task learning framework that incorporates the three sub-tasks into a end-to-end holistic network, to save a lot of parameters and exploit the underlying relevance among the three sub-tasks. The OM-Net proposed in [14] is described in Fig. 3, which is composed of the sharable parameters and task-specific parameters. Specially, the shared backbone model refers to the network layers outlined by the yellow dashed line in Fig. 2, while three respective branches for different sub-tasks are designed after the shared parts.



**Fig. 3.** Architecture of OM-Net. Data- $i$ , Feature- $i$ , and Output- $i$  denote training data, feature, and classification layer for the  $i$ -th task, respectively. The shared backbone model refers to the network layers outlined by the yellow dashed line in Fig. 2. This figure is reproduced from [14].

In addition, inspired by the curriculum learning theory proposed by Bengio et al. [12] that humans can learn a set of concepts much better when the concepts to be learned are presented by gradually increasing the difficulty level, we adopt the curriculum learning-based training strategy in [14] to train OM-Net more effectively. The training strategy of our framework is to start training the network on the first easiest sub-task, then gradually add the more difficult sub-tasks and their corresponding training data to the model. This is a process from easy to difficult, highly consistent with the thought of manual segmentation of the tumor. Besides, the training data conforming to the sampling strategy of the other sub-tasks can be transferred to achieve data sharing. Eventually, the OM-Net is a single deep model to solve three sub-tasks simultaneously in one-pass. It is also significantly smaller in the number of trainable parameters than model cascade strategy and can be trained end-to-end using stochastic gradient descent to achieve data sharing and parameters sharing in a holistic network.

### 3.2 Extended Networks

In this section, we extend and improve the MC-baseline and OM-Net from four aspects to further promote the performance. The four aspects are elaborated in the following.

**Deeper OM-Net.** We deepen the OM-Net by appending a residual block (the violet block in Fig. 2 right) after each existing residual block of OM-Net, which is the easiest and most direct way to boost the performance.

**Dense Connections.** Inspired by [17], the basic 3D network of MC-baseline is modified by adding a series of nested and dense skip connections to form a more powerful architecture. The purpose of the re-designed skip connections is to reduce the semantic gap between the feature maps of the encoder and decoder [17].

**Attention Mechanisms.** Attention mechanisms have been shown to improve performance across a range of tasks, which is attributed to their ability to focus on the more informative components and suppress less useful ones. Particularly, “Squeeze-and-Excitation” (SE) block is proposed to adaptively perform channel-wise feature recalibration by explicitly modelling interdependencies between channels in [16], in order to boost the representational power of CNNs.

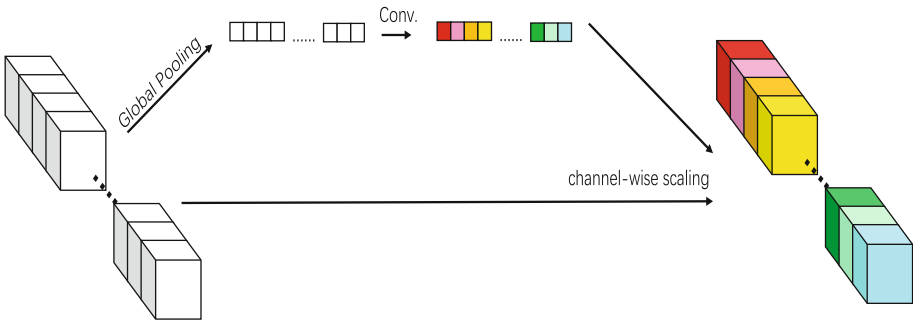
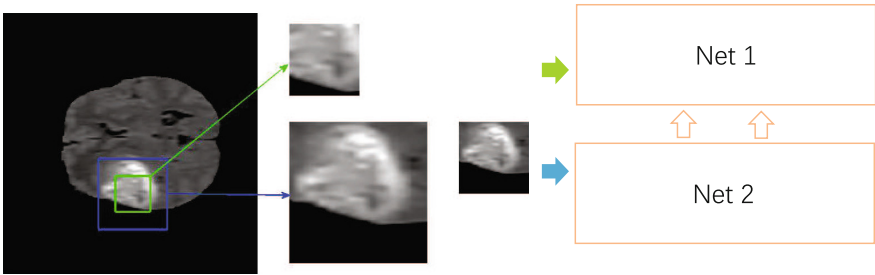


Fig. 4. The adopted “Squeeze-and-Excitation” (SE) block.

Inspired by it, we introduce SE blocks to OM-Net, in order to recalibrate the feature maps and further improve the learning and representational properties of OM-Net. The SE block is described in Fig. 4. Similar to [16], the SE block focuses on channels to adaptively recalibrate channel-wise feature responses in two steps, squeeze and excitation. It helps the network to increase the sensitivity to informative features and suppress less useful ones.

**Multi-scale Contextual Information.** To deal with the 3D medical scans, we employ the above 3D CNNs that process small 3D patches. However, small patches cause the network to learn the limited contextual information. It seems necessary to introduce larger patches, in order to provide larger receptive fields and more contextual information to the network. Therefore, inspired by [5], we design a two parallel pathway architecture that processes two scale input patches simultaneously. As shown in Fig. 5, we incorporate both local and larger contextual information to the model, which not only extracts semantic features at a higher resolution, but also considers larger contextual information from the lower resolution level. It can provide rich information to discriminate voxels that appear very similar when considering only local appearance, avoiding making wrong predictions.



**Fig. 5.** The proposed network architecture to introduce multi-scale contextual information.

**Table 1.** Mean values of Dice and Hausdorff95 measurements on BraTS 2018 validation set (submission id DL-86-61).

Method	Dice			Hausdorff95		
	Enh.	Whole	Core	Enh.	Whole	Core
MC-Net	0.7732	0.9006	0.8232	4.1647	4.4849	7.6216
OM-Net	0.7882	0.9034	0.8273	3.1003	6.5218	7.1974
MC-Net (Dense connections)	0.7768	0.9049	0.8358	3.3994	4.2390	6.8503
MC-Net (Multi-scale)	0.7751	0.9059	0.8181	2.8192	4.0085	6.3437
OM-Net (Attention)	0.7792	0.8986	0.8329	3.8949	6.1926	8.3459
Deeper OM-Net	0.7882	0.8991	0.8405	2.7649	8.0177	7.3671
Deeper OM-Net (Attention)	0.7925	0.8948	0.8333	2.8099	4.8093	6.7755
Ensembles	<b>0.8137</b>	0.9092	0.8530	2.7092	4.4519	7.1535
Ensembles + post-processing	0.8136	<b>0.9095</b>	<b>0.8651</b>	2.7162	4.1724	6.5445

**Table 2.** Mean values of Sensitivity and Specificity measurements on BraTS 2018 validation set.

Method	Sensitivity			Specificity		
	Enh.	Whole	Core	Enh.	Whole	Core
MC-Net	0.8082	0.9118	0.7971	0.9980	0.9943	0.9981
OM-Net	0.7953	0.9084	0.7883	0.9984	0.9948	0.9985
MC-Net (Dense connections)	0.8144	0.9127	0.8323	0.9979	0.9947	0.9973
MC-Net (Multi-scale)	0.8095	0.9065	0.7970	0.9980	0.9951	0.9981
OM-Net (Attention)	0.8245	0.9078	0.8103	0.9977	0.9943	0.9979
Deeper OM-Net	0.7962	0.9059	0.8176	0.9984	0.9945	0.9979
Deeper OM-Net (Attention)	0.7995	0.8972	0.8159	0.9983	0.9946	0.9975
Ensembles	0.8137	0.9148	0.8294	0.9983	0.9950	0.9981
Ensembles + post-processing	0.8135	0.9142	0.8683	0.9983	0.9951	0.9968

**Table 3.** The segmentation results of our proposed method on BraTS 2018 testing set.

	Dice			Hausdorff95		
	Enh.	Whole	Core	Enh.	Whole	Core
Mean	0.7775	0.8842	0.7960	2.9366	5.4681	6.8773
StdDev	0.2533	0.1127	0.2593	4.6894	7.6479	10.1779
Median	0.8498	0.9183	0.9030	1.7321	3.1623	3.0000
25quantil	0.7596	0.8725	0.8062	1.4142	2.0000	1.7321
75quantil	0.8997	0.9437	0.9376	2.7337	5.3852	7.2798

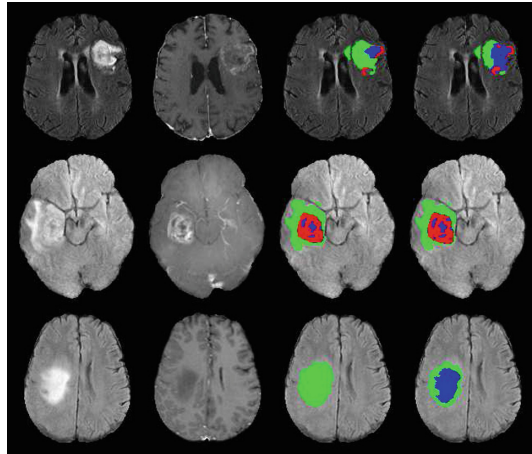
### 3.3 Ensembles of the Above Multiple Models

Model ensembling is an effective method to improve performance, e.g. Kamnitsas et al. [19] ensembled DeepMedic [5], 3D FCN [18], and 3D U-Net [9] into EMMA. In this paper, we also adopt model ensembling to obtain more robust segmentation results. Above multiple models, including MC-Net, OM-Net and their variants are trained separately, and the predicted probabilities are averaged at testing time. Additionally, a simple yet effective post-processing method [14] is adopted to improve segmentation performance.

## 4 Experiments and Results

**Pre-processing.** We adopt the minimal pre-processing operation to the BraTS 2018 data. That is, each sequence is individually normalized by subtracting its mean value and dividing by its standard deviation of the intensities within the brain area in that sequence.





**Fig. 6.** Example segmentation results on the validation set of BraTS 2018. From left to right: Flair, T1ce, segmentation results using MC-Net only overlaid on Flair image, and segmentation results using the proposed method overlaid on Flair image; edema (green), necrosis and non-enhancing (blue), and enhancing (red). (Color figure online)

**Segmentation Results.** Table 1 presents the mean values of Dice and Hausdorff95 measurements of the different models on BraTS 2018 validation set, meanwhile Table 2 presents the corresponding mean values of Sensitivity and Specificity measurements. We can see that the OM-Net is superior to MC-Net, despite the fewer training parameters of OM-Net. Besides, the extended networks including MC-Net (Dense connections), MC-Net (Multi-scale), OM-Net (Attention), Deeper OM-Net and Deeper OM-Net (Attention) improve the segmentation performance to some extent. Finally, it shows that the proposed method achieves promising performance with average Dice scores of 0.8136, 0.9095 and 0.8651 for enhancing tumor, whole tumor and tumor core, respectively. In addition, we also provide qualitative comparisons in Fig. 6. From Fig. 6, we can see that model ensembling is much better and the effectiveness of the proposed method is justified.

Table 3 presents the segmentation results of our proposed method on BraTS 2018 testing set. It shows that the proposed method yields excellent performance, winning the third position in the BraTS 2018 competition.

## 5 Conclusion

In this work, we employ the OM-Net to obtain strong basic results, and then extend and improve MC-baseline and OM-Net from multiple aspects to further promote the performance. Eventually, the predictions of these models are ensembled to produce robust performance for brain tumor segmentation. The proposed method yields promising results, winning third place in the final testing stage of the BraTS 2018 challenge.

**Acknowledgments.** Changxing Ding was supported in part by the National Natural Science Foundation of China (Grant No.: 61702193), Science and Technology Program of Guangzhou (Grant No.: 201804010272), and the Program for Guangdong Introducing Innovative and Entrepreneurial Teams (Grant No.: 2017ZT07X183). Dacheng Tao was supported by Australian Research Council Projects (FL-170100117, DP-180103424 and LP-150100671).

## References

1. Işın, A., Direkoğlu, C., Şah, M.: Review of MRI-based brain tumor image segmentation using deep learning methods. *Procedia Comput. Sci.* **102**, 317–324 (2016)
2. Menze, B.H., et al.: The multimodal brain tumor image segmentation benchmark (BRATS). *IEEE TMI* **34**(10), 1993–2024 (2015)
3. 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**, R97–R129 (2013)
4. Bakas, S., et al.: Advancing the cancer genome atlas glioma MRI collections with expert segmentation labels and radiomic features. *Nat. Sci. Data* **4**, 170117 (2017)
5. Kamnitsas, K., et al.: Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. *Med. Image Anal.* **36**, 61–78 (2017)
6. Shen, H., Wang, R., Zhang, J., McKenna, S.J.: Boundary-aware fully convolutional network for brain tumor segmentation. In: Descoteaux, M., Maier-Hein, L., Franz, A., Jannin, P., Collins, D.L., Duchesne, S. (eds.) *MICCAI 2017*. LNCS, vol. 10434, pp. 433–441. Springer, Cham (2017). [https://doi.org/10.1007/978-3-319-66185-8\\_49](https://doi.org/10.1007/978-3-319-66185-8_49)
7. Bakas, S., et al.: Segmentation labels and radiomic features for the pre-operative scans of the TCGA-GBM collection. In: *The Cancer Imaging Archive* (2017)
8. Bakas, S., et al.: Segmentation labels and radiomic features for the pre-operative scans of the TCGA-LGG collection. In: *The Cancer Imaging Archive* (2017)
9. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) *MICCAI 2015*. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). [https://doi.org/10.1007/978-3-319-24574-4\\_28](https://doi.org/10.1007/978-3-319-24574-4_28)
10. Quan, T.M., et al.: Fusionnet: a deep fully residual convolutional neural network for image segmentation in connectomics. *arXiv preprint arXiv:1612.05360* (2016)
11. Zhao, X., Wu, Y., Song, G., et al.: A deep learning model integrating FCNNs and CRFs for brain tumor segmentation. *Med. Image Anal.* **43**, 98–111 (2018)
12. Bengio, Y., Louradour, J., Collobert, R., and Weston, J.: Curriculum learning. In: *ICML*, pp. 41–48. ACM (2009)
13. Havaei, M., et al.: Brain tumor segmentation with deep neural networks. *Med. Image Anal.* **35**, 18–31 (2017)
14. Zhou, C., Ding, C., Lu, Z., Wang, X., Tao, D.: One-pass multi-task convolutional neural networks for efficient brain tumor segmentation. In: Frangi, A.F., Schnabel, J.A., Davatzikos, C., Alberola-López, C., Fichtinger, G. (eds.) *MICCAI 2018*. LNCS, vol. 11072, pp. 637–645. Springer, Cham (2018). [https://doi.org/10.1007/978-3-030-00931-1\\_73](https://doi.org/10.1007/978-3-030-00931-1_73)
15. Pereira, S., Pinto, A., Alves, V., Silva, C.A.: Brain tumor segmentation using convolutional neural networks in MRI images. *IEEE Trans. Med. Imag.* **35**(5), 1240–1251 (2016)
16. Hu, J., Shen, L., and Sun, G.: Squeeze-and-excitation networks. *arXiv preprint arXiv:1709.01507* (2016)

17. Zhou, Z., Siddiquee, M.M.R., Tajbakhsh, N., and Liang, J.: UNet++: A nested U-Net architecture for medical image segmentation. arXiv preprint [arXiv:1807.10165](https://arxiv.org/abs/1807.10165) (2018)
18. Long, J., et al.: Fully convolutional networks for semantic segmentation. In: CVPR, pp. 343–3440 (2015)
19. Kamnitsas, K., et al.: Ensembles of multiple models and architectures for robust brain tumour segmentation. In: Crimi, A., Bakas, S., Kuijff, H., Menze, B., Reyes, M. (eds.) BrainLes 2017. LNCS, vol. 10670, pp. 450–462. Springer, Cham (2018). [https://doi.org/10.1007/978-3-319-75238-9\\_38](https://doi.org/10.1007/978-3-319-75238-9_38)
20. Wang, G., Li, W., Ourselin, S., Vercauteren, T.: Automatic brain tumor segmentation using cascaded anisotropic convolutional neural networks. In: Crimi, A., Bakas, S., Kuijff, H., Menze, B., Reyes, M. (eds.) BrainLes 2017. LNCS, vol. 10670, pp. 178–190. Springer, Cham (2018). [https://doi.org/10.1007/978-3-319-75238-9\\_16](https://doi.org/10.1007/978-3-319-75238-9_16)
21. Bakas, S., Reyes, M., et al.: Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the BRATS challenge. arXiv preprint [arXiv:1811.02629](https://arxiv.org/abs/1811.02629) (2018)