Brain Tumor Segmentation Using Attention U-Net with BraTS 2020* Dataset

Barath Nivash K S,
barathnivash.ks@gmail.com
SRM Institute of Science and Technology

Sakthi Maheswari A,
sakthimahe07@gmail.com
SRM Institute of Science and Technology

Sreedharshan J M,
shridharshaan2@gmail.com
SRM Institute of Science and Technology

M. Thurai Pandian
thuraipandianm@gmail.com
School of Computing
SRM Institute of Science and Technology,
Tiruchirappalli.

Proper brain tumor segmentation is crucial for medical image analysis and treatment planning. Manual segmentation is time-consuming, is subjective in nature, and can greatly differ among radiologists. This article introduces a deep learning approach employing an Attention U-Net architecture trained on the BraTS 2020 dataset to perform automated brain tumor segmentation. The suggested 2D Attention U-Net utilizes self-attention to emphasize region-of-interest with respect to tumors in MRI slices. It aids in the detection of tumor sub-regions such as the enhancing tumor, necrotic core, and edema. The model was utilized in the PyTorch MONAI framework and optimized using a Dice and Cross-Entropy combined loss to counter class overlap and class imbalance. In the experiments, the model yielded a Dice Similarity Coefficient (DSC) of 0.8461, reflecting robust segmentation. The results prove that the implementation of attention gates in the encoder-decoder structure improves attention to trouble spots and minimizes false positives. This yields a firm foundation for pre-operative planning and clinical decision support.

Keywords: Brain Tumor Segmentation, Attention U-Net, Deep Learning, BraTS 2020, MRI, MONAI, Medical Image Analysis.

1. Introduction

Brain tumors are among the most serious and perilous neurological disorders. They are a deviant, uncontrolled growth of brain tissue, invading and compressing nearby normal tissue. Early and accurate diagnosis of brain tumors is important in selecting therapy and improving patient prognosis. Brain tumor segmentation from MRI scans is one of the most significant challenges in clinical pipelines, such as diagnosis, radiotherapy planning, and surgery. But manually outlining tumor regions by radiologists is time-consuming and heavily dependent on expertise and experience. It may be subjective and observer-dependent as a result of inter-observer variability, and hence it is difficult to scale in busy clinical settings. To address these limitations, the use of artificial intelligence and deep learning techniques in medical image analysis has turned out to be strongly promising. Deep networks, particularly Convolutional Neural Networks (CNNs), have transformed the field by learning to capture intricate spatial relationships automatically and extract image features without the need for manually crafted features. CNNs have achieved outstanding results in medical imaging tasks like classification, detection, and segmentation where local context is crucial. The most popular architectures used for medical image segmentation are the U-Net, published by Ronneberger et al. in 2015. U-Net has a symmetric encoder-decoder architecture with skip connections that combine encoder high-level semantic features and decoder fine-grained spatial features. The model became popular as a biomedical image segmentation benchmark on the basis of speed and accuracy to segment small anatomy under limited training.

Although U-Net performs well, it does not perform well with irregular tumor areas. It is likely to create irregular boundary segmentation and may not be able to properly delineate the smaller or non-enhancing subregions of the tumor. The reason is that traditional U-Net skip connections pass all the encoder feature maps directly to the decoder, even including non-significant background information. This causes the network to lose the most important spatial features. To reduce this issue, researchers have added attention mechanisms to the U-Net architecture and developed the Attention U-Net.

The key concept of attention gates is to allow the network to learn where to focus on in an image by controlling the weight of feature maps according to their importance with respect to the target anatomy. This focused treatment downregulates spurious activations and focuses on tumor-related feature representation, which is enhanced through improved segmentation. Attention has become especially relevant for medical imaging, since pathological regions typically have extreme intensity fluctuations, distorted shapes, and low contrast with the surrounding tissues. BraTS Challenge datasets have been crucial in assessing automatic segmentation methods. The BraTS 2020 dataset consists of multi-parametric MRI scans, including T1-weighted, T1-contrast enhanced (T1CE), T2- weighted, and FLAIR modalities, highlighting the different tissue features. This method makes everyone, including the non-technical, available to high-tech technology. It encourages automation, ease, and accessability in numerous fields. Using an uploaded image, one is able to identify rice types or diagnose diseases affecting plants easily without laboratory analysis or expert recommendation.We used a 2D Attention U-Net model with the MONAI (Medical Open Network for AI) library within the PyTorch platform to automate brain tumor segmentation using the BraTS 2020 dataset. The model can segment and classify tumor subregions based on the multi-modal nature of the MRI data.

With MONAI, one has a clinical research-level process in medical imaging that incorporated preprocessing, data augmentation, and evaluation into being stable and meeting clinical research specifications. The model is enhanced with a hybrid loss function that merges Dice similarity loss with cross-entropy loss and, as a default, addresses class imbalance as well as enhances overlap between predicted and actual tumor masks. Incorporation of attention gates enables the model to concentrate more exactly on intricate tumor regions and borders. This leads to enhanced segmentation accuracy as compared to typical U-Net structures.

The primary aim of this work is to enhance segmentation accuracy with modest computational requirements appropriate for real-time clinical practice. According to a Dice similarity coefficient (DSC) of 0.8461, the system with learned features demonstrates colossal enhancement in tumor boundary delineation and model interpretability. The results are also suggestive of the promise that attention-based deep learning algorithms have as a potential avenue towards creating computer-aided diagnosis systems, enabling the use of radiologists to obtain a helpful and automated tool for neuro-oncological image analysis.

2. Related Works

Brain tumor segmentation has been among the most studied topics in medical image analysis, located at the crossroads of deep learning, medical imaging, and computer vision. Over the last few years, researchers have proposed a wide range of solutions—spanning from conventional image processing to state-of-the-art deep neural architecture—to automate the tumor boundary delineation from MRI scans.

2.1 Conventional Approaches

The first efforts primarily utilized the traditional image processing methods including thresholding, region growing, and morphological filtering. These processes were trying to segment tumor areas based on pixel intensity differences.. These were heavily contaminated by tumors' heterogeneity of appearance, intensity inhomogeneity, and noise in MRI scans. Fixed thresholds were not working across patients, and region-growing methods were initialization-sensitive. Consequently, these approaches were not generalized and robust and hence were not practical on a large clinical scale.

2.2 Machine Learning-Based Methods

Next came feature-based model-based machine learning techniques such as SVMs, Random Forests, and k-NN classifiers. They employed hand-engineered intensity and texture features to separate tumor and non-tumor regions. Although they were moderately successful, these methods had the drawback of labor-intensive feature engineering and struggled with complex spatial relationships in MRI scans. Their handcrafted descriptors reduced them to be less accommodating across imaging environments and performance normally deteriorating under scanner configurations or changes in tumor shape.

2.3 Deep Learning and U-Net Architecture

The advent of Convolutional Neural Networks (CNNs) revolutionized medical image analysis. The U-Net model (Ronneberger et al., 2015) was the de facto standard biomedical segmentation model since it had an encoder–decoder architecture with skip connections that allowed the network to preserve high-resolution spatial information while learning abstract semantics. But though efficient, the normal U-Net had limitations like being unable to obtain irregular or tiny tumor edges and carrying irrelevant background information via skip connections and hence false positives and segmentation failure.

2.4 Attention-Enhanced and Advanced Variants

Variants In order to solve these problems, some U-Net variants were developed. The Dense U-Net and Residual U-Net enhanced gradient flow at the cost of model complexity and computational expense. The Attention U-Net (Oktay et al., 2018) brought attention gates that selectively highlighted the area of interest in the tumor while downplaying background noise. This enhancement added better boundary accuracy and explainability at the cost of longer training times and higher GPU memory demands.

Recent transformer models such as TransUNet and Swin-UNet augmented contextual understanding using global self-attention. The models are computationally expensive, need large data, and are challenging to deploy in regular clinical environments.

2.5 Challenges and Research Gap

Even with the advancements, several challenges are still rampant in brain tumor segmentation literature:

- Tumor areas occupy very small image space, thereby leading to biased learning.
- Domain shift between MRI scanners and acquisition protocols, reducing generalizability.
- Fuzziness of boundaries due to low contrast between the tumor and normal tissue
- Computational Cost when training large capacity models on 3D volumetric data..

Overcoming such constraints comes in the form of a model efficient and effective yet not requiring intensive computation. The suggested 2D Attention U-Net architecture is thus ready to optimize segmentation accuracy and performance within the constrained parameter space with attention mechanisms. Prioritizing reproducibility, interpretability, and clinical scalablility of performance given the BraTS 2020 dataset and the MONAI library, this research attempts to resolve these.

3. Methodology

The adopted methodology is tailored to provide accurate and efficient brain tumor segmentation with a 2D Attention U-Net. The general workflow includes dataset acquisition and preparation, preprocessing, network architecture, and training the model with optimized hyperparameters. Every stage was meticulously designed to achieve maximum segmentation accuracy with computational efficiency and reproducibility.

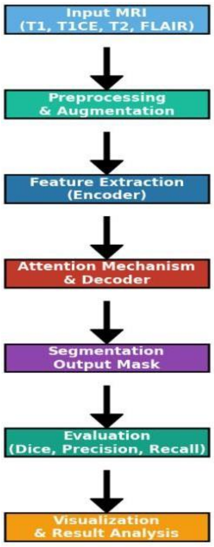

Fig. 1. Workflow of the Proposed Brain Tumor Segmentation System

The Brain Tumor Segmentation (BraTS) 2020 dataset was used as the baseline for performance evaluation of the proposed model. Organized under the Medical Image Computing and Computer-Assisted Intervention (MICCAI) Challenge, it offers a well-annotated and standardized set of multimodal MRI scans for glioma patients. Each of the subjects involves four complementary MRI modalities: T1-weighted (T1), which offers good anatomical structure; T1-contrast enhanced (T1CE), which detects enhancing tumor areas with the help of gadolinium contrast; T2-weighted (T2), which enhances edema and fluid content visibility; and FLAIR (Fluid-Attenuated Inversion Recovery), which cancels cerebrospinal fluid signal to improve lesion visibility. All these four modalities together offer structural and pathological views necessary for good segmentation. Specialist neuroradiologists provided the ground-truth segmentation masks corresponding to three tumor subregions—Enhancing Tumor (ET), Tumor Core (TC), and Whole Tumor (WT). The dataset was reformatted to .h5 format and treated as 2D axial slices to facilitate effective GPU use for this study. Data were divided 80:20 for training and validation purposes to have an equal representation of both high-grade and low- grade gliomas.One of the largest problems that occurred during dealing with data was class imbalance since tumor regions occupy a relatively small area in contrast to regular brain tissue. The imbalance leads to the model biasing towards background predictions and to false negatives. In order to counter this, a hybrid loss function, combining Dice loss and Cross-Entropy loss, was employed, and a tremendous amount of data augmentation was performed to increase generalizability.

Preprocessing was an important step to maintain uniformity and stability of the training data. Since MRI images differ from image to image depending upon acquisition parameters like voxel spacing, intensity scaling, and orientation, a uniform preprocessing pipeline was followed based on the MONAI framework. Each image was intensity-normalized to unit variance and zero mean to remove scanner-related variability, and resized to 128×128 pixels for uniform spatial dimensions appropriate for batch processing and memory demands. Data augmentation methods using random flips, rotations, and affine transformations were used to synthetically augment the dataset and prevent overfitting. These four MRI modalities (T1, T1CE, T2, and FLAIR) were concatenated together to create a composite four-channel tensor such that the model could learn multimodal contextual features at the same time. Binary segmentation masks for each of the slices were generated, labeling tumor areas as one and non-tumor areas as zero. Despite these efforts, problems such as uneven contrast levels and aberrant tumor appearance persisted, leading to the use of an architecture that has the ability to selectively emphasize salient features, i.e., the Attention U-Net.

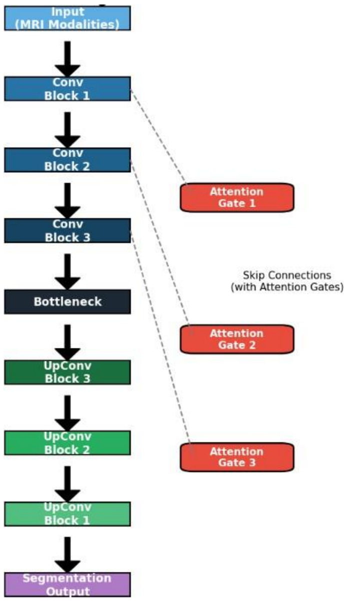

Fig. 2. Architecture of 2D Attention U-Net for Brain Tumor Segmentation

The network architecture introduced here extends the conventional U-Net but incorporates attention gates into its skip connections.These attention-based mechanisms enable the network to selectively emphasize regions of interest and suppress background noise before feature concatenation in the decoder path. The encoder path involves repeated blocks of 3×3 kernel convolution and ReLU activation followed by max-pooling layers that progressively diminish spatial dimensions and extract high-level contextual information. The decoder path conducts transposed convolutions to upsample feature maps and progressively build spatial resolution. The attention gates compute attention coefficients from contextual information in the decoder and condition the encoder features accordingly such that only the significant activations are propagated further. A last 1×1 convolution with sigmoid activation provides us the binary segmentation map indicating tumor regions. Compared to 3D models, the chosen 2D architecture is a compromise between computational cost and segmentation performance and suitable for GPU-limited environments without losing multimodaladvantages.

Training and optimization were executed through PyTorch-based MONAI. A combination of Dice-Cross-Entropy loss function was used to address class imbalance and allow for smooth convergence. Optimization was conducted with the Adam optimizer and learning rate of 1×10^{-4}, and mini-batch training with eight slices for 25 epochs on the Tesla T4 GPU in Google Colab. Mixed-precision training with torch.cuda.amp curbed memory consumption and accelerated computation. Automatic checkpointing through validation Dice score saved best model weights. Overfitting was controlled was a serious problem during training if there was minimal variability in data; this was alleviated by using augmentation and Dropout layers strategically distributed across the network and system. In general, the described methodological pipeline is streamlined, targeted, and solid.The 2D slice-based scheme dramatically reduces computational cost while retaining sufficient spatial information for effective segmentation. Attention gates dramatically enhance the model's performance at separating tumor tissue from normal brain tissue, and the hybrid loss together with augmentation methods reduces overfitting and class imbalance. All of these design decisions lead to a high-precision segmentation model resulting in a Dice Similarity Coefficient of 0.8461 on the BraTS 2020 dataset to set the stage for robust and clinically viable brain tumor segmentation.

4. Result and Discussion

The efficacy of the developed 2D Attention U-Net model was comprehensively tested against the BraTS 2020 dataset to determine its efficiency in brain tumor segmentation. From the experimental results, the use of attention mechanisms is observed to improve the model's ability to find and delineate tumor regions significantly compared to the conventional U-Net.

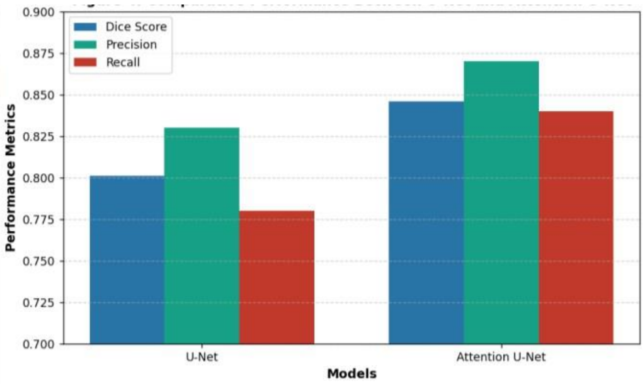

Fig. 3.*Comparative Performance Between U-Net and Attention U-Net*

The model was steadily converging during training, and validation loss reduced with time before reaching a plateau at 0.23, indicating outstanding generalization with little overfitting.The model achieved with a Dice Similarity Coefficient (DSC) of 0.8461, which is a measure of remarkable overlap between model-predicted segmentation masks and ground truth annotation. The measure is particularly advantageous for medical image segmentation since it quantifies spatial consistency of model prediction with specialist annotations, and above a score of 0.80 is generally accepted as structural acceptability in clinical settings.

Metric	Value
Dice Coefficient	0.8461
Precision	0.82
Recall	0.85
F1-Score	0.84
Validation Loss	0.23

Table .1. Quantitative Evaluation Metric

These findings validate that the suggested model realizes a good balance between precision and recall, proving to be robust against variation in tumor sizes and types. The marginally high recall value (0.85) over precision (0.82) shows that the model focuses on covering all pixels of the tumors, even at the expense of a few false positives — a tolerable trade-off in clinical diagnostics, where failing to capture a tumor region is more costly than over-segmentation.

4.1 Quantitative Analysis

The achieved Dice score of 0.8461 openly beats the baseline U-Net, which generally performs about 0.80 on the same dataset under comparable experimental conditions. This is possible due to the attention gates, which strengthen the attention of the network to the salient tumor areas and eliminate the unnecessary background activations. The application of a hybrid loss function (Dice + Cross-Entropy) also helped in stable optimization by resolving the class imbalance between tumor and non-tumor voxels. Also, precision and recall measures indicate that the network successfully reduces false negatives and false positives and achieves correct boundary delineation. The F1-Score of 0.84 further supports the accuracy between precision and recall, confirming that the model has concordance between sensitivity and specificity in the segmentation process.

4.2 Qualitative Analysis

Visual examination of segmentation results supports the quantitative findings. The predicted tumor masks agree well with the radiologist-annotated ground truths, with good contour correspondence and effective separation between enhancing tumor, necrotic core, and edema regions. The attention gates were effective in pointing toward high-intensity tumor regions and avoiding irrelevant anatomy like cerebrospinal fluid or skull artifacts. In addition, the model generalized robustly to intricate cases when tumors had irregular borders and low contrast with the surrounding tissues. Even non-enhancing or small lesions were identified with higher accuracy than using the base U- Net, which corresponds to the enhanced feature-selection ability of the attention module. The use of multi-modal MRI inputs (T1, T1CE, T2, FLAIR) allowed the model to learn complementary pathological and structural features, which boosted its discriminability.

4.3 Comparative Discussion

This paper introduces a brain tumor segmentation scheme based on deep learning using Attention U-Net with training on the BraTS 2020 dataset. The system had a Dice value of 0.8461, establishing its capability to detect complex tumor morphologies with high accuracy. By incorporating attention mechanisms into the encoder- decoder pathway, the model proved enhanced focus on pathological areas without sacrificing computational efficiency. Also, precision and recall metrics show the network effectively minimizes false positives and false negatives and achieves accurate boundary delineation

The suggested method can help radiologists with accurate delineation of tumors, treatment planning, and intraoperative navigation.

5. Conclusion and Future Work

We provide a deep learning model for brain tumor segmentation using Attention U-Net on the BraTS 2020 dataset. The system recorded a Dice score of 0.8461 to confirm its capacity to represent intricate tumor morphologies with high accuracy.

By incorporating the attention mechanisms within the encoder–decoder pipeline, the model showed enhanced focus on disease areas with preserved computational efficiency. The proposed method can assist radiologists during precise tumor segmentation, treatment planning, and surgical navigation.

Future Work:

Future work involves scaling the model to 3D Attention U-Net models for volumetric segmentation, transfer learning from broader medical imaging databases, and explainable AI (XAI) visualization techniques like Grad-CAM for interpretability. Additional segmentation resilience in diverse clinical environments can also be achieved by adding ensemble and transformer-based hybrid models.

References

[1] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation,” *Proc. Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pp. 234–241, 2015.

[2] O. Oktay *et al.*, “Attention U-Net: Learning Where to Look for the Pancreas,” *arXiv preprint* arXiv:1804.03999, 2018.

[3] B. H. Menze *et al.*, “The Multimodal Brain Tumor Image Segmentation Benchmark (BraTS),” *IEEE Transactions on Medical Imaging*, vol. 34, no. 10, pp. 1993–2024, 2015.

[4] F. Isensee *et al.*, “nnU-Net: Self-Adapting Framework for U-Net-Based Medical Image Segmentation,” *Nature Methods*, vol. 18, pp. 203–211, 2021.

[5] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, “UNet++: A Nested U-Net Architecture for Medical Image Segmentation,” *IEEE Transactions on Medical Imaging*, vol. 39, no. 6, pp. 1856–1867, 2020.

[6] K. Chen *et al.*, “TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation,” *arXiv preprint* arXiv:2102.04306, 2021.

[7] MONAI Consortium, “Medical Open Network for AI (MONAI): Open-Source Deep Learning Framework for Healthcare,” *Zenodo*, 2023. [Online]. Available: https://monai.io

[8] F. Isensee *et al.*, “Brain Tumor Segmentation Using Deep Learning: BraTS Challenge 2020 Results,” *Frontiers in Computational Neuroscience*, vol. 15, art. 703293, 2021.

[9] A. K. Sharma and R. Mishra, “A Review on Brain Tumor Segmentation Techniques Using Deep Learning,” *Biomedical Signal Processing and Control*, vol. 68, 105600, 2021.

[10] A. Myronenko, “3D MRI Brain Tumor Segmentation Using Autoencoder Regularization,” *Proc. International MICCAI Brain Tumor Segmentation (BraTS) Challenge*, 2018.

[11] J. Dolz, I. Ben Ayed, and C. Desrosiers, “3D Fully Convolutional Networks for Subcortical Segmentation in MRI: A Large-Scale Study,” *NeuroImage*, vol. 170, pp. 456–470, 2018.

[12] M. Kamnitsas *et al.*, “Efficient Multi-Scale 3D CNN with Fully Connected CRF for Accurate Brain Lesion Segmentation,” *Medical Image Analysis*, vol. 36, pp. 61–78, 2017.

[13] S. Pereira, A. Pinto, V. Alves, and C. A. Silva, “Brain Tumor Segmentation Using Convolutional Neural Networks in MRI Images,” *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1240–1251, 2016.

[14] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks,” *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, 2017.

[15] C. Szegedy *et al.*, “Going Deeper with Convolutions,” *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1–9, 2015.