Comparative Study of Automated Brain Tumor Segmentation Techniques

Koushik Rameshbabu, Nithish Krishna Shreenevasan, Sai Spandana Echambadi

1. Introduction

The rapid growth of medical imaging technologies has resulted in an excess of high-resolution volumetric data, such as CT and MRI scans, necessitating the development of reliable and accurate automated segmentation systems. Deep learning techniques, particularly convolutional neural networks (CNNs), have demonstrated promising results in solving this difficulty. U-Net architectures have become fundamental in medical image segmentation due to their encoder-decoder design, which successfully captures both spatial and contextual data.

Nonetheless, typical U-Net models frequently struggle to segment fine or complicated structures, especially in 3D data, because to their homogeneous treatment of features across spatial regions. To address these restrictions, architectural improvements such as attention techniques and multi-scale processing blocks have been proposed. Attention modules assist the model focus on the most significant regions, while inception blocks improve multi-scale feature extraction, both of which are critical for successful segmentation in medical imaging.

In this project, we study a variety of deep learning architectures developed for 3D medical image segmentation. Our goal is to determine how architectural modifications, such as attention methods, inception modules, and their combinations, affect model performance. We hope to uncover architectural options that significantly improve segmentation accuracy and robustness across complicated medical datasets by carefully analyzing various model configurations.

2. Motivation

3D medical image segmentation is crucial for clinical diagnoses, surgical planning, and treatment monitoring. However, the intrinsic complexity of volumetric data—such as noise, variable forms, and poor contrast—makes standard techniques difficult to perform effectively. Deep learning models, particularly convolutional architectures, have demonstrated significant promise, but their effectiveness is frequently dependent on their capacity to collect both local features and global context. This inspired us to investigate and compare various architectures, such as hybrid and attention-based models, in order to develop ways capable of delivering robust and accurate segmentation in difficult 3D medical imaging scenarios.

3. Problem Statement

Gliomas are among the most aggressive and diverse forms of brain tumors, necessitating precise and individualized treatment plans. The accurate segmentation of tumor sub-regions from multi-parametric magnetic resonance imaging (mpMRI) scans is crucial for diagnosis, therapy planning, and outcome monitoring. However, manual segmentation takes time, is vulnerable to inter-rater variability, and is not scalable in clinical practice.

The Brain Tumor Segmentation (BraTS) dataset includes preprocessed and co-registered mpMRI scans in T1, post-contrast T1-weighted (T1Gd), T2, and FLAIR modalities, as well as expert-annotated ground truth masks for four tumor sub-regions: enhancing tumor (ET), non-enhancing tumor core (NETC), surrounding non-enhancing FLAIR hyperintensity (SNFH), and resection cavity (RC). These annotations highlight significant anatomical and pathological locations of the tumor and surrounding tissue. The purpose of this work is to provide an automated system for segmenting these sub-regions in mpMRI scans. Specifically, the technique should precisely delineate:

- The enhancing tumor (ET): areas of active enhancement and tumor nodules.
- The non-enhancing tumor core (NETC) includes necrosis and cystic regions.
- The surrounding non-enhancing FLAIR hyperintensity (SNFH), which includes edema and infiltrative malignancy,
- The resection cavity (RC) contains fluid, blood, and post-surgical remains.

This segmentation problem is especially difficult due to inter-patient anatomical heterogeneity, the varying appearance of sub-regions across different modalities, and the presence of post-treatment modifications. As a result, the endeavor necessitates powerful deep learning models that can generalize across institutions, scanning techniques, and tumor shapes.

4. Dataset

The dataset for this project was obtained from the BraTS Lighthouse Challenge 2025, which contains a large collection of multi-institutional, clinically acquired multi-parametric MRI (mpMRI) images of patients with diffuse

gliomas. The data includes both pre- and post-treatment scans, which show the disease's temporal course and the consequences of various therapies. The mpMRI sequences used to image each person in the dataset are as follows:

- T1 (weighted)
- Post-contrast T1 weighted (T1Gd)
- T2 weighted (T2)
- T2 FLAIR: Fluid Attenuated Inversion Recovery

All scans underwent uniform preprocessing, including skull stripping, co-registration to a shared anatomical template, and resampling to an isotropic resolution of 1 mm^3 to ensure spatial consistency across modalities and subjects.

Ground truth annotations for tumor sub-regions were created manually by 1-4 expert raters and then approved by neuroradiologists. These annotations address four distinct tumor sub-regions:

- Label 1): Non-Enhancing Tumor Core (NETC)
- Label 2): Surrounding Non-Enhancing FLAIR Hyperintensity (SNFH)
- Label 3): Enhancing Tumor (ET)
- Label 4): Resection Cavity (RC)

Each data sample is saved in NIfTI format (.nii.gz) and uses a structured naming convention BraTS-GLI-¡SubjectID¿-¡TimepointCode¿, where:

- 000: Pre-operative baseline scan 1
- 001: Pre-operative baseline scan 2
- 100 - First post-operative/treatment scan
- 101: Second post-operative/post-treatment scan

Scans labeled with a leading 1 are from post-surgical instances, whilst others may represent pre-treatment or non-surgical procedures. The dataset is divided into training and validation subsets, and reference standard annotations are only accessible for the training set.

5. Literature Review

Recent advancements in deep learning have transformed medical image segmentation, allowing for accurate delineation of anatomical structures and pathologies. Convolutional Neural Networks (CNNs), particularly U-Net architectures, are central to contemporary approaches. This review highlights key developments in foundational CNN architectures, 3D segmentation models, and attention mechanisms that improve segmentation accuracy.

The development of deep convolutional neural networks (CNNs) has been crucial in the advancement of medical image analysis. [8] introduced Inception (GoogLeNet), a groundbreaking architecture that employed parallel convolutional filters of different sizes to efficiently capture multi-scale features. This innovation greatly enhanced feature extraction while preserving computational efficiency, paving the way for the creation of more complex architectures in the field of medical imaging.

Traditional 2D convolutional neural networks (CNNs) have limitations when it comes to processing volumetric medical data, such as MRI and CT scans. To overcome these challenges [5], which incorporates Inception modules within a 3D U-Net framework. This architecture enhances feature reuse through dense connections and utilizes multi-scale feature extraction to improve brain tumor segmentation. The model has shown superior performance in segmenting gliomas, multiple sclerosis lesions, and regions affected by stroke, demonstrating the effectiveness of 3D convolutions in medical imaging tasks.

Although U-Net variants have enhanced segmentation accuracy, they often face challenges when dealing with fine-grained structures in cluttered backgrounds. [2] introduced Attention U-Net, which incorporates soft attention gates to dynamically emphasize relevant regions, such as the pancreas in abdominal CT scans. This mechanism helps to suppress irrelevant features while enhancing diagnostically critical areas, resulting in more accurate segmentation. The success of Attention U-Net has inspired further research in multi-organ segmentation and lesion detection.

The transition from Inception-based CNNs to 3D U-Net variants and attention-augmented models illustrates the rapid evolution of medical image segmentation. While HI-Net excels at processing volumetric data, Attention U-Net offers a lightweight yet effective solution for refining segmentation masks. Future research may focus on developing hybrid architectures that combine 3D convolutions with attention mechanisms, as well as exploring self-supervised learning to reduce dependence on large annotated datasets.

The reviewed works highlight significant advancements in medical image segmentation, mainly due to innovations in multi-scale feature extraction (Inception), 3D convolutions (HI-Net), and attention mechanisms (Attention U-Net). These improvements have greatly enhanced diagnostic accuracy and are paving the way for more robust and efficient segmentation models in clinical practice.

6. Models Used

6.1. 3D UNet

The 3D U-Net (Figure 1) employed in this study is a fully convolutional neural network that was created primarily for volumetric segmentation of medical images like the

mpMRI scans in the BraTS dataset. The architecture uses a symmetric encoder-decoder structure, sometimes known as a U-shaped architecture, to enable fast learning of both low-level spatial features and high-level contextual information. The network operates on 3D volumes and employs 3D convolutions, allowing it to capture volumetric spatial relationships that are essential for segmenting complex brain tumor sub-regions. The network's contracting path consists of four

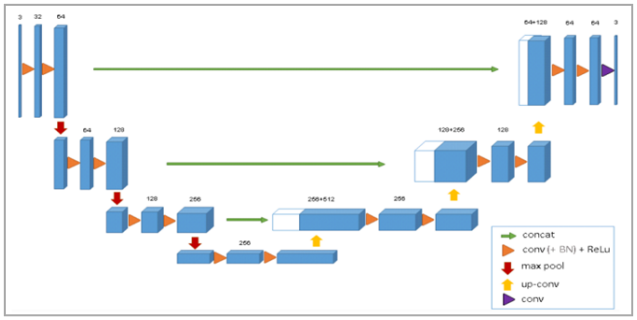

Figure 1. 3D UNet Architecture

double convolution blocks, each followed by 3D max pooling for downsampling. Each double convolution module consists of two consecutive 3D convolutional layers with kernel size 3 and padding 1, followed by batch normalization and dropout for regularization. The use of dropout rates scaled with feature map size helps to reduce overfitting, particularly in deeper layers. Following each downsampling step, the number of feature channels is doubled, allowing the network to gradually acquire more abstract representations of the input volume.

The expanding path mirrors the contracting path and is made up of a sequence of upsampling steps implemented with 3D transposed convolutions (also known as deconvolutions), followed by double convolution blocks. At each stage, the upsampled feature maps are concatenated with feature maps from the encoder path via skip connections. These skip connections preserve spatial resolution and detailed characteristics that would otherwise be lost during downsampling, considerably enhancing segmentation accuracy. Finally, a 1×1×1 convolution transfers the features to the required number of output channels, which usually match the number of segmentation classes. This design allows effective end-to-end learning for voxel-wise classification of tumor subregions throughout the 3D brain volume.

6.2. Inception UNet

The presented architecture (Figure 2) uses a modified 3D U-Net model with Inception-style blocks to segment brain tumors on the BraTS2025-GLI-PRE Challenge dataset. The model is based on an encoder-decoder structure with skip connections, which has been shown to be useful for medical segmentation of images applications. This approach stands out by substituting typical convolutional blocks with Inception blocks. These blocks use parallel convolutions of varied kernel sizes (1×1×1, 3×3×3, 5×5×5, and 7×7×7) to collect multi-scale information simultaneously. This multi-pathway strategy enables the network to learn both fine-grained local details and broader contextual patterns from volumetric MRI data, which is very useful for segmenting heterogeneous brain tumor locations of diverse sizes and textures.

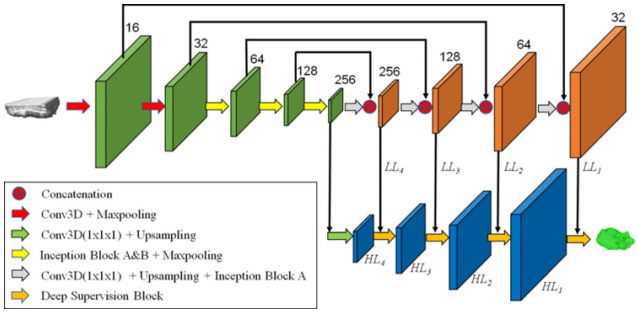

Figure 2. 3D Inception UNet Architecture

The encoder pathway comprises three Inception blocks followed by max pooling procedures, which reduce spatial dimensions while increasing feature depth (from 32 to 128 features). Each downsampling step doubles the feature channels, resulting in more abstract representations of the input volume. The bottleneck layer, which is implemented as an Inception block with 256 feature channels, connects the encoder and decoder while capturing the most complicated, high-level information. The decoder pathway is symmetrical to the encoder, with three upsampling steps utilizing transposed convolutions followed by Inception blocks. Skip connections from the encoder to the decoder enable the coupling of high-resolution spatial information with contextual understanding, preserving fine structural details in the final segmentation.

To address class imbalance difficulties that arise in brain tumor segmentation tasks, the model training adopts a combined loss function that incorporates cross-entropy loss and Dice loss. Cross-entropy loss gives pixel-level classification guidance, whereas Dice loss maximizes the overlap between predicted and ground truth segmentations. This dual-objective strategy allows the model to focus on reliably categorizing individual voxels while also generating coherent segmentation regions. In addition, the system contains advanced data handling procedures including per-channel intensity normalization via z-score standardization and center cropping to address potential dimension inconsistencies during feature map operations.

The model shows that it can process entire 3D brain volumes and segment them into separate tumour sub-regions (as demonstrated by the 4-class output indicating background, edema, enhancing tumor, and non-enhancing tu-

mor). This architecture efficiently combines computing economy with segmentation performance while still there for some room for improvement

6.3. Attention UNet

The Attention U-Net (Figure 3) is a more advanced version of the standard U-Net architecture that incorporates attention methods to increase segmentation performance. It employs an encoder-decoder architecture, with the encoder path capturing hierarchical characteristics via successive convolutional and pooling layers. Each level of the encoder employs a double 3D convolution block with batch normalization, ReLU activation, and dropout to extract rich spatial and contextual characteristics from the volumetric data. As the spatial resolution diminishes, the number of channels grows, allowing the model to learn deeper and more abstract representations of the input data. At the bottleneck,

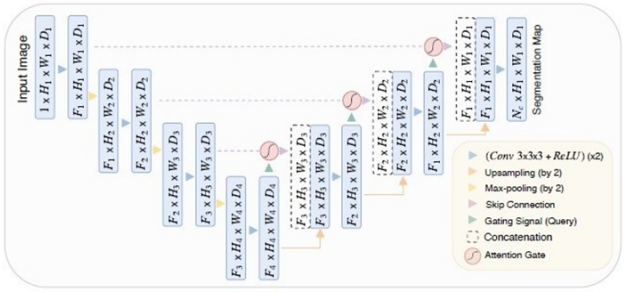

Figure 3. 3D Attention UNet Architecture

the most compressed version of the input is transmitted via additional convolutional layers that act as a bridge between the encoder and decoder. The attention gates put before each skip connection in the decoder are the Attention U-Net's main innovation. These attention blocks compare encoder features to gating signals from deeper layers in order to exclude irrelevant regions and emphasize important information. This selective filtering guarantees that only the most relevant spatial information passes through the skip connections, allowing the model to focus its attention on target structures during upsampling.

In the decoder or expanding path, the model gradually reconstructs the output by upsampling the feature maps and improving them via attention-weighted skip connections. To upsample the feature maps, transposed convolutions are utilized, followed by double convolution blocks that recover spatial details. This guided reconstruction produces more accurate and context-aware segmentation results. The final convolutional layer translates the revised feature representation to the necessary number of output classes, bringing the segmentation operation to completion. Overall, the architecture strikes a balance between spatial precision and semantic understanding, making it particularly helpful for medical picture analysis and other 3D data applications.

6.4. Attention-Inception (AI) UNet

The presented architecture (Figure 4) combines the previously established Inception-based U-Net framework and grid attention mechanisms for volumetric brain tumor segmentation. The model, like previous Inception U-Net implementations, has an encoder-decoder structure with three resolution levels, and Inception modules. Inception blocks utilize a four-pathway parallel convolutional approach with different kernel sizes (1×1×1, 3×3×3, 5×5×5, and 7×7×7), distributing output channels evenly over these branches. Following the established Inception paradigm, all branches' features are concatenated along the channel dimension and activated with ReLU. This architectural choice is still important for tumor segmentation tasks because it enables the simultaneous detection of tiny tumor boundaries as well as broader contextual relationships between tumor regions and nearby brain structures. The design, as previously im-

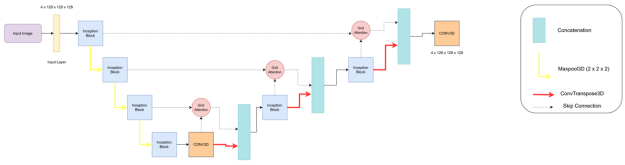

Figure 4. 3D AI UNet Architecture

plemented, includes GridAttentionBlock modules between matching levels of the encoder and decoder, which is an important feature that distinguishes it from ordinary U-Net implementation. These attention coefficients are upsampled to match the encoder feature dimensions and applied as multiplicative weights before being merged with the decoder path. This attention-guided feature fusion enables the network to focus on sensitive tumor borders while suppressing irrelevant background regions, correcting the inherent class imbalance in brain tumor segmentation, in which diseased regions account for just a small fraction of the overall volume.

The network architecture retains the core U-Net skip connection pattern while enhancing it with attention-modulated feature transfer. The encoder path is made up of three Inception blocks (enc1, enc2, and enc3) separated by max-pooling operations, which gradually reduce spatial dimensions while increasing feature channels from the input's four channels to 32, 64, and 128 features. The bottleneck Inception block processes features at 1/8 of the original resolution and 256 channels. The decoder uses three levels of transposed convolutions for upsampling, attention-modulated skip connections, and Inception blocks (dec3, dec2, dec1). Each decoder level receives features from the corresponding encoder level, but only after applying the attention mechanism, ensuring to capture the most relevant spatial locations

The model's optimization uses the same mixed loss func-

tion as before, combining cross-entropy loss with Dice loss to solve class imbalance, and the training procedure includes validation monitoring to prevent overfitting. The multi-scale processing capability of Inception blocks, together with the spatial selectivity of attention mechanisms, results in an integrated effect that allows the model to excel at segmenting complicated, heterogeneous tumor formations.

7. Metrics

The three metrics being calculated in this project are Mean IoU, Accuracy, and Dice Score each described in detailed below:

7.1. Mean IoU (Intersection over Union)

The Mean IoU measures the average overlap between expected and ground truth segmentations across all classes. It measures how well the projected areas match the actual labels, penalizing both over-segmentation and under-segmentation. IoU is calculated as follows:

$$IOU_c = \frac{Interesection_c}{Union_c} = \frac{|P_c \cap G_c|}{|P_c \cup G_c|}$$

where P_c is the predicted region for class c and G_c is the ground truth. IoU is computed per class and then averaged over classes present in the sample.

7.2. Mean Accuracy

This is the pixel-wise classification accuracy, or the proportion of correctly predicted voxel labels to the total number of voxels. It provides an overview of how many voxels were correctly classified, independent of class balance. The below equation shows how accuracy is computed:

$$\text{Accuracy} = \frac{\text{Number of correct voxels}}{\text{Total number of voxels}} = \frac{\sum 1(p_i = g_i)}{N}$$

where p_i is the predicted label for voxel i, g_i is the ground truth label, and N is the total number of voxels in the 3D volume.

7.3. Dice Score

The Dice Score (also known as the F1 score in segmentation) calculates the harmonic mean of precision and recall. It prioritizes accurate overlap over IoU and is particularly beneficial for reviewing medical photos where class imbalance is widespread. Dice score is calculated using the equation presented below:

$$\text{Dice}_c = \frac{2|P_c \cap G_c|}{|P_c| + |G_c|}$$

where $|P_c|$ and $|G_c|$ are the number of predicted and ground truth voxels for class c respectively. The mean dice score is computed across all valid classes.

Each of these metrics offer a different lens on performance: IoU is more strict on overlap, Accuracy gives overall correctness, and Dice balances precision and recall. Together, they give a holistic view of model quality.

8. Experiment Results

Figures 5, 6, 7, 8 show the sample masks predicted by the models on unseen MRI scans. The mask at the very end (right side) is the ground truth and the masks printed next to it (left) are the predicted masks.

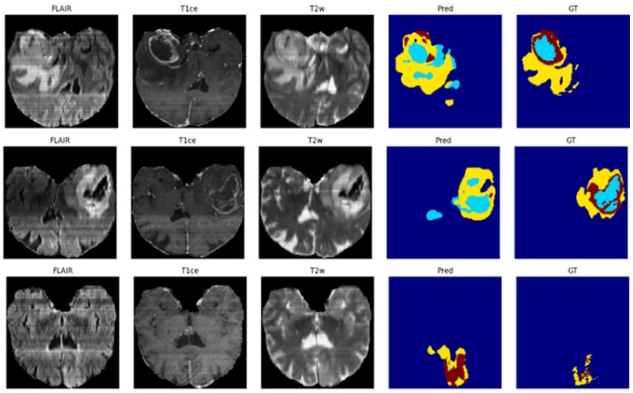

Figure 5. 3D UNet Inference Results

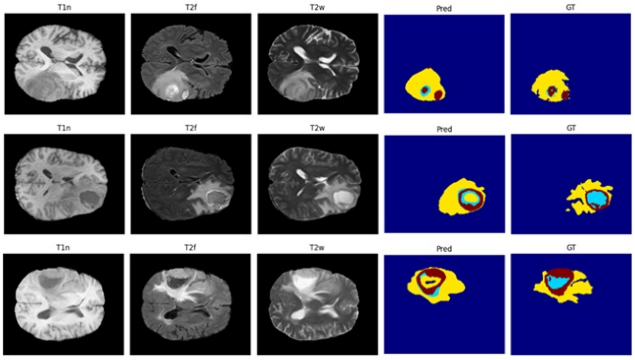

Figure 6. Inception UNet Inference Results

Table 1 summarizes the metrics obtained for each model on the test set.

9. Strengths and Weaknesses

The AI (Attention-Inception) U-Net emerged as the top-performing model, achieving the highest Dice coefficient (0.7047) and IoU (0.6153), which underscores its ability to integrate multi-scale feature extraction (Inception modules) with adaptive region weighting (attention gates). This hybrid design enables precise segmentation of heterogeneous

Model	IoU	Accuracy	Dice Score
3D UNet	0.4133	0.9495	0.4875
Inception UNet	0.5402	0.9939	0.6264
Attention UNet	0.6137	0.9401	0.7027
AI (Attention-Inception) UNet	0.6153	0.9950	0.7047

Table 1. Metrics for different models

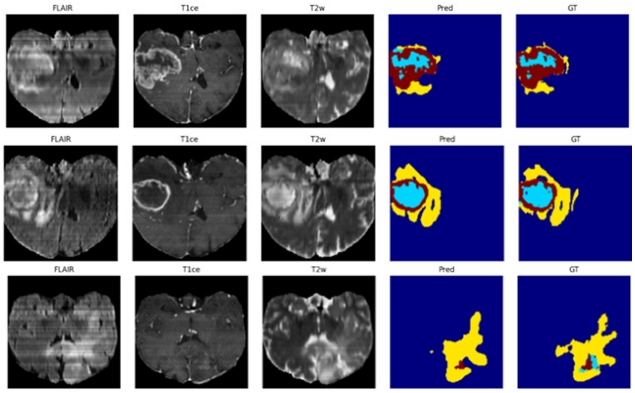

Figure 7. Attention UNet Inference Results

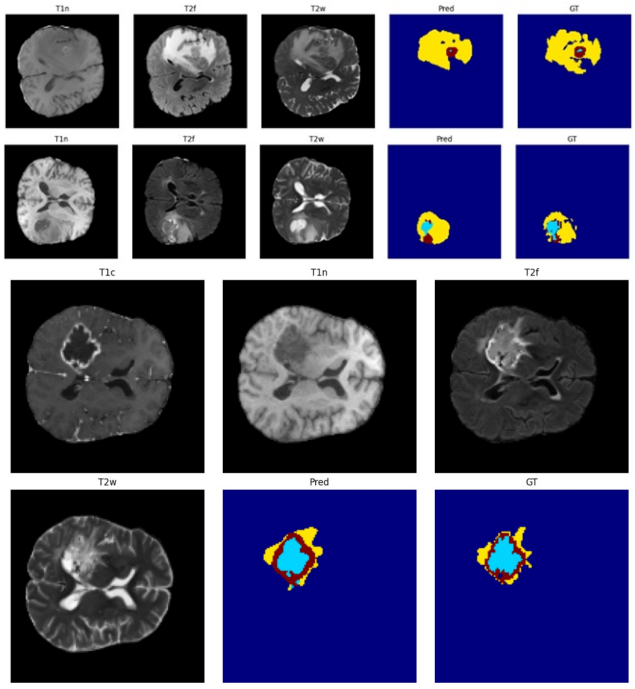

Figure 8. Attention UNet Inference Results

tumor sub-regions while maintaining near-perfect mean accuracy (0.9950), making it clinically reliable for tasks like surgical planning. Similarly, the Attention U-Net demonstrated notable gains (Dice: 0.7027) over simpler architectures, validating the efficacy of attention mechanisms in refining tumor boundaries. The Inception U-Net also showed significant improvement over the baseline 3D U-Net (Dice: 0.6264 vs. 0.4875), highlighting the advantage of multi-scale filters in capturing variable tumor sizes.

However, these advancements come with trade-offs. The AI U-Net's superior accuracy is offset by higher computational complexity, which may limit its deployment in resource-constrained settings. The Attention U-Net, while excelling in boundary detection, exhibited a slight dip in mean accuracy (0.9401 vs. 0.9939 in Inception U-Net), suggesting potential over-suppression of background regions. Meanwhile, the 3D U-Net's modest performance (Dice: 0.4875) reveals its limitations in handling small or diffuse tumors, despite its simplicity and efficiency. These findings emphasize a key design consideration: balancing precision with computational cost, where the optimal model depends on clinical priorities (e.g., accuracy vs. real-time processing).

10. Conclusion

Through our experiments, we observed that each architecture had distinct strengths and weaknesses: 3D U-Net tended to over-segment and missed accurate tumor boundaries, Inception U-Net captured shapes well but lacked fine detail, while Attention U-Net preserved fine structures but struggled with overall tumor morphology. By combining the strengths of both Inception and Attention mechanisms in a hybrid Attention Inception U-Net, we achieved significantly more accurate and balanced segmentation of glioma tumors.

11. Future Scope

Building on our current findings, future work will explore integrating transformer-based attention mechanisms, particularly multi-head attention within the attention gates to enhance contextual understanding. We also aim to fine-tune our hybrid model using parameter-efficient techniques like LoRA on additional brain tumor datasets or broader tumor segmentation tasks. Additionally, experimenting with residual variants of Inception blocks and introducing residual connections within the U-Net framework could further improve performance and generalization.

References

[1] J. Kugelman, J. Allman, S. A. Read, S. J. Vincent, J. Tong, M. Kalloniatis, F. K. Chen, M. J. Collins, and D. Alonso-Caneiro. A comparison of deep learning u-net architectures for posterior segment oct retinal layer segmentation. *Scientific reports*, 12(1):14888, 2022.

[2] O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz, et al. Attention u-net: Learning where to look for the pancreas. *arXiv preprint arXiv:1804.03999*, 2018.

[3] A. Paszke, A. Chaurasia, S. Kim, and E. Culurciello. Enet: A deep neural network architecture for real-time semantic segmentation. *arXiv preprint arXiv:1606.02147*, 2016.

[4] S. Qamar, P. Ahmad, and L. Shen. Hi-net: Hyperdense inception 3 d unet for brain tumor segmentation. In *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 6th International Workshop, BrainLes 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 4, 2020, Revised Selected Papers, Part II 6*, pages 50–57. Springer, 2021.

[5] S. R. Ravichandran, B. Nataraj, S. Huang, Z. Qin, Z. Lu, A. Katsuki, W. Huang, and Z. Zeng. 3d inception u-net for aorta segmentation using computed tomography cardiac angiography. In *2019 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)*, pages 1–4. IEEE, 2019.

[6] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*, pages 234–241. Springer, 2015.

[7] E. K. Rutoh, Q. Z. Guang, N. Bahadar, R. Raza, and M. S. Hanif. Gair-u-net: 3d guided attention inception residual u-net for brain tumor segmentation using multimodal mri images. *Journal of King Saud University-Computer and Information Sciences*, 36(6):102086, 2024.

[8] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9, 2015.

[9] Z. Zhang, C. Wu, S. Coleman, and D. Kerr. Dense-inception u-net for medical image segmentation. *Computer methods and programs in biomedicine*, 192:105395, 2020.

[10] Z. Zhou, M. M. Rahman Siddiquee, N. Tajbakhsh, and J. Liang. Unet++: A nested u-net architecture for medical image segmentation. In *Deep learning in medical image analysis and multimodal learning for clinical decision support: 4th international workshop, DLMIA 2018, and 8th international workshop, ML-CDS 2018, held in conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, proceedings 4*, pages 3–11. Springer, 2018.

[11] Z. Zhu, Y. Yan, R. Xu, Y. Zi, and J. Wang. Attention-unet: A deep learning approach for fast and accurate segmentation in medical imaging. *Journal of Computer Science and Software Applications*, 2(4):24–31, 2022.