Two-Stage Cascaded U-Net: 1st Place Solution to BraTS Challenge 2019 Segmentation Task

Chapter · May 2020

DOI: 10.1007/978-3-030-46640-4_22

4 authors, including:

Changxing Ding
South China University of Technology
45 PUBLICATIONS **1,878** CITATIONS

SEE PROFILE

Two-Stage Cascaded U-Net: 1st Place Solution to BraTS Challenge 2019 Segmentation Task

Zeyu Jiang[1], Changxing Ding[1(✉)], Minfeng Liu[2], and Dacheng Tao[3]

[1] School of Electronic and Information Engineering
South China University of Technology, Guangzhou, China
chxding@scut.edu.cn
[2] Nanfang Hospital, Southern Medical University,
Guangzhou, Guangdong 510515, China
[3] UBTECH Sydney AI Centre, SIT, FEIT, University of Sydney, Australia

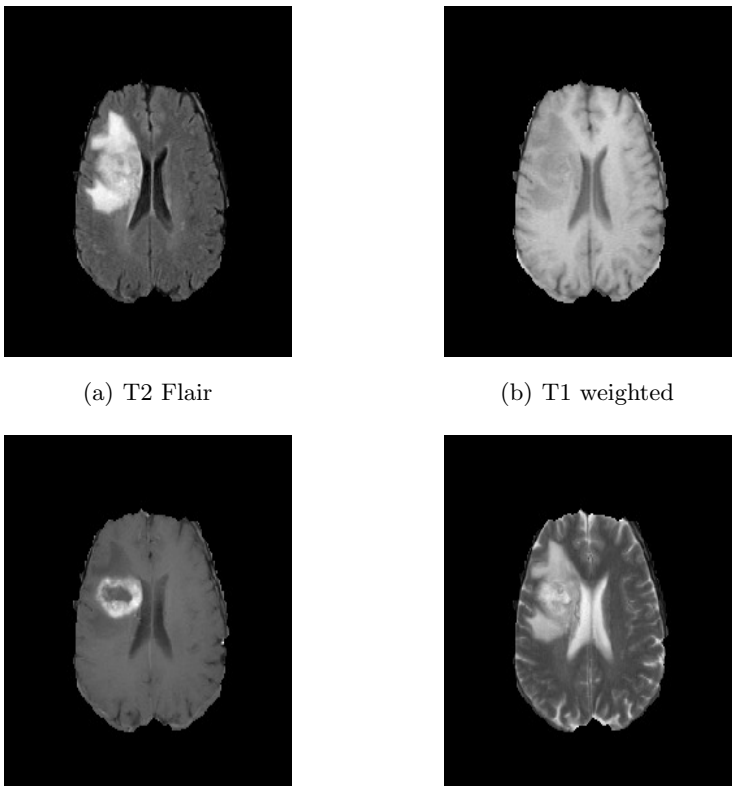
Abstract. In this paper, we devise a novel two-stage cascaded U-Net to segment the substructures of brain tumors from coarse to fine. The network is trained end-to-end on the Multimodal Brain Tumor Segmentation Challenge (BraTS) 2019 training dataset. Experimental results on the testing set demonstrate that the proposed method achieved average Dice scores of 0.83267, 0.88796 and 0.83697, as well as Hausdorff distances (95%) of 2.65056, 4.61809 and 4.13071, for the enhancing tumor, whole tumor and tumor core, respectively. The approach won the 1st place in the BraTS 2019 challenge segmentation task, with more than 70 teams participating in the challenge.

Keywords: Deep Learning · Brain Tumor Segmentation · U-Net

1 Introduction

Gliomas are the most common type of primary brain tumors. Automatic three-dimensional brain tumor segmentation can save doctors time and provide an appropriate method of additional tumor analysis and monitoring. Recently, deep learning approaches have consistently outperformed traditional brain tumor segmentation methods[6, 10, 17, 20, 24, 27].

The multimodal brain tumor segmentation challenge (BraTS) is aimed at evaluating state-of-the-art methods for the segmentation of brain tumors.[1–4, 13]. The BraTS 2019 training dataset, which comprises 259 cases of high-grade gliomas (HGG) and 76 cases of low-grade gliomas (LGG), is manually annotated by both clinicians and board-certified radiologists. For each patient, a native pre-contrast (T1), a post-contrast T1-weighted (T1Gd), a T2-weighted (T2) and a T2 Fluid Attenuated Inversion Recovery (T2-FLAIR) are provided. An example image set is presented in Fig. 1. Each tumor is segmented into enhancing tumor, the peritumoral edema, and the necrotic and non-enhancing tumor core. A number of metrics (Dice score, Hausdorff distance (95%), sensitivity and specificity) are used to measure the segmentation performance of the algorithms proposed by participants.



(a) T2 Flair (b) T1 weighted

(c) T1 post-contrast (d) T2 weighted

Fig. 1. Example of image modalities in the BraTS 2019 dataset.

In BraTS 2017, Kamnitsas et al. [9], who was the first-place winner of the challenge, proposed Ensembles of Multiple Models and Architectures (EMMA) for robust segmentation, which was achieved by combining several network architectures including DeepMedic[10], 3D U-Net[18] and 3D FCN[12]. These networks were trained with different optimization processes via diverse loss functions such as Dice loss[14] and cross-entropy loss. In BraTS 2018, Myronenko[15], who achieved the best performance on the testing dataset, utilized an asymmetrical U-Net with a larger encoder to extract image features, along with a smaller decoder to reconstruct the label. He fed a very large patch size ($160\times192\times128$ voxels) into the network, and also added a variational autoencoder (VAE) branch in order to regularize the shared encoder.

In this work, inspired by the cascaded strategy[19, 22, 25, 26],we propose a novel two-stage cascaded U-Net. In the first stage, we use a variant of U-Net as the first stage network to train a coarse prediction. In the second stage, we increase the width of the network and use two decoders so as to boost perfor-

mance. The second stage is added to refine the prediction map by concatenating a preliminary prediction map with the original input to utilize auto-context. We do not use any additional training data and only participate in the segmentation task in testing phase.

2 Methods

Myronenko [15] proposed an asymmetrical U-Net with a variational autoencoder branch [5, 11]. In this paper, we take a variant of this approach as the basic segmentation architecture. We further propose a two-stage cascaded U-Net. The details are illustrated as follows.

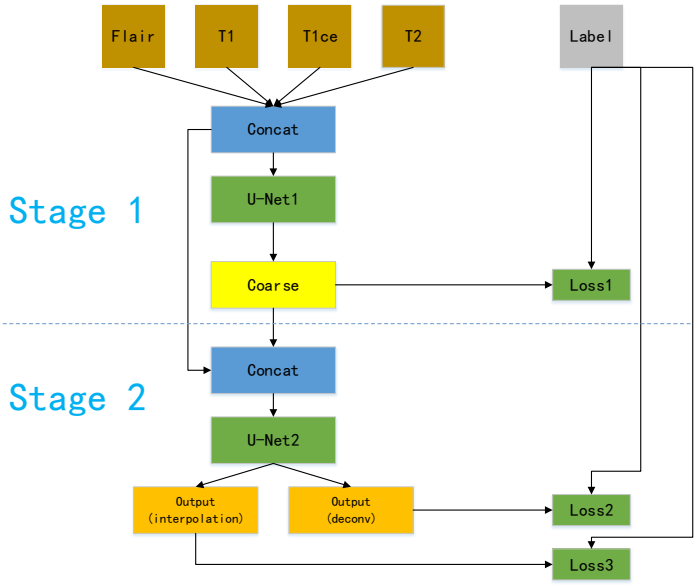


Fig. 2. Overview of the two-stage cascaded network.

2.1 Model Cascade

As can be seen in Fig. 2, in the first stage, multi-modal magnetic resonance images($4\times128\times128\times128$) are passed into the first stage U-Net and predict a segmentation map roughly. The coarse segmentation map is fed together with the raw images into the second stage U-net. The second stage can provide a more accurate segmentation map with more network parameters. The two-stage cascaded network is trained in an end-to-end fashion.

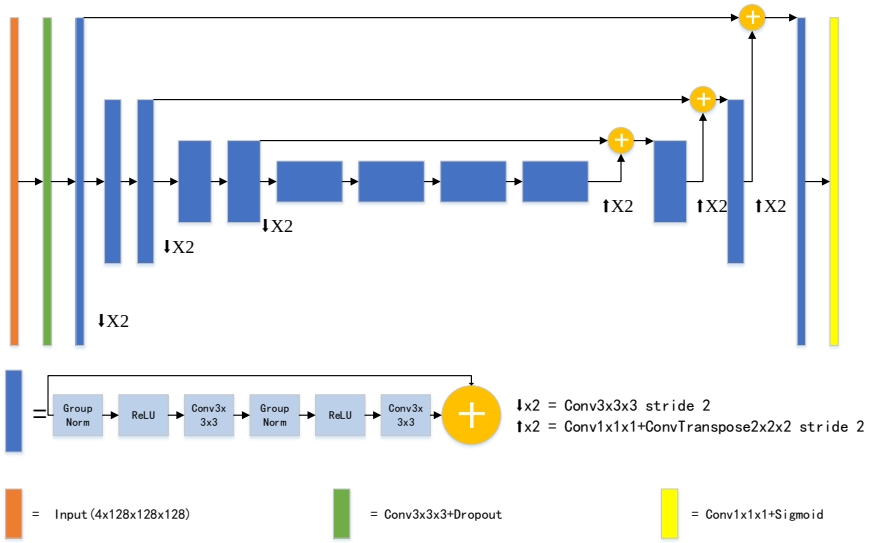


Fig. 3. The first stage network architecture.

2.2 The First Stage Network Architecture

Due to GPU memory limitations, our networks is designed to take input patches of size 128×128×128 voxels and to use a batch size of one. The network architecture consists of a larger encoding path, to extract complex semantic features, and a smaller decoding path, to recover a segmentation map with the same input size. The architecture of the first stage network is presented in Fig. 3.

The 3D U-Net has an encoder and a decoder path, each of which have four spatial levels. At the beginning of the encoder, patches of size 128×128×128 voxels with four channels are extracted from the brain tumor images as input, followed by an initial 3×3×3 3D convolution with 16 filters. We also use a dropout with a rate of 0.2 after the initial encoder convolution. The encoder part uses a pre-activated residual block[7, 8]. Each of these blocks consists of two 3x3x3 convolutions with Group Normalization [23] with group size of 8 and Rectified Linear Unit (ReLU) activation, followed by additive identity skip connection. The number of pre-activated residual blocks is 1, 2, 2, and 4 within each spatial level. Moreover, a convolution layer with a 3×3×3 filter and a stride of 2 is used to reduce the resolution of the feature maps by 2 and simultaneously increase the number of feature channels by 2.

Unlike the encoder, the decoder structure uses a single pre-activated residual block for each spatial level. Before up-sampling, we use 1×1×1 convolutions to reduce the number of features by a factor of 2. Compared with [15], we use a deconvolution with kernel size 2×2×2 and a stride of 2 rather than trilinear interpolation in order to double the size of the spatial dimension. The network features shortcut connections between corresponding layers with the same reso-

lution in the encoder and decoder by elementwise summation. At the end of the decoder, a 1×1×1 convolution is used to decrease the number of output channels to three, followed by a sigmoid function. The detail of the structure is shown in Table 1.

2.3 The Second Stage Network Architecture

Different from the network in the first stage, we double the number of filters in the initial 3D convolution in order to increase the network width. What's more, we use two decoders. The structure of the two decoders is the same except that one uses a deconvolution and the other uses trilinear interpolation. The interpolation decoder is used only during training. Because the performance of the decoder used deconvolution is better than used trilinear interpolation and add a decoder used trilinear interpolation to regularize the shared encoder can improve the performance in our experiment. The architecture of the second stage network is presented in Fig. 4 and the detail of the structure is shown in Table 2.

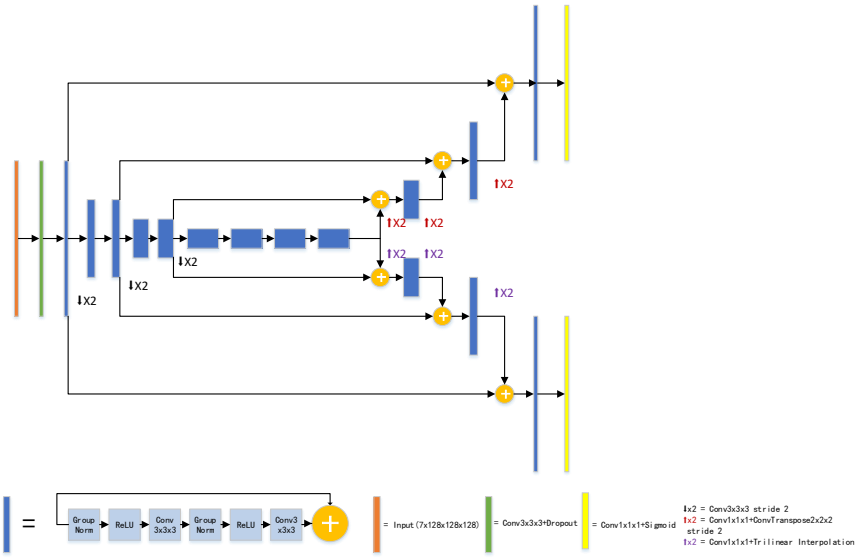


Fig. 4. The second stage network architecture.

2.4 Loss

The Dice Similarity Coefficient measures(DSC) the degree of overlap between the prediction map and ground truth. The DSC is calculated by Eq. 1, where S

is the output of network, R is the ground truth label and $|\cdot|$ denotes the volume of the region.

$$\mathrm{DSC} = \frac{2|S \cap R|}{|S| + |R|} \tag{1}$$

The soft Dice loss is designed as following:

$$\mathbf{L}_{dice} = \frac{2 * \sum S * R}{\sum S^2 + \sum R^2 + \epsilon} \tag{2}$$

Instead of learning the labels (e.g. enhancing tumor, edema, necrosis and non-enhancing), we directly optimize the three overlapping regions (whole tumor, tumor core and enhancing tumor) with the Dice loss, then simply add the Dice loss functions of each region together. We also add the loss of each stage together to arrive at the final loss.

Table 1. The first stage network structure,, where + stands for additive identity skip connection, Conv3 - 3×3×3 convolution, Conv1 - 1×1×1 convolution, GN - group normalization, ConvTranspose - deconvolution with kernel size 2×2×2.

U-Net 1				
	Name	Details	Repeat	Size
Encoder	Input			4×128×128×128
	InitConv	Conv3,Dropout	1	16×128×128×128
	EnBlock1	GN,ReLU,Conv3,GN,ReLU,Conv3,+	1	16×128×128×128
	EnDown1	Conv3 stride 2	1	32×64×64×64
	EnBlock2	GN,ReLU,Conv3,GN,ReLU,Conv3,+	2	32×64×64×64
	EnDown2	Conv3 stride 2	1	64×32×32×32
	EnBlock3	GN,ReLU,Conv3,GN,ReLU,Conv3,+	2	64×32×32×32
	EnDown3	Conv3 stride 2	1	128×16×16×16
	EnBlock4	GN,ReLU,Conv3,GN,ReLU,Conv3,+	4	128×16×16×16
Decoder	DeUp3	Conv1,ConvTranspose,+EnBlock3	1	64×32×32×32
	DeBlock3	GN,ReLU,Conv3,GN,ReLU,Conv3,+	1	64×32×32×32
	DeUp2	Conv1,ConvTranspose,+EnBlock2	1	32×64×64×64
	DeBlock2	GN,ReLU,Conv3,GN,ReLU,Conv3,+	1	32×64×64×64
	DeUp2	Conv1,ConvTranspose,+EnBlock1	1	16×128×128×128
	DeBlock1	GN,ReLU,Conv3,GN,ReLU,Conv3,+	1	16×128×128×128
	EndConv	Conv1	1	3×128×128×128
	Sigmoid	Sigmoid	1	3×128×128×128

3 Experiments

3.1 Data Pre-processing and Augmentation

Before feeding the data into the deep learning network, a preprocessing method is used to process the input data. Since the MRI intensity values are non-standardized, we apply intensity normalization to each MRI modality from each

Table 2. The second stage network structure, where + stands for additive identity skip connection, Conv3 - 3×3×3 convolution, Conv1 - 1×1×1 convolution, GN - group normalization, ConvTranspose - deconvolution with kernel size 2×2×2, Upsampling - trilinear interpolation, Decoder2 is used only during training.

U-Net 2				
	Name	Details	Repeat	Size
Encoder	Input			7×128×128×128
	InitConv	Conv3,Dropout	1	32×128×128×128
	EnBlock1	GN,ReLU,Conv3,GN,ReLU,Conv3,+	1	32×128×128×128
	EnDown1	Conv3 stride 2	1	64×64×64×64
	EnBlock2	GN,ReLU,Conv3,GN,ReLU,Conv3,+	2	64×64×64×64
	EnDown2	Conv3 stride 2	1	128×32×32×32
	EnBlock3	GN,ReLU,Conv3,GN,ReLU,Conv3,+	2	128×32×32×32
	EnDown3	Conv3 stride 2	1	256×16×16×16
	EnBlock4	GN,ReLU,Conv3,GN,ReLU,Conv3,+	4	256×16×16×16
Decoder1	DeUp3	Conv1,ConTranspose,+EnBlock3	1	128×32×32×32
	DeBlock3	GN,ReLU,Conv3,GN,ReLU,Conv3,+	1	128×32×32×32
	DeUp2	Conv1,ConTranspose,+EnBlock2	1	64×64×64×64
	DeBlock2	GN,ReLU,Conv3,GN,ReLU,Conv3,+	1	64×64×64×64
	DeUp2	Conv1,ConTranspose,+EnBlock1	1	32×128×128×128
	DeBlock1	GN,ReLU,Conv3,GN,ReLU,Conv3,+	1	32×128×128×128
	EndConv	Conv1	1	3×128×128×128
	Sigmoid	Sigmoid	1	3×128×128×128
Decoder2 (Used only during training)	DeUp3_1	Conv1,Upsampling,+EnBlock3	1	128×32×32×32
	DeBlock3_1	GN,ReLU,Conv3,GN,ReLU,Conv3,+	1	128×32×32×32
	DeUp2_1	Conv1,Upsampling,+EnBlock2	1	64×64×64×64
	DeBlock2_1	GN,ReLU,Conv3,GN,ReLU,Conv3,+	1	64×64×64×64
	DeUp2_1	Conv1,Upsampling,+EnBlock1	1	32×128×128×128
	DeBlock1_1	GN,ReLU,Conv3,GN,ReLU,Conv3,+	1	32×128×128×128
	EndConv_1	Conv1	1	3×128×128×128
	Sigmoid_1	Sigmoid	1	3×128×128×128

patient independently by subtracting the mean and dividing by the standard deviation of the brain region only.

Moreover, to prevent an overfitting issue from arising, we deploy three types of data augmentation. Firstly, we apply a random intensity shift between $[-0.1 - 0.1]$ of the standard deviation of each channel, as well as a random scaling intensity of the input between scales $[0.9 - 1.1]$. Secondly, we train our network by randomly cropping the MRI data from 240×240×155 voxels to 128×128×128 voxels due to memory limitation. Finally, we use random flipping along each 3D axis with a probability of 50%.

3.2 Training Details

The implementation of our network is based on PyTorch 1.1.0 [16]. The maximum number of training iterations is set to 405 epochs with 5 epochs of linear warmup. We use Adam optimizer to update the weights of the network, with a

batch size of 1 and an initial learning rate of $\alpha_0 = 1e-4$ at the very beginning and decays it as following:

$$\alpha = \alpha_0 \times \left(1 - \frac{e}{N_e}\right)^{0.9} \tag{3}$$

where e is an epoch counter, and N_e is a total number of epochs . We regularize using an l2 weight decay of $1e-5$. Training is performed on a Nvidia Titan V GPU with 12 Gb memory. However, our method requires slightly more than 12 Gb memory in our experiment. We utilize gradient checkpointing[21] by PyTorch to reduce the memory consumption.

3.3 Augmentation for Inference

At testing time, we segment the whole brain region at once instead of using a sliding window. The interpolation decoder is not used during the inference phase. To obtain a more robust prediction, we preserve eight weights of the model in the last time of the training progress for prediction. For each snapshot, the input images are used different flipping before being fed into the network. Finally, we average the output of the resulting eight segmentation probability maps.

3.4 Post-processing

We replace enhancing tumor with necrosis when the volume of predicted enhancing tumor is less than the threshold to post-process our segmentation results (The threshold is chosen for each experiment independently, depending on the performance of BraTS 2019 validation dataset).

4 Results

The variability of a single model can be quite high. We use total five networks from the 5-fold cross-validation as an ensemble to predict segmentation for BraTS 2019 validation dataset. Also, we use an ensemble of a set of 12 models, which are trained from scratch using the entire training dataset. The best single model is chosen from the set of 12 models.

We report the results of our approach on the BraTS 2019 validation dataset, which contains 125 cases with unknown glioma grade and unknown segmentation. All reported values are computed via the online evaluation platform (https://ipp.cbica.upenn.edu/) for evaluation of Dice score, sensitivity, specificity and Hausdorff distance (95%). Validation set results can be found in Table 3. The performance of the best single model is slightly better than ensemble of 5-fold cross-validation. The ensemble of 12 models results in a minor improvement compared with the best single model.

Testing set results are presented in Table 4. Our algorithm achieved the first place out of more than 70 participating teams.

Table 3. Mean Dice and Hausdorff measurements of the proposed segmentation method on BraTS 2019 validation set. DSC - dice similarity coefficient, HD95 - Hausdorff distance (95%), WT - whole tumor, TC - tumor core, ET - enhancing tumor core.

Method	DSC			HD95		
Validation	WT	TC	ET	WT	TC	ET
Ensemble of 5-fold	0.90797	0.85888	0.79667	4.35413	5.69195	3.12642
Best Single Model	0.90819	0.86321	0.80199	4.44375	5.86201	3.20551
Ensemble of 12 Models	0.90941	0.86473	0.80211	4.26398	5.43931	3.14581

Table 4. Mean Dice and Hausdorff measurements of the proposed segmentation method on BraTS 2019 testing set. DSC - dice similarity coefficient, HD95 - Hausdorff distance (95%), WT - whole tumor, TC - tumor core, ET - enhancing tumor core.

Method	DSC			HD95		
Testing	WT	TC	ET	WT	TC	ET
Ensemble of 12 Models	0.88796	0.83697	0.83267	4.61809	4.13071	2.65056

5 Conclusion

In this paper, we propose a two-stage cascaded U-Net. Our approach refines the prediction through a progressive cascaded network. Experiments on the BraTS 2019 validation set demonstrate that our method can obtain very competitive segmentation even though using single model. The testing results show that our proposed method can achieve excellent performance, winning the first position in the BraTS 2019 challenge segmentation task among 70+ participating teams.

Acknowledgements. 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. 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-GBM collection. the cancer imaging archive (2017) (2017). https://doi.org/10.7937/K9/TCIA.2017.KLXWJJ1Q
2. 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 **286** (2017)

3. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J.S., Freymann, J.B., Farahani, K., Davatzikos, C.: Advancing the cancer genome atlas glioma MRI collections with expert segmentation labels and radiomic features. Scientific data **4**, 170117 (2017). https://doi.org/10.1038/sdata.2017.117
4. Bakas, S., Reyes, M., Jakab, A., Bauer, S., Rempfler, M., Crimi, A., Shinohara, R.T., Berger, C., Ha, S.M., Rozycki, 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 (2018)
5. Doersch, C.: Tutorial on variational autoencoders. arXiv preprint arXiv:1606.05908 (2016)
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.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778 (2016)
8. He, K., Zhang, X., Ren, S., Sun, J.: Identity mappings in deep residual networks. In: European conference on computer vision. pp. 630–645. Springer (2016)
9. Kamnitsas, K., Bai, W., Ferrante, E., McDonagh, S., Sinclair, M., Pawlowski, N., Rajchl, M., Lee, M., Kainz, B., Rueckert, D., et al.: Ensembles of multiple models and architectures for robust brain tumour segmentation. In: International MICCAI Brainlesion Workshop. pp. 450–462. Springer (2017)
10. Kamnitsas, K., Ledig, C., Newcombe, V.F., 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)
11. Kingma, D.P., Welling, M.: Auto-encoding variational bayes. arXiv preprint arXiv:1312.6114 (2013)
12. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 3431–3440 (2015)
13. 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). https://doi.org/10.1109/tmi.2014.2377694
14. Milletari, F., Navab, N., Ahmadi, S.A.: V-net: Fully convolutional neural networks for volumetric medical image segmentation. In: 2016 Fourth International Conference on 3D Vision (3DV). pp. 565–571. IEEE (2016)
15. Myronenko, A.: 3D MRI brain tumor segmentation using autoencoder regularization. In: International MICCAI Brainlesion Workshop. pp. 311–320. Springer (2018)
16. Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., Lin, Z., Desmaison, A., Antiga, L., Lerer, A.: Automatic differentiation in pytorch (2017)
17. Pereira, S., Pinto, A., Alves, V., Silva, C.A.: Brain tumor segmentation using convolutional neural networks in mri images. IEEE transactions on medical imaging **35**(5), 1240–1251 (2016)
18. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. pp. 234–241. Springer (2015)

19. Roth, H.R., Shen, C., Oda, H., Sugino, T., Oda, M., Hayashi, Y., Misawa, K., Mori, K.: A multi-scale pyramid of 3d fully convolutional networks for abdominal multi-organ segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 417–425. Springer (2018)
20. Shen, H., Wang, R., Zhang, J., McKenna, S.J.: Boundary-aware fully convolutional network for brain tumor segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 433–441. Springer (2017)
21. Siskind, J.M., Pearlmutter, B.A.: Divide-and-conquer checkpointing for arbitrary programs with no user annotation. Optimization Methods and Software **33**(4-6), 1288–1330 (2018)
22. Tu, Z., Bai, X.: Auto-context and its application to high-level vision tasks and 3d brain image segmentation. IEEE transactions on pattern analysis and machine intelligence **32**(10), 1744–1757 (2009)
23. Wu, Y., He, K.: Group normalization. In: Proceedings of the European Conference on Computer Vision (ECCV). pp. 3–19 (2018)
24. Zhao, X., Wu, Y., Song, G., Li, Z., Zhang, Y., Fan, Y.: A deep learning model integrating fcnns and crfs for brain tumor segmentation. Medical image analysis **43**, 98–111 (2018)
25. Zhou, C., Chen, S., Ding, C., Tao, D.: Learning contextual and attentive information for brain tumor segmentation. In: International MICCAI Brainlesion Workshop. pp. 497–507. Springer (2018)
26. Zhou, C., Ding, C., Lu, Z., Wang, X., Tao, D.: One-pass multi-task convolutional neural networks for efficient brain tumor segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 637–645. Springer (2018)
27. Zhou, C., Ding, C., Wang, X., Lu, Z., Tao, D.: One-pass multi-task networks with cross-task guided attention for brain tumor segmentation. arXiv preprint arXiv:1906.01796 (2019)