Residual 3D U-Net with localization for Brain Tumor Segmentation

Marc Demoustier[1], Ines Khemir[1], Quoc Duong Nguyen[1], Lucien Martin-Gaffé[1], and Nicolas Boutry[2]

[1] EPITA Majeure Santé, Le Kremlin-Bicêtre 94270, France
[2] EPITA Research and Development Laboratory (LRDE), Le Kremlin-Bicêtre 94270, France

Abstract. Gliomas are brain tumors originating from the neuronal support tissue called glia, which can be benign or malignant. They are considered rare tumors, whose prognosis, which is highly fluctuating, is primarily related to several factors, including localization, size, degree of extension and certain immune factors. We propose an approach using a Residual 3D U-Net to segment these tumors with localization, a technique for centering and reducing the size of input images to make more accurate and faster predictions. We incorporated different training and post-processing techniques such as cross-validation and minimum pixel threshold.

Keywords: Brain Tumor Segmentation · Deep Learning · Convolutional Neural Networks · Residual 3D U-Net

1 Introduction

Gliomas or glial tumors are all brain tumors, benign or malignant, arising from the neuronal support tissue or glia. They are rare tumors, whose prognosis, which is extremely variable, is mainly related to several factors, including location, size, degree of extension and certain immune factors.

The average survival time is from 12 to 18 months. Brain tumor diagnosis and segmentation are difficult, particularly using manual segmentation.

In addition, medical image annotation experts have to manually annotate tumor segmentation, which is time consuming and difficult. Automatic segmentation of tumors allows for better diagnosis and treatment planning.

Nowadays, deep learning represents the most effective technology for many tasks such as segmentation, tracking and classification in medical image analysis. Many studies for brain tumor segmentation use deep learning techniques, especially convolutional neural networks *(CNN)*. Recent entries in the *Brain Tumor Segmentation Challenge (BraTS)* challenge are mostly based on these convolutional neural networks, specifically on the U-Net architecture [19] or similar, using an encoder and a decoder with skip-connections. They have shown very convincing performance in previous iterations of the challenge. [12].

The *BraTS* challenge provides the largest fully annotated, openly accessible database for model development and is the primary competition for objective comparison of segmentation methods [2–5,17]. The *BraTS 2021* dataset includes 1251 training cases and 219 validation cases. Reference annotations for the validation set are not provided to participants. Instead, participants can utilize the online evaluation platform to evaluate their models and compare their results with other teams on the online leaderboard. In parallel to the segmentation task, the *BraTS 2021* competition includes the task of predicting of the MGMT promoter methylation status in mpMRI scans. In this work, we only take part in the segmentation task.

To segment these tumors, the *BraTS* dataset contains 5 images in NIfTI format for each patient. These images come from MRI (Magnetic Resonance Imaging), each of the first four images coming from different moments of the MRI. These different modalities are named T1, T1ce, T2 and FLAIR. The last image corresponds to the ground truth, i.e. the tumor and its different regions. The pixel values of this image are:

- 4 for the GD-enhancing tumor
- 2 for the peritumoral edematous/invaded tissue
- 1 for the necrotic tumor core
- 0 for everything else

Using these pixel values, we can find the different tumor regions:

- Whole Tumor (WT): 1, 2, 4
- Tumor Core (TC): 1 and 4
- Enhanced Tumor (ET): 4

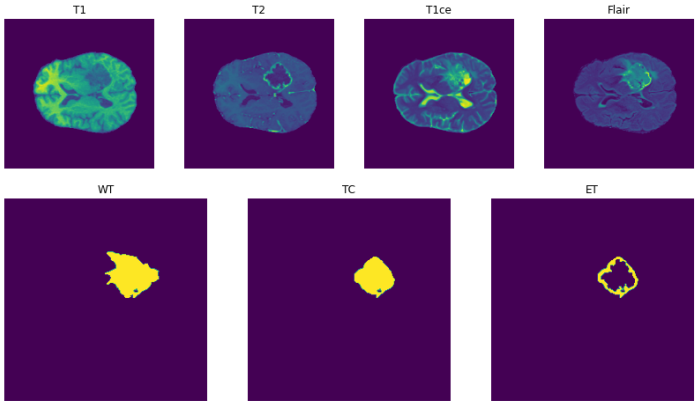

Fig. 1. Modalities and labels

In this paper, we use a residual 3D U-Net using localization with cross-validation for each region (WT, TC, ET).

2 Methods

The implementation used *PyTorch*. As a result, we describe the models with *PyTorch* keywords and methods.

2.1 Pre-processing

Images given in the BraTS dataset are in 240 (Width) × 240 (Height) × 155 (Number of slices) × 4 (FLAIR, T1, T1ce, T2), in the NIfTI format.

The goal in our approach was to keep the images as close as possible to the original data despite the limitations of GPU memory, that is, without too much pre-processing on the input images.

We chose to crop the images to 192 x 192 x 155 to remove the empty borders of the images, then added 5 empty slices to obtain a multiple of 8 on every dimension (except for the channels).

As a result, the images given as input of the model are left with as little modification as possible.

2.2 Residual 3D U-Net

The model we are using is a Residual 3D U-Net, based on Superhuman Accuracy on the SNEMI3D Connectomics Challenge [16]. Residual U-Nets have already been used for biomedical applications [18, 20]. Our model is a variant of the U-Net [19] in 3D [7].

The architecture inherits the main elements from U-Net: a contracting path with convolutions and downsampling, an expansive path with convolutions and upsampling, and skip connections from the contracting path to the expansive path.

Our model differs from the 3D U-Net on different aspects, such as the use of same convolution instead of valid convolution. We also added a residual skip connection to each convolution block, it helps to solve the vanishing gradient problem and to preserve information captured in the initial layers. As we are limited by VRAM, we have to use a small batch size. We used Group Normalization as it performs better than Batch Normalization on a small batch size and it improves the ability of the network to generalize and allows the model to converge rapidly.

In a residual block, the first two convolutions are preceded by group normalization and followed by the ReLU activation function. After the last convolution layer, there is a concatenation of the residual connection, and then activation is called to include the residual information.

On the contracting path of the U-Net, we use what we call an encoding residual block (ERB), which contains a `MaxPool3d` with a kernel size of 2 and a residual block.

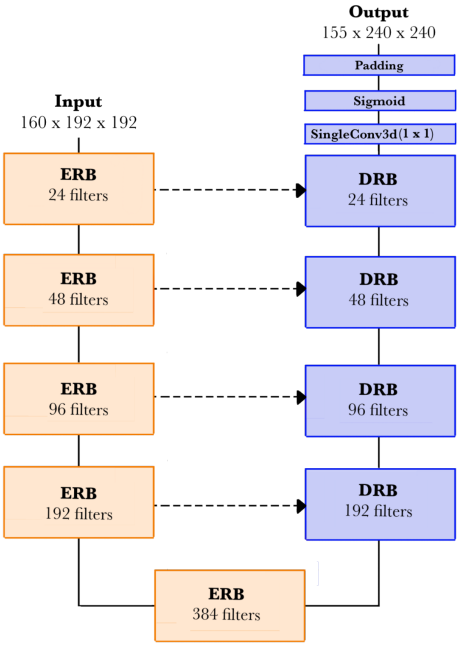

Fig. 2. Overview of the Residual 3D U-Net

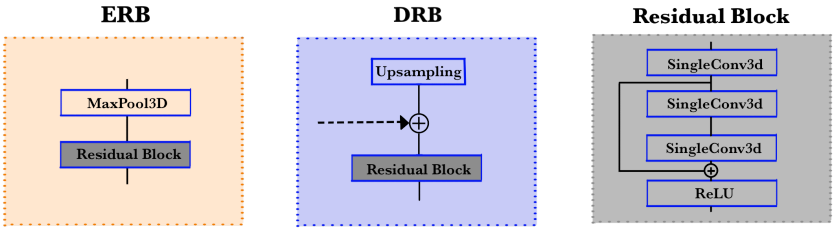

Fig. 3. Residual 3D U-Net blocks

On the expansive path of the U-Net, we use a decoding residual block (DRB), which contains a `ConvTranspose3d` layer with a kernel size of 3 and a scaling factor of size 2 to revert to the size of the encoding residual data from the same level skip connection. After the concatenation of the skip connection, the residual block is added.

At the end of this network, a 1×1 convolution is used to reduce the number of output channels to the number of labels. The number of labels will be 3 for a multi-class prediction and 1 in the case of a single-class prediction.

We have trained 3 separate single-class prediction models. One for each region WT, TC and ET which take 4 channels as input, again FLAIR, T1, T1ce and T2.

As the three models predict one label, a sigmoid has been used as final activation function.

The architecture of this network is built to recover the original shape of the data by using padding on heights and widths of the images. (see fig. 2).

As the MRI images are quite heavy, having only 16GB of VRAM on our GPUs, it was necessary to use 24 filters on the first layer of the network to avoid saturating the GPU memory. However, we are able to run the model with more filters using localization which we discuss in the next section.

2.3 Localization

Training the models on TC and ET did not give great results. These regions are particularly small and the models could not refine the predictions correctly. That is because the "base" model only uses 24 filters on the first level of the U-Net.

Increasing the number of filters was not possible because of our VRAM limitations. We thought about an interpolation to reduce the size of our input images but this technique is too destructive.

In order to make the best out of the VRAM limits, we use localization. It consists in using the predictions on WT, center the input images around the segmented tumors and crop the input images around these segmented tumors.

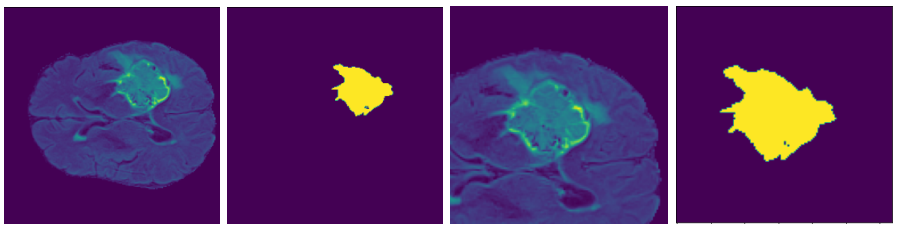

Fig. 4. FLAIR image of a brain with localization

Using this method, we are able to crop the input images into much smaller images of size 128 x 128 x 128 instead of 192 x 192 x 160. Whole tumors can fit inside these cropped images. As a result, the VRAM usage decreased and we were able to increase the number of filters from 24 to 64 on the first convolutional layer of the U-Net.

Once we have predicted the area of the tumor, we can run the models on WT, TC and ET with 64 filters using the cropped images as input. Note: We run the model on WT again with 64 filters to get the best results.

With a higher number of filters, the model is able to capture more complex features such as in the TC and ET regions.

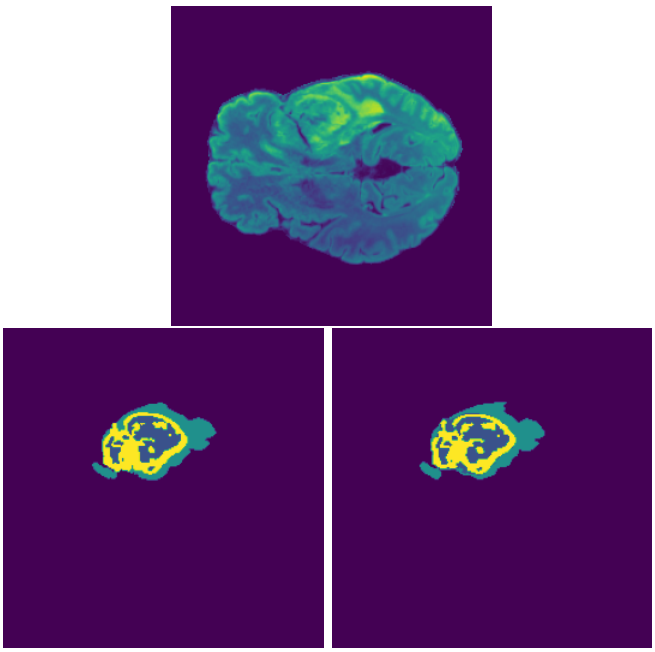

Fig. 5. Prediction example (label on the left and prediction on the right)

2.4 Loss function

The hybrid loss function is used to train the models for the WT and TC regions. This loss combines Dice loss with the standard binary cross-entropy (BCE) loss that is generally the default for segmentation models. Summing the two methods allows for some diversity in the loss while benefiting from the stability of BCE.

Both losses have the same coefficients in the hybrid loss.

$$BCE\_loss = \frac{1}{N}\sum_{n=1}^{N} H(p_n, q_n) = -\frac{1}{N}\sum_{n=1}^{N}[y_n log(\hat{y}_n) + (1-y_n) log(1-\hat{y}_n)] \quad (1)$$

$$Dice\_loss = \frac{2|X \cap Y|}{|X| + |Y|} \quad (2)$$

For the ET region, the standard binary cross-entropy (BCE) loss was used as it requires more stability in training.

2.5 Cross validation

Cross-validation is a method used to train multiple models and improve predictive performance. The test set is separated beforehand.

We chose the k-fold method which consists in dividing the dataset in 5 blocks. The k-fold method allows us to create 5 different models that each have one different validation block and the 4 remaining blocks as the training set.

In order to maximize our scores, we combined the predictions of all 5 models by doing an average (Regression Voting Ensemble) of the weights then binarized the outputs. We also tried a majority vote (Classification Voting Ensemble) after the binarization.

2.6 Post-processing

The analysis of our results obtained on the validation set of BraTS shows that our predictions contained a large number of false positives, on the TC and ET regions.

In order to decrease that number, we defined a threshold for the number of pixels on an image [12]. Each prediction containing a number of pixels below this threshold is considered an empty prediction because we know that a tumor does not necessarily contain an enhanced tumor (ET). Several threshold values were tested.

3 Experiments and results

3.1 Implementation details

In the *BraTS 2021 Segmentation Challenge*, the training data is composed of 1251 multimodal MRI cases.

The network is implemented with *PyTorch*. The models were trained on 4 NVIDIA Tesla V100 16 GB GPUs. Each model was trained for 40 to 60 epochs with a batch size of 4, Adam optimizer with a learning rate of 0.0001 (BCE-Dice loss), 0.00003 (BCE-Dice loss), 0.003 (BCE) respectively for WT, TC and ET.

The model used to crop the input images and center on the tumor has the same learning rate as the model trained for WT and also uses the BCE-Dice loss.

We reduced the learning rate with the callback `ReduceLROnPlateau` by a factor of 0.4, with a patience and a cooldown of 2 epochs.

3.2 Performance on the validation set of *BraTS 2021*

The Validation Dataset of BraTS 2021 contains 219 brains MRI. For each brain, the four modality (T1, T2, T1ce and FLAIR) are used in order to predict the multi-class prediction. Predictions are evaluated thanks to the Dice coefficient, the Hausdorff distance (Hausdorff95), the sensitivity (True Positive Rate) score and a specificity (True Negative Rate) score. They are defined as follows:

$$Dice = \frac{2TP}{FP + 2TP + FN} \tag{3}$$

$$Hausdorff(T, P) = max\{sup_{t \in T} inf_{p \in P} d(t, p), sup_{p \in P} inf_{t \in T} d(t, p)\} \tag{4}$$

$$Sensitivity = \frac{TP}{TP + FN} \tag{5}$$

$$Specificity = \frac{TN}{TN + FP} \tag{6}$$

where TP, FP, TN and FN denote respectively the number of true positive, false positive, true negative and false negative voxels.

The Hausdorff distance computes the distance between the predicted regions and the ground truth regions. t and p denote respectively the pixels in the ground truth regions T and the predicted regions P. *d(t,p)* is the function that computes the distance between the points t and p.

Table 1. Performance comparison using the dice coefficient on the BraTS 2021 Validation set using the online tool

Methods	WT (%)	TC (%)	ET (%)
Baseline (BCE loss)	90.98	80.24	74.07
Classification Voting Ensemble [1]	91.42	80.96	77.07
Regression Voting Ensemble (RVE) [1]	91.45	80.98	77.58
BCE-Dice loss (WT & TC) [2]	91.34	82.71	77.58
RVE [1] + Localization [4]	91.64	82.71	78.26
RVE [1] + ET threshold 100 [3]	91.45	80.98	78.91
RVE + Localization [4] + ET threshold 400 [3]	91.64	82.71	78.71
RVE + Localization [4] + ET threshold 600 [3]	**91.64**	**82.71**	**80.22**

Table 2. Submission result on validation set

Tumor Region	WT (%)	TC (%)	ET (%)
Dice	91.64	82.71	80.22
Hausdorff95	4.35	12.50	25.13
Sensitivity	93.61	85.74	79.14
Specificity	99.90	99.95	99.97

3.3 Performance on the test set of *BraTS 2021*

	Dice_WT (%)	Dice_TC (%)	Dice_ET (%)	HD95_WT	HD95_TC	HD95_ET
Mean	89.21	81.30	80.64	11.81	26.57	32.18
StdDev	16.87	28.79	26.52	47.15	86.49	98.98
Median	94.42	93.63	90.86	2.24	2.0	1.41
25quantile	89.92	86.17	80.48	1.41	1.0	1.0
75quantile	96.72	96.54	95.03	5.10	4.97	3.0

[1] Cross-Validation evaluation method (see 2.5)
[2] BCE-Dice loss function (see 2.4)
[3] Post processing using thresholding (see 2.6)
[4] Second network with reframing around the WT (see 2.3)

4 Conclusion

In this paper, we propose a segmentation method using the Residual 3D U-Net as the skeleton of the network, which uses the four modalities on an area where the tumor has been predicted. The localization method allows us to exploit the limitations of VRAM to the fullest by cropping and centering on the whole tumor without any performance loss. The evaluation of our method on the BraTS 2021 test set gives dice scores of 89.21, 81.30, 80.64 for the whole tumor, the tumor core and enhancing tumor, respectively.

Acknowledgement

We would like to thank Arnaud Renard and his team, for the access to their supercomputer ROMEO, the Regional Super Computer Center hosted by the University of Reims Champagne-Ardenne. We also would like to thank Christian Chabrerie for his support throughout the project.

References

1. Abraham, N., Khan, N.: A novel focal tversky loss function with improved attention u-net for lesion segmentation. 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019) pp. 683–687 (2019)
2. Baid, U., Ghodasara, S., Bilello, M., Mohan, S., Calabrese, E., Colak, E., Farahani, K., Kalpathy-Cramer, J., Kitamura, F.C., Pati, S., Prevedello, L.M., Rudie, J.D., Sako, C., Shinohara, R.T., Bergquist, T., Chai, R., Eddy, J., Elliott, J., Reade, W., Schaffter, T., Yu, T., Zheng, J., Annotators, B., Davatzikos, C., Mongan, J., Hess, C., Cha, S., Villanueva-Meyer, J.E., Freymann, J.B., Kirby, J.S., Wiestler, B., Crivellaro, P., Colen, R.R., Kotrotsou, A., Marcus, D.S., Milchenko, M., Nazeri, A., Fathallah-Shaykh, H.M., Wiest, R., Jakab, A., Weber, M., Mahajan, A., Menze, B.H., Flanders, A.E., Bakas, S.: The RSNA-ASNR-MICCAI brats 2021 benchmark on brain tumor segmentation and radiogenomic classification. CoRR **abs/2107.02314** (2021), `https://arxiv.org/abs/2107.02314`
3. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J., Freymann, J., Farahani, K., Davatzikos, C.: Advancing the cancer genome atlas glioma mri collections with expert segmentation labels and radiomic features. Scientific Data **4** (09 2017). https://doi.org/10.1038/sdata.2017.117
4. 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 (07 2017). https://doi.org/10.7937/K9/TCIA.2017.KLXWJJ1Q
5. 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 (07 2017). https://doi.org/10.7937/K9/TCIA.2017.GJQ7R0EF
6. Caesar, H., Uijlings, J.R.R., Ferrari, V.: Region-based semantic segmentation with end-to-end training (2016), `http://arxiv.org/abs/1607.07671`

7. Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O.: 3d u-net: Learning dense volumetric segmentation from sparse annotation. In: Ourselin, S., Joskowicz, L., Sabuncu, M.R., Unal, G., Wells, W. (eds.) Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016. pp. 424–432. Springer International Publishing, Cham (2016)
8. Eelbode, T., Bertels, J., Berman, M., Vandermeulen, D., Maes, F., Bisschops, R., Blaschko, M.B.: Optimization for medical image segmentation: Theory and practice when evaluating with dice score or jaccard index. IEEE Transactions on Medical Imaging **39**(11), 3679–3690 (2020). https://doi.org/10.1109/TMI.2020.3002417
9. Fidon, L., Li, W., Garcia-Peraza-Herrera, L.C., Ekanayake, J., Kitchen, N., Ourselin, S., Vercauteren, T.: Generalised wasserstein dice score for imbalanced multi-class segmentation using holistic convolutional networks. In: Crimi, A., Bakas, S., Kuijf, H., Menze, B., Reyes, M. (eds.) Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. pp. 64–76. Springer International Publishing, Cham (2018)
10. Guo, D., Wang, L., Song, T., Wang, G.: Cascaded global context convolutional neural network for brain tumor segmentation. In: Crimi, A., Bakas, S. (eds.) Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. pp. 315–326. Springer International Publishing, Cham (2020)
11. Isensee, F., Kickingereder, P., Wick, W., Bendszus, M., Maier-Hein, K.H.: No newnet. In: International MICCAI Brainlesion Workshop. pp. 234–244. Springer International Publishing (2018)
12. Isensee, F., Jäger, P.F., Full, P.M., Vollmuth, P., Maier-Hein, K.H.: nnu-net for brain tumor segmentation. In: Crimi, A., Bakas, S. (eds.) Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. pp. 118–132. Springer International Publishing, Cham (2021)
13. Islam, M., Vibashan, V.S., Jose, V.J.M., Wijethilake, N., Utkarsh, U., Ren, H.: Brain tumor segmentation and survival prediction using 3d attention unet. In: Crimi, A., Bakas, S. (eds.) Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. pp. 262–272. Springer International Publishing, Cham (2020)
14. Jiang, Z., Ding, C., Liu, M., Tao, D.: Two-stage cascaded u-net: 1st place solution to brats challenge 2019 segmentation task. In: International MICCAI Brainlesion Workshop. pp. 231–241. Springer International Publishing (2019)
15. Lee, C.Y., Xie, S., Gallagher, P., Zhang, Z., Tu, Z.: Deeply-Supervised Nets. In: Lebanon, G., Vishwanathan, S.V.N. (eds.) Proceedings of the Eighteenth International Conference on Artificial Intelligence and Statistics. Proceedings of Machine Learning Research, vol. 38, pp. 562–570. PMLR, San Diego, California, USA (09–12 May 2015), `http://proceedings.mlr.press/v38/lee15a.html`
16. Lee, K., Zung, J., Li, P., Jain, V., Seung, H.: Superhuman accuracy on the snemi3d connectomics challenge. ArXiv **abs/1706.00120** (2017)
17. Menze, B.H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., Lanczi, L., Gerstner, E., Weber, M.A., Arbel, T., Avants, B.B., Ayache, N., Buendia, P., Collins, D.L., Cordier, N., Corso, J.J., Criminisi, A., Das, T., Delingette, H., Demiralp, , Durst, C.R., Dojat, M., Doyle, S., Festa, J., Forbes, F., Geremia, E., Glocker, B., Golland, P., Guo, X., Hamamci, A., Iftekharuddin, K.M., Jena, R., John, N.M., Konukoglu, E., Lashkari, D., Mariz, J.A., Meier, R., Pereira, S., Precup, D., Price, S.J., Raviv, T.R., Reza, S.M.S., Ryan, M., Sarikaya, D., Schwartz, L., Shin, H.C., Shotton, J., Silva, C.A., Sousa, N., Subbanna, N.K., Szekely, G., Taylor, T.J., Thomas, O.M., Tustison,

N.J., Unal, G., Vasseur, F., Wintermark, M., Ye, D.H., Zhao, L., Zhao, B., Zikic, D., Prastawa, M., Reyes, M., Van Leemput, K.: 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

18. Milletari, F., Navab, N., Ahmadi, S.A.: V-net: Fully convolutional neural networks for volumetric medical image segmentation (2016)
19. 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.) Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. pp. 234–241. Springer International Publishing, Cham (2015)
20. Yu, L., Yang, X., Chen, H., Qin, J., Heng, P.A.: Volumetric convnets with mixed residual connections for automated prostate segmentation from 3d mr images (01 2017)
21. Zhao, Y.X., Zhang, Y.M., Liu, C.L.: Bag of tricks for 3d mri brain tumor segmentation. In: International MICCAI Brainlesion Workshop. pp. 210–220. Springer International Publishing (2019)