



Brain Tumor Segmentation on Multimodal MR Imaging Using Multi-level Upsampling in Decoder

Yan Hu¹, Xiang Liu², Xin Wen², Chen Niu², and Yong Xia¹(✉)

¹ National Engineering Laboratory for Integrated Aero-Space-Ground-Ocean Big Data Application Technology, School of Computer Science and Engineering, Northwestern Polytechnical University, Xi'an 710072, People's Republic of China
yxia@nwpu.edu.cn

² The First Affiliated Hospital of Xi'an Jiao Tong University, Xi'an 710061, People's Republic of China

Abstract. Accurate brain tumor segmentation plays a pivotal role in clinical practice and research settings. In this paper, we propose the multi-level up-sampling network (MU-Net) to learn the image presentations of transverse, sagittal and coronal view and fuse them to automatically segment brain tumors, including necrosis, edema, non-enhancing, and enhancing tumor, in multimodal magnetic resonance (MR) sequences. The MU-Net model has an encoder-decoder structure, in which low level feature maps obtained by the encoder and high level feature maps obtained by the decoder are combined by using a newly designed global attention (GA) module. The proposed model has been evaluated on the BraTS 2018 Challenge validation dataset and achieved an average Dice similarity coefficient of 0.88, 0.74, 0.69 and 0.85, 0.72, 0.66 for the whole tumor, core tumor and enhancing tumor on the validation dataset and testing dataset, respectively. Our results indicate that the proposed model has a promising performance in automated brain tumor segmentation.

Keywords: Magnetic resonance imaging · Brain tumor segmentation · Encoder-decoder · Multi-level upsampling · Global attention

1 Introduction

Glioma is a type of tumors that starts in the glial cells of the brain or the spin, comprising about 30% of all brain tumors and central nervous system tumors, and 80% of all malignant brain tumors [1]. Shape and localization of tumors are crucial for diagnosis, treatment planning and follow-up observation in clinical, while the manual segmentation of brain tumor in magnetic resonance (MR) images requires a high degree of skills and concentration, and is time-consuming, expensive and prone to operator bias. Thus, a fully automated and reliable segmentation algorithm is of great significance. However, despite considerable research efforts being devoted to this task [2], automated segmentation of brain tumors remains a challenge, largely due to the variable shapes and locations, diffusion and poor contrast of brain tissues in MR images.

In recent years, deep learning techniques, especially deep convolutional neural networks (DCNNs), have led to significant breakthroughs in computer vision, since they provide an ‘end-to-end’ framework for simultaneous presentation learning and image segmentation and thus free users from the troublesome extraction of handcrafted features. Such breakthroughs have prompted many researchers to use DCNNs for brain tumor segmentation. The solutions published in the literature can be roughly divided into two groups. One group of solutions are based on the classification of image patches. Pereira et al. [3] designed an 11-layer CNN and a 9-layer CNN to classify the patches extracted from high grade gliomas (HGG) and low grade gliomas (LGG), respectively. To simultaneously learn the presentation of both fine details and coarse structures from input images, Zhao et al. [4] proposed a three-convolutional-pathway network, in which the input patches for three pathways have a size of 48×48 , 28×28 and 12×12 , respectively, and concatenated these three outputs for classification. Kamnitsas et al. [5] adopted a 3D CNN architecture, i.e. DeepMedic, with multiple input image resolutions, residual connections and fully connected conditional random field. Castillo et al. [6] developed a neural network with four contracting pathways and residual connections that receive patches centered on the same voxel, but with different spatial resolutions. Lopez et al. [7] removed max pooling layers in dilated residual network [8] to avoid loss of upsampling the prediction by interpolation, but at the same time enlarge the receptive field through dilated convolutional operations. McKinley et al. [9] also replaced max pooling layers by dilated convolutions without influencing the receptive field of the classifier in Densenet. The other group of solutions are based on fully convolutional networks (FCNs). Pereira et al. [3] employed two U-Nets, one for the localization of tumors and the other for the segmentation of intra-tumor structures. Li et al. [10] used three parallel end-to-end networks for three views and generated the segmentation results using majority voting. Kamnitsas et al. [11] trained seven end-to-end networks and used ensemble learning to produce robust segmentation results. Wang et al. [12] proposed a cascade of fully convolutional neural networks to decompose the multi-class segmentation problem into a sequence of three binary segmentation problems according to the subregion hierarchy. In our previous work [13], we used a cascaded U-Net model and a patch-wise CNN to detect and segment brain tumors.

In this paper, we propose a FCN called the multi-level upsampling network (MU-Net) to segment brain tumor structures, including necrosis, edema and enhancing tumor from multimodality MR. Our main contributes are: (a) we designed a global attention (GA) module to combine the low level feature from encoder and high level feature from decoder; (b) we designed a multi-level decoding architecture. The proposed algorithm has been evaluated on the BraTS 2018 Challenge validation dataset and achieved a promising result.

2 Dataset

The proposed MU-Net model was evaluated on the Brain Tumor Segmentation 2018 (BraTS 2018) Challenge dataset [14–16]. There are 285 cases for training, including 210 HGG and 75 LGG cases. Each case has four multimodal MR scans, including the

T1, T1c, T2, and FLAIR. All these scans were co-registered to the same anatomical template, interpolated to the same dimension of $240 \times 240 \times 155$ and the same voxel size of $1.0 \times 1.0 \times 1.0 \text{ mm}^3$ and skull-stripped. Each case has been segmented manually, by up to four raters, following the same annotation protocol, and their annotations were approved by experienced neuro-radiologists. Annotations of tumor tissues comprise the enhancing tumor (ET-label 4), the peritumoral edema (ED-label 2), and the necrotic and non-enhancing tumor core (NCR/NET-label 1). The validation and testing datasets consist of 66 and 191 cases, respectively, but their grade and ground truth are unseen.

3 Methods

The 3D brain MR sequences are resliced from three views, transverse, sagittal and coronal respectively. Three probability maps of these three views are learned by three identical MU-Nets, respectively, and concatenated together as the input of a multi-view fusion network. The pipeline of proposed algorithm is shown in Fig. 1.

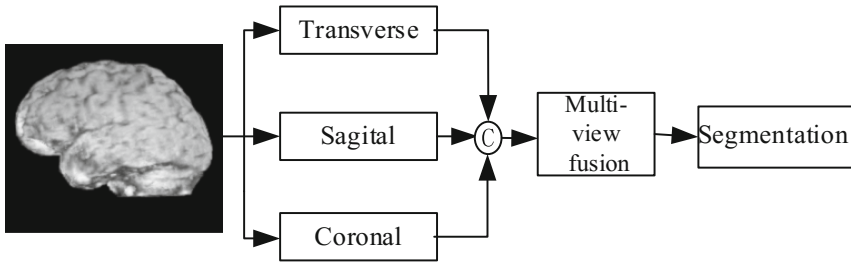


Fig. 1. Pipeline of proposed algorithm.

3.1 MU-Net

The proposed MU-Net model adopts the encoder-decoder structure, consisting of five convolutional blocks, a spatial pyramid pooling (SPP) module [17], five global attention (GA) modules, and nine upsampling feature (UF) modules. The architecture of this model is shown in Fig. 2.

The encoder branch is a variants of ResNet-101. The convolutional layer with $64 \times 7 \times 7$ kernels and a stride of 2 in the root block (i.e. Block 1) is replaced with five convolutional layers, each consisting $64 \times 3 \times 3$ kernels. The stride of the third convolutional layer is 2, and the stride of other convolutional layers is 1. Other blocks in this branch is the same as those in ResNet-101 [18].

Between the encoder and decoder, we add a SPP module, in which there are five parallel operators, including three 3×3 dilated convolution with a dilation rate of 6, 12, and 18, respectively, a 1×1 convolution and a global pooling (see Fig. 3(a)). The input of the SPP module is processed by these operators simultaneously, and the feature maps generated by these operators are concatenated as the output of the SPP module.

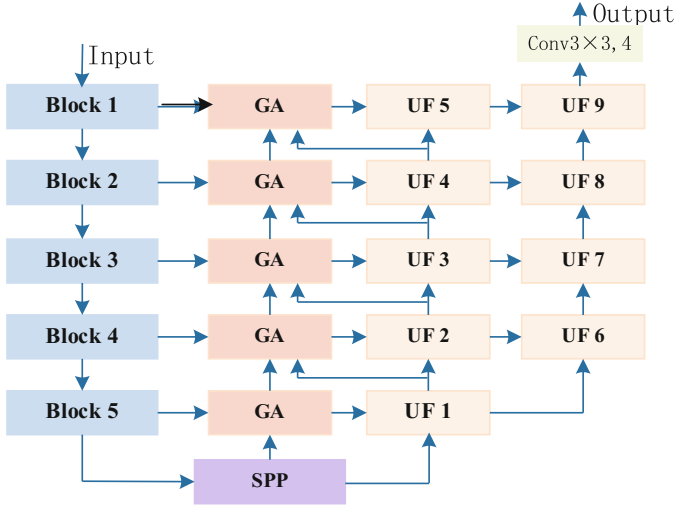


Fig. 2. Architecture of the proposed MU-Net model

The major part of the decoder branch contains five decode modules (i.e. UF 1 – UF 5), which are designed to recover the size of feature maps. Usually, there are two 3×3 convolutions and a bilinear interpolation between them in each UF module (see Fig. 3 (c)). However, since there is no down-sampling operation in the encoder block 3–5, the interpolation operation is omitted in UF 1, UF 2, and UF 5 modules such that the output feature maps have the same size as the input of MU-Net. Meanwhile, to combine low-level feature maps and high-level feature maps in the decoding process, we add five GA modules to the MU-Net model. Each GA module takes two groups of inputs - low-level feature maps from the corresponding encoder block and high-level feature maps from the UF module at the previous level. Two 3×3 convolutions are applied to low-level feature maps, respectively. High-level feature maps are also processed by two operations – one is the global average pooling followed by a 1×1 convolution as, and the other is a 3×3 convolution. The processed high-level feature maps are then used as the element-wise weighting mask of the processed low-level feature maps (see Fig. 3(b)). In addition, the output of each of UF 2 – UF 5 are fed simultaneously to the UF module (UF 6 – UF 9) at the next level. Eventually, the output of the UF 6 and the output of UF1 are concatenated and fed to a 3×3 convolution another UF module to produce the segmentation results.

3.2 Multi-view Fusion

Three views are fused by a shallow encoder-decoder network. The encoder consists of three convolutional layers with 64, 128 and $256 \times 3 \times 3$ kernels, followed by three max pooling layers respectively. The decoder comprises three deconvolutional layers with 256, 128 and 64 kernels of size 3×3 . Then, we convolve the output of the decoder by four 3×3 kernels and predict by max possibility.

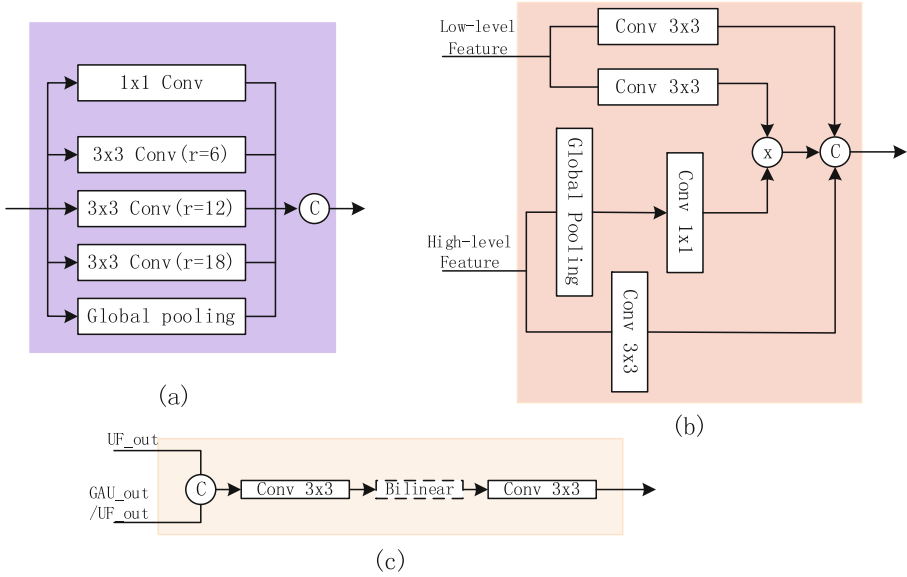


Fig. 3. Architecture of modules used in segmentation model. (a) shows the SPP module; (b) shows GA module; (c) shows the UF module.

3.3 Implementation

With the proposed MU-Net model, brain tumor segmentation can be performed on a slice-by-slice basis. The slices in each training dataset were cropped and padded to 224×224 , 224×160 , 224×160 for transverse, sagittal, and coronal view, respectively, and the voxel values of each modality were normalized by the min-max normalization. The encoding branch was initialized by the pre-trained ResNet-101 [19]. The positive slices (with tumor) and negative slices (without tumor) were randomly selected at a rate of 5:1. The cross entropy was used as the loss function, and the adaptive moment estimator (Adam) with an exponentially descending learning rate of 0.001–0.00001 was adopted as the optimizer. It took about twenty hours to train each MU-Net model with a batch size of 8 and epochs of 30 on two GPUs (NVIDIA 1080 Ti, 12 GB RAM) four hours to train the fusion network with a batch size of 16 and epochs of 20.

4 Experiments and Results

Following the request of the challenge, four intra-tumor structures have been grouped into three mutually inclusive tumor regions: (a) whole tumor (WT) that consists of all tumor tissues, (b) tumor core (TC) that consists of the enhancing tumor and necrotic and non-enhancing tumor core, and (c) enhancing tumor (ET). The performance of segmenting each tumor region was quantitatively evaluated through an online system

by using three metrics, including the average Dice similarity coefficient, sensitivity and Hausdorff distance.

Preliminary results for the BraTS 2018 Training dataset have been obtained by hold-out using 80% of the data (228 cases) for training and the remaining 20% for validation (57 cases). Table 1 shows the quantitative evaluation and Fig. 4 presents some examples of the predictions against the ground truth on predicted cases from BraTS 2018 training data. It appears that this proposed model works well when the edge is relatively smooth, as the first three examples shown in Fig. 4. However, similarly to other semantic image segmentation task, our deep model works weakly on pixels distributed near the edge as th last two examples shown in Fig. 4. Tables 2 and 3 give the quantitative evaluation of our algorithm on 66 validation and 191 testing unseen subjects. We can observe that performance on training data, validation data and testing data are consistent, which indicates that this model generalizes well to unseen examples. Figure 5 shows the visualization of segmentation result from validation dataset.

Table 1. Quantitative result of validation on BraTS 2018 training set.

	Dice			Sensitivity			Hausdorf-95		
	ET	WT	TC	ET	WT	TC	ET	WT	TC
Mean	0.61	0.83	0.73	0.83	0.89	0.75	41.48	47.23	41.14
StdDev	0.27	0.11	0.17	0.17	0.09	0.21	37.49	23.49	28.90
Median	0.72	0.86	0.77	0.88	0.91	0.82	41.69	46.70	44.77
25quantile	0.50	0.78	0.64	0.80	0.85	0.62	5.12	30.36	12.37
75quantile	0.80	0.90	0.86	0.95	0.96	0.93	61.25	59.84	59.67

5 Discussion

5.1 Multi-level Upsampling

To demonstrate the performance improvement resulted from using the GA module, we trained a similar network but without using multi-level upsampling on the BraTS 2018 training dataset and tested it on the validation dataset. Table 4 gives the performance of both models measured by the average Dice similarity coefficient, sensitivity, specificity and Hausdorff-95. It reveals that multi-level upsampling connection is able to improve the performance.

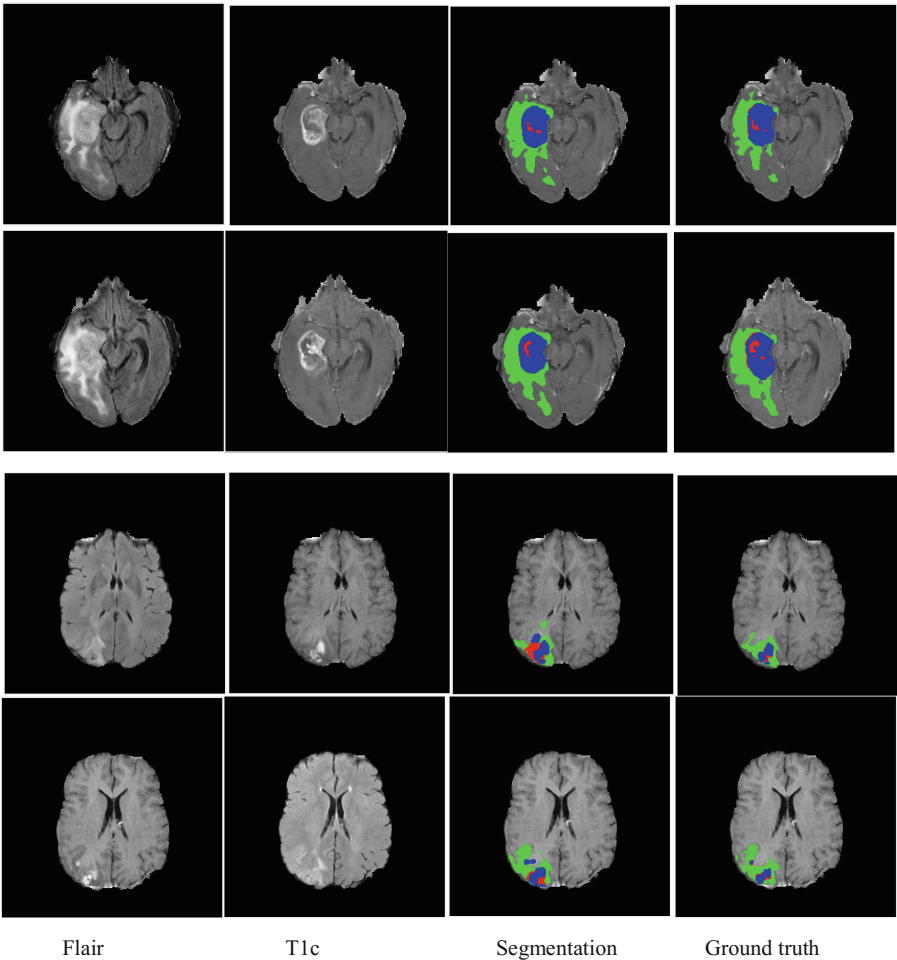


Fig. 4. Segmentation examples from the validation set. From top to bottom: the 55th, 57th, slices from the subject Brats18_TCIA01_147_1 and the 55th, 57th slices from the subject Brats18_TCIA10_629_1. Red - NCR&NET, Blue - ET, Green – ED. (Color figure online)

Table 2. Quantitative result on BraTS 2018 validation set.

	Dice			Sensitivity			Hausdorf-95		
	ET	WT	TC	ET	WT	TC	ET	WT	TC
Mean	0.69	0.88	0.74	0.71	0.87	0.77	6.69	4.76	10.67
StdDev	0.27	0.10	0.24	0.28	0.15	0.25	12.43	4.04	9.87
Median	0.80	0.91	0.84	0.82	0.93	0.88	2.83	3.00	6.78
25quantile	0.66	0.88	0.69	0.63	0.84	0.68	1.73	2.24	4.36
75quantile	0.86	0.94	0.90	0.90	0.95	0.93	5.39	5.74	14.73

Table 3. Quantitative result on BraTS 2018 testing set.

	Dice			Hausdorf-95		
	ET	WT	TC	ET	WT	TC
Mean	0.66	0.85	0.72	5.94	6.29	9.04
StdDev	0.29	0.15	0.27	9.04	9.25	11.47
Median	0.77	0.90	0.82	2.83	4.00	5.39
25quantile	0.60	0.84	0.65	2.00	2.34	3.16
75quantile	0.85	0.93	0.90	5.15	6.28	11.32

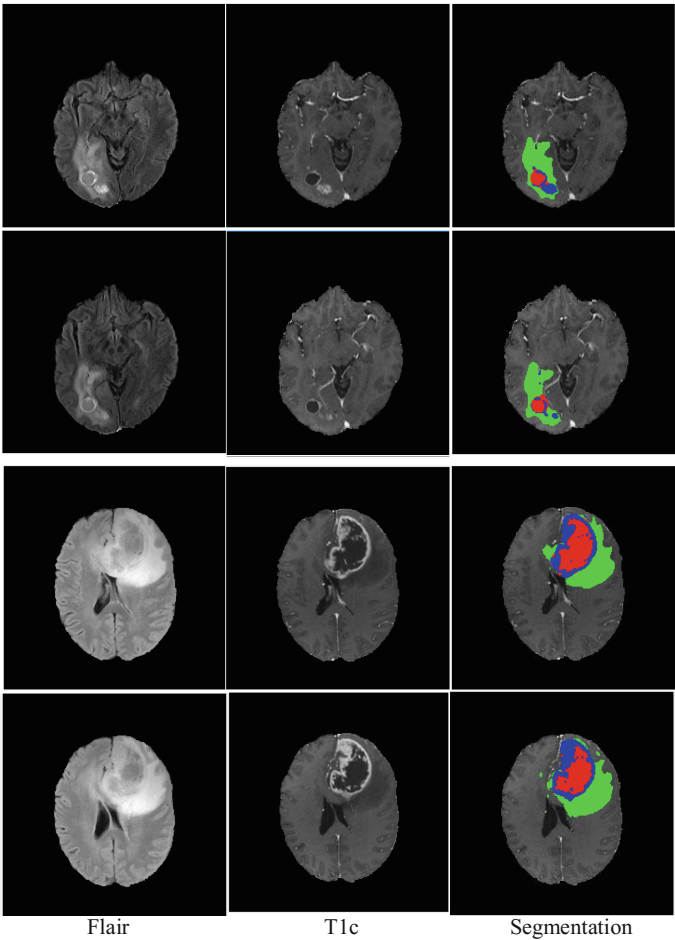


Fig. 5. Segmentation examples from the validation set. From top to bottom: the 54th and 57th slices from the subject Brats18_CBICA_ANK_1 and 87th and 91th, slices from the subject Brats18_CBICA_ANK_1. Red - NCR&NET, Blue - ET, Green – ED (Color figure online)

Table 4. Comparison of model with multi-level upsampling and without multi-level upsampling on BraTS 2018 validation dataset

	Dice			Sensitivity			Hausdorf-95		
	ET	WT	TC	ET	WT	TC	ET	WT	TC
With MU	0.69	0.88	0.74	0.71	0.87	0.77	6.69	4.76	10.67
Without MU	0.65	0.84	0.69	0.74	0.84	0.70	29.82	34.51	38.73

6 Conclusion

In this paper, we proposed a novel end-to-end segmentation model called MU-Net to segment brain tumors and their intra structures from multimodal MR scans, which learns the presentation of MR scans in transverse, sagittal and coronal views and fused them through a convolutional neural network for image segmentation. This model has been evaluated on the BraTS 2018 Challenge online system and achieved an average Dice similarity coefficient of 0.88, 0.74, 0.69 and 0.85, 0.72, 0.66 for whole tumor, core tumor, and enhancing tumor on the validation dataset and testing dataset, respectively.

Acknowledgement. This work was supported in part by the National Natural Science Foundation of China under Grants 61471297 and 61771397.

References

1. Goodenberger, M.L., Jenkins, R.B.: Genetics of adult glioma. *Cancer Genet.* **205**, 613–621 (2012). <https://doi.org/10.1016/j.cancergen.2012.10.009>
2. Menze, B.H., et al.: The multimodal brain tumor image segmentation benchmark (BRATS). *IEEE Trans. Med. Imaging* **34**, 1993–2024 (2015)
3. Pereira, S., Oliveira, A., Alves, V., Silva, C.A.: On hierarchical brain tumor segmentation in MRI using fully convolutional neural networks: a preliminary study. In: 2017 IEEE 5th Portuguese Meeting on Bioengineering (ENBENG), pp. 1–4. IEEE (2017)
4. Zhao, L., Jia, K.: Multiscale CNNs for brain tumor segmentation and diagnosis. *Comput. Math. Methods Med.* **2016** (2016)
5. Kamnitsas, K., et al.: 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*, pp. 138–149. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-55524-9_14
6. Castillo, L.S., Daza, L.A., Rivera, L.C., Arbeláez, P.: Brain Tumor segmentation and parsing on MRIs using multiresolution neural networks. In: Crimi, A., Bakas, S., Kuijf, H., Menze, B., Reyes, M. (eds.) *BrainLes 2017. LNCS*, vol. 10670, pp. 332–343. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-75238-9_29
7. Moreno Lopez, M., Ventura, J.: Dilated convolutions for brain tumor segmentation in MRI scans. In: Crimi, A., Bakas, S., Kuijf, H., Menze, B., Reyes, M. (eds.) *BrainLes 2017. LNCS*, vol. 10670, pp. 253–262. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-75238-9_22

8. Yu, F., Koltun, V., Funkhouser, T.A.: Dilated residual networks. In: Computer Vision and Pattern Recognition, pp. 636–644 (2017)
9. McKinley, R., Jungo, A., Wiest, R., Reyes, M.: Pooling-free fully convolutional networks with dense skip connections for semantic segmentation, with application to brain tumor segmentation. In: Crimi, A., Bakas, S., Kuijff, H., Menze, B., Reyes, M. (eds.) BrainLes 2017. LNCS, vol. 10670, pp. 169–177. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-75238-9_15
10. Li, Y., Shen, L.: Deep learning based multimodal brain tumor diagnosis. In: Crimi, A., Bakas, S., Kuijff, H., Menze, B., Reyes, M. (eds.) BrainLes 2017. LNCS, vol. 10670, pp. 149–158. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-75238-9_13
11. 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
12. 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
13. Hu, Y., Xia, Y.: 3D deep neural network-based brain tumor segmentation using multimodality magnetic resonance sequences. In: Crimi, A., Bakas, S., Kuijff, H., Menze, B., Reyes, M. (eds.) BrainLes 2017. LNCS, vol. 10670, pp. 423–434. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-75238-9_36
14. Bakas, S.: Advancing the cancer genome atlas glioma MRI collections with expert segmentation labels and radiomic features. *Sci. Data* **4**, 170117 (2017)
15. Bakas, S., et al.: Segmentation labels and radiomic features for the pre-operative scans of the TCGA-GBM collection. The Cancer Imaging Archive (2017). <https://doi.org/10.7937/K9/TCIA.2017.KLXWJJ1Q>
16. Bakas, S., et al.: Segmentation labels and radiomic features for the preoperative scans of the TCGA-LGG collection. The Cancer Imaging Archive (2017). <https://doi.org/10.7937/K9/TCIA.2017.GJQ7R0EF>
17. Chen, L.C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H.: Encoder-decoder with atrous separable convolution for semantic image segmentation. arXiv preprint [arXiv:1802.02611](https://arxiv.org/abs/1802.02611) (2018)
18. 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. LNCS, vol. 9908, pp. 630–645. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46493-0_38
19. Pre-trained Resnet_v2_101 model. http://download.tensorflow.org/models/resnet_v2_101_2017_04_14.tar.gz
20. Bakas, S., Reyes, M., Jakab, A., Bauer 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)