Segmentation of Brain Tumors in 3D-MRI Data and Patient Survival Prediction: Methods for the BraTS 2018 Challenge

Leon Weninger, Oliver Rippel, Simon Koppers, and Dorit Merhof

Institute of Imaging & Computer Vision
RWTH Aachen University
Aachen, Germany
{last name}@lfb.rwth-aachen.de

Abstract. Brain tumor localization and segmentation is an important step in the treatment of brain tumor patients. It is the base for later clinical steps, e.g., a possible resection of the tumor. Hence, an automatic segmentation algorithm would be preferable, as it does not suffer from inter-rater variability, and results could be available directly after the brain imaging procedure. Using this automatic tumor segmentation, it could also be possible to predict the survival of patients. The BraTS 2018 challenge consists of these two tasks: tumor segmentation in 3D-MRI images of brain tumor patients and survival prediction based on these images. For the tumor segmentation, we utilize a two-step approach: First, the tumor is located using a 3D U-net. Second, another 3D U-net – more complex, but with a smaller field-of-view – detects subtle differences in the tumor volume, i.e., it segments the located tumor into tumor core, enhanced tumor, and peritumoral edema. The survival prediction of the patients is done with a rather simple, yet accurate algorithm which outperforms other tested approaches.

Keywords: BraTS 2018, Brain Tumor, Automatic Segmentation, Survival Prediction, Deep Learning.

1 Introduction

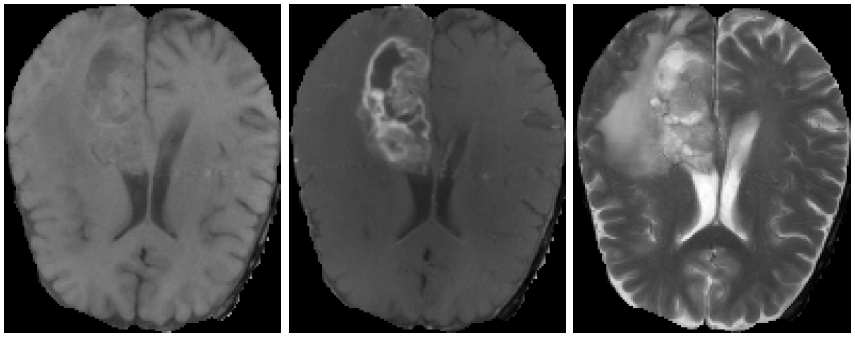
Brain tumors can appear in different forms, shapes and sizes and can grow to a considerable size until they are discovered. They can be distinguished into glioblastoma (GBM/HGG) and low grade glioma (LGG). A common way of screening for brain tumor is with MRI-scans, where even different brain tumor regions can be determined. In effect, MRI scans of the brain are not only the basis for tumor screening, but are even utilized for pre-operative planning. Thus, an accurate, fast and reproducible segmentation of brain tumors in MRI scans is needed for several clinical applications.

HGG patients have a poor survival prognosis, as metastases often develop even when the initial tumor was completely resected. Whether patient overall

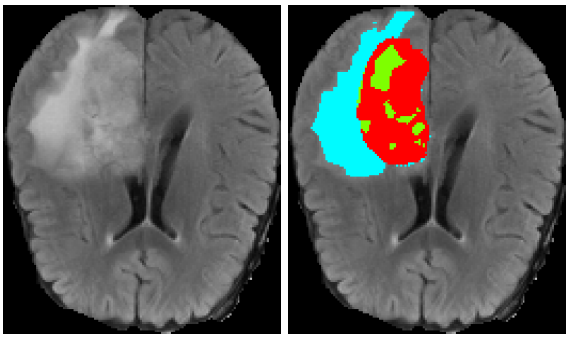
survival can be accurately predicted from pre-operative scans, i.e., knowing factors such as radiomics features, tumor location and tumor shape, remains an open question.

The BraTS challenge [6] addresses these problems, and is one of the biggest and best-known machine learning challenges in the field of medical imaging. Last year around 50 different competitors from around the world took part. The challenge is divided in two parts: First, tumor segmentation based on 3D-MRI images, and second, survival prediction of the brain tumor patients based on only the pre-operative scans and the age of the patients.

Similar to the BraTS 2017 dataset, the BraTS 2018 training dataset consists of images of 285 brain tumor patients from 19 different contributers. The dataset includes T1, T1 post-contrast, T2, and T2 Fluid Attenuated Inversion Recovery (Flair) volumes, as well as hand-annotated expert labels for each patient [3] [2] [1].



(a) T1 weighted **(b)** T1 post-contrast **(c)** T2 weighted



(d) T2 Flair **(e)** Flair with labels

Fig. 1: Different MRI-modalities and groundtruth-labels in the BraTS 2018 dataset. Blue indicates the peritumoral edema, green the necrotic and non-enhancing tumor, and red the GD-enhancing core, as described in the BraTS paper [6].

Motivated by the success of the U-net [9] in biomedical image segmentation, we choose the 3D-adaptation [4] of this network to tackle the segmentation part of the BraTS challenge. Two different versions of this architecture are used, a first one to coarsely locate the tumor, and a second one to accurately segment the located tumor into different areas.

Concerning the survival prediction, we found that complex models using different types of radiomics features such as shape and texture of the tumor and the brain could not outperform a simple linear regressor based on just a few basic features. Using only the patient age and tumor region sizes as features, we achieve competitive results.

The code developed for this challenge will be available online after the final deadline: `https://github.com/weningerleon/BraTS2018`

2 Methods

2.1 Segmentation

We tackle the segmentation task in a two-step approach: First, the location of the brain tumor is determined. Second, this region is segmented into the three different annotations: *peritumoral edema (ed)*, *necrotic tumor (nec)*, and *GD-enhancing core (gde)*.

Preprocessing We first define a brain mask based on all voxels unequal to zero, on which all preprocessing is carried out. Before we calculate the mean and standard deviation of the brain, we clip the values of intensities at 2.5% and 97.5%. Using this preprocessing technique, very high and very low values, often occurring due to imaging artifacts, have less influence on the mean and standard deviation. Since different MRI-scanners and sequences are used, we independently normalize each image and modality based on the obtained values. Non-brain regions remain zero.

The whole tumor is strongly visible in T1, T2 and Flair MRI-images, so we discard all other MRI-modalities for the tumor localization step. We construct a cuboid bounding box around the brain, and crop the non-brain regions to facilitate training. The training target is constructed by merging the three different tumor classes of the groundtruth labels.

For training of the tumor segmentation step, the 3D-images are cropped around a padded tumor area, which is defined as the area of 20 voxels in every direction around the tumor.

Network architectures and employed hardware For both steps, a 3D U-net [4] with a depth of 4 is employed.

The first U-net uses padding in every convolutional layer, such that the input size corresponds exactly to the output size. Every convolutional layer is followed by a ReLU activation function. 16 feature maps are used in the first layer, and the number of feature maps doubles as the depth increases. For normalization

between the different layers, instance-norm layers [10] are used, as they seem to be better suited for normalization in segmentation tasks and for small batch sizes. Testing different training hyperparameters, the Adam optimizer [5] with an initial learning rate of 0.001 together with a binary cross entropy loss was chosen for the tumor localization step. An L2-regularization of 1e-5 is applied to the weights, and the learning rate was reduced by a factor of 0.015 after every epoch. One epoch denotes a training step over every brain.

The U-net utilized in the second step has a similar architecture as the previous one, but with double as many feature maps per layer. To counteract the increased memory usage, no padding is used, which drastically reduces the size of the output as well as the memory consumption of later feature maps.

Here, we apply a multiclass dice loss to the output of our 3D U-net and the labels for training, as described in [7]. A learning rate of 0.005 was chosen, while weight decay and learning rate reduction remain the same as in step 1.

Our contribution to the BraTS challenge was implemented using pyTorch [8]. Training and prediction is carried out on a Nvidia 1080 Ti GPU with a memory size of 11 Gb.

Training In the first step, we train with complete brain images cropped to the brain mask. The brain mask is determined by all voxels not equal to zero. Using a rather simple U-net, a training pass with a batch-size of one fits on a GPU even for larger brains. Due to the bounding box around the brain, different sizes need to be passed through the network. In practice this is possible using a fully convolutional network architecture and a batch size of one.

For the second step, we choose the input to be fixed to 124x124x124. Due to the unpadded convolutions, this results in an output shape of 36x36x36. Hence, the training labels are the 36x36x36 sized segmented voxels in the middle of the input. Here, a batch-size of two was chosen.

During training, 25 such patches are chosen at random from inside the padded tumor bounding box for each patient. Having 286 training datasets, this gives us 7125 training patches.

Inference Similar to the training procedure, the first step is carried out directly on a complete 3-channel (T1, T2, Flair) 3D image of the brain.

Before the tumor / non-tumor segmentation of this step is used as basis in the second step, only the largest connected area is kept. Based on the assumption that there is only one tumorous area in the brain, we can suppress false positive voxels in the rest of the brain with this method.

We then predict 36x36x36 sized patches with the trained unpadded U-net. Patches are chosen so that they cover the tumorous area, the distance between two neighboring patches was set to 9 in each direction. Several predictions per voxel result. Accordingly, a majority vote over these predictions gives the final result.

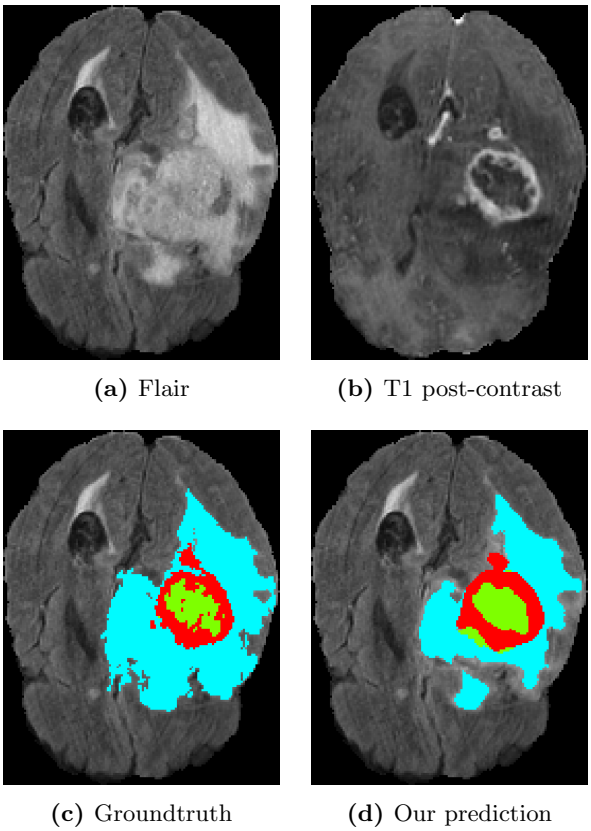
2.2 Survival Prediction

According to the information given by the segmentation labels, we count the number of voxels of the tumor segmentation. This volume information about the necrotic tumor core, the GD-enhancing tumor and peritumoral edema as well as the distance between the centroids of tumor and brain and the age of the patient are used as input for a linear regressor.

3 Results

3.1 Segmentation

For evaluation on the training dataset, we split the training dataset randomly into 245 training images and 40 test images to evaluate our approach with ground truth labels. No external data was used for training or pre-training.



(a) Flair **(b)** T1 post-contrast

(c) Groundtruth **(d)** Our prediction

Fig. 2: Comparison of our segmentation result with the groundtruth labels.

Based on our experience with the training dataset, we choose 200 epochs as an appropriate training duration for the first step, and 60 epochs as an appropriate training duration for the second step. We thus train from scratch on all training images for the determined optimal number of epochs, and use the obtained networks for evaluation on the validation set. The results obtained by this method can be seen in Table 1.

Dataset	Dice			Sensitivity			Specificity			Hausdorff Dist.		
	ET	WT	TC	ET	WT	TC	ET	WT	TC	ET	WT	TC
Train set	0.763	0.860	0.817	0.747	0.784	0.787	0.998	0.998	0.998	5.63	7.01	7.88
Val set	0.712	0.889	0.758	0.757	0.887	0.735	0.998	0.995	0.998	6.28	6.97	10.91

Table 1: Results for the segmentation challenge. Training and test set errors according to the online submission system.

3.2 Survival Prediction

For evaluating our approach on the training dataset, we fit and evaluate our linear regressor with a 50-fold cross-validation on the training images. We compare the results obtained by solely using the age of the patient versus using the age with a subset of the tumor region sizes as features. On top, we consider the distance between the centroid of the tumor and the centroid of the brain as a feature. Our finding is that all features other than the age of the patient increase the error on left-out images. In Tables 2 and 3, we show the exact results for the different input features.

Features	MSE	Median Err.
Age	**87089**	**206**
Age + gde	93599	215
Age + ed	91767	212
Age + nec	92320	207
Age + dist	95070	207
Age+gde+ed+nec	98053	222

Table 2: Training Data: Mean Squared Error and Median Error for 50-fold cross-validation of the linear regressor. The different features considered are the age of the patient, the volume in voxels of the enhancing tumor (*gde*), of the necrotic tumor (*nec*), of the edema (*ed*) as well as the distance between the centroid of the tumor and the centroid of the brain (*dist*).

Features	Accuracy	MSE	Median SE	stdSE	SpearmanR
Age	0.5	**97759.5**	**46120.5**	**139670.7**	**0.267**
Age+gde+ed+nec	**0.536**	101012.0	51006.5	140511.5	0.258

Table 3: Validation Data: Accuracy metrics according to the online portal.

4 Discussion & Conclusion

Our contribution submitted to the BraTS challenge 2018 was summarized in this paper. We used a two-step approach for tumor segmentation, which already gives promising results. In the near future we will evaluate a broader variety of different network architectures, and will also include 3D data-augmentation techniques into our framework.

Our algorithm for the survival analysis task is a straight-forward approach. We considered other, more complex approaches, which were however not able to beat this baseline algorithm.

So far, our survival prediction algorithm ranks among the top submissions, e.g., the age-only approach achieves the lowest MSE and second highest accuracy on the validation set. From these observations, it can be concluded that pre-operative scans are not well suited for survival prediction. Here, other datasets could be better suited for survival prediction, e.g., post-operative or follow-up scans of the patient.

References

1. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J., Freyman, J., Farahani, K., Davatzikos, C.: Segmentation labels and radiomic features for the pre-operative scans of the tcga-lgg collection. The Cancer Imaging Archive (2017)
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-gbm collection. The Cancer Imaging Archive (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 EP – (09 2017), `http://dx.doi.org/10.1038/sdata.2017.117`
4. Çiçek, Ö., Abdulkadir, A., Lienkamp, S., Brox, T., Ronneberger, O.: 3d u-net: Learning dense volumetric segmentation from sparse annotation. In: Medical Image Computing and Computer-Assisted Intervention (MICCAI). LNCS, vol. 9901, pp. 424–432. Springer (Oct 2016), `http://lmb.informatik.uni-freiburg.de/Publications/2016/CABR16`, (available on arXiv:1606.06650 [cs.CV])
5. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. CoRR **abs/1412.6980** (2014), `http://arxiv.org/abs/1412.6980`
6. 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., Leemput, K.V.: The multimodal brain tumor image segmentation benchmark (brats). IEEE Transactions on Medical Imaging **34**(10), 1993–2024 (Oct 2015). https://doi.org/10.1109/TMI.2014.2377694

7. Milletari, F., Navab, N., Ahmadi, S.: V-net: Fully convolutional neural networks for volumetric medical image segmentation. CoRR **abs/1606.04797** (2016), `http://arxiv.org/abs/1606.04797`
8. Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., Lin, Z., Desmaison, A., Antiga, L., Lerer, A.: Automatic differentiation in pytorch. In: NIPS-W (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.) Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. pp. 234–241. Springer International Publishing, Cham (2015)
10. Ulyanov, D., Vedaldi, A., Lempitsky, V.S.: Instance normalization: The missing ingredient for fast stylization. CoRR **abs/1607.08022** (2016), `http://arxiv.org/abs/1607.08022`