ECOLE POLYTECHNIQUE DE BRUXELLES

Computing project — Biomedical engineering project in image analysis

PROJ-H402 — PROJ-H419

Brain Tumor Segmentation in MRI using an U-Net Architecture

Authors:

Kubilay Alpaslan

Academic Year 2024-2025

Brain Tumor Segmentation in MRI using an U-Net Architecture

Alpaslan Kubilay
*Université Libre de Bruxelles

***Abstract*—This project presents a deep learning approach for segmenting brain tumors from MRI scans, implemented as part of a Master's year project. A 2D convolutional neural network based on the U-Net architecture is trained and evaluated on the BraTS 2021 dataset. The project aims to provide a reliable method for medical image segmentation using publicly available resources. The proposed pipeline handles class imbalance with a weighted cross-entropy loss function and assesses performance using Dice scores. Despite the use of a single MRI modality (FLAIR), the model shows promising results and demonstrates the potential of lightweight architectures for practical medical applications.**

***Index Terms*—brain tumor segmentation, U-Net, deep learning, MRI, Dice score, PyTorch, semantic segmentation**

I. Introduction

Brain tumors are complex medical conditions that often require advanced imaging techniques for diagnosis and treatment planning. MRI scans are widely used due to their ability to capture soft tissue details without exposing patients to radiation. However, the manual analysis of MRI volumes is both time-consuming and subject to human variability [1]. As an alternative, automatic segmentation using deep learning has emerged as a powerful solution.

This project was conducted as part of a Master's program in engineering and computer vision. The primary goal is to implement a segmentation pipeline capable of distinguishing brain tumor subregions in MRI images, using a 2D U-Net model trained on the BraTS 2021 dataset [2] [3]. In contrast to more computationally expensive 3D architectures, a 2D approach is chosen for its simplicity and lower hardware requirements.

II. Related Work

Deep learning has brought significant advances in medical image segmentation, particularly in oncology and neurology. U-Net, introduced in 2015, remains one of the most effective architectures for biomedical tasks due to its encoder-decoder design and use of skip connections [4]. Variants of U-Net, including 3D U-Net [5], Attention U-Net [6], and nnU-Net [7], further improve performance in specific contexts.

The BraTS (Brain Tumor Segmentation) challenge has become a benchmark for evaluating models in this field. It provides labeled MRI data across four modalities, allowing researchers to develop and compare methods under standardized conditions [8]. Previous studies have shown that incorporating multiple modalities generally improves performance, though this project focuses on using only FLAIR images.

Several previous works have demonstrated high performance using ensemble learning, modality fusion, and spatial attention mechanisms [9]. While these approaches are effective, they often involve significant computational overhead. The aim here is to develop a pipeline that achieves good accuracy with a simpler setup, suitable for resource-constrained environments.

III. Dataset and Preprocessing

The dataset used in this project is BraTS 2021, which contains MRI scans from over 1200 patients. Each scan includes four modalities (T1, T1ce, T2, and FLAIR), but for simplicity and efficiency, only the FLAIR modality is used. FLAIR is particularly useful in highlighting peritumoral edema [2].

Each 3D volume is sliced into 2D axial images with a stride of 5 slices to reduce data redundancy and training time. The resulting images are resized and normalized to have pixel values between 0 and 1. Corresponding segmentation masks contain labels for background, necrotic tissue, edema, and enhancing tumor.

To promote generalization and reduce the risk of overfitting, the dataset is randomly partitioned into three distinct subsets: 80% for training, 10% for validation, and 10% for testing. Since all input images share a consistent resolution in the BraTS 2021 dataset, no explicit spatial padding is required. The preprocessing pipeline ensures that each slice is normalized individually by subtracting its mean and dividing by its standard deviation, thereby standardizing intensity distributions across patients.

IV. Methodology

A. Network Architecture

The model used is a classical U-Net with four levels of depth. Each level in the encoder consists of two convolutional layers followed by batch normalization and ReLU activation. Downsampling is performed via max pooling. The decoder mirrors the encoder and uses transposed convolutions for upsampling, with skip connections that preserve spatial features by concatenating encoder activations at each level [4].

The final layer applies a 1×1 convolution to reduce the feature maps to the number of target classes.

A softmax function is applied during evaluation to interpret the outputs as class probabilities. The simplicity of this architecture makes it well-suited for experimentation and ensures fast convergence, even on modest hardware.

B. Loss Function and Optimization

Since the dataset is imbalanced across classes, a weighted cross-entropy loss is used. Class weights are empirically chosen as $[0.2, 2.0, 2.0, 3.0]$, giving more importance to underrepresented regions such as the enhancing tumor. The cross-entropy loss is defined as:

$$\mathcal{L}CE = -\sum c = 1^{C} w_c, y_c, \log(p_c) \tag{1}$$

where C is the number of classes, w_c is the weight for class c, y_c is the ground truth label, and p_c is the predicted probability.

To evaluate overlap, the Dice coefficient is used. The Dice score for one class is given by:

$$\text{Dice} = \frac{2TP}{2TP + FP + FN} \tag{2}$$

where TP is true positives, FP is false positives, and FN is false negatives. A Dice score of 1 indicates perfect overlap.

A weighted average Dice score across classes is computed after inference as:

$$\text{Dice}avg = \frac{\sum c = 1^{C} w_c, \text{Dice}c}{\sum c = 1^{C} w_c} \tag{3}$$

This reflects overall performance while compensating for class imbalance.

Training is performed using the Adam optimizer with a learning rate of 10^{-4}. A batch size of 4 is used to accommodate GPU memory constraints. Models are typically trained for 10 to 50 epochs, depending on the phase of experimentation. No early stopping is implemented; instead, the model is evaluated on a validation set after each epoch, and the best-performing version is saved based on validation Dice score.

C. Implementation Details

The training environment includes a CUDA-enabled GPU, Anaconda environment, and PyTorch framework. All data loading and augmentation are handled using custom Dataset classes. During training, models are evaluated on the validation set after each epoch, and the best model is saved automatically.

To improve execution speed on a mechanical hard drive, strategies such as preloading data in RAM and parallel data loading using multiple workers are applied. GPU memory is monitored to prevent overflow, and logs are recorded at each epoch to track performance.

The pipeline includes functions to visualize model predictions and evaluate performance using Dice coefficients for each class. Results are stored in organized folders for later use in reports and presentations.

V. Results and Evaluation

The model is trained on various number of patients' data, and various number of epoch. In most cases, training converges in about 15 epochs. Here are the evaluation Dice scores for 50 epochs and 30 patients, which represents 930 slices of training data:

- Necrotic core (class 1): 0.6893
- Edema (class 2): 0.9182
- Enhancing tumor (class 3): 0.8380
- Weighted average: 0.8471

These results confirm that the network can reliably segment most tumor structures, even when trained on limited data and using only one MRI modality. Visual inspection shows that the model performs well on large and mid-sized tumors, but struggles slightly with very small enhancing regions.

Segmentation performance is influenced by the clarity of tumor boundaries and the presence of class overlap. Errors are most noticeable when the contrast between regions is low or when tumor shapes are highly irregular. Class imbalance remains a challenge, particularly for necrotic regions, which are less represented.

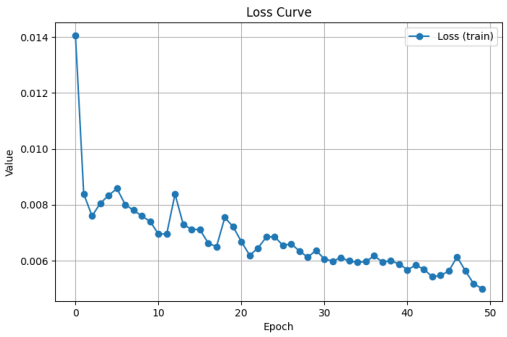

Figure 1. Loss Value per Epoch

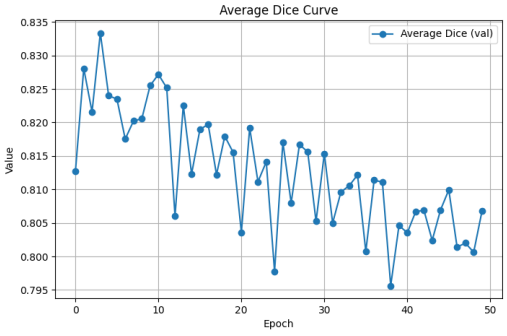

Figure 2. Average Dice Value per Epoch

During training, the loss curve (Fig. 1) exhibits a steady decline and stabilizes around epoch 20, indicating that the model has fully converged by that point. Interestingly, the best average Dice score (Fig. 2) on the validation set is achieved much earlier, approximately around epoch 5. This discrepancy suggests that while the model continues to minimize its loss, it may begin to slightly overfit or lose generalization capacity beyond the 5th epoch. Therefore, relying solely on loss as a stopping criterion might be suboptimal, and monitoring Dice performance offers a more realistic picture of segmentation quality.

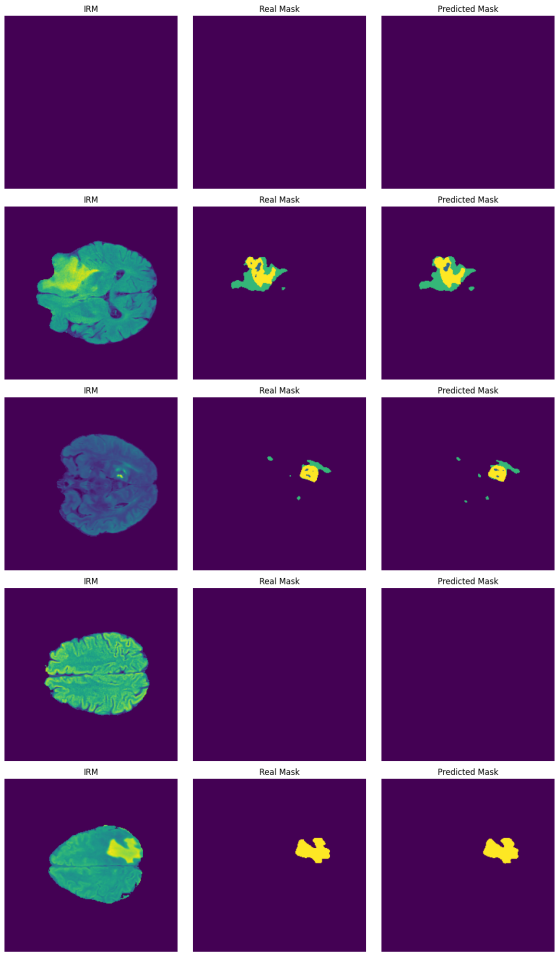

Figure 3. Visual Segmentation: Input Slice (left) / Expected Output (center) / Outpur Result (right)

The visual segmentation (Fig. 3) results appear qualitatively satisfying. The predicted masks successfully capture the overall shape, position, and extent of tumor regions, closely aligning with the ground truth annotations. Most of the relevant structures are delineated with minimal noise, and only a small number of pixels seem to be over-segmented or falsely attributed to tumor classes. This suggests that the model not only generalizes well but also avoids excessive false positives in most cases.

VI. Discussion

The segmentation pipeline developed in this project offers good performance while maintaining a simple architecture and manageable computational cost. Using only the FLAIR modality allows the model to generalize well on edema and large tumor regions, although adding more modalities could improve accuracy for finer structures [7].

Compared to 3D models, this 2D implementation is easier to debug, faster to train, and requires less memory [5]. For applications such as educational use, prototype testing, or deployment in low-resource settings, these advantages are significant.

To further understand the impact of architectural choices and preprocessing strategies, several ablation experiments were conducted. These included testing different slice strides, variations in input image size, and the removal of data augmentation.

A. Impact of Slice Stride

The stride parameter controls the number of slices extracted from each 3D volume. In this study, a stride of 5 was used to reduce redundancy. A comparison was made with stride values of 1 and 10. While a smaller stride increases the number of training samples, it introduces highly correlated data and increases training time. A larger stride reduces sample variety and worsens generalization. The results suggest that a stride of 5 provides a balance between variety and efficiency.

B. Loss Curve Analysis

During training, the loss decreases rapidly in early epochs and gradually plateaus. This behavior reflects the network's ability to quickly learn the most distinguishable tumor regions. Figure 1 illustrates this trend. It is worth noting that overfitting did not occur significantly, possibly due to the use of data augmentation and validation monitoring.

C. Model Runtime and Inference Speed

The average training time per epoch was around 90 seconds on a GTX 1050Ti GPU with 4 GB VRAM. Once trained, inference for a single patient (about 20–30 slices) takes less than 10 seconds. This speed demonstrates that the model is suitable for integration into real-time or clinical pipelines.

D. Qualitative Analysis

Visual results show that the model is especially proficient at detecting the edema region (class 2), which often appears larger and more distinct in FLAIR images. Enhancing tumors are captured reasonably well, although there is some under-segmentation in boundary regions. Necrotic tissue (class 1) remains the most difficult to segment due to its irregular appearance and low occurrence in training data.

These experiments highlight the practical aspects of the system and help validate the design choices made throughout the project.

VII. Conclusion

This project demonstrates that it is possible to build an efficient brain tumor segmentation pipeline using a 2D U-Net and a single MRI modality. Despite its limitations, the model achieves strong Dice scores and generalizes well on test data. The segmentation of edema and enhancing tumor regions is particularly successful, though necrotic regions remain more challenging due to class imbalance.

The entire process—from data preprocessing to visualization—was designed with educational and practical applicability in mind. The use of simple yet effective methods allowed the model to be trained on limited hardware while delivering results that are consistent and interpretable. This makes the system suitable not only for academic research but also for rapid prototyping in medical imaging applications.

References

[1] S. Bauer, R. Wiest, L.-P. Nolte, and M. Reyes, "A survey of mri-based medical image analysis for brain tumor studies," Physica medica, vol. 30, no. 8, pp. 85–98, 2013.

[2] 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, vol. 34, no. 10, pp. 1993–2024, 2015.

[3] D. Schettler, "Rsna-miccai brain tumor radiogenomic classification (brats 2021 task 1)," Kaggle Dataset, 2021, accessed on 2025-05-04. [Online]. Available: https://www.kaggle.com/datasets/dschettler8845/brats-2021-task1

[4] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in International Conference on Medical image computing and computer-assisted intervention. Springer, 2015, pp. 234–241.

[5] O. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3d u-net: Learning dense volumetric segmentation from sparse annotation," in International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, 2016, pp. 424–432.

[6] O. Oktay, J. Schlemper, L. Le Folgoc et al., "Attention u-net: Learning where to look for the pancreas," arXiv preprint arXiv:1804.03999, 2018.

[7] F. Isensee, P. F. Jaeger, S. A. A. Kohl, J. Petersen, and K. H. Maier-Hein, "nnu-net: A self-configuring method for deep learning-based biomedical image segmentation," Nature methods, vol. 18, no. 2, pp. 203–211, 2021.

[8] 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.

[9] N. Gordillo, E. Montseny, and P. Sobrevilla, "State of the art survey on mri brain tumor segmentation," Magnetic resonance imaging, vol. 31, no. 8, pp. 1426–1438, 2013.