3D Dual Residual U-Net with Attention Gate and Spatial Attention Mechanisms (3D-DRUwAS) for Brain Tumor Segmentation

Anonymous CVPR submission

Submission Number: 16582

Abstract

Glioma is a harmful brain tumor that requires early detection to ensure better health results. there is a challenging task to find tumors due to tumor characteristics like location and size. A reliable method to accurately separate tumor zones from healthy tissues is deep learning models, which have shown promising results over the last few years. In this research, a 3D Dual Residual U-Net with Attention Gate and Spatial Attention Mechanisms (3D-DRUwAS) is introduced. This model is an innovative combination of dual residual networks, attention gate, and spatial attention mechanisms. The dual residual network architecture captures high-level semantic and intricate low-level details from brain images, ensuring precise segmentation of different tumor parts, types, and hard regions. The integrated attention gate and spatial attention mechanisms assign variable weights to image regions and preserve the spatial context, enhancing the focus on informative features related to tumor regions and improving segmentation accuracy. We initially trained the model for 100 epochs using the ReLU activation function, achieving substantial results. Then, we trained our model for an additional 60 epochs, utilizing pre-trained weights and the swish function. These adjustments have improved our model's accuracy in tumor detection and segmentation. The 3D-DRUwAS model is evaluated based on BraTS 2018 and BraTS 2019 and BraTS 2020 dataset. The results demonstrate a higher Dice score in comparison to state-of-the-art techniques.

1. Introduction

A brain tumor, an abnormal growth of cells, poses a significant threat to the patient's health and survival. Early detection of tumors is essential in effective treatment, requiring the precise segmentation of tumorous areas. There are various glioma imaging techniques; among them, MRI is considered the standard imaging method for diagnosing brain tumors [1]. There are four modalities in MRI images for Glioma: fluid-attenuated inversion recovery (FLAIR), T1 weighted (T1), T2 weighted (T2), and contrast-enhanced T1 weighted (T1ce). Fig 1 shows different modalities along with their ground truth.

Figure 1. Different Modalities of Brain MRI with their Ground Truth

Traditionally, manual annotation of these tumors by specialists has been the norm, but it is time-consuming, costly, and sometimes inconsistent due to human error. This has led to the rise of automated segmentation approaches powered by deep-learning-based methods, particularly using Convolutional Neural Networks (CNNs), which have high efficiency and strong generalizability in the detection of brain tumors. In addition to CNN, Deep learning techniques have demonstrated significant promise in enhancing the accuracy of tumor segmentation [2]. However, these methods typically employ additional convolutional layers and pooling layers, which can lead to a network degradation problem. This often results in reducing segmentation accuracy and the overall performance of the model.

To address this challenge, various techniques have been explored, such as skip connections [3], residual connections [4], and attention mechanisms [5]. These methods aim to alleviate the negative effects of network degradation and maintain or even enhance segmentation precision. Residual networks, for instance, were introduced as a solution to increase the performance of deep neural networks [4]. Alongside these developments, the U-Net Architecture has become a widely adopted framework in this field [3]. Despite the progress made in brain tumor segmentation, a significant challenge remains in dealing with the issue of unequal representation between healthy and tumorous tissue. This paper proposes a novel 3D-Net model to segment a brain tumor. This enhanced model integrates a new architecture of attention gates, spatial attention, and ResNet blocks into the traditional U-Net framework. Our experiments aim to demonstrate that this model offers improved performance over existing architectures in brain tumor segmentation.

2. Related Work

2.1. U-Net Architecture

U-Net, a CNN optimized for biomedical image segmentation, offers fast and precise results [3]. Its unique feature is a symmetrical expanding path that complements the contracting path, reducing input size while increasing depth for context capture. The expanding path then upsamples feature maps for detailed pixel predictions. U-Net's skip connections transfer information between layers, enhancing local and global context use in segmentation. Despite these innovative features, U-Net exhibits certain limitations for more complex tasks such as brain tumor segmentation, particularly when the regions of interest in an image are small or subtle [6].

To overcome these challenges, several enhancements have been proposed to address U-Net's limitations. One of the significant adaptations is the development of a 3D U-Net, which extends U-Net's 2D operations into 3D, thereby enabling more effective handling of volumetric images [7]. Furthermore, the integration of residual connections and attention mechanisms has shown promising results. Residual U-Net, for instance, merges the benefits of U-Net and Residual Network (ResNet), enabling deeper networks without the issue of vanishing gradients [4].

On the other hand, Attention U-Net uses attention gates to selectively focus on relevant features and suppress irrelevant ones, potentially improving segmentation performance in complex tasks [8]. While U-Net has established a significant foundation for biomedical image segmentation, there still exists potential for further improvements. By embracing approaches such as residual learning, attention mechanisms, and complex connection structures, the proficiency and versatility of biomedical image segmentation can steadily progress.

2.2. Residual Networks

Residual Networks, or ResNets [4], address the problem of training deep neural networks by using skip connections or shortcuts, known as "residual blocks" , to jump over some layers. These links enable the model to establish a function that guarantees each subsequent layer will operate at least as effective as the preceding one, helping to reduce the issue of diminishing gradients. The main idea is to redefine the underlying mapping to be learned by the network, making optimizing and enabling the training of deeper networks easier. Adding these residual connections to U-Net can help learn more complex mappings and make optimization easier, thus potentially improving the model's performance on complex segmentation tasks.

Despite the significant achievements of ResNets, they are not flawless. In particular, their performance on segmentation tasks, including brain tumor segmentation, can be improved further. In [9], a system utilizing an 18-layer ResNet architecture is proposed for identifying brain tumor types. Another study [10], introduces a multipathway architecture built upon the U-Net with residual connections, which efficiently predicts multiple tumor types in a single pass. Furthermore, a paper [11] presents a model that combines a modified LinkNet structure with the ResNet152 for advanced tumor segmentation. While ResNets provides a powerful tool for deep learning tasks, their potential can be further unleashed through the right combination of its methods and architectures.

2.3. Attention Mechanisms

Attention mechanisms initially presented in Natural Language Processing (NLP) and have gained popularity in many other areas, including computer vision [5]. The principal concept of attention mechanisms enables the model to concentrate on specific input sections. Among the various types of attention mechanisms, Attention Gate and Spatial Attention are chosen for this work due to their potential ability to extract important feature maps in the proposed model.

Attention Gate (AG) helps the network focus on specific areas of an input image by using gating feature maps from the encoder path in the U-Net [8]. This approach enhances model performance by highlighting relevant features and suppressing irrelevant ones. Its integration into U-Net has been particularly successful in brain tumor segmentation. For instance, the GCAUNet model emphasizes tumor details and key features [12], while the Attention Res-UNet with Guided Decoder (ARU-GD) combines a guided decoder with attention gates for advanced feature activation [13]. the Multimodal Attention-gated Cascaded U-Net (MAC U-Net) is tailored for early-stage low-grade brain tumors, using group normalization and attention gates for precision [14]. Additionally, the AGResU-Net merges residual modules with attention gates, proving effective for small tumor segmentation [15].

Spatial attention, on the other hand, allows the network to focus on pertinent sections of the image. However, this attention mechanism works by assigning varying importance levels to different spatial locations [16]. A variety of novel architectures have been proposed with the integration of Spatial attention mechanism. The Efficient Spatial Attention Network (ESA-Net) is an example [17], optimizing U-Net with Efficient Spatial Attention (ESA) blocks for enhanced efficiency and accuracy. Another innovative model, the SCAR U-Net, combines channel and spatial attention in a 3D residual U-Net architecture [18]. These advancements highlight the growing role of attention mechanisms in refining image segmentation models.

3. Dataset Description

Our research employs the BraTS 2020 dataset, designed for the Multimodal Brain Tumor Segmentation Challenge (BraTS) [19]. It includes 3D MRI scans from 369 patients with Glioma, across four modalities, each with a dimension of 240 x 240 x 155 slices. These modalities provide a detailed view of the brain's structure.

The BraTS 2020 dataset includes four labels: Label 0 (Background), Label 1 (Necrosis and non-enhancing tumor), Label 2 (Edema), and Label 4 (Enhancing tumor). Each of these labels corresponds to unique characteristics of the brain scans, which are individually marked by one to four neuroradiologists. For our study, we used labels 1, 2, and 4 for one-hot encoding, creating a binary representation to simplify the network's prediction process and reduce segmentation complexity.

Additionally, we extended our research to include the BraTS 2018 and 2019 datasets [6, 20, 21]. The 2018 dataset contains data from 285 patients and a validation set of 66 cases, while the 2019 dataset includes 335 glial tumor cases and a validation set of 125 cases. Both datasets provide the same four MRI modalities and ground truth labels for comprehensive analysis.

4. Proposed Method

Our proposed architecture, 3D Dual Residual U-Net with Attention Gate and Spatial Attention Mechanisms (3D-DRUwAS), combines the proven strengths of the 3D U-Net architecture along with a series of enhancements tailored for the effective handling of three-dimensional data. Specifically, it integrates residual connections, spatial attention, and attention gates to construct an intricate and potent network capable of performing volumetric segmentation or classification tasks with high efficiency.

The significant innovation in the 3D-DRUwAS is the integration of dual residual connections in the form of modified Residual Blocks (ResBlocks). These ResBlocks form the backbone of the encoder part of the model, replacing the traditional convolutional layers used in U-Net models. Each ResBlock incorporates two 3D convolutional layers, utilizing ReLU activations and Group Normalization for effective feature extraction. The residual connection within each block allows the model to bypass the main layers if necessary, alleviating potential issues of vanishing or exploding gradients and enabling the model to learn identity functions. Consequently, the ResBlocks helps capture both simple and complex patterns in the data across different levels of abstraction.

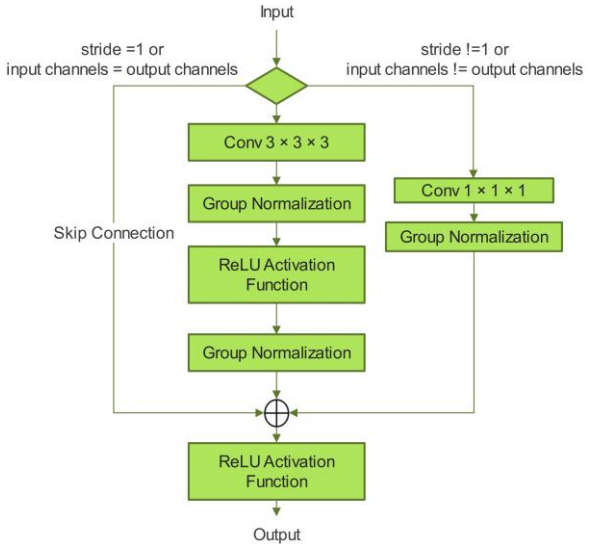


Figure 1: Residual Block of the proposed architecture

In this architecture, the input of the residual block choose one path based on these conditions:

- If stride = 1 or input channels = output channels, this path directly goes down as a "Skip Connection," bypassing all the intermediate layers and connecting directly to the end.
- If stride = 1 or input channels ≠ output channels, the data flows through a convolutional layer "Conv 3 × 3 × 3" followed by "Group Normalization." After this normalization, the "ReLU Activation Function" is applied. The output from the ReLU function undergoes another "Group Normalization." The flow then proceeds through a "Conv 1 × 1 × 1" convolutional layer followed by its own "Group Normalization."

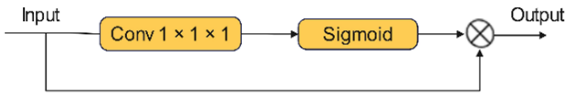


Figure 2. Spatial Attention Block of 3D-DRUwAS architecture.

The 3D-DRUwAS 's main strength comes from its advanced attention mechanisms, which include both spatial attention and attention gates. The spatial attention blocks apply a 3D convolution to the input and use a sigmoid activation function to generate a spatial attention map. This map is then multiplied with the original input to direct the model's attention toward the regions of extracted features that hold the most significant information. Meanwhile, the attention gates use gating and input signals to compute attention coefficients for the input features. This gate guides the network to give more emphasis to certain channels and suppress others, facilitating the model's

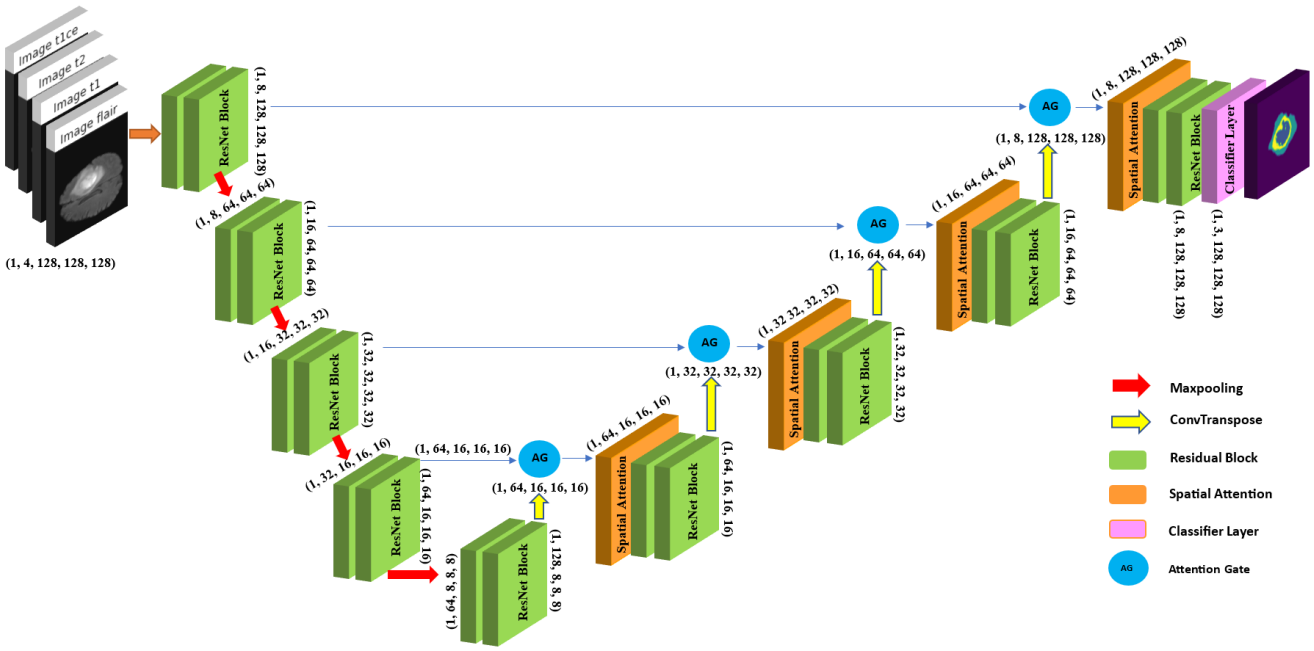


Figure 3: The architecture of 3D Dual Residual U-Net with Attention Gate and Spatial Attention Mechanisms (3D-DRUwAS)

ability to identify intricate and varied dependencies in the 3D input data. The proposed architecture is illustrated in Fig. 3.

The proposed residual block incorporates "Group Normalization" in place of the traditional Batch Normalization. It also adds an additional "Conv 1 × 1 × 1" convolutional layer, which is followed by Group Normalization, before it connects with the skip connection. Additionally, in this design, ReLU activation functions are positioned after the group normalization layers, while in the original design [4], they are placed after Batch Normalization and before the convolutional operations. The residual block is shown in Fig. 1.

In Fig. 4, Attention gate mechanism is depicted which employs Group normalization rather than batch normalization. Group normalization offers benefits in situations where batch size is small or inconsistent. In addition of attention mechanism, a spatial attention is integrated to the model to increase the model's performance. It involves processing 3D Convolution block outputs into attention maps using a Sigmoid Activation block. These maps are then element-wise multiplied with the input tensor, resulting in a Weighted Tensor with spatially focused features. Finally, you get the Weighted Tensor as the output, which contains the spatially attended features. Fig. 2 presents the Spatial Attention Block, where spatial attention involves calculating attention scores, normalizing them (often with softmax), and applying them to the input, typically through multiplication [5]. Our approach uses a 3D convolutional layer with a 1-size kernel for linear voxel transformation, followed by a sigmoid activation to generate spatial attention weights, which are then multiplied with the input.

The model begins with an input of shape (1, 4, 128, 128, 128), where 1 represents the batch size, 4 represents the channels (the image modalities), and the last three dimensions are for the epth, height, and width.

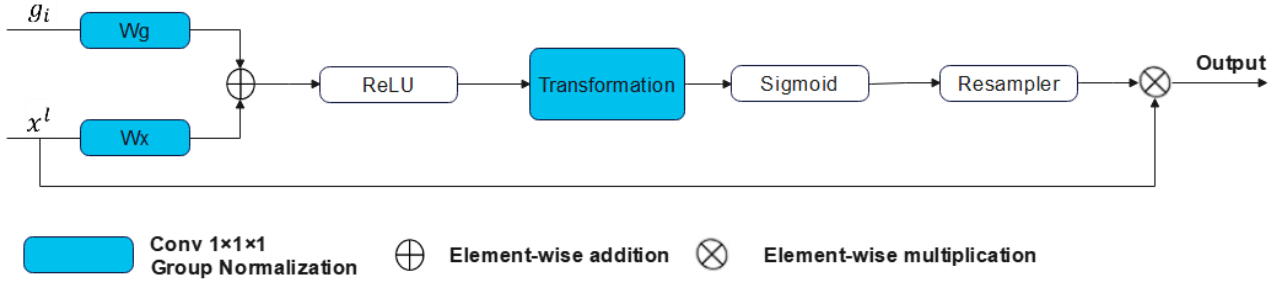


Figure 4: Spatial Attention Block of 3D-DRUwAS architecture.

Table 1: Comparison of the proposed model with other approaches on the BraTS 2020 dataset.

Architecture	Year	Dice Score			HD95		
		WT	TC	ET	WT	TC	ET
Dual-Path attention U-Net [22]	2021	0.8780	0.7790	0.7520	6.3000	11.0200	30.6500
3D self-ensemble ResUNet [23]	2021	0.89	0.81	0.76	5.28	7.74	33.26
modified 3D U-net [24]	2021	0.90	0.84	0.74	5.08	8.69	36.04
TransBTS [25]	2021	0.8900	0.8136	0.7850	6.4690	10.4680	16.7160
MENet [26]	2021	0.8800	0.7400	0.7000	6.9500	30.1800	38.6000
ensembles of CNN [27]	2021	0.90	0.81	0.77	6.16	7.55	21.80
Swinbts [28]	2022	0.8906	0.8030	0.7736	8.56	15.78	26.84
3D PSwinbts [29]	2022	0.9076	0.8420	0.7948	5.573	7.252	19.437
SwinUnet [30]	2023	0.8934	0.7760	0.7895	7.855	14.5940	11.0050
Pre-operative three-dimensional MRI [31]	2023	0.88	0.79	0.77	7.79	13.86	32.69
A deep CNN model for glioma tumor segmentation [32]	2023	0.85	0.90	0.74	6.33	5.75	5
HMNet [33]	2023	0.901	0.823	0.781	5.954	7.055	21.340
Selective Deeply Supervised Multi-Scale Attention Network [34]	2023	90.24	86.93	80.64	4.27	6.32	5.87
Self-supervised hybrid fusion network [35]	2023	86.84	74.12	73.49	6.416	23.439	23.439
Segmentation via pixel-level and feature-level image fusion [36]	2023	0.8950	0.8178	0.7745	5.3117	9.4285	4.4715
Large Kernel Attention [37]	2023	90.68	84.82	78.94	3.65	5.02	25.22
dResU-Net [38]	2023	0.8660	0.8357	0.8004	-	-	-
Multimodal Transformer of Incomplete MRI Data [39]	2023	90.64	87.41	81.55	5.65	6.70	5.83
MM-UNet [40]	2023	0.850	0.765	0.762	8.243	10.766	6.389
3D-DRUwAS(Relu) (ours)	**2023**	**0.9096**	**0.8869**	**0.8385**	**1.5853**	**2.5981**	**2.6794**
3D-DRUwAS(Relu+Swish) (ours)	**2023**	**0.9115**	**0.9137**	**0.8584**	**1.4195**	**1.9541**	**2.2840**

In the model's downsampling path, each encoder block has two residual blocks and a max-pooling layer, with feature maps doubling from an initial count of eight. This captures the input image's context. The output then passes through a bottleneck comprising two residual blocks, processing the data further before the decoding pathway.

In the decoder, a transposed 3D convolution upsamples feature maps. An attention gate merges these with encoder outputs, focusing on relevant features. A spatial attention mechanism, using sigmoid-activated 3D convolution, assigns weights to different feature map areas, highlighting key regions. This process repeats in each decoding stage, gradually enhancing the spatial resolution of the features.

In the final step, a convolution layer of 1×1×1 dimensions is utilized to transform the multi-channel feature maps to the requisite number of classes, producing the ultimate segmentation map. In this case, the output size is (1, 3, 128, 128, 128), where 3 represents the number of segmentation classes.

5. Evaluation measures

Our methodology, built using Python programming language, leverages the capabilities of the PyTorch library [45]. The optimization process utilized the ADAM

Table 2: Comparison of the proposed model with other approaches on the BraTS 2018 dataset.

Architecture	Dice Score			HD95		
	WT	TC	ET	WT	TC	ET
3D U-Net with Augmentation [41]	0.873	0.783	0.751	5.90	8.03	4.53
Context Aware 3D UNet [42]	0.872	0.795	0.741	5.04	9.59	5.58
MCCNN and CRFs [43]	0.8824	0.7481	0.7178	12.6069	9.6223	5.6864
ResU-Net [15]	0.870	0.802	0.760	-	-	-
AG-ResUNet [15]	0.872	0.808	0.772	-	-	-
Attention-aware fusion [44]	0.861	0.871	0.789	6.2	5.2	3.1
3D-DRUwAS (ours)	**0.8968**	**0.8524**	**0.7968**	**2.6901**	**4.6084**	**5.0269**

Table 3: Comparison of the proposed model with cutting-edge techniques utilizing the BraTS 2029 dataset

Architecture	Dice Score			HD95		
	WT	TC	ET	WT	TC	ET
dual supervision guided attentional network [46]	0.882	0.771	0.727	8.09	10.3	6.6
heuristic approach for segmentation [47]	0.8598	0.7728	0.7153	-	-	-
RAAGR2-Net [48]	0.884	0.814	0.763	-	-	-
Swinbts **[28]**	0.8975	0.7928	0.7443	-	-	-
Cascaded 3D U-Net and 3D U-Net++ [49]	0.867	0.834	0.802	-	-	-
Multiscale lightweight 3D segmentation with attention mechanism [50]	**0.8994**	0.8349	0.7791	5.45	6.56	**4.03**
3D-DRUwAS (ours)	0.8939	**0.8502**	**0.8117**	**2.0616**	**3.5858**	4.3453

optimizer, beginning with a learning rate of 5e-4.
To further enhance the training, a learning rate scheduler based on the ReduceLROnPlateau function with a patience of 4 was employed. This effectively adjusted the learning rate downwards whenever the model's performance reached a plateau. In terms of architectural composition,

we utilized the ReLU activation function in conjunction with group batch normalization. This combination has been proven to enhance our model's stability and performance by normalizing the network. In spite of the constraints related to computational resources, we succeeded in training the model over a cumulative of 100 epochs, using a batch size of 4. Additionally, the input image was constructed using four modalities stacked together, each with a size of 128*128*128. The data was divided so that 70% was allocated for training, 20% for testing, and 10% for validation. This computational task was executed concurrently on two Tesla T4 GPUs, each equipped with 16 GB of RAM, made available by Kaggle.

Our research expands upon the successes of the conventional U-Net model that has proven effective in the domain of brain tumor segmentation. We've innovatively refined this model by embedding a residual network, an attention gate, and spatial attention, enhancing its depth with ResNet blocks in our architectural framework.

5.1. Dice Coefficient Score

The Dice Coefficient Score (DSC) is a metric employed to assess the likeness between two sets. Within the scope of image segmentation, for example, in brain tumor segmentation, these "sets" indicate the anticipated segmentation and the actual or 'ground truth' segmentation. DSC is particularly beneficial in medical imaging because it provides a measure of overlap that is easy to interpret. It evaluates how close the predicted segmentation is to the actual segmentation by quantifying the spatial overlap accuracy. Therefore, a higher DSC suggests a higher match between the predicted and actual values, leading to better segmentation results.

DSC Equation is calculated as follows in Equation 1:

$$DSC = 2 \times \frac{(X \cap Y)}{(|X| + |Y|)} \qquad (1)$$

In this context, 'X' represents the anticipated segmentation, while
'Y' stands for the actual truth. The symbol '∩' implies intersection, and '|' signifies the magnitude of the set. Each voxel of the tumor is tagged as 1, and those not part of the tumor are tagged as 0.

5.2. Hausdorff Distance

The Hausdorff Distance (HD) is a well-liked metric for assessing the similarity between two sets of points [33]. It's particularly effective in quantifying the 'proximity' of the predicted segmentation to the actual, especially in medical image analysis. A variant of this metric, the 95th percentile Hausdorff Distance (HD95), excludes the top 5% of distance values.

The value of HD95 lies in its robustness as a measure of the spatial distance between the boundaries of the segmented and true regions, making it less susceptible to outliers than the traditional Hausdorff distance. HD95 can be determined as shown in Equation 2:

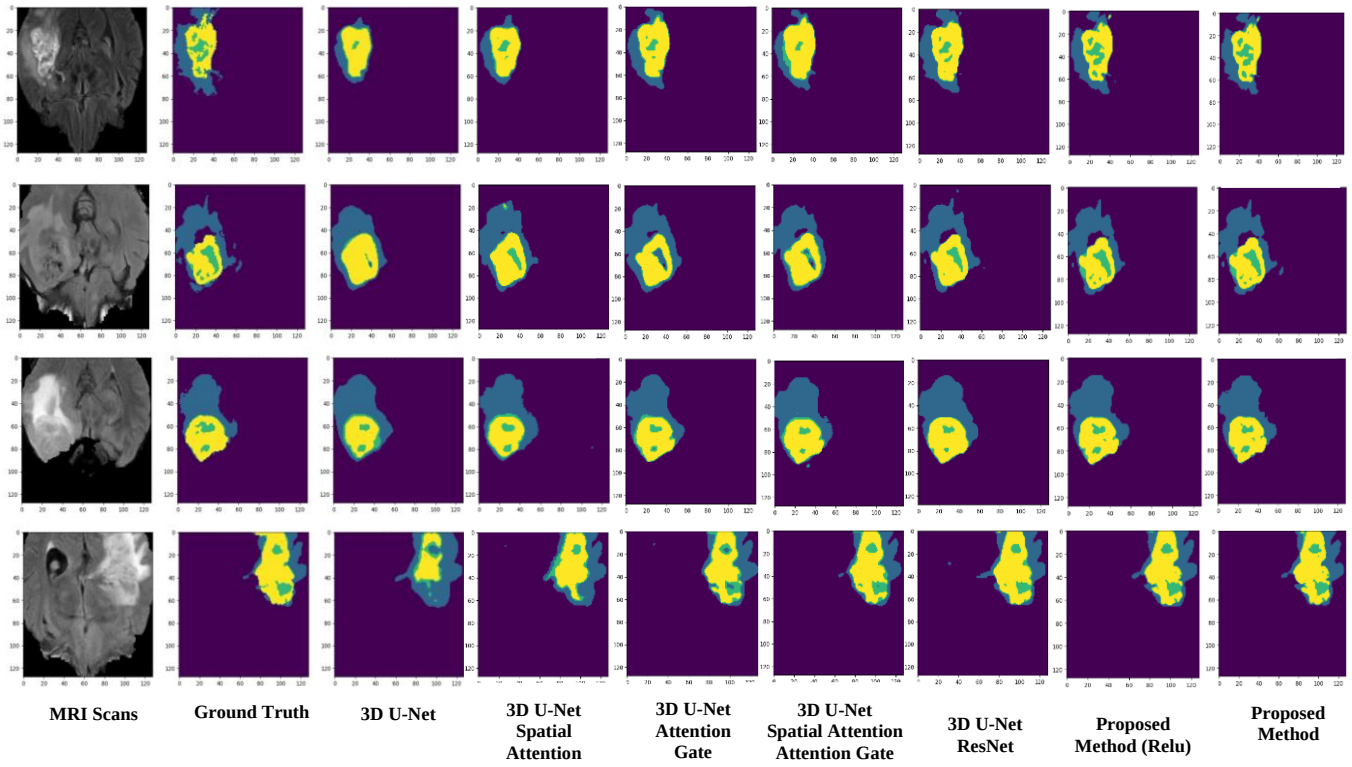
$$HD95\ (X,Y) = \max\big(hd95(X,Y), hd95(Y,X)\big) \qquad (2)$$

In the Equation, X and Y denote two sets of points. The function hd95(X,Y) determines the 95th percentile of the shortest distances from any point in set X to any point in set Y. Conversely, hd95(Y,X) does the same, but from set Y to set X.

The combination of DSC and HD95 provides a comprehensive assessment of the performance of segmentation models, considering both the overlap of segmented regions (DSC) and the spatial distance between their boundaries (HD95).

We also experimented with various alternatives to the ReLU activation function to boost the predictive accuracy, including LeakyReLU, GeLU (Gaussian error Linear Unit), and Swish, utilizing our pre-trained model.

Figure 5: The Model's performance and its ablation study on segmentation of Brain tumor.



Importantly, the Swish activation function resulted in enhanced model performance. Figure 4 depicts an additional 60 epochs of training for the model using the Swish activation function to attain its peak performance.

The Dice Score is utilized to gauge the overlap between two instances. A higher Dice Score indicates better overlap, suggesting the better performance of the model in accurately segmenting the tumor area. Regarding the Dice Score, our proposed model provides a good performance in segmenting Whole Tumor (WT), Tumor Core (TC), and Enhanced Tumor (ET), attaining scores of 0.9121, 0.9121, and 0.8583, respectively. Also, we used HD95 to provide an estimate of the worst-case segmentation error. Lower values imply better performance in terms of segmentation accuracy. Our proposed model 3D-DRUwAS demonstrates a better average performance compared to all the other models, with HD95 scores for Whole Tumor, Tumor Code, and Enhanced Tumor being 1.2261, 1.33527, and 1.6837, respectively.

From these results, the proposed models not only demonstrate high accuracy in tumor segmentation, as evidenced by the high Dice scores, but they also show high precision by achieving low Hausdorff distances. It is apparent that our proposed models, especially 3D-DRUwAS (ReLU + Swish), provide better performance in terms of both segmentation overlap and accuracy in regard to the existing state-of-the-art models. Table 1 presents a comparative analysis between the 3D-DRUwAS model and the latest models using the BraTS 2020 dataset. The bolded values show the model's performance in compared to the other methods. Also, the proposed method has been tested on BraTS 2018 and 2019, which is shown in Table 2 and 3, respectively.

6. Discussion

Brain tumor segmentation is an essential task in medical diagnostics. With the help of precise automatic segmentation tools, doctors can detect the tumor and follow up treatment plans immediately. Recently, U-Net models have shown significant promise in these complex segmentation tasks. However, there are computational limitations when dealing with high-resolution 3D images like brain MRIs. Furthermore, the precision of U-Net segmentation is highly dependent on the amount of training data.

Our study proposed an innovative 3D Dual Residual U-Net with Attention Gate and Spatial Attention Mechanisms (3D-DRUwAS). The dual residual network helps the model learn features at different levels of abstraction, thereby capturing the intricate and diverse nature of gliomas.

Table 4: Ablation study of our proposed method on BraTS 2020 dataset.

Architecture	Dice Score			HD95		
	WT	TC	ET	WT	TC	ET
3D-Unet (Baseline)	0.8489	0.7181	0.5915	3.2305	6.0357	18.084
Baseline + Spatial Attention	0.8380	0.7465	0.6793	3.4882	4.8240	8.1997
Baseline + Attention Gate	0.8779	0.7710	0.7551	3.5241	3.7701	2.7929
Baseline + Spatial Attention and Attention Gate	0.8693	0.8026	0.7792	3.7875	3.8907	2.9142
Baseline + ResNet	0.8986	0.8586	0.7896	3.1693	2.5136	3.2973
Proposed Method (Relu)	0.9096	0.8869	0.8385	1.5853	2.5981	2.6794
Proposed Method (Relu + Swish)	0.9115	0.9137	0.8584	1.4195	1.9541	2.2840

The attention gate mechanism reduces the impact of irrelevant regions, allowing the model to focus on the more informative tumor regions. The spatial attention mechanism maintains the spatial context of images, ensuring the accurate segmentation of smaller or hard-to-detect tumor regions. This approach has made an advancement in brain tumor segmentation by enhancing segmentation performance and surpassing previous methods in terms of precision and proficiency.

The effectiveness of these modules on the model's segmentation performance, particularly their impact on Dice scores and HD95 values, is detailed in Table 4. Additionally, Figure 1 depicts the model's performance and its predictions, demonstrating the enhanced accuracy achieved through the integration of these modules.

While our novel 3D-DRUwAS model has shown promising results in detecting brain tumors, there are certain limitations that need to be addressed. It uses similar convolution operations throughout, possibly overlooking the varied computational needs for different image details. Also, the model's performance highly depends on extensive datasets for training. Limited datasets might reduce the model's robustness and generalizability in detecting a wide range of brain tumors. Future directions in tumor segmentation involve using longitudinal data for insights into tumor growth, multimodal imaging for robustness, and generative models like GANs for training with limited data. Transfer learning could also enhance performance where data is scarce. We aim to continue advancing in medical imaging segmentation with these developments.

7. Conclusion

Our proposed model, with its dual residual network structure, successfully captures both high-level semantic features and low-level details from brain images. The comprehensive feature learning allows the model to effectively distinguish between hard regions, tumor regions, and healthy tissues, demonstrating improved precision and accuracy in isolating different tumor types and parts. The integration of the attention mechanism assigns varying weights to different image regions, enabling the model to concentrate on the more informative tumor and hard regions while lessening the impact of non-tumor areas. This mechanism improves overall segmentation performance and enhances boundary delineation. Furthermore, the spatial attention mechanism maintains intricate spatial structures, facilitating the accurate segmentation of smaller or hard regions of the tumor, which might otherwise be overlooked or lost during the segmentation process. Our experimental results demonstrate the proposed 3D-DRUwAS method significantly outperforms contemporary state-of-the-art methods. It validates our innovative combination of dual residual networks, attention mechanisms, and spatial attention as a robust architecture for brain tumor segmentation.

8. References

[1] M. Weller *et al.*, "European Association for Neuro-Oncology (EANO) guideline on the diagnosis and treatment of adult astrocytic and oligodendroglial gliomas," *The lancet oncology*, vol. 18, no. 6, pp. e315-e329, 2017.

[2] M. Havaei *et al.*, "Brain tumor segmentation with deep neural networks," *Medical image analysis*, vol. 35, pp. 18-31, 2017.

[3] 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*, 2015: Springer, pp. 234-241.

[4] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770-778.

[5] A. Vaswani *et al.*, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.

[6] 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, 2014.

[7] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-Net: learning dense volumetric segmentation from sparse annotation," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2016: 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part II 19*, 2016: Springer, pp. 424-432.

[8] J. Schlemper *et al.*, "Attention gated networks: Learning to leverage salient regions in medical images," *Medical image analysis*, vol. 53, pp. 197-207, 2019.
[9] J. P. V. de Mello *et al.*, "Deep Learning-based Type Identification of Volumetric MRI Sequences," in *2020 25th International Conference on Pattern Recognition (ICPR)*, 2021: IEEE, pp. 1-8.
[10] A. Saha, Y.-D. Zhang, and S. C. Satapathy, "Brain tumour segmentation with a muti-pathway ResNet based UNet," *Journal of Grid Computing*, vol. 19, pp. 1-10, 2021.
[11] G. Ramasamy, T. Singh, and X. Yuan, "Multi-Modal Semantic Segmentation Model using Encoder Based Link-Net Architecture for BraTS 2020 Challenge," *Procedia Computer Science*, vol. 218, pp. 732-740, 2023.
[12] Z. Huang, Y. Zhao, Y. Liu, and G. Song, "GCAUNet: A group cross-channel attention residual UNet for slice based brain tumor segmentation," *Biomedical Signal Processing and Control*, vol. 70, p. 102958, 2021.
[13] D. Maji, P. Sigedar, and M. Singh, "Attention Res-UNet with Guided Decoder for semantic segmentation of brain tumors," *Biomedical Signal Processing and Control*, vol. 71, p. 103077, 2022.
[14] S. K. R. Chinnam, V. Sistla, and V. K. K. Kolli, "Multimodal attention-gated cascaded U-Net model for automatic brain tumor detection and segmentation," *Biomedical Signal Processing and Control*, vol. 78, p. 103907, 2022.
[15] J. Zhang, Z. Jiang, J. Dong, Y. Hou, and B. Liu, "Attention gate resU-Net for automatic MRI brain tumor segmentation," *IEEE Access*, vol. 8, pp. 58533-58545, 2020.
[16] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "Cbam: Convolutional block attention module," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 3-19.
[17] I. Mazumdar and J. Mukherjee, "Fully automatic MRI brain tumor segmentation using efficient spatial attention convolutional networks with composite loss," *Neurocomputing*, vol. 500, pp. 243-254, 2022.
[18] M. Chi, H. An, X. Jin, K. Wen, and Z. Nie, "SCAR U-Net: A 3D Spatial-Channel Attention ResU-Net for Brain Tumor Segmentation," in *Proceedings of the 3rd International Symposium on Artificial Intelligence for Medicine Sciences*, 2022, pp. 497-501.
[19] S. Jadon, "A survey of loss functions for semantic segmentation," in *2020 IEEE conference on computational intelligence in bioinformatics and computational biology (CIBCB)*, 2020: IEEE, pp. 1-7.
[20] S. Bakas *et al.*, "Advancing the cancer genome atlas glioma MRI collections with expert segmentation labels and radiomic features," *Scientific data*, vol. 4, no. 1, pp. 1-13, 2017.
[21] S. Bakas *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.
[22] W. Jun, X. Haoxiang, and Z. Wang, "Brain tumor segmentation using dual-path attention U-net in 3D MRI images," 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 I 6*, 2021: Springer, pp. 183-193.
[23] L. Pei, A. Murat, and R. Colen, "Multimodal brain tumor segmentation and survival prediction using a 3D self-ensemble ResUNet," 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 I 6*, 2021: Springer, pp. 367-375.
[24] B. Parmar and M. Parikh, "brain tumor segmentation and survival prediction using patch based modified 3D U-Net," 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*, 2021: Springer, pp. 398-409.
[25] W. Wang, C. Chen, M. Ding, H. Yu, S. Zha, and J. Li, "Transbts: Multimodal brain tumor segmentation using transformer," in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part I 24*, 2021: Springer, pp. 109-119.
[26] W. Zhang *et al.*, "ME-Net: multi-encoder net framework for brain tumor segmentation," *International Journal of Imaging Systems and Technology*, vol. 31, no. 4, pp. 1834-1848, 2021.
[27] S. R. González, I. Zemmoura, and C. Tauber, "3D brain tumor segmentation and survival prediction using ensembles of convolutional neural networks," 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*, 2021: Springer, pp. 241-254.
[28] Y. Jiang, Y. Zhang, X. Lin, J. Dong, T. Cheng, and J. Liang, "SwinBTS: A method for 3D multimodal brain tumor segmentation using swin transformer," *Brain Sciences*, vol. 12, no. 6, p. 797, 2022.
[29] J. Liang, C. Yang, and L. Zeng, "3D PSwinBTS: An efficient transformer-based Unet using 3D parallel shifted windows for brain tumor segmentation," *Digital Signal Processing*, vol. 131, p. 103784, 2022.
[30] H. Cao *et al.*, "Swin-unet: Unet-like pure transformer for medical image segmentation," in *Computer Vision–ECCV 2022 Workshops: Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part III*, 2023: Springer, pp. 205-218.
[31] G. Kaur, P. S. Rana, and V. Arora, "Deep learning and machine learning - based early survival predictions of glioblastoma patients using pre - operative three - dimensional brain magnetic resonance imaging modalities," *International Journal of Imaging Systems and Technology*, vol. 33, no. 1, pp. 340-361, 2023.
[32] W. Ayadi, W. Elhamzi, and M. Atri, "A deep conventional neural network model for glioma tumor segmentation," *International Journal of Imaging Systems and Technology*, 2023.
[33] R. Zhang, S. Jia, M. J. Adamuand, W. Nie, Q. Li, and T. Wu, "HMNet: Hierarchical Multi-Scale Brain Tumor Segmentation Network," *Journal of Clinical Medicine*, vol. 12, no. 2, p. 538, 2023.
[34] A. Rehman, M. Usman, A. Shahid, S. Latif, and J. Qadir, "Selective Deeply Supervised Multi-Scale Attention

Network for Brain Tumor Segmentation," *Sensors*, vol. 23, no. 4, p. 2346, 2023.

[35] L. Zhao, C. Jia, J. Ma, Y. Shao, Z. Liu, and H. Yuan, "Medical image segmentation based on self-supervised hybrid fusion network," *Frontiers in Oncology*, vol. 13, 2023.

[36] Y. Liu, F. Mu, Y. Shi, J. Cheng, C. Li, and X. Chen, "Brain tumor segmentation in multimodal MRI via pixel-level and feature-level image fusion," *Multimodal Brain Image Fusion Methods Eval. Appl*, vol. 16648714, p. 62, 2023.

[37] H. Li, Y. Nan, J. Del Ser, and G. Yang, "Large-kernel attention for 3d medical image segmentation," *Cognitive Computation*, pp. 1-15, 2023.

[38] R. Raza, U. I. Bajwa, Y. Mehmood, M. W. Anwar, and M. H. Jamal, "dResU-Net: 3D deep residual U-Net based brain tumor segmentation from multimodal MRI," *Biomedical Signal Processing and Control*, vol. 79, p. 103861, 2023.

[39] H. Ting and M. Liu, "Multimodal Transformer of Incomplete MRI Data for Brain Tumor Segmentation," *IEEE Journal of Biomedical and Health Informatics*, 2023.

[40] L. Zhao, J. Ma, Y. Shao, C. Jia, J. Zhao, and H. Yuan, "MM-UNet: A multimodality brain tumor segmentation network in MRI images," *Frontiers in oncology*, vol. 12, p. 950706, 2022.

[41] G. Wang, W. Li, S. Ourselin, and T. Vercauteren, "Automatic brain tumor segmentation using convolutional neural networks with test-time augmentation," in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 4th International Workshop, BrainLes 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16, 2018, Revised Selected Papers, Part II 4*, 2019: Springer, pp. 61-72.

[42] S. Chandra *et al.*, "Context aware 3D CNNs for brain tumor segmentation," in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 4th International Workshop, BrainLes 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16, 2018, Revised Selected Papers, Part II 4*, 2019: Springer, pp. 299-310.

[43] K. Hu *et al.*, "Brain tumor segmentation using multi-cascaded convolutional neural networks and conditional random field," *IEEE Access*, vol. 7, pp. 92615-92629, 2019.

[44] T. Zhou and S. Zhu, "Uncertainty quantification and attention-aware fusion guided multi-modal MR brain tumor segmentation," *Computers in Biology and Medicine*, p. 107142, 2023.

[45] A. Paszke *et al.*, "Pytorch: An imperative style, high-performance deep learning library," *Advances in neural information processing systems*, vol. 32, 2019.

[46] T. Zhou, S. Canu, P. Vera, and S. Ruan, "A dual supervision guided attentional network for multimodal mr brain tumor segmentation," in *Proceedings of 2021 International Conference on Medical Imaging and Computer-Aided Diagnosis (MICAD 2021) Medical Imaging and Computer-Aided Diagnosis*, 2022: Springer, pp. 3-11.

[47] A. Di Ieva *et al.*, "Application of deep learning for automatic segmentation of brain tumors on magnetic resonance imaging: a heuristic approach in the clinical scenario," *Neuroradiology*, vol. 63, pp. 1253-1262, 2021.

[48] M. U. Rehman, J. Ryu, I. F. Nizami, and K. T. Chong, "RAAGR2-Net: A brain tumor segmentation network using parallel processing of multiple spatial frames," *Computers in Biology and Medicine*, vol. 152, p. 106426, 2023.

[49] P. Li, W. Wu, L. Liu, F. M. Serry, J. Wang, and H. Han, "Automatic brain tumor segmentation from Multiparametric MRI based on cascaded 3D U-Net and 3D U-Net++," *Biomedical Signal Processing and Control*, vol. 78, p. 103979, 2022.

[50] H. Liu, G. Huo, Q. Li, X. Guan, and M.-L. Tseng, "Multiscale lightweight 3D segmentation algorithm with attention mechanism: Brain tumor image segmentation," *Expert Systems with Applications*, vol. 214, p. 119166, 2023.