BRAIN TUMOR SEGMENTATION USING DEEP LEARNING WITH U-NET ARCHITECTURE

Nurkeldi Iznat (30261427) Justice Nsafoah (30076935)
Osama Kashif (30037753) Junqi Li(30270860)

ABSTRACT

This study presents an automated approach for brain tumor segmentation using a deep learning framework, employing a U-Net architecture on the publicly available brain tumor dataset. The motivation behind this work arises from the pressing need for accurate tumor detection methods in the field of medical imaging. Manual segmentation, being labor-intensive and error-prone, often results in suboptimal and inconsistent identification of tumor boundaries. By leveraging the power of deep convolutional neural networks and innovative architectural designs such as U-Net, our methodology is aimed at delivering competitive segmentation performance with acceptable computational efficiency. In our experiments, we observe that, while the approach reliably captures the key features of the brain tumors, challenges remain in handling variations in tumor morphology, image noise, and class imbalance. Overall, our study highlights the potential of deep learning approaches for critical clinical applications, while also outlining several avenues for future research and model optimization.

Index Terms— brain tumor, U-net, segmentation

1. INTRODUCTION

Brain tumor segmentation is a critical component of the medical imaging workflow that directly influences treatment planning and disease monitoring. In a clinical setting, accurate segmentation facilitates the quantification of tumor volume, localization of abnormal tissue, and determination of the best course of intervention. However, manual delineation by experts is not only time-consuming but also subject to inter-observer variability, thereby introducing inconsistencies in diagnosis and treatment planning. For example, Diwanji, T., et al. [1] evaluated the inter-observer variability of atlas-based manual hippocampal segmentation in patients with brain metastases. The results indicated that significant inter-observer variability remains even with atlas guidance. Weltens, C., et al. [2]investigated the inter-observer variability in gross tumor volume (GTV) delineation for brain tumors. The results showed considerable differences among observers in GTV delineation, and the addition of MRI to CT did not significantly reduce this variability.

In recent years, machine learning (ML) techniques have been increasingly applied to medical image segmentation, enabling automated delineation of anatomical structures and pathological regions in modalities such as MRI and CT. Early approaches using traditional ML methods—such as feedforward neural networks (FNNs), decision trees, and support vector machines—relied heavily on handcrafted features and domain-specific knowledge[3]. While these methods offered initial improvements in segmentation performance due to their simplicity and low computational cost, they often struggled to generalize across different imaging modalities and patient populations. Moreover, their limited capacity to capture complex spatial relationships and contextual information in high-dimensional medical images posed significant challenges, particularly in tasks like tumor segmentation where precise boundary detection is critical[4].

In contrast, modern deep learning approaches—especially convolutional neural networks (CNNs)—have revolutionized medical image segmentation by learning hierarchical feature representations directly from raw image data. Among these, U-Net has become one of the most influential and widely adopted architectures for biomedical segmentation tasks[5]. Its encoder-decoder structure, combined with skip connections, enables the model to retain fine-grained spatial information while integrating high-level semantic context. This makes U-Net particularly effective in segmenting tumors with irregular shapes and ambiguous boundaries. Variants such as U-Net++, Attention U-Net, and nnU-Net have further enhanced performance and adaptability across datasets[6, 7]. Numerous studies have shown that U-Net-based models consistently outperform traditional ML methods in terms of accuracy, robustness, and generalization, even when trained on limited annotated data[8].

In this project, we build upon recent advances in medical image analysis by proposing a comprehensive approach that combines a detailed data preprocessing pipeline, advanced augmentation techniques, and an efficient model training strategy. Specifically, we adopt the widely used U-Net architecture for segmentation tasks, integrating a patch-based training method to

effectively handle high-resolution brain images and improve local feature learning. Furthermore, we leverage mask images as ground truth annotations to supervise the learning process. We use an MRI dataset with T1-weighted images and tumor masks for segmentation. By applying our method to this dataset, we aim to demonstrate that an optimal combination of established deep learning practices can yield robust and accurate segmentation results, even on challenging data with limited variability in image scale and contrast.

2. RELATED WORK

Recent advances in deep learning-based medical image segmentation have been driven primarily by the U-Net architecture and its adaptations. Since its introduction by Ronneberger et al[5], U-Net has emerged as the most widely adopted architecture for biomedical image segmentation.

Several studies have extended the original U-Net to better address the challenges of medical data. For instance, Çiçek et al. [8] proposed 3D U-Net, which adapts the architecture to volumetric data such as 3D MRI or CT scans. Their model demonstrated high performance in segmenting organs and tumors from 3D biomedical images, with significantly improved boundary delineation. Zhou et al. [6] introduced UNet++, which incorporates nested and dense skip connections to enhance feature fusion between encoder and decoder paths, achieving superior performance in multi-class segmentation tasks.

In addition to architectural modifications, patch-based training strategies have emerged as effective techniques for handling large medical images that exceed GPU memory limitations. For example, Kamnitsas et al. [9] used a patch-based 3D CNN (DeepMedic) for brain lesion segmentation, demonstrating that training on small sub-volumes not only reduced memory load but also improved the model's ability to focus on local context. Similarly, Isensee et al. [10]employed patch-wise training in their nnU-Net framework, which automatically configures preprocessing, patch size, and training parameters to adapt to various segmentation tasks. These patch-based approaches typically involve pairing image patches with their corresponding ground truth masks, allowing the model to learn pixel-level anatomical features under full supervision.

Patch-based methods are particularly useful when working with high-resolution images, such as histopathology slides or full-brain MRIs, where full image input is computationally infeasible. In digital pathology, HookNet utilizes both low- and high-resolution image patches to capture global context and fine structural details for accurate whole-slide image segmentation[11]. Similarly, Jabeen et al. applied a patch-based CNN to brain MRI tumor segmentation, showing that training on smaller patches improves local feature learning while reducing computational cost[12].

These studies demonstrate that combining the U-Net architecture with intelligent patching strategies and supervised training using segmentation masks can effectively tackle the computational and structural challenges inherent in medical image segmentation. Our work builds upon these foundations by adopting U-Net and integrating a patch-based training pipeline tailored to the segmentation of brain tumors in T1-weighted MRI scans.

3. MATERIALS AND METHODS

3.1 Dataset and preprocessing

The brain tumor dataset forms the cornerstone of this project. We use a brain tumor MRI dataset[13], originally derived from the Figshare Brain Tumor Dataset[14], which includes axial T1-weighted 3064 images (512×512 pixels) along with manually annotated binary masks (512×512 pixels) for tumor regions. Our dataset was systematically partitioned as follows: 80% of the images were used for training, 10% for validation, and the remaining 10% reserved for testing. Every image in our dataset was first patched into four images with 256×256 pixels. This allowed the dataset to be 4 times as large. On top of that, because of the lower resolution, it also allowed for a training process with a higher batch number. In addition, pixel intensity normalization was applied to standardize the images, which is crucial for speeding up convergence during training. To further increase the robustness of the model and prevent overfitting, a series of data augmentation techniques were introduced. These transformations included random horizontal flips, rotations (up to 10 degrees), and minor color jittering, which collectively help simulate variations in imaging conditions.

3.2 Model architecture and implementation

The implementation is based on an evolution of the U-Net architecture. In our version, the encoder comprises four convolutional blocks with filter sizes that double at each step—from 64 up to 512 filters—allowing the network to capture features at multiple scales. Each block consists of two convolutional layers followed by ReLU activations, and each block's output is passed through a max-pooling layer for spatial downsampling.

At the network's bottleneck, further convolutional layers are applied to extract the most abstract features from the input, setting the stage for the decoder. The decoder mirrors the encoder: it begins with upsampling through transposed convolutions and uses

skip-connections to concatenate corresponding feature maps from the encoder. These concatenated features help recover spatial details lost during the pooling operations, leading to more precise segmentation maps. The final layer makes use of a 1×1 convolution to reduce the output to the desired number of classes—in this case, a binary segmentation with the classes being background and tumor.

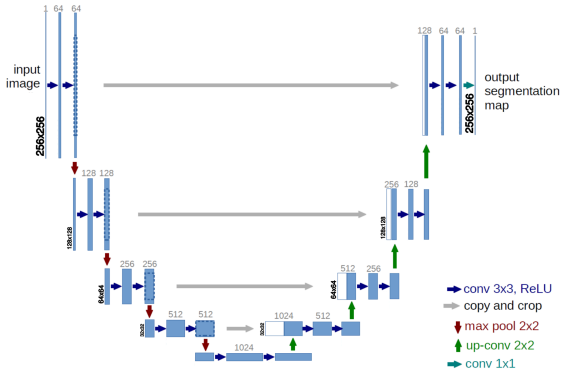

Figure 1. Model architecture

The training was configured with a batch size of 16, and we employed the Adam optimizer with an initial learning rate of 1e-4. The learning process was driven by the binary cross entropy loss function, which is well-suited for the pixel-wise binary classification inherent in segmentation tasks. In addition, a scheduler that reduces the learning rate upon a plateau in validation loss was utilized to further fine-tune the training process.

3.3 Training and evaluation setup

The training loop was designed to provide detailed progress updates at regular intervals, ensuring that every mini-batch's loss was recorded and aggregated. The model was trained for 50 epochs, during which performance was monitored using the validation set (see Figure 2). When the validation loss improved, a checkpoint was saved, guaranteeing that the best-performing model was retained. For evaluating the overall performance, we employed a combination of quantitative metrics such as the Intersection over Union (IoU), as well as qualitative visual comparisons between the predicted masks and the ground truth.

The entire process was executed on a GPU-enabled TALC cluster node, leveraging parallel computing capabilities to accelerate both training and inference. In addition, a custom data loader was built to efficiently manage the dataset during training, emphasizing correct pairing between images and their corresponding masks through systematic filename matching.

Figure 2. Training Process

4. RESULTS AND DISCUSSION

The lowest loss for the validation set was after 17 epochs, the weights of which were saved and used for the following results. The trained U-Net model demonstrated competitive segmentation performance on the test set. Quantitative evaluation using the dice coefficient revealed that most tumor pixels were correctly classified. In particular, the model achieved an encouraging accuracy, with good overlap between predicted tumor boundaries and the ground truth with a dice coefficient of 0.68. Visual inspection corroborated these findings, where side-by-side comparisons of input images, ground truth masks, and the model's output showed clearly defined tumor borders that closely matched expert annotations (see Figure 3).

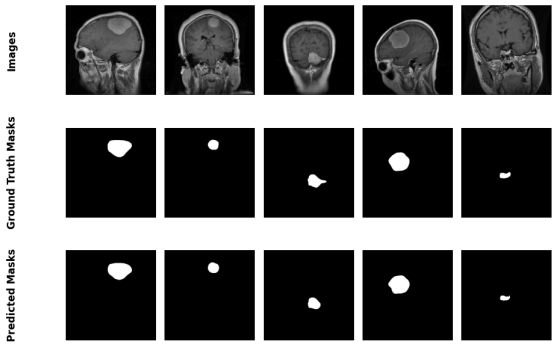

Figure 3. Prediction and Ground Truth Comparison

4.1 Strengths of the approach

One of the notable strengths of our implementation is the comprehensive data preprocessing pipeline, which played a critical role in harmonizing the diverse set of MRI images. The use of aggressive data augmentation techniques not only enriched the training dataset but also improved the generalization capability of the model

by enabling it to handle variations in image orientation, brightness, and contrast. Furthermore, the use of a validation-based checkpointing system ensured that the optimal model parameters were retained, reducing overfitting and leading to smoother learning curves.

The inherent architecture of the U-Net, with its effective skip connections and balanced encoder-decoder structure, proved to be a robust design choice for the segmentation task. Moreover, leveraging GPU acceleration significantly reduced training times and allowed for quick iterations over the model design, making it feasible to explore various hyperparameter configurations.

4.2 Limitations and areas for improvement

Despite the advantages, several limitations were observed during the experimentation. First, the approach was restricted to binary segmentation, which may not fully capture the complexity of multi-class problems that arise in more advanced clinical applications. Moreover, the fixed input size of 512×512 pixels, while practical from a computational standpoint, might limit the model's ability to resolve very fine details present in larger images. Additionally, the relatively small batch size, dictated by memory constraints on the GPU, might have affected the stability of gradient updates during training.

The basic U-Net architecture used in the study does not incorporate several recent advancements such as attention mechanisms or residual connections, both of which have been shown in the literature to further enhance performance. Finally, class imbalance remains an unresolved issue, as most of the images contain considerably more background pixels compared to tumor pixels. Future investigations might benefit from exploring advanced loss functions or sampling strategies to mitigate this problem.

5. CONCLUSIONS

In summary, our study demonstrates that the U-Net architecture can be effectively harnessed for the task of brain tumor segmentation in MRI images. By integrating rigorous data preprocessing, carefully tuned augmentation techniques, and a robust training regime, the trained model was able to deliver competitive performance in identifying tumor regions within the brain. The findings of this study underscore the immense potential of deep learning methodologies in the healthcare field, particularly for automating tasks that traditionally rely on manual interpretation.

At the same time, the project highlights several challenges that continue to motivate further research. The limitations—such as the model's binary segmentation constraint, handling of varying image sizes, and inherent class imbalance—point toward opportunities for methodological improvements. Future work could explore enhanced variants of U-Net, more balanced loss functions, and even ensemble methods to improve segmentation accuracy. Ultimately, the integration of sophisticated deep learning models into clinical workflows holds great promise for faster, more reliable diagnostic processes, which may lead to improved patient outcomes.

6. REFERENCES

[1] Diwanji, T. et al.,(2016). Interobserver Variability in Atlas-Based, Manual Segmentation of the Hippocampus in Patients With Brain Metastases, International Journal of Radiation Oncology, Biology, Physics, 96 (2), p. E130

[2] Weltens, C. et al.,(2001). Interobserver variations in gross tumor volume delineation of brain tumors on computed tomography and impact of magnetic resonance imaging, Radiotherapy and Oncology, 60 (1), pp. 49-59

[3] Bauer, S., Wiest, R., Nolte, L. P., and Reyes, M.,(2013). A survey of MRI-based medical image analysis for brain tumor studies, (in eng), Phys Med Biol, 58 (13), pp. R97-129

[4] Xu, Y., Quan, R., Xu, W., Huang, Y., Chen, X., and Liu, F.,(2024). Advances in Medical Image Segmentation: A Comprehensive Review of Traditional, Deep Learning and Hybrid Approaches, (in eng), Bioengineering (Basel), 11 (10),

[5] Ronneberger, O., Fischer, P., and Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation, in Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, Cham

[6] Zhou, Z., Rahman Siddiquee, M. M., Tajbakhsh, N., and Liang, J. UNet++: A Nested U-Net Architecture for Medical Image Segmentation, Cham

[7] Isensee, F., Jaeger, P. F., Kohl, S. A. A., Petersen, J., and Maier-Hein, K. H.,(2021). nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation, Nature methods, 18 (2), pp. 203-211

[8] Çiçek, Ö., Abdulkadir, A., Lienkamp, S. S., Brox, T., and Ronneberger, O. 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation, Cham

[9] Kamnitsas, K. et al.,(2017). Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation, Medical Image Analysis, 36 pp. 61-78

[10] Isensee, F., Jaeger, P. F., Kohl, S. A., Petersen, J., and Maier-Hein, K. H.,(2021). nnU-Net: a Self-configuring Method for Deep Learning-based

Biomedical Image Segmentation, Nature methods, 18 (2), pp. 203-211
[11] Van Rijthoven, M., Balkenhol, M., Siliņa, K., Van Der Laak, J., and Ciompi, F.,(2021). HookNet: Multi-resolution Convolutional Neural Networks for Semantic Segmentation in Histopathology Whole-slide Images, Medical Image Analysis, 68 p. 101890
[12] Ullah, F., Salam, A., Abrar, M., and Amin, F.,(2023). Brain Tumor Segmentation Using a Patch-Based Convolutional Neural Network: A Big Data Analysis Approach, Mathematics, 11 (7), p. 1635
[13] Tomar, N. Brain Tumor Segmentation, Kaggle. [Online]. Available: https://www.kaggle.com/datasets/nikhilroxtomar/brain-tumor-segmentation
[14] Cheng, J. Brain Tumor Dataset, Figshare. [Online]. Available: https://doi.org/10.6084/m9.figshare.1512427