Tyree McCloud

Dhruvil Prafulkumar kachchhi

MRI Segmentation Report

The objective of this project was to develop a machine learning model capable of performing binary semantic segmentation on brain MRI images to identify the location of tumors. Unlike classification tasks that only determine whether a tumor is present, semantic segmentation produces a mask that outlines the exact tumor shape and location. By implementing a full pipeline that includes data preprocessing, model design, training, and metric evaluation, this project demonstrates how convolutional neural networks can be applied to medical imaging tasks.

The dataset used in this project consisted of MRI brain scans paired with binary ground-truth segmentation masks. Because the dataset is organized in COCO-style JSON format, preprocessing involved parsing the JSON file to map each image ID to its tumor annotations. In addition to mask generation, preprocessing included normalization and data augmentation. All MRI images were loaded, resized to the model input dimension, normalized to the range [0,1], and converted to PyTorch tensors. When augmentation was enabled, random vertical flips, horizontal flips, and 90-degree rotations were applied simultaneously to both the image and the mask. These techniques increased dataset variability and helped the model generalize better to unseen MRI scans.

The core model used for this project was a customized U-Net architecture, chosen because it is highly effective for biomedical segmentation tasks where spatial details must be preserved. U-Net follows an encoder–decoder structure with skip connections.
The encoder consisted of four levels of convolutional feature extraction blocks that progressively reduced spatial resolution using max pooling. The number of feature channels expanded from 32 up to 256, enabling the encoder to capture detailed contextual information about tumor shapes and more. At the center of the network, a 512-channel bottleneck performed the deepest level of feature processing before reconstruction began. The decoder then gradually restored spatial resolution using transposed convolutions. Importantly, skip connections concatenated encoder

outputs with decoder inputs at matching scales. These connections allowed the model to recover the fine-grained spatial details lost during down sampling, to attempt to ensure that tumor boundaries and small structures remained sharp in the final segmentation mask. The final output layer was a single 1 x 1 convolution producing a single-channel logits map representing the pixel-wise tumor probability. This raw output was passed through a sigmoid function during evaluation to convert logits to probabilities. Post-processing techniques were also implemented to improve mask quality

To optimize accuracy and handle class imbalance between tumor and non-tumor pixels, multiple loss functions were experimented with. Dice Loss was implemented because it directly maximizes the overlap between prediction and ground truth. A combined BCE + Dice loss was also introduced to stabilize training by balancing pixel-level accuracy and overlap. Model training was performed using a batch size of 1 due to GPU memory constraints. The learning rate was set to .0004. And training ran for 25 epochs.

To measure segmentation quality, multiple evaluation metrics were computed during validation, including Dice Coefficient, Intersection over Union (IoU), and pixel accuracy. Dice and IoU directly measure the overlap between predicted and ground-truth tumor regions, making them strong indicators for medical segmentation tasks. After thresholding the sigmoid output, additional post-processing cleanup was applied before computing metrics to ensure fair evaluation of the model's true structural performance. Over the training epochs, these metrics were monitored to detect overfitting and ensure the model learned to generalize to new images.

The final model produced mixed results across the three-evaluation metrics. While the pixel accuracy reached 0.97, indicating that the model correctly classified many pixels, this metric alone can be misleading due to the heavy class imbalance between tumor and non-tumor regions. More meaningful metrics for segmentation were Dice coefficient ≈ 0.53 and Intersection over Union ≈ 0.50, these show that the model successfully identified the tumor regions but struggled with being precise about complete region coverage. This means that the network learned the general shape and location of tumors but had difficulty segmenting. To improve performance toward an IoU of 0.70 or higher, several enhancements could be implemented, using a larger batch size or gradient accumulation for more stable updates, adding attention

mechanisms to help the model focus on subtle tumor regions, strengthening data augmentation to improve generalization, or using more advanced loss functions to handle class imbalance more effectively. With these improvements, the segmentation quality could become significantly more accurate.