CS 484 Term Project Final Report: Semantic Segmentation of Tumors in MRI Brain Scan Images Using A Pre-Trained Convolutional Neural Network

Alkım Ege Akarsu
Electrical and Electronics Engineering
Bilkent University
Ankara, Türkiye
ege.akarsu@ug.bilkent.edu.tr

Ateş Fettahoğlu
Electrical and Electronics Engineering
Bilkent University
Ankara, Türkiye
ates.fettahoglu@ug.bilkent.edu.tr

Abstract—This report presents a comprehensive study on the application of deep learning techniques, specifically Convolutional Neural Networks, for the semantic segmentation of brain tumors in MRI scans. We utilized a pre-trained Mask R-CNN model with a ResNet-50 backbone integrated with a Feature Pyramid Network for this task. The model was trained and validated on the "TumorSeg Computer Vision Project" dataset, achieving an AP50 score of 76.1% on the test set. Our findings demonstrate the potential of deep learning in medical imaging, offering a promising approach for early and accurate diagnosis of brain tumors. We believe that deep learning computer vision models can aid healthcare professionals in lessening their workloads, and as a result, achieve better patient outcomes.

I. INTRODUCTION

The discipline of computer vision (CV) is considered to be one of the most challenging in the computer science space. This is because of many reasons, one of which is a high degree of variability of input for the same output. Differences in viewing angles, lighting, scale, obstructions, and deformations are just a few factors that increase the difficulty of many CV tasks. While hand-crafted methods are able to excel under controlled conditions, when faced with the realities of outside the laboratory, a more flexible approach has proved to be more beneficial [1].

Artificial neural networks (ANN) are universal function approximators. That is, given enough input and output data for training, ANNs are able to approximate (learn) the relationships between the input and the output no matter the mathematical nature of the relationship. These algorithms also mimic the structure of the human brain, containing "neurons" and connections between them.

Deep learning (DL) is the name given to the usage of ANNs that have one or more hidden layers. Very generally, an increase in neuron and hidden layer count increases the ability of the model to learn more complex patterns. In recent years, CV methods that employ DL have surpassed the performance of hand-crafted methods in many tasks [1]. The DL structure that is the most popular in CV use cases is the convolutional neural network (CNN). The reason for the popularity of CNNs is their ability to retain spatial information in images using their convolutional layers.

One of the fields in which CV has a large potential impact is medicine. Two examples of CV use cases in medicine can be given as face recognition for hospital security and patient identification, and noise removal from any imaging operation. The specific task we investigate in this project is semantic segmentation of brain tumors in MRI scans. Semantic segmentation is the task of labeling every pixel separately in an image. As a result, the objects of interest in an image and their location are detected.

Early diagnosis of brain tumors is vital for increasing the probability of survival in patients. However, factors such as high medical expert workload increase the probability of mistakes in diagnosis of brain tumors from MRI scan images. We hope that the CV methods investigated in this project assist medical experts with diagnosing brain tumors more accurately and with less workload.

This final report provides a detailed overview of the work done by the team over the course of the semester regarding the CS 484 term project. The purpose of this report is to present and discuss the findings related to the project topic.

II. METHODOLOGY

A. The Dataset

We use the "TumorSeg Computer Vision Project" dataset [2] for our project. It consists of 2,146 MRI brain scan images in total, 1,502 (70%) of which are for training, 429 (20%) of which are for validation, and 215 (10%) of which are for testing. All images in the dataset have a resolution of 640x640. The tumors in the images are marked with bounding boxes and the annotation data is in COCO segmentation [3] format. All similar images have the same orientations. For example, in MRI scan images taken from the side profile, the nose is always pointed upwards, or if the image is taken from the top, the nose is pointed left etc.

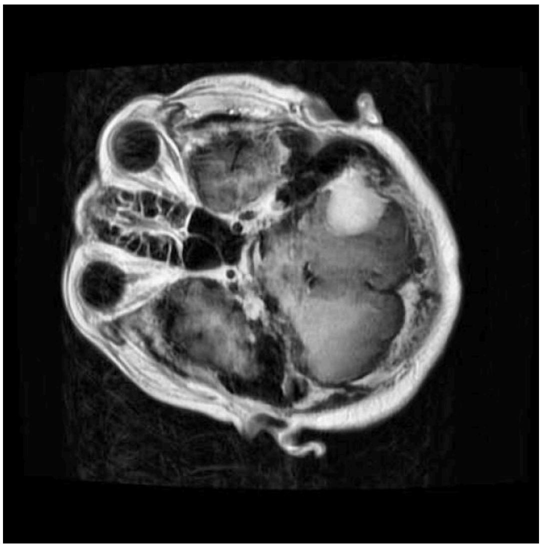

Figure 1: A sample image from the dataset.

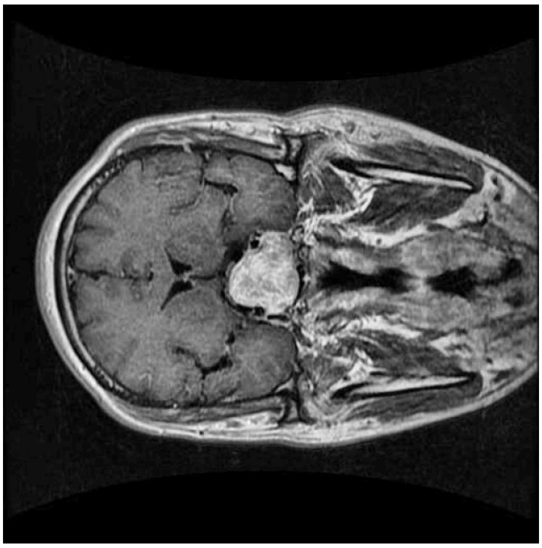

Figure 2: A sample image from the dataset.

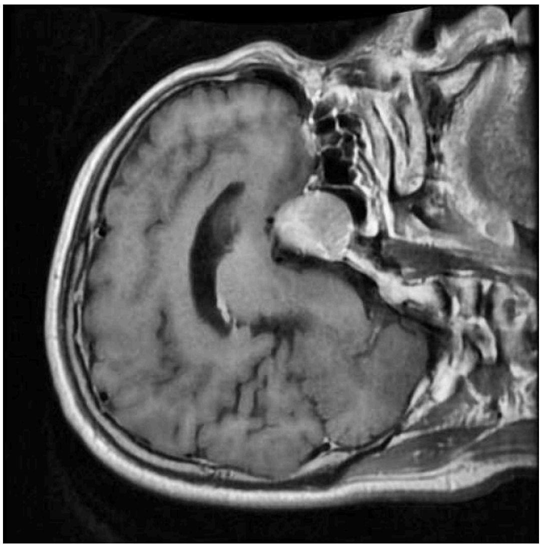

Figure 3: A sample image from the dataset.

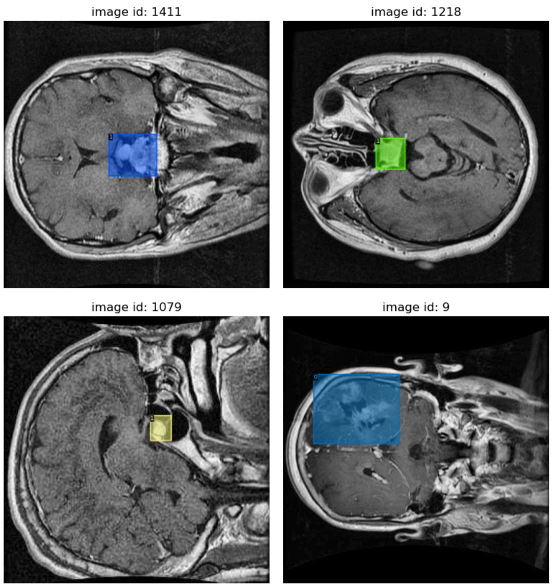

Figure 4: Some sample images with bounding boxes.

B. The Semantic Segmentation Framework

We utilize a pre-trained Mask R-CNN [4] with a ResNet-50 [5] backbone integrated with a Feature Pyramid Network (FPN) [6], leveraging the robust feature extraction capabilities of ResNet-50 and the multi-scale object detection of FPN. This combination is particularly suited for detailed semantic segmentation tasks like ours, where precision in delineating tumor boundaries in brain MRI images is crucial. The model was initially trained on the COCO 2017 dataset [7], providing a diverse feature base, which we fine-tune for our specific task of tumor segmentation.

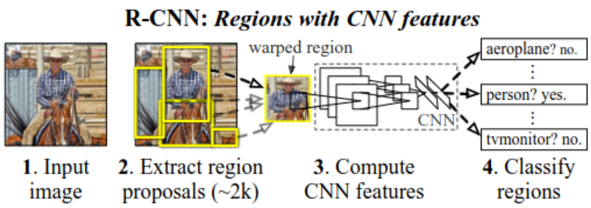

Figure 5: Overview of R-CNN [8]

To understand Mask R-CNN, we must first examine CNNs. As opposed to fully connected ANNs, CNNs apply two dimensional convolutional filters before neuron connections, which enables each new neuron weight to be affected by neighboring values. After CNNs, R-CNNs improve performance by applying CNNs to selected region proposals [8]. By classifying each region using CNNs, accuracy is greatly improved. Next, the Fast R-CNN method vastly improves the speed and accuracy of R-CNNs. Rather than applying CNNs to feature proposals like in R-CNN, Fast R-CNN employs a CNN first, eliminating the need to pass many region proposals to a CNN [9]. Next, the Faster R-CNN method greatly reduces the cost of region proposals apparent in the previous methods. This is achieved by a network called the region proposal network (fully convolutional) combined with Fast R-CNN [10]. Finally, following Faster R-CNN, Mask R-CNN is proposed. This method adds another branch to Faster R-CNN for predicting the object segmentation mask simultaneously with the bounding box [11].

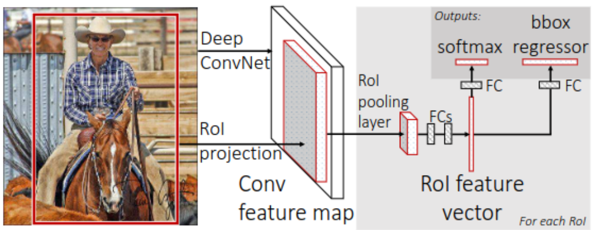

Figure 6: Overview of Fast R-CNN [9]

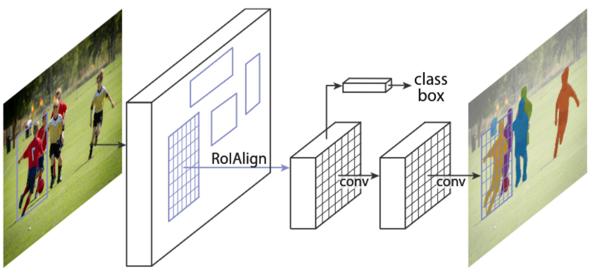

Figure 7: Overview of Mask R-CNN [11]

ResNet-50 is the CNN architecture we chose to use in the Mask R-CNN method described above. The architecture contains 50 layers and residual connections. The residual connections are identity weight links that connect every other layer. These residual connections enable the training of deeper models by avoiding the gradient vanishing/explosion problems [5].

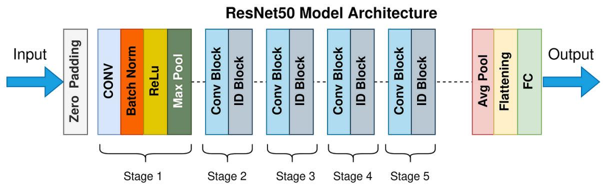

Figure 8: Overview of ResNet-50 [5]

Finally, the FPN provides the method with feature extraction capabilities at different scales, improving object detection performance for varying sizes of tumors.

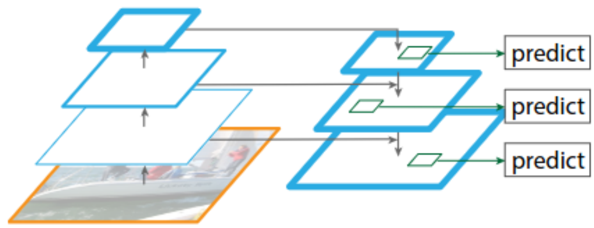

Figure 9: Overview of FPN [6]

For transfer learning, we use the ADAM optimizer [12] and do not freeze any weights.

III. RESULTS

A. Hyperparameter Tuning

To maximize the performance of our model, we tune the number of training epochs and learning rate hyperparameters. The values we try are:

- Number of training epochs: 1,000, 1,500, 2,000, 2,500, 3,000.
- Learning rates: $5\text{x}10^{-4}$, 10^{-4}, $5\text{x}10^{-5}$.

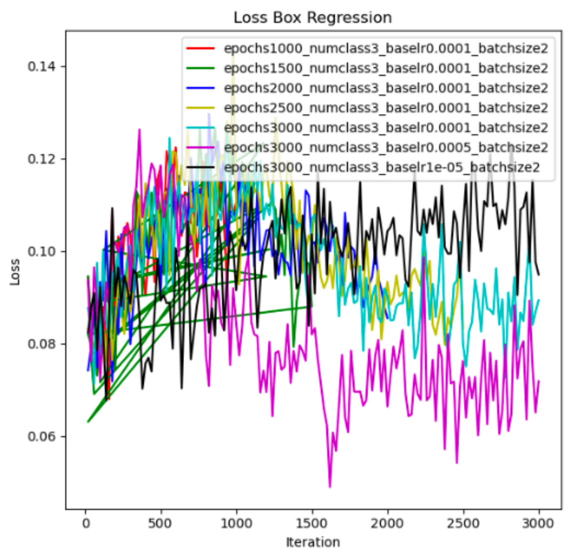

Figure 10: Loss box regression for various hyperparameters.

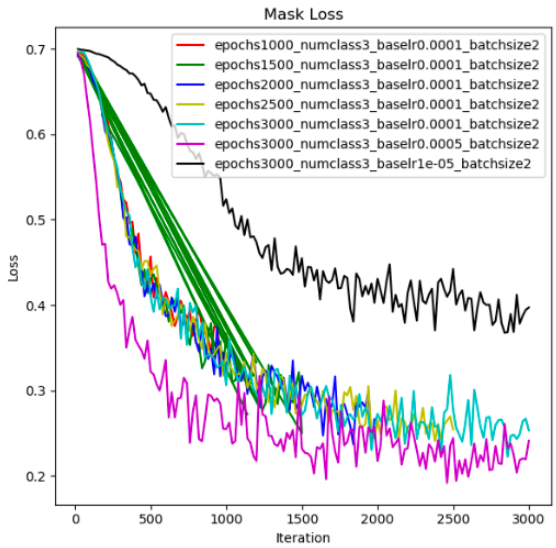

Figure 11: Mask loss for various hyperparameters.

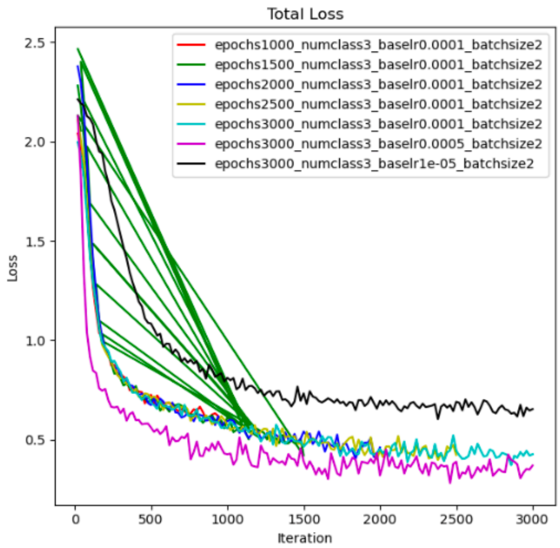

Figure 12: Total loss for various hyperparameters.

As seen from figures 10, 11, and 12 the combination of 3,000 training epochs and a learning rate of $5x10^{-4}$ provides the best performance.

B. Test Set Performance

The test set performance of the model is measured by the mean precision at various intersection over union levels.

Intersection over union is calculated by dividing the area of the intersection of the ground truth bounding box and the predicted bounding box by the area of the union of said bounding boxes.

AP	AP50	AP75	APs	APm	APl
37.8%	73.1%	33.7%	NaN	32.3%	42.3%

Table 1: Test scores of the finalized model.

The scores in table 1 are in line with other models found in Kaggle [13].

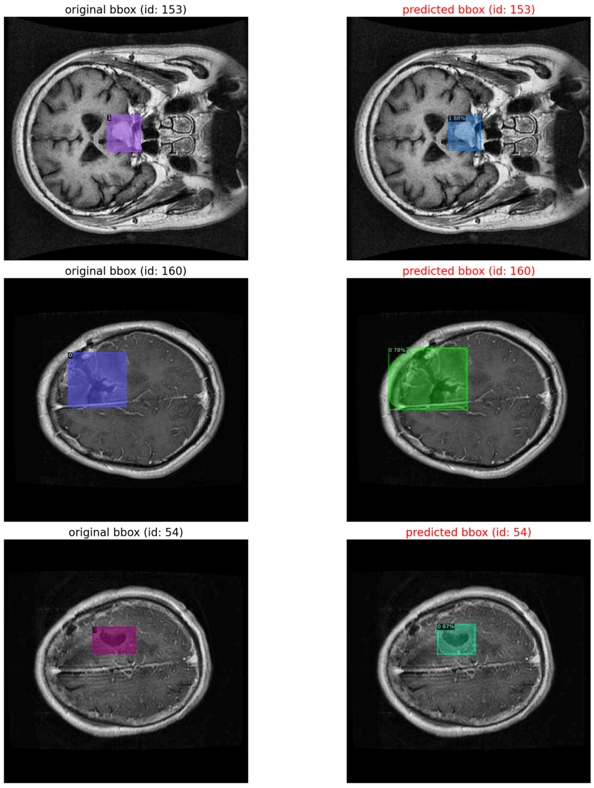

Figure 13: Sample True and predicted bounding boxes

IV. DISCUSSION

The results of our project demonstrate the potential of deep learning, specifically convolutional neural networks, in the field of medical imaging. Our model, a pre-trained Mask R-CNN with a ResNet-50 backbone integrated with a Feature Pyramid Network, was able to achieve an AP50 score of 76.1% on the test set. This performance is comparable to other models found on Kaggle.

However, it's important to note that while our model shows promise, there are several areas for improvement. The variability in the size and shape of tumors presents a significant challenge in accurately segmenting them from MRI scans. Additionally, the quality and resolution of the images can greatly affect the performance of the model. Future work could explore the use of more sophisticated image preprocessing techniques.

V. CONCLUSION

In conclusion, this project has shown that deep learning methods can be effectively applied to the task of semantic segmentation of brain tumors in MRI scans.

While the results are promising, it is important to remember that the ultimate goal is to provide a tool that can assist medical professionals in diagnosing brain tumors. With continued refinement and validation, we believe that CV models have the potential to become a valuable tool in early brain tumor detection, ultimately contributing to improved patient outcomes.

VI. REFERENCES

[1] Z. Li, F. Liu, W. Yang, S. Peng, and J. Zhou, "A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects," IEEE Transactions on Neural Networks and Learning Systems, vol. 33, no. 12, pp. 1–21, 2021, doi: https://doi.org/10.1109/tnnls.2021.3084827.

[2] "TumorSeg Semantic Segmentation Dataset and Pre-Trained Model by Seg," Roboflow. https://universe.roboflow.com/seg-aokuq/tumorseg

[3] "COCO - Common Objects in Context," cocodataset.org. https://cocodataset.org/#format-data

[4] K. He, G. Gkioxari, P. Dollar, and R. Girshick, "Mask R-CNN," IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 1–1, 2018, doi: https://doi.org/10.1109/tpami.2018.2844175.

[5] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," arXiv (Cornell University), Dec. 2015, doi: https://doi.org/10.48550/arxiv.1512.03385.

[6] T.-Y. Lin, Piotr Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature Pyramid Networks for Object Detection," arXiv (Cornell University), Dec. 2016, doi: https://doi.org/10.48550/arxiv.1612.03144.

[7] "COCO - Common Objects in Context," cocodataset.org. https://cocodataset.org/#download

[8] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," arXiv (Cornell University), Nov. 2013, doi: https://doi.org/10.48550/arxiv.1311.2524.

[9] R. Girshick, "Fast R-CNN," 2015 IEEE International Conference on Computer Vision (ICCV), pp. 1440–1448, Dec. 2015, doi: https://doi.org/10.1109/iccv.2015.169.

[10] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 6, pp. 1137–1149, Jun. 2017, doi: https://doi.org/10.1109/tpami.2016.2577031.

[11] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," arXiv.org, Jan. 24, 2018. https://arxiv.org/abs/1703.06870v3

[12] D. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," Computer Science, 2014, doi: https://doi.org/10.48550/arXiv.1412.6980.

[13] "Brain Tumor Image DataSet : Semantic Segmentation," www.kaggle.com. https://www.kaggle.com/datasets/pkdarabi/brain-tumor-image-dataset-semantic-segmentation/code (accessed May 16, 2024).