








Ensemble of Fully Convolutional Neural Network for Brain Tumor Segmentation from Magnetic Resonance Images

Avinash Kori , Mehul Soni , B. Pranjali , Mahendra Khened, Varghese Alex , and Ganapathy Krishnamurthi 

Indian Institute of Technology Madras, Chennai 600036, India
gankrish@iitm.ac.in

Abstract. We utilize an ensemble of the fully convolutional neural networks (CNN) for segmentation of gliomas and its constituents from multimodal Magnetic Resonance Images (MRI). The ensemble comprises of 3 networks, two 3-D and one 2-D network. Of the 3 networks, 2 of them (one 2-D & one 3-D) utilize dense connectivity patterns while the other 3-D network makes use of the residual connection. Additionally, a 2-D fully convolutional semantic segmentation network was trained to distinguish between air, brain, and lesion in the slice and thereby localize the lesion the volume. Lesion localized by the above network was multiplied with the segmentation mask generated by the ensemble to reduce false positives. On the BraTS validation data ($n=66$), the scheme utilized in this manuscript achieved a whole tumor, tumor core and active tumor dice of 0.89 0.76, 0.76 respectively, while on the BraTS test data ($n=191$), our scheme achieved the whole tumor, tumor core and active tumor dice of 0.83 0.72, 0.69 respectively.

Keywords: Brain tumor · MRI · CNN · 3-D · Ensemble

1 Introduction

Manual tracing, detection of organs and tumor structure from medical images is considered as one of the preliminary step in diseases diagnosis and treatment planning. In a clinical setup this time-consuming process is carried out by radiologists, however, this approach becomes infeasible as the number of patients increases. This necessitates the scope of research in automated segmentation methods.

Diffused boundaries of the lesion and partial volume effects in the MR images makes automated segmentation of gliomas from MR volumes a challenging task. In the recent year's convolutional neural networks (CNN) have produced state of the art results for the task of segmentation of gliomas from MR images [6, 9]. Typically, medical images are volumetric, organs being imaged are 3-D entities and henceforth we exploit the nature of 3-D CNN based architectures for segmentation task.

The segmentation generated by a trained network has an associated bias and variance. Ensembling the predictions generated by multiple models or networks aids in the reduction of the variance in the generated segmentation. In this manuscript, we make use of 3 networks (two 3-D networks and one 2-D network) for the task of segmentation of gliomas from MR volumes. Additionally, a 2-D fully semantic segmentation network was trained to delineate the air, brain, and lesion in a slice of the brain. The aforementioned network was used to reduce the false positive generated by the ensemble. The predictions were further processed by conditional random fields (CRF) & 3-D connected components analysis.

2 Materials and Methods

An ensemble of fully convolutional neural network were utilized to segment gliomas and its constituents from multi modal MR volume. The ensemble comprises of 3 networks (two 3-D networks and one 2-D network). Two networks (a 3-D and a 2-D network) utilizes dense connectivity patterns while the other 3-D network comprises of residual connection. The networks with dense connectivity pattern were semantic segmentation networks and predicts the class associated with all pixels or voxels that form the input to the network. The network with residual connectivity pattern was composed of inception modules so as to learn multi-resolution features. This multi-resolution network unlike the other networks in the ensemble classifies only a subset of voxels.

A 2-D fully convolutional semantic segmentation (Air-Brain-Lesion Network) was trained to delineate air, brain and lesion from axial slice of the MR volumes and thereby localize the lesion in the volume. The predictions generated by the ensemble were smoothed by using Conditional random fields. The smoothed prediction and the output generated by the Air-Brain-Lesion network were used in tandem to reduce the false positives in the prediction. The false positives in the predictions were further reduced by incorporating a class-wise 3-D connected component analysis in the pipeline. The pipeline utilised for segmentation of glioma is illustrated in Fig. 1.

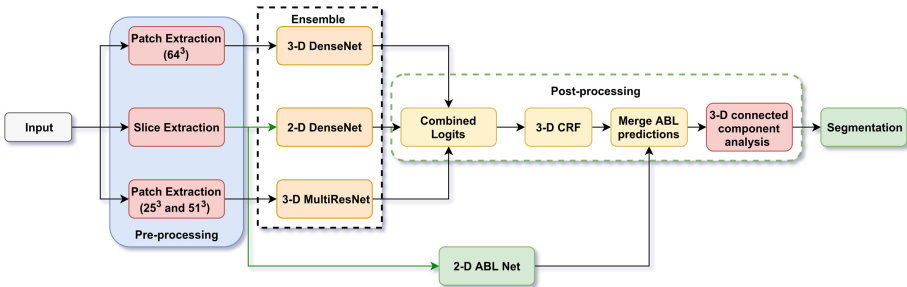


Fig. 1. Proposed pipeline for segmentation of Brain tumor and its constituents from Magnetic Resonance Images.

2.1 Data

Brats 2018 challenge data was used to train the networks [1–4, 8] was used in this manuscript for segmentation task. The training dataset comprises 210 high-grade glioma volumes and 75 low-grade gliomas along with expert annotated pixel level ground truth segmentation mask. Each subject comprises 4 MR sequences, namely FLAIR, T2, T1, T1 post contrast.

2.2 Data Pre-processing

As a part of pre-processing, the volumes were normalized to have zero mean and unit standard deviation.

2.3 Segmentation Network

The 3-D networks used in ensemble accepts 3-D patches as input while the 2-D network accepts an axial slice of the brain as the input. The architecture, training and testing regime associated with each network in the ensemble is explained in the following paragraphs.

3-D Densely Connected Semantic Segmentation Network

Architecture: The network is a fully convolutional semantic segmentation network. The network accepts input cubes of size 64^3 and predicts the class associated with all the voxels in the input cube fed to the network. The network is composed of an encoding and decoding section. The encoding section is composed of Dense blocks and Transition Down blocks. The Dense blocks are composed of a series of convolutions followed by non-linearity (ReLU) & each convolutional layer receives input from all the preceding convolutional layers in the block. This connectivity pattern leads to the explosion of a number of feature maps with the depth of the network which was circumvented by setting the number of output feature maps per convolutional layer to a small value ($k = 4$). The Transition down blocks are utilized in the network to reduce the spatial dimension of the feature maps.

The decoding or the up-sampling pathway in the network comprises of the Dense blocks and Transition Up blocks. The Transition Up blocks are composed of transposed convolution layers to upsample feature maps. The features from the encoding section of the network are concatenated with the up-sampled feature maps to form the input to the Dense block in the decoding section. The architecture of the network is given in Fig. 2.

Patch Extraction: Patches of size 64^3 were extracted from the brain. The class imbalance among the various classes in the data was addressed by extracting relatively more number of patches from lesser frequent classes such as necrosis. Figure 3 illustrates the number of patches extracted for each class.

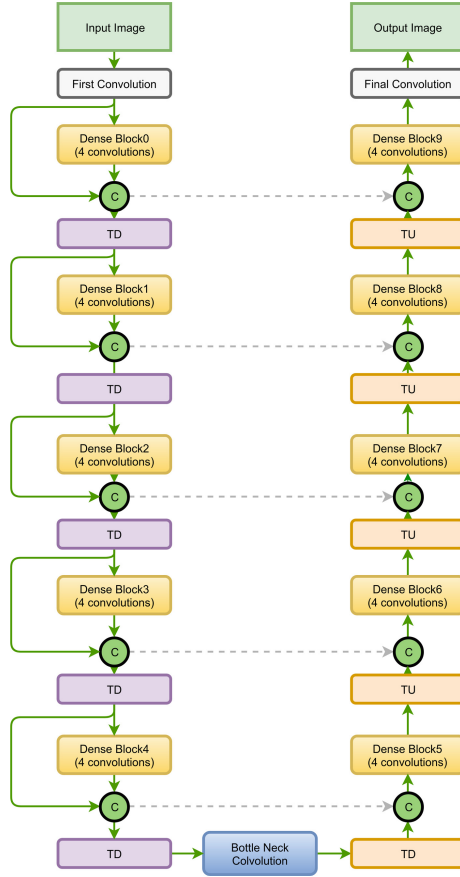


Fig. 2. Densely connected convolutional network used for segmentation task. TU: Transition Up block; TD: Transition Down block; C: Concatenation block

The 3-D dense fully connected network accepts an input of dimension 64^3 and predicts the class associated to all the voxels in the input. The network comprises 77 layers. The dense connection between the various convolutional layers in the network aids in the effective reuse of the features in the network. The presence of dense connections between layers increases the number of computations. This bottleneck was circumvented by keeping the number of convolutions to a small number say 4. Figure 2 shows the network architecture used in semantic segmentation task.

Training: Stratified sampling based on the grade of the gliomas was done to split the dataset into training, validation, and testing in the ratio 70: 20: 10. The network was trained and validated on 182 and 63 HGG & LGG volumes respectively. To further address the issue of class imbalance in the network, the parameters of the network were trained by minimizing weighted cross entropy.

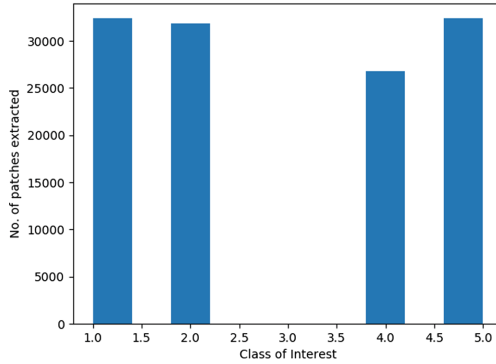


Fig. 3. Histogram of patches sampled surrounding certain class

The weight associated with each class was equivalent to the ratio of the median of the class frequency to the frequency of the class of interest [5]. The number of samples per batch was set at 4, while the learning rate was initialized to 0.0001 and decayed by a factor of 10% every-time the validation loss plateaued.

Testing: During inference, patches of the dimension of 64^3 were extracted from the volume and fed to the network with the stride of 32. CNN's being a deterministic technique is bound to generate predict the presence of the lesion in physiologically impossible place.

2-D Semantic Segmentation Network

Architecture: The architecture of this network is similar to that of the architecture of the 3-D network. The only difference between the networks is the usage of 2-D convolutions rather than 3-D convolutions. The network comprises 77 layers. The network accepts inputs of dimension 240×240 and predicts the class associated with all the pixels in the input.

Slice Extraction: In the given dataset, apart from the T1 post contrast, sequences such as FLAIR, T2 & T1 were 2-D sequences. Majority of the 2-D sequences in the given dataset were acquired axially and thus had good resolution along the axial plane. The 2-D network was trained on the axial slices of brain. The class imbalance in the dataset was addressed by extracting slices which comprise of at least one pixel of the lesion in it.

Training: The parameters of the network were initialized using Xavier initialization and the parameters of the network were learned by reducing the hybrid loss (cross entropy & dice loss). The imbalance among the various classes was further reduced by using weighted cross entropy rather than vanilla cross entropy. The weights assigned to each class were determined as explained earlier. Hyper-parameters such as batch size, learning rate, and learning rate decay etc. were similar to the ones used to train the 3-D network.

Testing: During inference, axial slices from the 3-D volume were fed to the trained network to generate the segmentation maps.

3-D Multi-resolution Segmentation Network

Architecture: The architecture comprises of the two pathways viz high-resolution pathway and low resolution like [6]. 3-D patches of size 25^3 were input to the high-resolution pathway while 51^3 resized to 19^3 were input to the low-resolution path in the network. The network predicts the class of the center 9^3 voxels of the input. The feature maps in the low resolution pathway were upsampled using transposed convolutions, to match the dimension with the feature maps from high-resolution path. This network, unlike the previously explained two other networks, differs by:

1. Predicting the class associated to a subset of voxels in the input 3-D patch.
2. Making use of dual pathway to captures associated global and local features.
3. Making use of inception module [10] (3×3 , 5×5 & 7×7) so as to learn multi-resolution features.

The architecture of the network is given in Fig. 4(a) and the building block of each unit in the network is illustrated in Fig. 4(b).

Patch Extraction: Patches of sizes 25^3 and 51^3 centered around voxels were extracted to form the training data to the network. The degree of class imbalance was reduced by extracting more patches from under-represented classes.

Training: Parameters in the network were initialized with Xavier initialization technique. The network was trained using the similar hyper-parameters that were used for the other two other networks proposed in the ensemble. The network was trained for 50 epochs and model that yielded lowest validation error was utilized for inference.

Testing: For testing, the stride was set to 9^3 and patches of 25^3 and 51^3 were extracted from the MR volume and input to the trained network to produce the segmentation mask.

2.4 Post-processing

Air-Brain-Lesion Network. The Air-Brain-Lesion (ABL Net) network was 2-D network densely connected the fully convolutional network. The network was trained to delineate lesion, air and the brain in a volume. The prediction made by this network was used to reduce the false positives generated by the segmentation network.

Architecture: The architecture of the network is similar to the 2-D network utilized in the segmentation ensemble model.

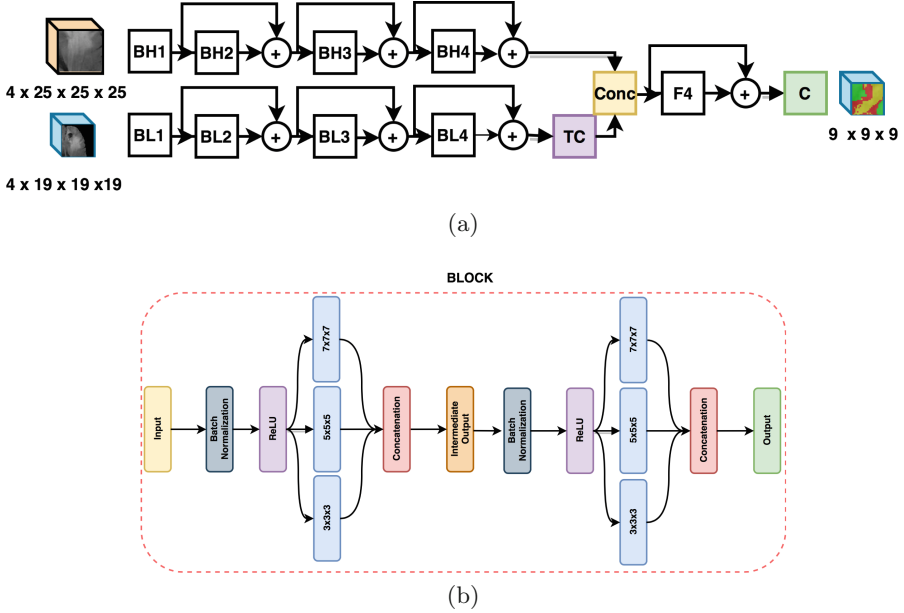


Fig. 4. 3-D Multi-Resolution Network for segmentation of gliomas from MR volumes. (a) The architecture of the network. The top portion of the network accepts high-resolution patches (25^3) while the bottom pathway accepts low-resolution input (51^3 patches resized to 19^3) as input. Both the high and low-resolution pathway is composed to inception modules so as learn multi-resolution features. TC in the network stands for transposed convolution and is used to match the spatial dimension of the features in low-resolution pathway with those learned in the high-resolution path. (b) The building block of the network. In the block, the dimension of the feature map in an inception module was maintained by setting the padding to 0, 1, 2 for 3×3 , 5×5 & 7×7 respectively.

Slice Extraction: The Network was trained using axial slices as they correspond to the highest resolution. Various constituents of the lesion were clubbed to form the lesion while air and brain class labels were determined using a threshold on the volume Fig. 5 illustrates the slice of the brain with the aforementioned classes.

Training and Testing: The training & testing regime were similar to the ones used for the 2-D Densely connected segmentation network.

CRF. To the smoothen the segmentation predicted by the models a fully connected conditional random fields with Gaussian edge potentials as proposed by Krähenbühl et al. [7] was utilized. The posterior probabilities generated by each model in the ensemble were averaged to form the unary potentials for the CRF. The CRF was implemented by using open source code from the pydensecrf¹. The

¹ pydensecrf: <https://github.com/lucasb-eyer/pydensecrf>.

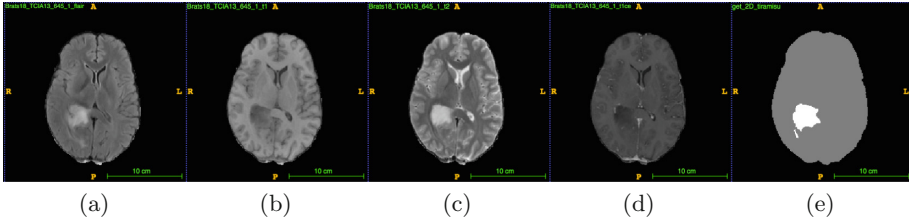


Fig. 5. Data for training the Air-Brain-Lesion network. (a) FLAIR, (b) T1, (c) T2, (d) T1ce, (e) Modified Ground truth. In image (e), black, gray and white represent Air, Brain and lesion respectively.

output obtained after smoothening using CRF and the output predicted from air-brain-lesion model were multiplied to reduce false positives in the generated segmentation mask.

Connected Components. False positives in the segmentation mask were further reduced by performing class-wise 3-D connected component analysis. All components within each class which composed more than 12,000 voxels were retained while the rest were discarded.

3 Results

The performance of the network was tested on 3 different namely: held out test data ($n = 40$), BraTS validation data ($n = 66$) & BraTS testing data ($n = 191$) (Table 1).

3.1 Performance of the Segmentation Networks on the Held Out Test Data

On the held out test data ($n = 40$), the performance of each of the network in the segmentation ensemble is given in Table 2(a, b, c). Table 2(d) showcases the performance on the held out test data post ensembling the networks. Comparing the whole tumor, tumor core and active tumor core dice score it was observed that ensembling of networks aided in reducing the variance and increasing the overall performance of the network. Figure 6 illustrates the segmentation generated by a trained network.

The post-processing which included CRFs & 3-D class-wise connected components aid in reducing the false positives generated by the networks. Figure 7 illustrates the effect post-processing on segmentation. The contribution of the various the components in the post processing pipeline (CRF, ABL Net, & Connected Components) are illustrated in Table 2.

Table 1. Performance of individual networks and ensemble on held out test data ($n = 40$). In the table WT, TC, AT stand for the whole tumor, tumor core & active tumor respectively.

3-D Densely connected network

	WT	TC	AT
Mean	0.88	0.78	0.72
Std	0.10	0.26	0.31
Median	0.92	0.92	0.86

2-D Densely connected network

	WT	TC	AT
Mean	0.88	0.72	0.69
Std	0.11	0.30	0.28
Median	0.91	0.88	0.82

3-D Multi-Resolution network

	WT	TC	AT
Mean	0.88	0.74	0.73
Std	0.21	0.26	0.32
Median	0.89	0.81	0.83

Ensemble of network

	WT	TC	AT
Mean	0.89	0.78	0.78
Std	0.08	0.21	0.20
Median	0.92	0.89	0.86

Table 2. The contribution of all the components used in post processing pipeline. (CC: 3-D Connected Components)

No post-processing

	WT	TC	AT
Mean	0.85	0.76	0.71
Std	0.10	0.18	0.29
Median	0.88	0.86	0.84

CRF post-processing

	WT	TC	AT
Mean	0.86	0.77	0.73
Std	0.09	0.19	0.28
Median	0.89	0.85	0.84

CRF + ABL Network

	WT	TC	AT
Mean	0.86	0.79	0.74
Std	0.12	0.19	0.28
Median	0.91	0.85	0.84

CRF + ABL + CC

	WT	TC	AT
Mean	0.89	0.78	0.78
Std	0.08	0.19	0.28
Median	0.92	0.89	0.86

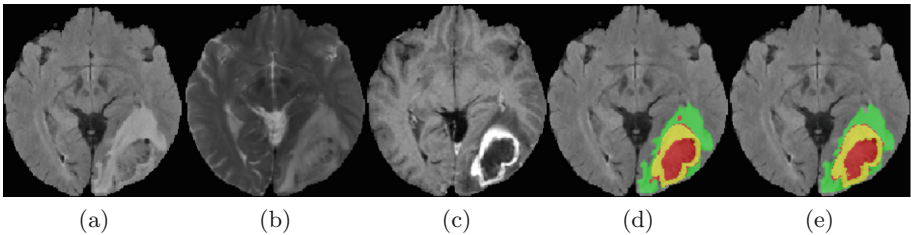


Fig. 6. (a) FLAIR, (b) T2, (c) T1c, (d) Prediction, (e) Segmentation. In images d and e, Green, Yellow & Red represent Edema, Enhancing Tumor and Necrosis present in the lesion. (Color figure online)

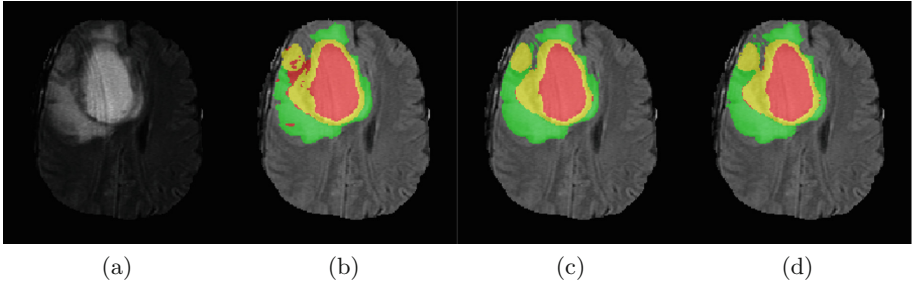


Fig. 7. (a) FLAIR, (b) Without Post-processing, (c) With Post-processing, (d) Ground truth. In images b, c and d, Green, Yellow & Red represent Edema, Enhancing Tumor and Necrosis present in the lesion. (Color figure online)

3.2 Performance on the BraTS Validation Data

On the BraTS validation data ($n = 66$), the performance of each of the networks that form the ensemble is listed in Table 3 respectively. Similar to the observation seen in the held out test data, it was observed that ensembling prediction from multiple networks helped in achieving better segmentation results by lowering variance in the predictions.

Table 3. Performance on validation data ($n = 66$)

3-D Densely connected network

	WT	TC	AT
Mean	0.85	0.74	0.71
Std	0.11	0.22	0.28
Median	0.88	0.82	0.83

2-D Densely connected network

	WT	TC	AT
Mean	0.87	0.73	0.71
Std	0.10	0.27	0.27
Median	0.90	0.86	0.82

3-D Multi-Resolution network

	WT	TC	AT
Mean	0.85	0.73	0.71
Std	0.17	0.26	0.30
Median	0.90	0.83	0.85

Ensemble of network

	WT	TC	AT
Mean	0.89	0.76	0.76
Std	0.07	0.23	0.25
Median	0.91	0.86	0.86

3.3 Performance on BraTS Test Data

The performance of the proposed scheme on the BraTS test data ($n = 191$) is illustrated in Table 4. It was observed that the network achieved good segmentation on unseen data.

Table 4. Performance of the Ensemble of Segmentation on the test data (n = 191)

	Whole tumor	Tumor core	Active tumor
Mean	0.83	0.72	0.69
Std	0.19	0.29	0.29
Median	0.90	0.87	0.80

4 Conclusion

We made use of an ensemble of convolutional neural networks for segmentation of gliomas. From the experiments carried out it was observed that the ensemble aids in reducing the variance associated in the prediction and also helped in increasing quality of the segmentation generated. The false positives generated by the network were minimized by using multiplying the predictions with network trained to delineate lesion from MR volumes. The segmentation was further post-processed by utilizing CRF & 3-D connected component analysis. On the BraTS 2018 validation data (n = 66), the network achieved a competitive dice score of 0.89, 0.76 and 0.76 for the whole tumor, tumor core and active tumor respectively. On the BraTS test data, the network used in the manuscript achieved a mean whole tumor, tumor core and active tumor dice of 0.83, 0.72 and 0.69 respectively.

References

1. Bakas, S., et al.: Segmentation labels and radiomic features for the pre-operative scans of the TCGA-GBM collection. *Cancer Imaging Arch.*, 286 (2017)
2. Bakas, S.: Segmentation labels and radiomic features for the pre-operative scans of the TCGA-LGG collection. *Cancer Imaging Arch.* (2017)
3. Bakas, S., et al.: Advancing the cancer genome atlas glioma MRI collections with expert segmentation labels and radiomic features. *Sci. Data* **4**, 170117 (2017)
4. 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)
5. Eigen, D., Fergus, R.: Predicting depth, surface normals and semantic labels with a common multi-scale convolutional architecture. In: *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2650–2658 (2015)
6. 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)
7. Krähenbühl, P., Koltun, V.: Efficient inference in fully connected CRFs with Gaussian edge potentials. In: *Advances in Neural Information Processing Systems*, pp. 109–117 (2011)
8. Menze, B.H., et al.: The multimodal brain tumor image segmentation benchmark (BRATS). *IEEE Trans. Med. Imaging* **34**(10), 1993 (2015)

9. Pereira, S., Pinto, A., Alves, V., Silva, C.A.: Brain tumor segmentation using convolutional neural networks in MRI images. *IEEE Trans. Med. Imaging* **35**(5), 1240–1251 (2016)
10. Szegedy, C., Ioffe, S., Vanhoucke, V., Alemi, A.A.: Inception-v4, inception-resnet and the impact of residual connections on learning. In: *AAAI*, vol. 4, p. 12 (2017)