
Brain Tumor Segmentation

Shaina Reji, Erika Earley, Manisha Basak

Department of Computer Science

Stanford University

sreji@stanford.edu, eearley@stanford.edu, mbasak@stanford.edu

Abstract

Using deep learning to perform segmentations on medical images is a fast growing field of research. This paper explores segmenting brain tumors using two methods. The first method uses a cascade of a WNet and a UNet and the second method uses a Mask R-CNN framework to classify tumors in the brain. The experiments use the BraTS 2018 dataset. Based on the results, the cascaded network seems to perform better at segmenting brain tumors than the Mask R-CNN.

1 Introduction

Gliomas are types of tumors that occur in the brain and spinal cord. Despite recent research on gliomas, patient diagnosis relies on images being evaluated either based on qualitative criteria only or by relying on quantitative measures. There are two main challenges with automatic brain segmentation. First, the size, shape, and localization of brain tumors vary between patients, which significantly reduces the reusability of prior information to train networks. Second, the boundaries between parts of a tumor structure are often unclear, which leads to disagreements between medical professionals when doing image classification.

This paper will explore two different methods of multimodal brain tumor segmentation on the Brain Tumor Image Segmentation Benchmark (BraTS) 2018 dataset. The first method used was Mask R-CNN, an instance segmentation framework, which identifies objects at a pixel level [2]. The input to this framework included two datasets of images. One was the slices of the brain images, with segmentations of the whole tumor, tumor core, and the enhanced tumor core. The second was their corresponding masks identifying only the enhanced tumor core. The output was the segmented image.

The second method method was implementing UNet in conjunction with the first network (WNet) used by the model created by Wang et al., 2017 [1]. The input to this model was a MRI scan of a whole brain with a tumor in it. We then used the WNet which outputted the bounding box of the whole tumor. The image of the whole tumor was then fed into a UNet to output the segmented brain tumor.

2 Related work

We began research by reading two sources provided by the BraTs competition: *The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)* by Menze et al., 2014 [3] and *Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features* by Bakas et al., 2017. [4] These provided background knowledge.

The methodology we pursued was based mainly off of four papers. Two of them discuss the use of UNets for image segmentation: *U-Net: Convolutional Networks for Biomedical Image*

Segmentation by Ronneberger et al., 2015 [5] and *Radio frequency interference mitigation using deep convolutional neural networks* by Akeret et al., 2017. [6] The former was very useful for gaining intuition about various components of the UNet architecture.

Automatic Brain Tumor Segmentation using Cascaded Anisotropic Convolutional Neural Networks by Wang et al., 2017 [1] explains the CNN implementation that we cloned as a base for our model. This paper included a discussion which was particularly illustrative for understanding some advantages of using a cascade of fully convolutional neural networks given the hierarchical nature of the tumor sub-regions that we would like to label for segmentation.

In order to learn more about the possible advantages and disadvantages of using Mask R-CNN for image segmentation, we consulted *Mask R-CNN* by He et al. 2018. [2] In comparing this to Faster R-CNN we found that Mask R-CNN is the state-of-the-art model for object detection and segmentation.

3 Dataset and Features

The dataset was taken from the MICCAI BraTs Challenge from the University of Pennsylvania [3], [4]. The images were split up into two categories, HGG and LGG. HGG tumors are malignant glioblastomas and LGG tumors are lower grade gliomas. Each image was 240x240 and had 155 slices. Each patient was scanned using four different MRI modes: T1, T1c, T2, Flair. The images were also broken up into three different views of the brain- axial, coronal, and sagittal. The ground truth segmentation files were manually created by clinicians. Each tumor in the segmentation files were broken up into three subregions- the whole tumor, the tumor core, and the enhanced tumor core. Originally we had 324,900 images in the training set and we had 39,900 in the test set. However, due to complications with AWS, we had to train locally and therefore had to reduce the number of examples to 10,455 images for training and 1,260 images for testing. Since we were training on images, the features were naturally represented as multi-dimensional numpy arrays.

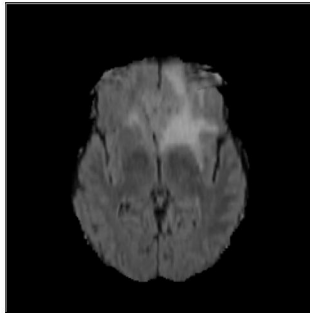


Figure 1: Axial view of HGG Flair image



Figure 2: Sagittal view of HGG T1c image and label

4 Methods

4.1 Mask R-CNN

Mask R-CNN is a leading object segmentation framework, which follow from Faster R-CNN, an object detection framework. In the first stage of Mask R-CNN, images are scanned and proposals, areas likely to contain an object, are generated. In the second stage, proposals are classified and

The first network of this model acts like a normal CNN, taking in the image of the whole brain and outputting a bounding box for the whole tumor. As seen in the figure above, this network uses 10 residual blocks with dilated convolution, residual connections and multi scale fusion. The dilated convolutions use batch normalization and PReLU. This reduces the impact of the earlier layers within the network and also helps with regularization. The Residual connections help smooth propagation. The multi-scale fusion allows us to represent both low level and high level features.

The second network of our model is the UNet. The whole network has 23 layers. The first half of the UNet is a contracting, downsampling path, meaning it acts like a regular CNN. However, the second half of the network is an expansive, upsampling path. Both halves use convolutions and ReLU. The first half also has max pooling and the second half has up-convolutions. Having a network such as this allows us to capture the context of the image while simultaneously allowing the localization to be more precise.

For both of these networks, a softmax with a cross entropy loss was used:

$$p_k(x) = \frac{e^{\alpha_k(x)}}{\sum_{k'=1}^K e^{\alpha_{k'}(x)}} + \sum_{x \in \Omega} w(x) \log(p_{t(x)}(x))$$

During training for both networks, the Adaptive Moment Estimation (Adam) optimizer was used.

5 Experiments/Results/Discussion

In order to use the 3D images given in the form of .nii files for the Mask R-CNN, they had to be converted to .png files. To work with the framework, each image frame had to have an accompanying binary mask segmenting the enhanced tumor core in a COCO-style annotation. This was done based on the ground truth segmentation images the dataset provides.

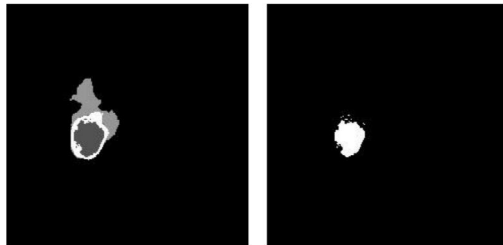


Figure 5: Binary mask segmenting the enhanced tumor core

The Mask R-CNN paper uses a learning rate of 0.02 but since this causes the weights to explode on TensorFlow, a learning rate of 0.001 and a learning momentum of 0.9 was used in our implementation. For Mask R-CNN, the images had to be resized to 256 by 256 to work with the code and to ensure precision and granularity, 256 anchors per image were used for RPN training. Batch size is calculated as the product of the number of GPU's used and images per GPU. Since we were not able to train using GPU, the batch size was 1.

Due to computation power limitations, we had to use transfer learning. Instead of training the model from scratch, we started training with a weights file from the trained Microsoft COCO (Common Objects in Context) dataset. Since this dataset has pretty general image classifications, such as animals and cars, training the model using these weights on much more specific images of brain tumors did not translate well. Therefore, the accuracy was very low for image segmentation using Mask R-CNN.

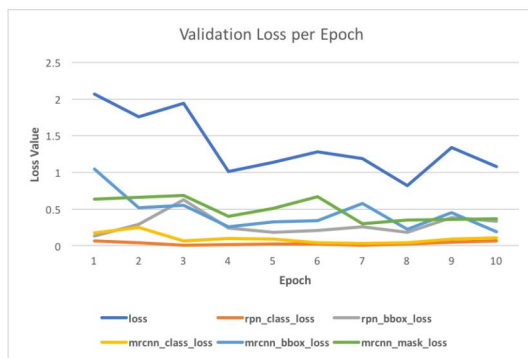


Figure 6: Loss value on validation data shows overall decrease in loss

For the UNet, 1 was passed in as the number of channels because we were using gray scale and 2 was used as the number of of classes since we were predicting the enhanced tumor core versus no enhanced tumor core. The learning rate used was 0.2 and the batch size was 1. For this method we did not perform any hyperparameter tuning because the model we built off of already had tuned hyperparameters. The metric that was used for this model was accuracy. Below was the training accuracies for the three different views:

View	Training Accuracy
Axial	88.117
Coronal	86.627
Sagittal	97.317

Table 1: Training Accuracy

Since we only able to train on a smaller subset of the data due to issues with using a GPU, we overfit to our training set. If we had been able to run the whole training set, we would have been able to mitigate the overfitting issue. Also, due to computational power limitations, we had to cut the test set and therefore only tested on the saggital view. We chose this view because the output seemed to be the most interpretable. The test set error for this was 47.57 percent. Below is the output we get from running the UNet.



Figure 7: Input vs. segmented pixel values during testing

6 Conclusion/Future Work

Overall, we found that the cascaded WNet to UNet model performed better in segmenting brain tumors than the Mask R-CNN. We imagine that incorporating UNet into the model is beneficial for a learning task like ours because the combination of symmetric contracting and expanding paths allow it to localize and make use of context at the same time. In the future, we would use pre-existing medical segmentation weights instead of COCO weights as the COCO dataset is trained on images of everyday objects such as cars and people, so using these weights for classifying an object as specific as a brain tumor did not yield good results. We would also try ResNet50 as the backbone of RCNN instead of ResNet101 as a first step in limiting variance of the model. We would also consider implementing UNet for the whole model and not just one network as part of the Wnet-to-Unet cascade to try a fully end-to-end network. Lastly, pre-trained models for segmentation during pre-processing could be used instead of using the ground truth to limit over-fitting.

7 Contributions

Basak worked on the Mask RCNN model while Earley and Reji worked on adapting implementations of WNet and UNet to the project.

References

- [1] Wang, Guotai, Li, Wenqi, Ourselin, Sebastien & Vercauteren, Tom (2017) *Automatic Brain Tumor Segmentation using Cascaded Anisotropic Convolutional Neural Networks*, In Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. Pages 179-190. Springer, 2018.
- [2] He, Kaiming & Gkioxari, Georgia & Dollár, Piotr & Girshick, Ross (2017) *Mask R-CNN*. International Conference on Computer Vision (ICCV) 2017
- [3] Menze BH, Jakab A, Bauer S, Kalpathy-Cramer J, Farahani K, Kirby J, Burren Y, Porz N, Slotboom J, Wiest R, Lanczi L, Gerstner E, Weber MA, Arbel T, Avants BB, Ayache N, Buendia P, Collins DL, Cordier N, Corso JJ, Criminisi A, Das T, Delingette H, Demiralp T, Durst CR, Dojat M, Doyle S, Festa J, Forbes F, Geremia E, Glocker B, Golland P, Guo X, Hamamci A, Iftekharuddin KM, Jena R, John NM, Konukoglu E, Lashkari D, Mariz JA, Meier R, Pereira S, Precup D, Price SJ, Raviv TR, Reza SM, Ryan M, Sarikaya D, Schwartz L, Shin HC, Shotton J, Silva CA, Sousa N, Subbanna NK, Szekely G, Taylor TJ, Thomas OM, Tustison NJ, Unal G, Vasseur F, Wintermark M, Ye DH, Zhao L, Zhao B, Zikic D, Prastawa M, Reyes M, & Van Leemput K. (2015) *The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)* IEEE Transactions on Medical Imaging 34(10), 1993-2024
- [4] Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby JS, Freymann JB, Farahani K, & Davatzikos C. (2017) *Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features* Nature Scientific Data, 2017
- [5] Ronneberger O., Fisher P., Brox T. (2015) *U-Net: Convolutional Networks for Biomedical Image Segmentation* In Navab N., Hornegger J., Wells W., Frangi A. (eds) Medical Image Computing and Computer Assisted Intervention, MICCAI 2015. Lecture Notes in Computer Science, vol 9351. Springer, Cham.
- [6] Akeret J., Chihway C., Lucchi, A., & Refregier A. (2017) *Radio frequency interference mitigation using deep convolutional neural networks* Astron. Comput., vol. 18, pp. 3539, (2017)
- [7] Ren, Shaoqing, et al. (2017) *Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks*. Facebook AI Research (FAIR) 2017