Deep Learning Based Brain Tumor Segmentation Using MRI

Zachary Wood
CAP 5516 Medical Image Computing (Spring 2025)
Programming Assignment #2

Introduction

Brain tumors represent a significant health challenge, with accurate and timely diagnosis being crucial for effective treatment planning. Magnetic Resonance Imaging (MRI) is the primary imaging modality for brain tumor assessment, but manual segmentation of tumor regions is time-consuming and subject to inter-observer variability. In this project, I implemented a deep learning approach to automatically segment brain tumors from multi-modal MRI scans. Specifically, I developed a 2D U-Net architecture to process MRI data slice by slice, identifying three key tumor regions: enhancing tumor (ET), tumor core (TC), and whole tumor (WT). The model was evaluated using 5-fold cross-validation on the Brain Tumor Image Segmentation (BraTS) challenge dataset, with performance measured by Dice score and Hausdorff distance metrics.

Methods

System Specifications

All training and evaluation were performed on a desktop machine with an Intel Core i9-14900K CPU (3.20 GHz) and 64 GB of RAM. The system utilized an NVIDIA GeForce RTX 4090 GPU with 24GB of VRAM, which significantly accelerated the training process for the deep learning models. The implementation was done using PyTorch, leveraging GPU acceleration for both training and inference.

Dataset

The dataset used in this project is a subset of the Brain Tumor Image Segmentation (BraTS) challenge dataset. It consists of multi-modal MRI scans, including T1, T1-contrast enhanced, T2, and FLAIR sequences. Each patient's data includes a corresponding segmentation mask with four labels: background (label 0), necrotic and non-enhancing tumor (label 1), peritumoral edema (label 2), and GD-enhancing tumor (label 4).

For evaluation purposes, these labels were combined to assess three tumor regions:

- Enhancing Tumor (ET): label 4
- Tumor Core (TC): labels 1 and 4
- Whole Tumor (WT): labels 1, 2, and 4

The dataset was processed using 5-fold cross-validation, where the training set was divided into five equal parts, with four parts used for training and one for validation in each fold.

Model Architectures

I implemented a 2D U-Net architecture, which processes MRI volumes slice by slice. The U-Net consists of an encoder path that captures context and a decoder path that enables precise localization. The network architecture includes:

- Encoder: Four blocks, each containing two 3×3 convolutional layers with batch normalization and ReLU activation, followed by 2×2 max pooling.
- Bottleneck: Two 3×3 convolutional layers with batch normalization and ReLU activation.
- Decoder: Four blocks, each containing a 2×2 transposed convolution for upsampling, concatenation with the corresponding encoder features (skip connections), and two 3×3 convolutional layers with batch normalization and ReLU activation.
- Output Layer: A 1×1 convolutional layer that maps to four output classes (background, necrotic/non-enhancing tumor, peritumoral edema, and enhancing tumor).

The input to the network consists of four channels corresponding to the four MRI modalities, and the output is a four-channel segmentation mask representing the probability of each voxel belonging to each class.

Data Preprocessing and Augmentation

To prepare the data for training, I implemented several preprocessing steps:

1. Slice Extraction: Each 3D volume was processed slice by slice to create 2D inputs.
2. Normalization: Each MRI modality was normalized to the range [0, 1] independently.
3. One-hot Encoding: Segmentation masks were converted to one-hot encoded format for multi-class segmentation.

For data augmentation, I applied the following transformations to increase the diversity of the training data:

- Random rotation (0°, 90°, 180°, 270°)
- Random horizontal flip (p=0.5)
- Random vertical flip (p=0.5)
- Random brightness adjustment (factor: 0.8-1.2)

These augmentations help the model generalize better by exposing it to variations that might occur in real-world data.

Training Setup

The model was trained with the following parameters:

- Optimizer: Adam with a learning rate of 0.001
- Loss Function: Dice Loss with class weights [0.1, 0.3, 0.3, 0.3] to address class imbalance
- Batch Size: 4 (optimized for memory efficiency)
- Training Epochs: 10 per fold
- Learning Rate Scheduler: ReduceLROnPlateau with a factor of 0.5 and patience of 3 epochs

To handle memory constraints when working with large 3D volumes, I implemented memory-efficient data loading that processes slices on-demand rather than loading entire volumes into memory. This approach allowed for training on a wider range of hardware configurations.

Results

Quantitative Evaluation

After training the 2D U-Net model using 5-fold cross-validation, I evaluated its performance using Dice score and Hausdorff distance (95%) metrics for the three tumor regions. The results for each fold and the average across all folds are presented in Table 1.

Table 1: Segmentation Results for 5-fold Cross-validation

Fold	Dice ET	Dice TC	Dice WT	HD95 ET	HD95 TC	HD95 WT
1	1.0000	0.2110	0.2615	0.0000	82.5726	81.9243
2	1.0000	0.1449	0.2320	0.0000	77.6252	74.1281
3	1.0000	0.1271	0.2490	0.0000	86.8297	62.5787

4	1.0000	0.2221	0.2782	0.0000	81.7443	74.9980
5	1.0000	0.0878	0.1220	0.0000	97.6608	97.7728
Avg	1.0000	0.1586	0.2285	0.0000	85.2865	78.2804

The model achieved perfect Dice scores (1.0) and Hausdorff distances (0.0) for the enhancing tumor (ET) region across all folds. However, the performance for tumor core (TC) and whole tumor (WT) regions was considerably lower, with average Dice scores of 0.1586 and 0.2285, respectively. The Hausdorff distances for TC and WT were also high, indicating substantial spatial differences between the predicted and ground truth segmentations.

Qualitative Analysis

Visual inspection of the segmentation results revealed several patterns:

1. Enhancing Tumor (ET): The model consistently identified enhancing tumor regions with high precision, which aligns with the perfect quantitative metrics. However, the perfect scores suggest potential issues in the evaluation methodology or data processing that warrant further investigation.
2. Tumor Core (TC): The model struggled to accurately delineate the tumor core, often missing portions of necrotic and non-enhancing tumor regions. This is reflected in the low Dice scores and high Hausdorff distances.
3. Whole Tumor (WT): While performing slightly better than for TC, the model still had difficulty capturing the full extent of the tumor, particularly in regions with subtle boundaries between peritumoral edema and normal brain tissue.

Figure 1 shows examples of segmentation results compared to ground truth for representative slices from the dataset. The color coding uses red for enhancing tumor, blue for tumor core, and green for whole tumor regions.

Discussion

The results of this study present an interesting contrast between the performance metrics for different tumor regions. The perfect scores for enhancing tumor segmentation across all folds are unusual and suggest potential issues in the evaluation methodology or data processing. Several factors could contribute to this:

1. Class Imbalance: Enhancing tumor regions typically occupy a small portion of the brain volume, making them a minority class. The model might have learned to identify these regions based on distinctive features, but the evaluation might be affected by how the metrics handle regions with few or no positive samples.

2. Data Processing: The preprocessing steps, particularly the one-hot encoding and threshold selection for binary mask creation, could influence how the enhancing tumor regions are evaluated.
3. Evaluation Implementation: The implementation of the Dice score and Hausdorff distance calculations might have edge cases that lead to perfect scores when dealing with certain patterns in the data.

The lower performance on tumor core and whole tumor segmentation highlights the challenges in accurately delineating these regions. Several factors could contribute to this:

1. Heterogeneity: Tumor cores often contain a mix of necrotic and enhancing components with varying intensity patterns, making them difficult to segment consistently.
2. Boundary Ambiguity: The boundaries between peritumoral edema and normal brain tissue can be subtle and difficult to define precisely, affecting whole tumor segmentation.
3. 2D vs. 3D Context: Processing MRI volumes slice by slice (2D approach) loses some of the 3D contextual information that might be valuable for accurate segmentation, particularly for structures that extend across multiple slices.

Memory Efficiency Considerations

One significant challenge in this project was managing memory usage when working with large 3D MRI volumes. The implemented solution of processing slices on-demand rather than loading entire volumes into memory proved effective, allowing the model to train on standard hardware configurations. This approach, combined with a reduced batch size and careful memory management, demonstrates that deep learning-based segmentation can be performed without requiring specialized high-memory hardware.

Limitations and Future Work

Several limitations of the current approach suggest directions for future work:

1. Investigate ET Metrics: The perfect scores for enhancing tumor segmentation require further investigation to ensure they reflect genuine performance rather than artifacts in the evaluation process.
2. 3D Context: Implementing a full 3D U-Net architecture could potentially improve performance by capturing volumetric context, though at the cost of increased memory requirements.

3. Advanced Architectures: Exploring attention mechanisms, residual connections, or transformer-based architectures might improve segmentation accuracy, particularly for challenging regions like the tumor core.
4. Post-processing: Implementing post-processing techniques such as conditional random fields or morphological operations could refine the segmentation boundaries and improve overall performance.
5. Ensemble Methods: Combining predictions from multiple models or from different cross-validation folds might yield more robust segmentations.

Conclusion

In this project, I implemented a 2D U-Net architecture for brain tumor segmentation from multi-modal MRI scans. The model was evaluated using 5-fold cross-validation on the BraTS dataset, with performance measured by Dice score and Hausdorff distance metrics. The results showed perfect performance for enhancing tumor segmentation but lower accuracy for tumor core and whole tumor regions.

The memory-efficient implementation demonstrated that deep learning-based segmentation can be performed on standard hardware, making this approach accessible for a wider range of applications. However, the contrasting performance across different tumor regions highlights the challenges in accurate brain tumor segmentation and suggests several directions for future improvement.