Brain Tumor Segmentation Project Final Report

Jeong-Wan Choi
Boston College
choiqz@bc.edu

Steven Kim
Boston College
kimbqf@bc.edu

Andrew Yang
Boston College
yangns@bc.edu

1. Introduction

Brain cancer affected approximately 24,810 adults in the United States in 2023. Accurate detection and monitoring of brain tumors are crucial for diagnosis, treatment planning, and follow-up. Magnetic Resonance Imaging (MRI) is a widely used imaging modality for visualizing brain tumors, but manual segmentation of tumors in MRI scans is time-consuming, labor-intensive, and prone to human error. To address this challenge, researchers have been exploring the use of deep learning techniques, particularly convolutional neural networks (CNNs), for automated brain tumor segmentation. By training neural networks on large datasets of annotated MRI scans, it is possible to develop models that can accurately and efficiently segment brain tumors, reducing the workload of medical experts and improving the consistency of tumor delineation. In this study, we propose to investigate and compare the performance of several state-of-the-art CNN architectures, including U-Net, ResNet, DenseNet, and the Segment Anything Model (SAM), for brain tumor segmentation in MRI scans. By conducting a comprehensive evaluation of these architectures on a dataset of brain MRI scans, we aim to identify the most effective model for automated brain tumor segmentation. The findings of this study have the potential to streamline the tumor segmentation process, improve the accuracy of tumor delineation, and ultimately enhance the management and treatment of patients with brain cancer.

2. Related Work

Machine learning techniques for segmenting brain tumors with MRIs have had many previous developments. Generative Adversarial Networks (GANs) have been used to generate high quality synthetic images. However, there are issues with complex training requirements as well as the risk of mode collapse. ResNets and DenseNets have also been popular for efficiently training deeper networks. We incorporate ResNet and DenseNet ideas into U-Net architecture to capture some of the efficiency. Multimodal approaches have also been tried in the past, but it seems that U-Net is currently the most preferred model. We attempt to leverage the pretrained power of Meta's Segment Anything Model, to compare how general segmentation models can stand up against more specialized architectures.

3. Methods

Various methods were used in our project:

1. U-Net: Baseline model we used as a comparison to the ResNet and DenseNet architectures. The Attention U-Net is an extension of the popular U-Net architecture that incorporates attention gates. These attention gates allow the model to focus on relevant features at each level of the encoder-decoder structure. The attention mechanism helps the network highlight important regions and suppress irrelevant areas, leading to improved segmentation accuracy, especially for medical image segmentation tasks.

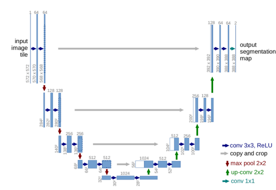


Figure 1. U-Net Architecture [5]

2. ResNet: We used ResNet because it introduces the concept of residual connections to enable training of very deep CNNs. The core idea is to add skip connections that bypass one or more layers, allowing the network to learn residual functions with reference to the layer inputs. This helps alleviate the vanishing gradient problem and allows for much deeper networks (e.g., 50, 101, 152 layers) to be trained effectively. ResNets have also previously achieved state-of-the-art performance on various computer vision tasks.

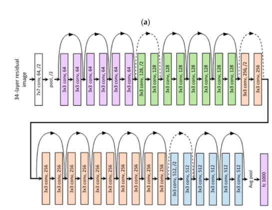


Figure 2. ResNet Architecture [1]

3. DenseNet: DenseNet is an architecture that focuses on feature reuse by connecting each layer to every other layer in a feed-forward fashion. In DenseNets, each layer receives the feature maps of all preceding layers as input, allowing for efficient information flow and gradient propagation. This

dense connectivity pattern promotes feature reuse, reduces the number of parameters, and mitigates the vanishing gradient problem. DenseNets have also shown excellent performance on image classification tasks.

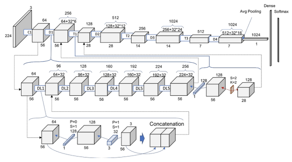


Figure 3. DenseNet Architecture [4]

4. SAM (Segment Anything Model): SAM is a recently proposed model by Meta AI that aims to segment any object in an image using a promptable image embedding and a lightweight mask decoder. The key idea is to train the model on a large dataset of images and masks, allowing it to learn a generic image embedding that can be used for segmenting objects based on simple prompts like points or boxes. SAM consists of an image encoder, a prompt encoder, and a mask decoder. The image encoder generates a feature map, the prompt encoder processes the input prompt, and the mask decoder predicts the segmentation mask based on the image features and the prompt embedding.

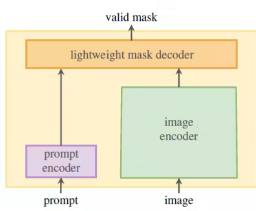


Figure 4. SAM Model Components [2]

5. MedSAM (Medical Segment Anything Model): MedSAM is an adaptation of the Segment Anything Model (SAM) specifically tailored for medical image segmentation tasks. It leverages the powerful segmentation capabilities of SAM and fine-tunes the model on medical imaging datasets to capture domain-specific features and anatomical structures. MedSAM maintains the key components of SAM, including the image encoder, prompt encoder, and mask decoder, but is trained on a diverse set of medical images and corresponding segmentation masks. The image encoder in MedSAM is typically a convolutional neural network (CNN) backbone pre-trained on large-scale medical imaging datasets to learn meaningful representations. The prompt encoder processes user-provided prompts, such as points or bounding boxes, to guide the segmentation process. The mask decoder takes the image features and prompt embeddings as input and generates the final segmentation mask. MedSAM has shown promising results in segmenting various anatomical structures, lesions, and abnormalities across different medical imaging modalities, such as MRI, CT, and X-ray. By leveraging the power of promptable segmentation, MedSAM enables efficient and accurate segmentation of medical images with minimal user input, making it a valuable tool for clinical applications and research.

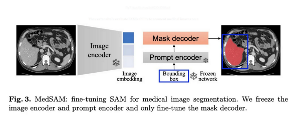


Figure 5. MedSAM Model Components [3]

4. Experiments

These are some experiments performed to improve model performance:

1. Data: Our dataset was taken from Kaggle and consists of around 2000 images of 110 different patients. These images comprise brain MR (Magnetic Resonance) images and their corresponding manual segmentation masks for FLAIR (Fluid-Attenuated Inversion Recovery) abnormality regions. The data was sourced from The Cancer Imaging Archive (TCIA) and is associated with patients included in The Cancer Genome Atlas (TCGA) lower-grade glioma collection.

2. Comparing U-Net to ResNet and DenseNet: We fine-tuned and experimented with different hyperparameters in our project. Different learning rates, such as 1e-3, 1e-4, and 1e-5, to find the optimal value. We also evaluated different batch sizes, including 16, 32, and 64, to balance memory constraints and training efficiency. We trained the models for a standard 50 epochs to observe the impact on performance. To evaluate the performance of the segmentation models, we used the Dice Coefficient, which measures the overlap between the predicted segmentation mask and the ground truth mask. It ranges from 0 to 1, with 1 indicating perfect overlap. We also used Intersection over Union (IoU), which calculates the ratio of the intersection between the predicted and ground truth masks to their union. It provides a measure of segmentation accuracy.

   Below is a table detailing the performance of the ResNet and DenseNet with the Attention U-Net model. All models were trained with the same batch sizes, learning rate $1e-3$, number of epochs (50), and Adam optimizer. We compare the mean training DICE, mean validation DICE, and mean IOU for the three architectures.

Table 1. Performance of U-Net compared to ResNet and DenseNet

Architecture	**Train DICE**	**Val. DICE**	**Mean IOU**
Attention U-Net	0.7902	0.8344	91.0%
ResNet	0.8691	0.8994	94.0%
DenseNet	0.8671	0.8938	94.0%

3. SAM: For the hyperparameters, the batch size is 6 and the learning rate is 0.005. We tested on 1, 5, and 10 epochs to observe the progression and accuracy of the model. It uses DICE Loss to evaluate the performance of the segmentation masks. Although SAM is pre-trained with fantastic general segmentation abilities, it is not the most useful for medical domain specific tasks. Perhaps because its original training dataset completely differs from the form that MRIs are in, it doesn't compare well against more popular methods of MRI segmentation, such as U-Net.

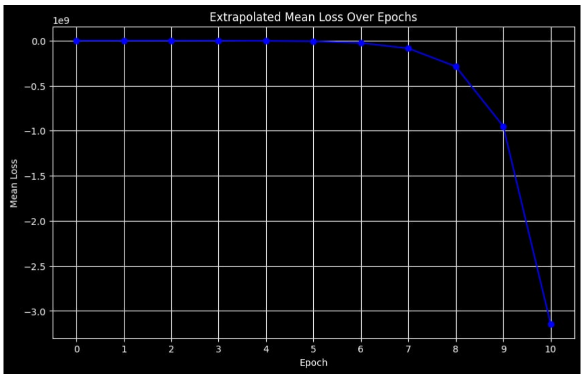
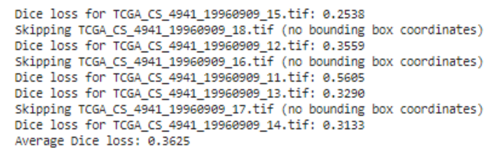


Figure 6. SAM Loss

4. MedSAM: Since MedSAM is already fine-tuned, we just want to check the average mean iou loss for pictures to compare with SAM and the other architectures. We used selections of the TGZ dataset that is consistent with what SAM and UNet uses. The data being used consists of tif files containing medical images (presumably from the TCGA dataset, based on the file names) and their corresponding ground truth mask files. The specific files mentioned are:
TCGA_CS_4941_19960909_11.tif,
TCGA_CS_4941_19960909_11_mask.tif
TCGA_CS_4941_19960909_12.tif,
TCGA_CS_4941_19960909_12_mask.tif
TCGA_CS_4941_19960909_13.tif,
TCGA_CS_4941_19960909_13_mask.tif
TCGA_CS_4941_19960909_14.tif,
TCGA_CS_4941_19960909_14_mask.tif
CGA_CS_4941_19960909_15.tif,
TCGA_CS_4941_19960909_15_mask.tif
Suppose that there were more files put in, we would have a more accurate score for our loss. However, we would have to find a way to automate the bounding box because the job manually was tedious. For Hyperperameters, the medsam model just uses the same structure as SAM but with its checkpoint, which likely has its own set of hyperparameters that are determined during training. Evaluation - the model is evaluated by comparing its segmentation output (medsam_seg) with the ground truth mask (gt_mask) using the Dice loss metric. This is to be consistent in measuring accuracy across all three models.



```
Dice loss for TCGA_CS_4941_19960909_15.tif: 0.2538
Skipping TCGA_CS_4941_19960909_18.tif (no bounding box coordinates)
Dice loss for TCGA_CS_4941_19960909_12.tif: 0.3559
Skipping TCGA_CS_4941_19960909_16.tif (no bounding box coordinates)
Dice loss for TCGA_CS_4941_19960909_11.tif: 0.5605
Dice loss for TCGA_CS_4941_19960909_13.tif: 0.3290
Skipping TCGA_CS_4941_19960909_17.tif (no bounding box coordinates)
Dice loss for TCGA_CS_4941_19960909_14.tif: 0.3133
Average Dice loss: 0.3625
```

Figure 7. MedSAM Loss

5. Conclusion

In this study, we investigated and compared the performance of several respectable CNN architectures, including Attention U-Net, ResNet, DenseNet, SAM, and MedSAM, for brain tumor segmentation in MRI scans. Our experiments demonstrated that the ResNet and DenseNet architectures, when incorporated into the U-Net framework, outperformed the baseline Attention U-Net model in terms of both validation and training Dice coefficients. ResNet achieved the highest mean validation Dice score of 0.8994, followed closely by DenseNet with 0.8938, while Attention U-Net obtained 0.8344. The results highlight the effectiveness of residual connections and dense connectivity patterns in capturing relevant features and improving segmentation accuracy. However, the SAM and MedSAM models, despite their success in general segmentation tasks, did not perform as well as the specialized U-Net-based architectures in the domain-specific task of brain tumor segmentation. This suggests that further fine-tuning and adaptation of these models to the medical imaging domain may be necessary. Future work should explore techniques for effectively leveraging the pre-trained knowledge of SAM and MedSAM while incorporating domain-specific features and anatomical priors. Additionally, the reproducibility of our work can be enhanced by providing detailed documentation of the dataset, preprocessing steps, and hyperparameter settings used in our experiments. Overall, this study contributes to the ongoing research on automated brain tumor segmentation and highlights the potential of deep learning techniques in assisting medical professionals in the diagnosis and treatment planning of brain cancer.

6. Contribution

List of our individual contributions:

Jeong-Wan Choi: Wrote related work and conclusion, evaluated and worked on MedSAM and SAM
Steven Kim: Worked and wrote on methods/experiments for SAM and MedSAM
Andrew Yang: Worked and wrote on methods/experiments for UNet, ResNet, and DenseNet, wrote introduction

GitHub Repo: https://github.com/AndrewY7/CSCI3397FP

References

[1] Bangar, Siddhesh. "Resnet Architecture Explained." Medium, Medium, 5 July 2022,

medium.com/@siddheshb008/resnet-architecture-explained-47309ea9283d.

[2] "Meta Ai's Segment Anything Model (Sam) Explained: The Ultimate Guide." Encord, encord.com/blog/segment-anything-model-explained/. Accessed 12 May 2024.

[3] Pandey, Shubham. "Revolutionizing Medical Image Segmentation: Unveiling Medsam and the Future of Diagnosis." Medium, Medium, 24 Jan. 2024, medium.com/@spandey8/revolutionizing-medical-image-segmentation-unveiling-medsam-and-the-future-of-diagnosis-de444649694c.

[4] Ruiz, Pablo. "Understanding and Visualizing DenseNets." Medium, Towards Data Science, 18 Oct. 2018, towardsdatascience.com/understanding-and-visualizing-densenets-7f688092391a.

[5] Zhang, Jeremy. "UNet Line by Line Explanation." Medium, Towards Data Science, 18 Oct. 2019, towardsdatascience.com/unet-line-by-line-explanation-9b191c76baf5.