MRI-Based Brain Tumor Segmentation using Mask R-CNN

Monica Del Valle
College of Engineering and Computer Science
University of Central Florida
Orlando, Florida USA
md15z@knights.ucf.edu

ABSTRACT

Accurate brain tumor segmentation from Magnetic Resonance Imaging (MRI) is critical part of improving precision oncology and the treatment of brain tumors. Automatic, deep learning methods have been shown to perform image segmentation faster than traditional machine learning methods, like support vector machines or conditional random fields. This paper presents a performance evaluation of Mask R-CNN in comparison to previously tested convolutional neural network architectures and implementations. Based on the reported results, Mask R-CNN has the potential to rival previously studied deep learning methods used for brain tumor segmentation with further experimentation and implementation improvements.

CCS CONCEPTS

• Computing methodologies ~ Machine learning approaches
• Computing methodologies ~ Neural networks

KEYWORDS

Deep learning, convolutional neural network, brain tumor segmentation, MRI, oncology

1 Introduction

Cancer is the result of uncontrolled cell division that leads to net cell growth and cell destruction [16]. A brain tumor forms when this uncontrolled cell growth occurs in brain tissue. Brain cancer is not common, making up only 1.4% of new cancer cases in 2019, but the long-term survival rate is quite low, resulting in an estimated 17,760 deaths in 2019 [1].

Brain tumor segmentation research has primarily focused on gliomas, tumors that form from glial cells, or the cells that surround neurons [1]. Gliomas are often more challenging to segment because the core tumor region diffuses into the outer edema region [16]. The ability to accurately segment gliomas allows oncologists to treat tumors precisely, minimizing harm to surrounding healthy cells. Low grade gliomas are the less aggressive malignant brain tumor classification and are characterized by slower cell division. These tumors tend to have better prognoses. High grade gliomas are the more aggressive form of brain tumor characterized by rapid cell growth and are more difficult to treat effectively [16].

MRI is the standard imaging technique used by oncologists to diagnose and treat tumors because it is widely available, and the images produce distinct contrast between healthy soft-tissue and diseased tissue [17]. There are four MRI sequences commonly used when segmenting brain tumors: T1, T1-contrast enhanced (T1-c), T2, and fluid-attenuated inversion recovery (FLAIR). Each sequence produces images with different properties to aid tumor region segmentation. T1 clearly distinguishes healthy brain tissue, T2-weighted distinguishes the edema region, or swelling caused by the tumor, and FLAIR provides a clear outline of the tumor [17]. Figure 1 provides a comparison of the four common MRI sequencing methods.

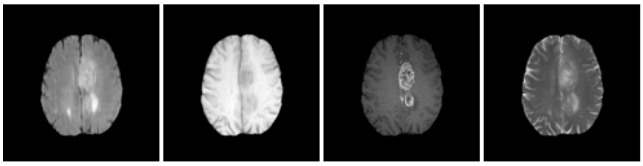

Figure 1: Comparison of MRI sequencing methods on the same sample. From left to right: T1, FLAIR, T1-c, T2

The need for different MRI sequences when analyzing brain tumors rests on the fact that brain tumors are irregular in shape, size, and position, from patient to patient. This also means that hundreds of MRI scans are required from each sequence to provide a clear picture of the brain, which adds to the complexity of the brain tumor segmentation task.

2 Related Work

2.1 Manual Segmentation

Manual segmentation is the process of segmenting the tumor from the MRI scans by hand. This process does not require any computer assistance but is very time consuming due to the large amount of MRI scans that need to be evaluated. Although manual segmentation can be extremely accurate, it is prone to inconsistent segmentation from person to person as one radiologist may segment the same tumor differently than another radiologist. These two problems inspired the need for automatic brain tumor segmentation methods.

2.2 Semi-Automatic Methods

Semi-automatic tumor segmentation methods typically rely on machine learning algorithms such as support vector machines (SVMs) and random forest classifiers [13]. In some supervised

implementations the user is required to define a region of interest (ROI) containing the tumor and set certain parameters for the classifiers to train on [2]. The user is also expected to evaluate results and provide feedback on the model's performance to aid in the learning process. These methods are able to segment tumors fairly well and reduce the amount of time to complete segmentation when compared to manual segmentation. One notable implementation is the use of concatenated random forests [14]. However, semi-automatic methods still require the input of an expert to define ROIs and provide model feedback, which still introduces inconsistent tumor segmentation.

2.3 Automatic Methods

2.3.1 Clustering. Most recent brain tumor segmentation research has been focused on improving pre-existing automatic methods or designing new ones. Traditional clustering methods such as k-means, fuzzy c-means, and region growing based clustering have been implemented and tested against manual segmentation, ground truth FLAIR images by Singh V. et al [15]. K-means clustering produced the most false-negative classifications, although it was the fastest. Fuzzy c-means and region growing-based clustering methods were comparable in terms of accuracy and speed. Automatic clustering methods reduce computational complexity, but do not perform as well when the tumor region is dispersed and highly irregular in shape. These methods tend to have higher variability in the results produced since the dependence of the tumor region and the MRI sequence features is not considered.

2.3.2 Deep Learning. A deep learning approach to brain tumor segmentation has proven to be the most promising in terms of accuracy and consistency. Convolutional neural networks (CNNs) have become a popular deep learning method for image segmentation tasks in both medical and non-medical applications. CNNs are able to learn complex features from input images. CNNs assign weights to different characteristics of the images that correspond to the importance of that characteristic. After training, CNNs are able to distinguish objects in an image based on these weighted characteristics and features.

Ali Işın et al. [12] proposed two, fully automatic CNN architectures that accurately segmented tumors into the four main tumor regions**:** edema, enhanced tumor region, tumor core (necrosis), and non-enhanced tumor region. The top performing architectures included a cascaded input architecture and a two-pathway architecture. In this implementation the channels of the CNN input data consisted of the different MRI sequences (as opposed to the typical RGB input channels), which improved the models' ability to segment the different tumor regions accurately. Both architectures maintained a specificity and sensitivity above 0.80. The models were also able to segment the tumors in 25 seconds to 3 minutes, which is much faster than manual segmentation or semi-automatic methods.

3 Method: Mask R-CNN

This paper focuses on the implementation of Mask R-CNN, which is an instance segmentation, region-based CNN [10]. Mask R-CNN was presented by the Facebook AI Research group as an extension of Faster R-CNN. The Mask R-CNN instance segmentation is performed in two stages: the bounding box proposal and the parallel prediction of a bounding box and binary mask. Figure 2 provides a high-level overview of the flow of data in Mask R-CNN.

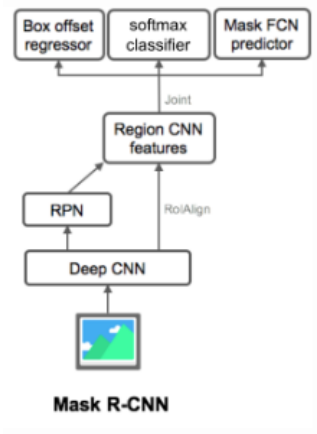

Figure 2: Overview of Mask R-CNN architecture [18]

In the first stage low-level features, like edges and corners, are initially extracted from the input image by a pretrained CNN. Mask R-CNN supports ResNet50 and ResNet101 as this backbone CNN [10]. To improve initial feature extraction, Mask R-CNN uses a Feature Pyramid Network (FPN) in addition to ResNet50 or ResNet101. Each layer in the bottom-up pathway of the FPN, the first pyramid, undergoes dimensionality reduction using 1x1 convolutions and is then used as input to the next layer [9]. The second pyramid layers receive the high-level features extracted from the first pyramid and undergo a 3x3 convolution to pass the final feature maps to lower layers. The advantage of using an FPN is that it is capable of producing feature maps for multi-scale objects [11].

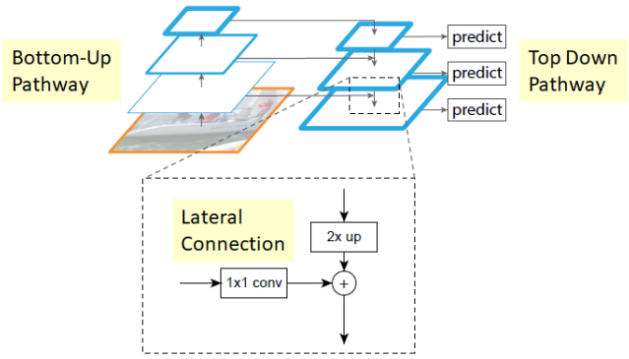

Figure 3: Feature Pyramid Network (FPN) used by Mask R-CNN [11]

Once the features have been extracted, a Region Proposal Network (RPN) scans the feature map to find ROIs. The RPN, a lightweight neural network, scans multiple square areas of varying sizes, called anchors, on the feature map in parallel, to speed up the process. The RPN scores which boxes are likely to contain an

object. The list of ROI predictions is refined using non-maximum suppression [10].

In the second stage a SoftMax ROI classifier is used to label each ROI object with a class name, completing the object detection task. The predicted bounding boxes are also refined using a bounding box regressor. Finally, the distinguishing feature of Mask R-CNN, a "soft mask" is generated for each ROI labeled as an object. A soft mask is represented by floating point values rather than integers, like in a binary mask, to produce a more detailed pixel-by-pixel mask [10].

Mask R-CNN seemed like a promising solution to accurate brain tumor segmentation due to its pixel-by-pixel masking and computationally complex ROI filtering. The use of an FPN to extract high-level and low-level features was also critical feature considered.

3.1 Implementation Details

The implementation used in this paper is an extension of the original open-source Mask R-CNN implementation on Python3, TensorFlow, and Keras. It uses a ResNet101 and FPN backbone [8]. Table 1 details some of the hyperparameters used during training of the final model.

Parameter	Value	Parameter	Value
Learning Rate	0.001	**Learning Momentum**	0.9
Batch Size	1.0	**Class Loss**	1.0
Images per GPU	1.0	**B-Box Loss**	1.0
Validation Steps	5.0	**Mask Loss**	1.0
Weight Decay	0.0001	**Anchors per Image**	64
FPN Layer Size	1024	**ROIs per Image**	32

Table 1: Hyperparameters used for final model training

The original implementation used a learning rate of 0.02, but for a small batch size of 1, this value was too large and caused weights to dramatically increase, so a learning rate of 0.001 was used instead. The batch size and number of images per GPU had to be kept small due to GPU limitations.

In an effort to reduce computation time, pretrained weights were used for model training. The weights were produced by a CNN trained on the ImageNet database, which is an object recognition database that contains millions of images [8]. Previous applications of Mask R-CNN also used these pretrained weights with some success which made them a viable option for testing.

Training took approximately 5 hours and 46 minutes on an Intel HD Graphics 4000 1536 MB. Mask predictions, or tumor segmentation, took approximately 30 seconds per image.

4 Evaluation

The experiments performed used MRI scans of real patients from the Brain Tumor Segmentation challenge 2019 (BraTS 2019) dataset. The BraTS 2019 dataset includes thousands of training instances with ground truth segmentation masks and validation instances [3, 4, 5, 6, 7]. The dataset also provides all four MRI sequences (T1, T1-c, T2, FLAIR) for the same brain sample. Figure 5 gives an example of a training instance provided by BraTS.

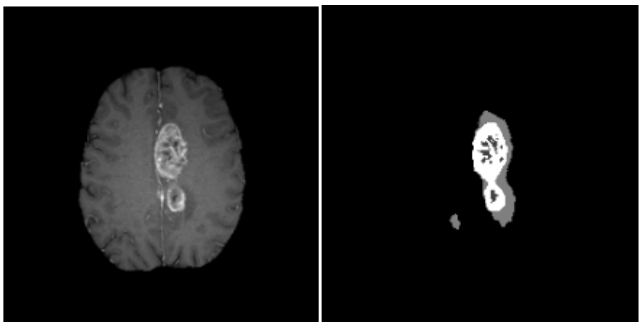

Figure 5: The image on the left is a T1-c MRI with the corresponding ground-truth segmentation mask on the right

Before model evaluation the data provided by BraTS needed to be parsed and processed to be converted to a usable format. The original image format is the Neuroimaging Informatics Technology Initiative data format (NIfTI-1) with teh file extension '.nii'. This format provides volumetric data, but for this implementation only 2D images were required. Only the axial scans (seen in Figure 1 and 5) were used and converted to portable network graphic (PNG) files with the extension '.png'. In addition to formatting, some of the training instances provided were poorly segmented or not segmented at all, hence why the data needed to be parsed. The poor quality of initially selected training data was reflected in the model's initial predictions. The model was producing an extremely large number of false positives with high confidence scores as seen in Figure 6. The masks were not refined or remotely close to segmenting the tumors after 30 epochs of training lasting about 3 hours.

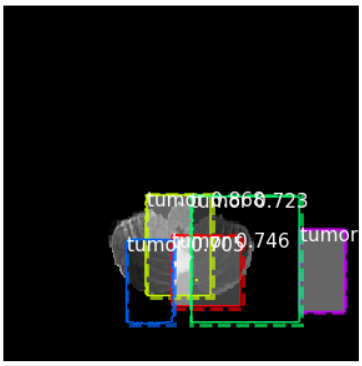

Figure 6: Initial model performance on poorly labeled training set

As a result of these initial predictions, the model needed to be retrained on a smaller and more refined dataset. Due to the limitations of the GPU available, only 23 images (20 high grade gliomas and 3 low grade gliomas) were used to train the model. The dataset was primarily represented by high-grade gliomas as they are usually more distinct in size. The T1-c sequence was chosen as input as it gave the best contrast between tumor regions and healthy tissue.

The model was trained in two stages, the first being 10 epochs training the layer heads, or top layers, and the second being 20 epochs training all layers. Both training stages used 100 steps per epoch. For this implementation the model was training to identify two classes: the tumor or background. The segmentation of the tumor into 4 distinct regions did not seem like a feasible task given the limited training resources. Having the model only segmenting the whole tumor region seemed like a more reasonable initial task. The hyperparameters used to train and test the model were modified with each experiment to produce the best results. Figure 7 is an example of the predictions produced with the default model parameters after training. The bounding box is nowhere near the tumor (upper left) and the mask produced is oddly distributed.

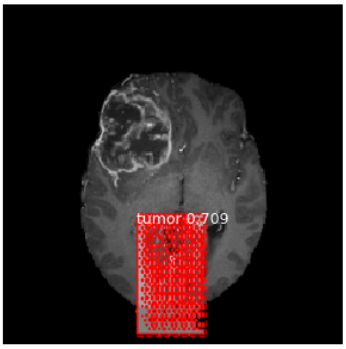

Figure 7: Default model parameter segmentation predictions miss the tumor (upper left) completely and instead attempts to mask normal brain tissue

After attempting to tune the parameters through trial and error, the model was able to better detect the tumors by placing bounding boxes around a large portion of the tumors, however the model was predicting multiple tumors in areas of the brain scans that had similar shapes and color patterns. An example is shown in Figure 8. The model predicted with 78% percent confidence that the lower right bounding box contained a tumor while the ground truth manual segmentation shows there was only one tumor in the upper left part of the brain.

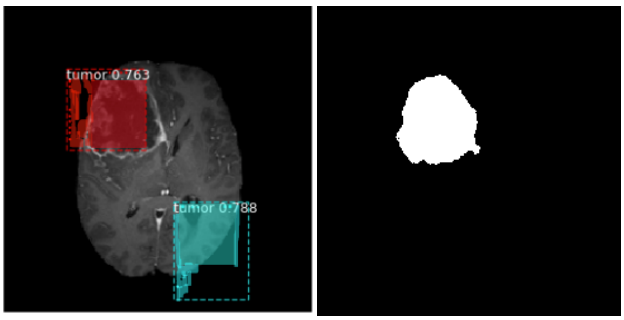

Figure 8: The model's segmentation predictions (left) versus the ground truth manual segmentation mask of the whole tumor region

4.1 Discussion

Overall, the final model did not perform well even after refining the dataset and tuning the hyperparameters. The accuracy of the model was quite low, at less than 10%. This value does not come close to the 80% accuracy score of previous deep learning methods [12]. The model's strength lies in its speed, being able to segment the images in approximately 30 seconds. However, the model was also unable to produce precise pixel-by-pixel masks around the predicted tumor regions. There are a few reasons as to why this might be the case. The model was trained on a very small dataset compared to the number of MRI scans available in the original dataset. It is possible that the model was overfitted to the training set and due to the variability of tumor shape, size, and location in the test set, the model was unable to generalize tumor features and make accurate predictions. Another reason could be the hyperparameters chosen during training. The Mask R-CNN architecture had 48 different configurable parameters available. The parameters I chose to modify and test, were only a small subset of the 48. With more testing, optimal parameters could have been identified for all 48 parameters.

The use of only one input channel also may have affected the model's performance. CNNs typically extract and learn features using the red, blue, and green channels of an image. In this case, the input images are grayscale and contain only one channel to extract features from. It is possible that using each MRI sequence as an input channel, as seen in Havaei, M. et al [12], would have helped the model learn more valuable tumor feature weights.

5 Conclusions

5.1 Mask R-CNN Performance

Brain tumor segmentation is a difficult task that has been tackled by many researchers in order to determine the most accurate

automatic methods currently available. The importance of improving brain tumor segmentation methods rests on the fact that brain cancer has a low survival rate that could be mitigated by more precise treatment. Deep learning has proven to be a promising solution as CNNs with varying architectures have successfully segmented brain tumors in previous studies [12, 13]. In this paper, I explored the performance capabilities of Mask R-CNN when being used for brain tumor segmentation. Mask R-CNN has been used to segment nuclei in cells, balloons, and geometric shapes from input images in the past, so an application in medical imaging seemed plausible [8].

The widely used brain tumor segmentation BraTS dataset was used to train and test the Mask R-CNN model. The model was able to segment gliomas with less than 10% accuracy in 30 seconds. Based on these metrics, this model does not perform as well as the current state-of-the-art deep learning methods presented by Havaei, M. et al. or the traditional supervised machine learning approaches [12, 13].

5. 2 Future Work

In order to obtain a better representation of Mask R-CNN's performance capabilities, more experimentation needs to be conducted. Future implementations should be trained on a larger dataset with a more robust GPU to account for the computational and memory intensive image processing. A cloud computing resource, like Google Colab's Tesla K80 GPU would be a viable option. A different training schedule could also be considered to better suit the larger dataset, like having more epochs and/or more steps per epoch.

In order to test how well Mask R-CNN could segment each of the four tumor regions (edema, enhanced tumor region, tumor core (necrosis), and non-enhanced tumor region) as opposed to only segmenting the whole tumor region, each MRI sequence could be used as an input channel as opposed to the single-sequence channel used in this paper. Incorporating the properties of each MRI sequence could give the model a more accurate representation of each component of the tumor. In addition, using custom initial weights rather than the ImageNet database weights is another solution to explore.

Finally, to get a more accurate evaluation of Mask R-CNN with these improvements, the model should be submitted to the official Brain Tumor Segmentation Challenge [5] to be tested and ranked among all other brain tumor segmentation methods.

REFERENCES

[1] Jovčevska, I., Kočevar, N., & Komel, R. (2013). Glioma and glioblastoma - how much do we (not) know?. *Molecular and clinical oncology*, *1*(6), 935–941. doi:10.3892/mco.2013.172

[2] Foo JL. A survey of user interaction and automation in medical image segmentation methods. Tech rep ISUHCI20062, Human Computer Interaction Department, Iowa State Univ; 2006.

[3] B. H. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, et al. "The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)", IEEE Transactions on Medical Imaging 34(10), 1993-2024 (2015) DOI: 10.1109/TMI.2014.2377694

[4] S. Bakas, H. Akbari, A. Sotiras, M. Bilello, M. Rozycki, J.S. Kirby, et al., "Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features", Nature Scientific Data, 4:170117 (2017) DOI: 10.1038/sdata.2017.117

[5] S. Bakas, M. Reyes, A. Jakab, S. Bauer, M. Rempfler, A. Crimi, 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 (2018)

[6] S. Bakas, H. Akbari, A. Sotiras, M. Bilello, M. Rozycki, J. Kirby, et al., "Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-GBM collection", The Cancer Imaging Archive, 2017. DOI: 10.7937/K9/TCIA.2017.KLXWJJ1Q

[7] S. Bakas, H. Akbari, A. Sotiras, M. Bilello, M. Rozycki, J. Kirby, et al., "Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-LGG collection", The Cancer Imaging Archive, 2017. DOI: 10.7937/K9/TCIA.2017.GJQ7R0EF

[8] Abdulla, Waleed, "Mask R-CNN for object detection and instance segmentation on Keras and TensorFlow", Github, Github repository, matterport_maskrcnn_2017, (2017), URL: https://github.com/matterport/Mask-RCNN

[9] Tsang, Sik-Ho. "Review: FPN - Feature Pyramid Network (Object Detection)." *Medium*, Towards Data Science, 20 Mar. 2019, towardsdatascience.com/review-fpn-feature-pyramid-network-object-detection-262fc7482610.

[10] He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision* (pp. 2961-2969).

[11] Lin, T. Y., Dollár, P., Girshick, R., He, K., Hariharan, B., & Belongie, S. (2017). Feature pyramid networks for object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2117-2125).

[12] Havaei, M., Davy, A., Warde-Farley, D., Biard, A., Courville, A., Bengio, Y., ... & Larochelle, H. (2017). Brain tumor segmentation with deep neural networks. *Medical image analysis*, *35*, 18-31.

[13] Işın, A., Direkoğlu, C., & Şah, M. (2016). Review of MRI-based brain tumor image segmentation using deep learning methods. Procedia Computer Science, 102, 317-324.

[14] Tustison N. et al. Optimal symmetric multimodal templates and concatenated random forests for supervised brain tumor segmentation (simplified) with antsr. Neuroinformatics 2015;13(2): 209–225.

[15] Tunga, P. P., & Singh, V. (2018, August). Brain Tumor Extraction from MRI Using Clustering Methods and Evaluation of Their Performance. In 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA) (pp. 1-5). IEEE.

[16] "Gliomas." *Johns Hopkins Medicine*, www.hopkinsmedicine.org/health/conditions-and-diseases/gliomas. (2019)

[17] Gaillard, F., Mullen, L. (2019). "MRI Sequences (Overview)". *Radiopaedia.* https://radiopaedia.org/articles/mri-sequences-overview?lang=us

[18] Weng, L. (2019). Object Detection for Dummies Part 3: R-CNN Family. [online] Lil'Log. Available at: https://lilianweng.github.io/lil-log/2017/12/31/object-recognition-for-dummies-part-3.html#mask-r-cnn [Accessed 1 Dec. 2019].