Journal of Innovative Image Processing (ISSN: 2582-4252)
www.irojournals.com/iroiip/

ResAtt-NASFPN: A Residual Attention Driven NAS-FPN Framework for Robust 3D Brain Tumor Segmentation

Ronak R Patel[1], Miral Patel[2], Deep Kothadiya[3]

[1,3]U and P U. Patel Dept. Of Computer Engineering, Chandubhai S. Patel Information and Technology (CSPIT) Faculty of Technology and Engineering (FTE), Charotar University of Science and Technology (CHARUSAT), Changa, Anand, Gujarat, India.
[2]G H Patel College of Engineering and Technology, CVM University, V V Nagar- Anand, Gujarat, India.

E-mail: [1]ronakrpatel.ce@charusat.ac.in, [2]2025miru@gmail.com, [3]deepkhothadiya.ce@charusat.ac.in

Abstract

Multi-parametric MRI scans are used to accurately identify the morphological variability of glioblastoma sub-regions of the brain tumor. For this purpose, T1, T1CE, T2, and FLAIR modalities are used to identify essential information of about ET, TC and WT for treatment planning. Small ET regions struggle to be identified by traditional architecture because the area occupied by ET is less than 2% of brain volume, and even spatial resolution adoption is one of the challenges. The paper proposes an Enhanced Neural Architecture Search-based Feature Pyramid Network UNet3D (NAS-FPN-UNet3D) that incorporates autonomous operation selection inside multi-scale feature pyramids. In NAS-FPN-UNet3D, adaptive feature reconfiguration is managed by Convolutional Block Attention Module (CBAM). To maintain image quality and spatial accuracy, gradient stability is managed through residual blocks. NAS is one of the important components for effectively refining long-range information through five parallel operations: Standard Convolution, Dilated Convolution, Depth-wise Separable Convolution (with 94.7% parameter reduction), Asymmetric Convolution and identify connection. Operations are automatically selected based on the learnable parameters; for contextual modelling, dilated convolutions are preferred at coarse levels, while for efficiency at finer levels, depth-wise operations are employed. When evaluated on the BraTS2020 dataset, the proposed model uses parameters and yields Dice scores of 0.8406 ± 0.05 (ET), 0.868 ± 0.06 (TC), and 0.9157 ± 0.03 (WT) with corresponding Hausdorff distances of 4.57mm ± 0.05 mm, 6.75 mm ± 0.04 mm, and 6.85 mm ± 0.02mm, respectively. The proposed model requires 34% fewer parameter than fixed-convolution baselines while achieving greater accuracy.

Keywords: Residual Attention, Neural Architecture Search, Feature Pyramid Network, CBAM, Brain Tumor.

1. Introduction

The National Cancer Centre recently acknowledged that there were 109,000 new brain tumor cases identified and 59,000 deaths in 2016, which needs consideration [1]. Abnormal growth of the tissues in the brain causes major health issues. The WHO has classified tumors

DOI: https://doi.org/10.36548/jiip.2026.1.003
Received: 24.11.2025, received in revised form: 22.12.2025, accepted: 06.01.2026, published: 17.01.2026

into high-grade and low-grade categories. Early detection is helpful for the prognosis of tumor types such as gliomas, meningiomas, and pituitary tumors. Gliomas, in particular, are the most prevalent primary brain tumors in adult patients, and because of their extensive and challenging treatment, they are a major area of scientific interest [2]. Gliomas are diagnosed by magnetic resonance imaging (MRI), and accurate tumor segmentation is essential for successful treatment planning. MRI images exhibit high complexity, including variations in size and shape, low contrast, and blurred boundaries [3]. Manual diagnoses of the MRI are difficult and time-consuming, so efficient and accurate segmentation methods are needed to improve accuracy and treatment planning [4].

Deep learning techniques have made significant progress in medical segmentation over the last couple of years. Ronneberger et al. introduced the U-Net architecture for medical image segmentation [5]. Encoder-decoder-based structures with the skip connection approach are state-of-the-art methods for better region identification. A variety of innovative hybrid segmentation methods based on transformers, U-Net, and convolutional neural networks (CNNs) have recently been implemented and validated on a variety of datasets [5]. Even more challenging are limited datasets with different modalities that face the issue of accuracy. To address this, numerous attention-based models have been introduced, which help deal with small regions but still face issues in the identification of long-range regions. MRI scans contain heterogeneous information, and processing capabilities need to be more advanced so significant progress can be made through attention mechanisms, multi-scale feature extraction, and hybrid modelling.

The proposed approach addresses key challenges in brain tumor segmentation, such as the complex and heterogeneous regions of the enhanced tumor (ET) and tumor core (TC). To effectively handle the intricate tumor structure, a Residual Convolution [6] Block is employed to improve gradient flow and preserve critical features. The CBAM [7] module enables the network to dynamically recalibrate feature maps, allowing precise recognition of small regions and delicate tumor boundaries. To suppress irrelevant background noise, an Attention Gate [8] is incorporated. For adaptive learning in the decoder stage and to ensure robustness across diverse images, a NAS [9] block is utilized. Finally, multi-scale feature fusion through the FPN [10] enhances the model's ability to accurately segment tumors. By addressing the fundamental challenges of heterogeneity, small-target detection, and variability in imaging modalities, this approach supports accurate tumor delineation—which is crucial for clinical diagnosis and treatment planning.

Key Contribution of this Research

1. Proposed a 3D brain tumor segmentation model that combines CBAM attention with residual convolutional blocks to improve localization and feature extraction

2. Improved resilience across tumor sizes and imaging modalities by adaptively fusing multi-scale volumetric information using a NAS-driven feature pyramid network.

3. Increased segmentation accuracy by emphasizing pertinent features, lowering noise, and speeding up training convergence through the use of an attention-gated decoder and deep supervision.

2. Related Work

2.1 U-Net & Its Variants

Deep learning techniques, especially encoder-decoder architectures like U-Net and its variations, are now the standard for brain tumor segmentation because they can understand complex spatial information from multimodal MRI data. For instance, 3D U-Net extensions use volumetric data effectively to accurately define tumor sub-regions, and they perform better on common datasets like BraTS [11].

With Dice scores of 88.95% for WT, 85.06% for TC, and 82.03% for ET, Isensee et al. [11]'s nnU-Net won first place in the BraTS 2020[12] competition. This model stands out for its baseline setup, which produces competitive performance without the need for task-specific modifications like region-based training or post-processing. However, nnU-Net's application to a wider range of datasets may be limited due to its lack of specific adaptations for various tumor types and imaging settings.

In order to address voxel imbalance, Zhang et al. [13] suggested a multi-encoder model for 3D MRI segmentation that included a unique categorical Dice loss function. With Dice scores of 70.24% for WT, 88.26% for TC, and 73.86% for ET, their approach produced noteworthy outcomes. The model's performance on entire tumor segmentation is still lower than that of other approaches, despite the encouraging results, suggesting that segmentation accuracy might still be improved.

2.2 Feature Pyramid Networks (FPN) and Variants

In object detection and semantic segmentation tasks, FPN [10] plays an important role. The main challenge of extracting information from the multiscale scans is addressed by FPN, which constructs a hierarchical pyramid approach. FPN is effective when contextual information is required across different scales. It utilizes both the top-down and bottom-up pathways for effective feature fusion at different levels of abstraction.

In the bottom-up approach, FPN typically makes use of pre-trained networks. Using a ResNet backbone to extract hierarchical features from input images, Lin et al. [14] employed FPN for object detection. A conventional ResNet architecture that has been pre-trained on the ImageNet dataset is the first step in the FPN's bottom-up process [14]. A hierarchy of feature maps at different sizes is produced by ResNet. After that, these feature maps are processed using the top-down method, which upsamples geographically coarser but semantically stronger feature maps from the higher layers of the pyramid to produce higher-resolution features.

BiFPN [15] improves feature fusion by introducing a bidirectional flow, in which data moves from top to bottom as well as from bottom to top at each layer, resulting in a more interconnected structure. BiFPN can effectively capture multiscale information without adding to the computational cost of the original FPN thanks to its bidirectional method and a straightforward add-and-multiply fusion process.

2.3 Attention Mechanisms: Attention GATES & CBAM

By adding bottleneck units and a shuffle attention mechanism to the U-Net architecture, Magadza et al. [16] improved the model's performance on BraTS 2020, achieving Dice scores

of 91.2% for WT, 84.8% for TC, and 79.2% for ET. Even with these improvements, the technique's ability to increase tumor detection is still subpar, indicating that attention processes still require enhancement in order to properly handle this problem.

Hengxin Liu et al. [17] introduced ADHDC-Net; which utilizes expansive dilated convolutions, hierarchical decoupled convolutions, and an integrated attention mechanism. By carefully fine-tuning the network's focus on certain tumor locations and their connections, the embedded attention mechanism increases segmentation fidelity. The approach was evaluated on BraTS 2020 with a Dice score of 77.91% for ET, 89.94% for WT, and 83.89% for TC.

Zhang et al.'s [18] introduced a self-attention-based U-Net architecture for BraTS 2020. This module enables selective focus on crucial tumor regions and healthy brain structure. This approach strengthens the identification of hierarchical features on both local and global scales. In terms of performance, it achieves a Dice score of 0.86 for WT.

2.4 Fusion Convolutional Blocks & Hybrid Architecture

Daimary et al.[19] proposed a hybrid CNN architecture with fusion convolution. The key component is multiple convolutional blocks, which extract both local and global features, likely assembled using a cascaded CNN. A feature selection technique helps to identify region-based details. The overall results are 0.812, 0.871, and 0.842 for ET, WT, and TC, respectively.

Nizamani et al [20] proposed an approach that emphasizes feature fusions. The core component of the model is integrating the feature extraction module and fusion methods. The hybrid model combines information from different modalities. This approach achieves a Dice score similarity of 0.89 for WT.

Munir et al. [21] proposed a U-Net based architecture with multiple down-sampling blocks followed by convolution blocks, batch normalization and ReLU activation. The bottleneck layer applies an additional convolutional block to maximize feature extraction. The approach used Grad-CAM [22] for better identification. The evaluation was done on the BraTS2020 Dataset [12].

3. Methodology

The proposed architecture enhances the encoder-decoder model. These enhancements significantly target local features and long-range boundary regions for accurately identifying the size and location of the affected areas.

3.1 ResAtt-NASFPN Architecture

Figure 1 represents the proposed architecture, which is evaluated on the BraTS2020[13] dataset, having four different modalities for training. The proposed architecture enhances U-Net to make it more effective in finding spatial information; a residual convolution 3D block is incorporated. MRI images have low intensity, and the reduction of noise is important, as well as the identification of long-range regions identified by the CBAM attention block on the encoder side.

The deconvolution block helps identify spatial information on the decoder side through upsampling. The attention gate concatenates the features identified by the CBAM and targeted

by the deconvolution block. The proposed model still struggles with small regions and boundaries of ET, TC and WT, which will be addressed by the FPN-NAS block. Fusion convolution blocks are used for final upsampling.

The systematic channel configuration 16→32→64→96→160 follows on both sides. Encoders enforce the reduction of spatial resolution by a factor of two at every level through max-pooling, while feature channels are increased. Trilinear upsampling is used to restore spatial resolution on the decoder side based on the reverse order of channel dimensions. The final prediction size B x C x D x H x W generated using the 1x1x1 convolution. The flow of execution of the proposed approach is shown in Algorithm 1.

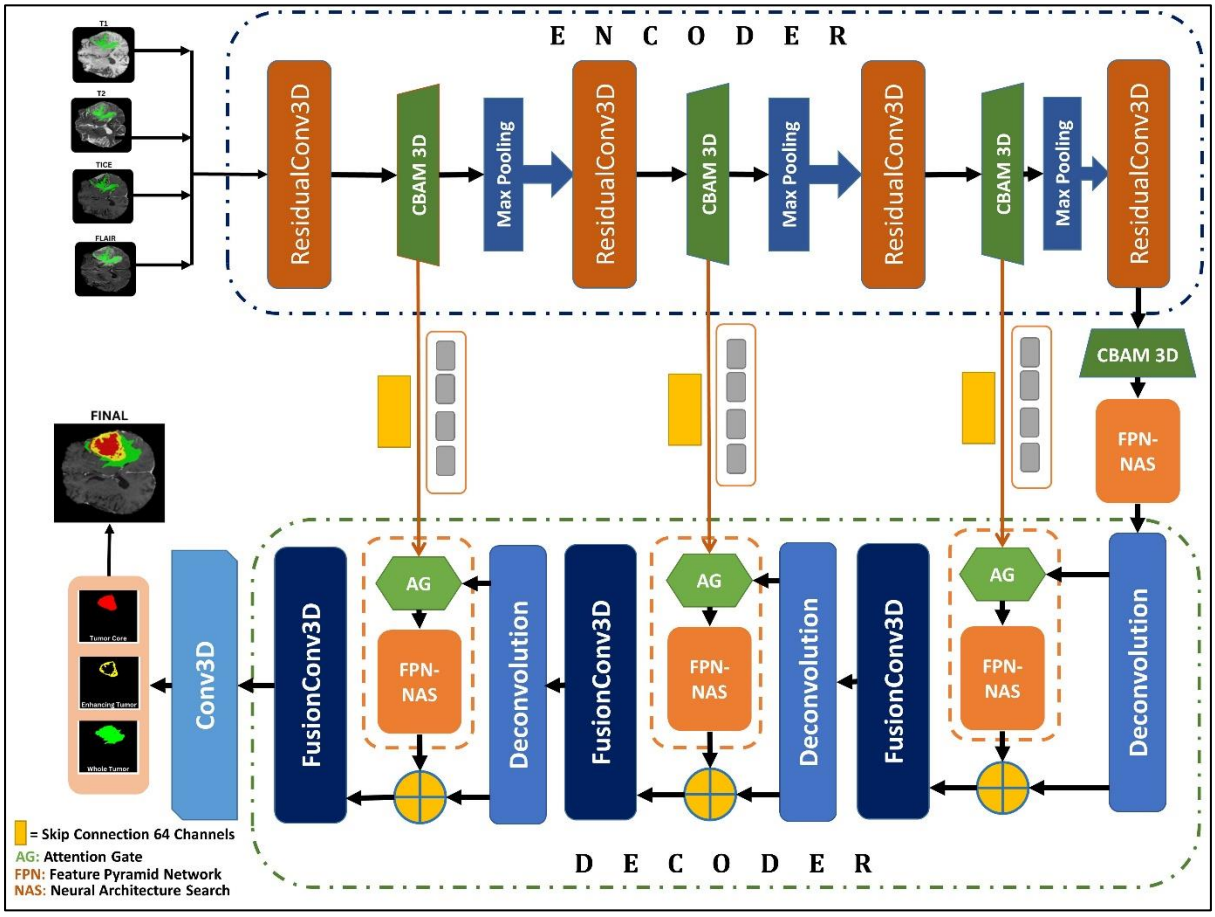

Figure 1. Proposed System Architecture

Algorithm 1 Brain Tumor Segmentation Training Pipeline

Require: 3D MRI volume *X*, Ground-truth *Y*, batch size, epochs
Ensure: Segmentation output for WT, TC, ET

procedure Train_Model (*X* ,*Y* ,*epochs* ,*batch_size)*
  Initialize model weight θ,optimizer Adam (lr = 0.001)
  for $e = 1$ to *epochs* **do**
    for each batch $(x_b, y_b) \in DataLoader(X, Y, batch\_size)$ **do**
      $x_b \Leftarrow Augment(x_b) + Normalize(x_b)$
      $\hat{Y} \Leftarrow Forward\_Pass(x_b, \theta)$
      $Loss \Leftarrow DiceLoss(\hat{Y}, y_b) + FocalLoss(\hat{Y}, y_b)$

8. $\theta \Leftarrow \theta - Adam(\nabla Loss)$
9. **end for**
10. Validate model and save best checkpoint
11. **end for**
12. return θ
13. end procedure
14. procedure $Forward\_Pass(x, \theta)$
Encoder : 4 levels with ResConv + CBAM
15. for $i = 1$ *to* 4 **do**
16. $c_i \Leftarrow ResConv(x, channels_i) + CBAM(c_i)$
17. $x \Leftarrow MaxPool(c_i)$
18. end for
Decoder : FPN with NAS + Attention Gate
19. $f_4 \Leftarrow NAS\_Conv(c_4)$
20. for $\boldsymbol{i = 3}$ down to 1 **do**
21. $u_i \Leftarrow Upsample(f_{i+1})$
22. $c_i' \Leftarrow Attention(u_i, c_i)$
23. $f_i \Leftarrow NAS\_Conv(c_i') + Fusion(u_i)$
24. end for
25. $\hat{Y} \Leftarrow Conv3D(f_1, out = 3)$
26. return $\hat{Y}$
27. end procedure
Component Definitions
28. $ResConv$: $Conv3D \rightarrow BN \rightarrow ReLU \rightarrow Conv3D \rightarrow BN \rightarrow Skip\ Connection \rightarrow ReLU$
29. CBAM : Channel Attention + Spatial Attention
30. NAS_Conv: $Softmax(\alpha) \times [StdConv, DilatedConv, DepthwiseConv, AnisoConv]$
31. $AttentionGate$: $c \times \sigma(\varphi\left(ReLU\left(w_g(u) + w_x(c)\right)\right))$
32. $DiceLoss$: $1 - \frac{2|\hat{Y} - Y|}{|\hat{Y}| + |Y|}$
33. $FocalLoss$: $-\alpha(1-p)^r \log(p)\,, \gamma = 2$

3.2 Residual Block + CBAM

The residual convolution block on the encoder side improves gradient flow and accurately retains fine-grained features. The internal structure of the residual block maintains continuous 3-dimensional convolution layers through normalization and ReLU. Stability of the convolution blocks is achieved through the skip connection, and providing depth. The transformation is evaluated based on equation 1 , which has input $X \in R^{C \times D \times W \times H}$.

$$y = ReLU(F(x, \theta) + T(x)) \tag{1}$$

Where $F(.)$ denotes the stacked convolution transformation, and $T(.)$ is linear transformation aligned with channel dimensions for the skip connections. Global context of whole tumor, effectively identify by this block.

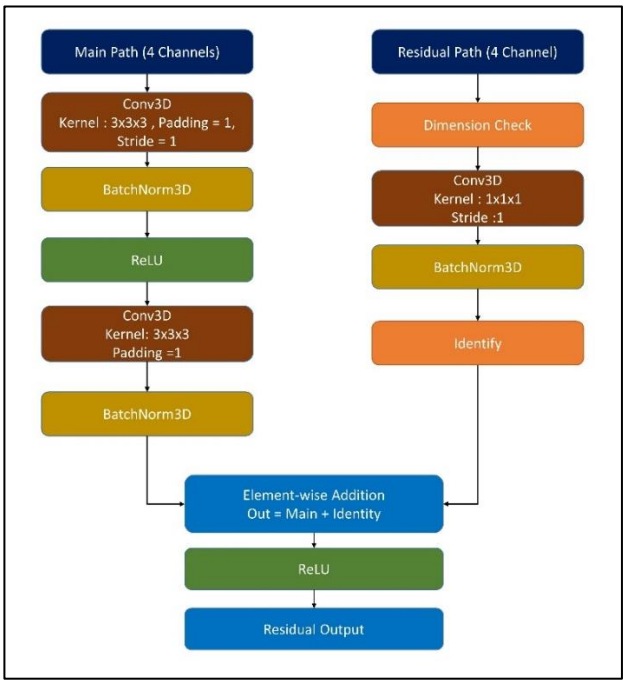

Figure 2. Component stack of Residual Block

A combination of channel and spatial attention helps to target informative regions through CBAM. After average and max-pooling, a channel-wise attention block is employed for better identification. Equation 2 represents the calculation of the feature tensor X of shape (B, C, D, H, W).

$$M_c = \sigma(FC_2\left(ReLU\left(FC_1(Z)\right)\right)) \tag{2}$$

Where Z denotes pooled features and denotes sigmoid activation applied channel-wise.

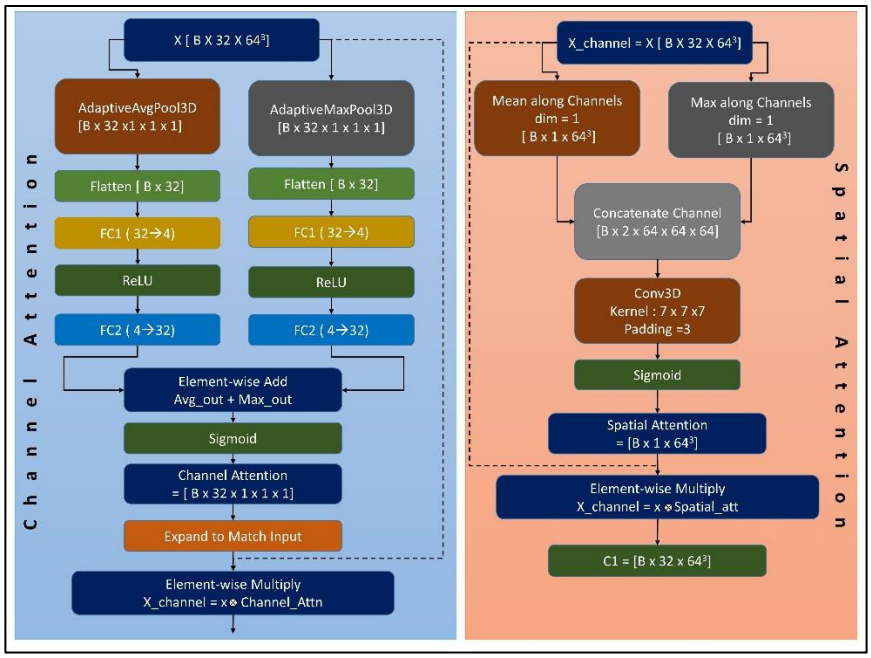

Figure 3. Internal Structure of Channel Attention and Block Attention

Spatial-attention highlights informative part, in the proposed model, it deals with the boundary delineation in enhanced tumor and tumor core sub-regions based on equation 3:

$$M_c = f([AvgPool(X); MaxPool(X)]) \quad (3)$$

3.3 Attention Gate on Decoder Side

Attention gates are incorporated to refine the fusion of the encoder and decoder features as skip connections. The spatial feature, which is identified by the CBAM attention block from the encoder, is considered as x_1. After every deconvolution block, the feature that is identified on the decoder side is considered as g. Attention block calculates the coefficient based on equation 4 and both provided features are denoted as α.

$$\alpha = \sigma(W^T \left(ReLU\left(W_g * g + W_x * x\right)\right)) \quad (4)$$

This attention block helps to filter irrelevant background information and more effectively target the small regions such as ET and TC where regions are discontinued.

3.4 FPN + NAS Block

To construct a hierarchical multi-scale representation, FPN (Feature Pyramid Network) is used on the decoder side, establishing a connection between the encoder features and progressively mapping from the decoded features at four levels $l \in \{1,2,3,4\}$. All pyramid levels are projected to a fixed channel dimension of 256 to ensure consistent feature alignment across scales. This design simplifies multi-scale feature fusion, stabilizes optimization, and provides a balanced trade-off between representational capacity and computational efficiency. At each level encoder features are represented with the channel dimension $C_1 = R^{B \times C_1 \times D_1 \times H_1 \times W_1}$, the channel dimension which is considered, $C_1 = \{32,64,128,256\}$ at spatial resolution they are considered as $\{64^3, 32^3, 16^3, 8^3\}$ for the refinement before lateral processing. For the prior feature transformation, an attention gate is used for encoder activation and decoder context gating signals based on equation 5:

$$C_l^{gate} = C_l \odot \sigma(\varphi \left(ReLU\left(W_g^{(l)} \times g_{l+1} + W_x^{(l)} \times C_l\right)\right)) \quad (5)$$

Where, $W_g^{(l)}, W_x^{(l)} \in R^{(F_{int} \times C_l \times 1^3)}$ represents a learnable $1 \times 1 \times 1$ convolutional kernel, $g_{(l+1)}$ represents the up-sampled features from the previous level, φ is a $1 \times 1 \times 1$ convolution reducing F_{int} to a single attention map, and σ is a sigmoid activation function. The gated features undergo a three-stage transformation represent as equation 6:

$$P_l = Conv_{1\times1}(NAS\left(Conv_{1\times1}\left(C_l^{gated}, 256\right); \alpha_l\right), 256) \quad (6)$$

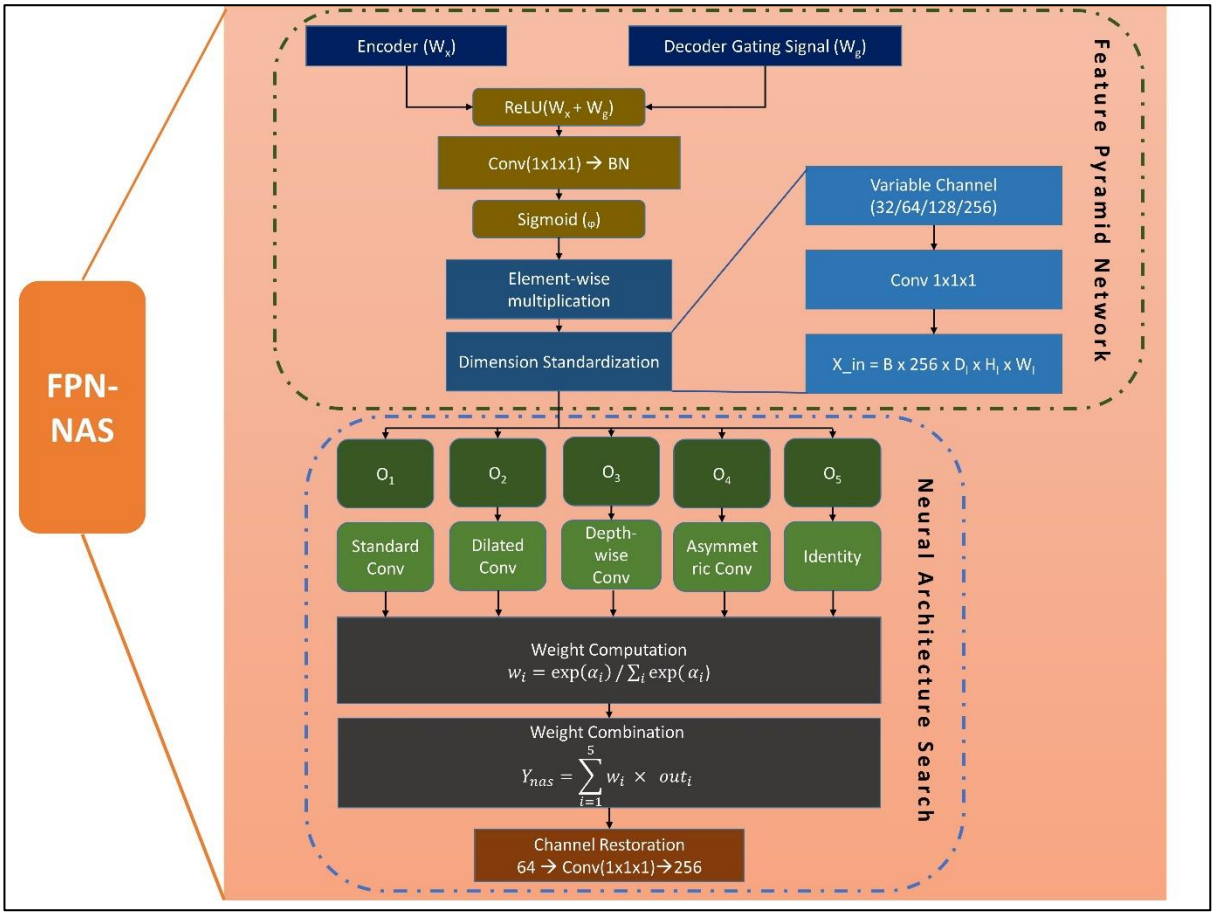

Figure 4. Feature Pyramid Network Integrate with Neural Architecture Search for High Resolution Regions

where all pyramid levels are standardized to 256 channels before and after NAS processing. Multi-scale fusion combines lateral features with up-sampled decoder features:

$$P_l^{fused} = Conv_{1\times1}([P_l || Upsample_2 \times (P_{l+1}^{fused})]) \quad (7)$$

Equation 7 propagates semantic information from coarse to fine resolutions for accurate tumor boundary delineation. This gating process reduces the impact of spatially unrelated features and highlights areas that match the decoder's predictions. This is essential for identifying small enhancing tumor (ET) regions among noisy high-resolution backgrounds.

The NAS block is explicitly designed to address morphological variability through a heterogeneous search space consisting of five parallel operations with learnable mixing weights: standard convolution, dilated convolution (for expanded receptive fields), depth-wise convolution (for computational efficiency), and asymmetric convolution (for anisotropic features). Identity operation dominance is prevented in fine-level NAS blocks through channel dimension mismatch. Because all lateral connections perform dimensional reduction (256→64 channels), the identity operation is architecturally incompatible and automatically excluded from the search space. During execution, depth-wise convolutions were selected most frequently, followed by dilated and asymmetric convolutions, indicating their relative importance in feature extraction and contextual modelling.

$$O_1(X) = Conv3D(X; K = 3^3, P = 1) \quad (8)$$

$$O_2(X) = Conv3D(X; K = 3^3, D = 2) \quad (9)$$

$$O_3(X) = Conv3D(DWConv3D(X)) \quad (10)$$

$$O_4(X) = Conv3D(X; K = (3,1,1)) \quad (11)$$

$$O_5(X) = X \quad (12)$$

Architecture parameters $a_l = [a_{(l,1)}, \dots\dots\dots\dots, a_{(l,5)}]$ are transformed via softmax. The softmax temperature τ was fixed to 1.0, corresponding to the standard softmax formulation. This choice ensures stable and unbiased weighting of operations, allowing the NAS framework to balance exploration and exploitation without enforcing premature hard selection. The NAS block output is computed as a weighted superposition of all candidate operations:

$$Y_l = \sum_{i=1}^{5} w_{(l,i)} \times O_i(X_l) \quad (13)$$

Here $X_{(l)} \in R^{(B\times 256\times D_l\times H_l\times W_l)}$ represents the input feature tensor and $Y_{(l)} \in R^{(B\times 64\times D_l\times H_l\times W_l)}$ denotes the aggregate output. Each pyramid level independently learns optimal operation combinations tailored to its spatial resolution and feature abstraction requirements, with finer levels typically converging to computationally efficient operations (depth-wise separable, asymmetric) for high-resolution boundary refinement, while coarser levels favor operations with larger receptive fields (dilated convolutions) for capturing global tumor context.

3.5 Fusion Convolution Block

The Fusion Conv Block is used to concatenate feature maps and apply 3D convolution to combine and compress the information into a uniform 3-channel format. This process is essential after concatenation, as it allows the model to learn optimal combinations of spatial and semantic information from different network stages, enabling richer feature representation for subsequent decoding and segmentation steps. Since the architecture is deterministic and does not involve stochastic operation selection, learned operations are inherently stable with respect to random seed variations.

4. Simulation Setup and Evaluation Parameters

4.1 Implementation and Training Details

MONAI 0.6 and PyTorch 2.4.0 were used to develop and assess our suggested architecture on NVIDIA Tesla T4 x 2 GPUs with 32 GB of available RAM. This allocation facilitates the pre-processing of the high-resolution 3D MRI scans. Since the learning rate has a direct impact on both model training and validation loss, it is set to a mean of 0.01 from the interval [0.0001, 0.01]. The batch size is 1, and additional hyperparameters, such as weight decay (0.0001), are employed to address overfitting scenarios. 50 epochs are used for processing to train the architecture. To make the model further, the AdamW [23] optimizer is utilized.

Table 1. Summary of Training Hyperparameters

Parameter	Value
Framework	MONAI 0.6 / PyTorch 2.4.0
GPU	NVIDIA Tesla V100, P100
Batch Size	1
Epochs	50
Optimizer	Adam
Learning Rate	0.01
Weight Decay	1e-5
Loss Function	Dice Loss
Precision Mode	FP16

To train more quickly and use less GPU memory, evaluation is done on a 3D MRI scan and FP16 precision mode is employed. Additionally, the results from every epoch are saved for the validation performance; no early stopping criteria are applied.

4.2 Evaluation Matrics

We employed the HD95 (95th Percentile Hausdorff Distance) [24] and the DSC (Dice Similarity Coefficient) [25] to evaluate the proposed architecture. These two evaluation parameters are primarily used to assess the performance of medical image segmentation models. Together, these metrics offer a thorough evaluation of boundary precision and volumetric overlap.

1. Dice Similarity Coefficient (DSC)

The similarity between the generated segmented output and the ground truth is used to calculate the dice score [25]. In class-imbalanced situations, like tumor segmentation, where target structures (like enhancing tumors) might only take up a small percentage of the scan, it is especially reliable. Equation 14 is used to calculate it:

$$Dice\ Score = \frac{2TP}{2TP+FP+FN} \tag{14}$$

where the numbers for true positives, false positives, and false negatives are denoted by TP, FP and FN, respectively. A higher DSC indicates better segmentation quality, especially in terms of volumetric agreement.

2. Hausdorff Distance (95^{th} Percentile – HD95)

The Hausdorff distance captures the largest boundary error between the expected and actual segmentation in extreme cases where even small boundary changes may have clinical

significance [24]. We employ the 95th percentile formulation to lessen our sensitivity to outliers based on equation 15:

$$HD_{95}(A,B) = \max\{percentile_{95}(\min_{b\in B}||a-b||), percentile_{95}(\min_{a\in A}||b-a||)\} \quad (15)$$

Where the sets of boundary points from the ground truth and anticipated segmentations are represented by the letters A and B, respectively. Lower HD95 values indicate better boundary alignment.

4.3 Dataset Description

The BraTS 2020 dataset, which was made publicly available as part of the MICCAI 2020 Brain Tumor Segmentation Challenge [26], is used in this study. To facilitate the development and evaluation of automated brain tumor segmentation algorithms, the dataset was carefully selected using 3D multimodal MRI data.

- **Whole Tumor (WT):** Consisting of all observable abnormalities associated with the tumor.
- **Tumor Core (TC):** This includes necrotic and enhancing areas but excludes peritumoral edema.
- **Enhancing Tumor (ET):** Areas that exhibit hyper-intensity in contrast-enhanced T1-weighted scans, which usually denote active tumor tissue.

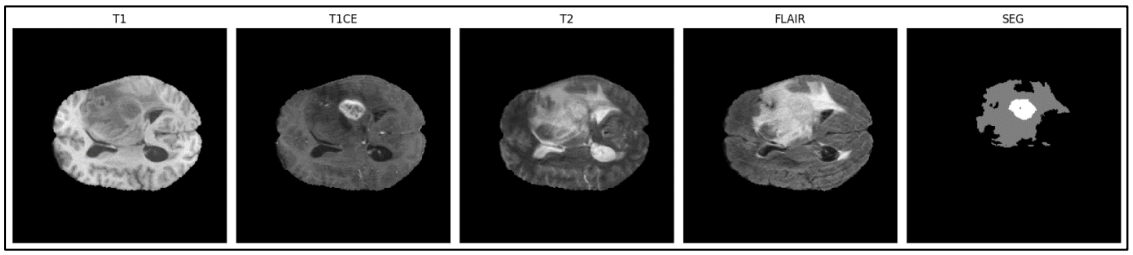

Figure 5. An Overview of BraTS2020 Dataset used for the Proposed Approach

The dataset's diversity and practical relevance are enhanced by the inclusion of 369 MRI scans collected from various institutions using different acquisition protocols. To guarantee a consistent experimental setup throughout research studies, the data is divided into 280 training, 50 validation, and 39 testing volumes. Four co-registered 3D MRI modalities are included in each subject scan:

- **T1-weighted (T1):** Helpful for identifying the general anatomy of the brain and tissue borders.
- **T1-contrast enhanced (T1CE):** A sequence to detect tumor cells that are active or infiltrating.
- **T2-weighted (T2):** Helpful for identifying tumor-associated edema and distinguishing tumor tissue from surrounding CSF.
- **Fluid-Attenuated Inversion Recovery (FLAIR):** Emphasizes peritumoral edema.
- **Ground Truth(SEG):** A neurologist identifies the tumor.

Skull-stripped, pre-aligned, and resampled to a consistent pixel density of $240 \times 240 \times 155$ voxels, all MRI volumes are supplied in NIfTI (.nii.gz) format. The raw data characteristics are maintained by not normalizing the voxel intensities across patients. High-quality annotations were ensured by manually delineating ground truth segmentations by skilled radiologists and confirming them by consensus. Tumor segmentation algorithms can be tested on this multimodal, high-resolution dataset, which offers a variety of appearance patterns, minute intensity variations, and intricate tumor boundaries. Figure 5 illustrates modality-specific contrasts for the same anatomical region in a representative view of the four modalities.

5. Analysis of Results

The ResAtt-NASFPN evaluates the segmentation performance based on multi-regional tumors from the MRI of BraTS2020. Validation is also done through BraTS2020 sample data.

5.1 Insight Results through Visualization

Progressive segmentation of the three classes is shown in Figure 6. The proposed architecture perfectly detects the whole tumor (WT, green) compared to the ground truth geographical context. Regarding the Enhanced tumor (ET, red) and Tumor core (TC, yellow), the proposed architecture is still struggling with boundary mismatches due to scattered affected tissues.

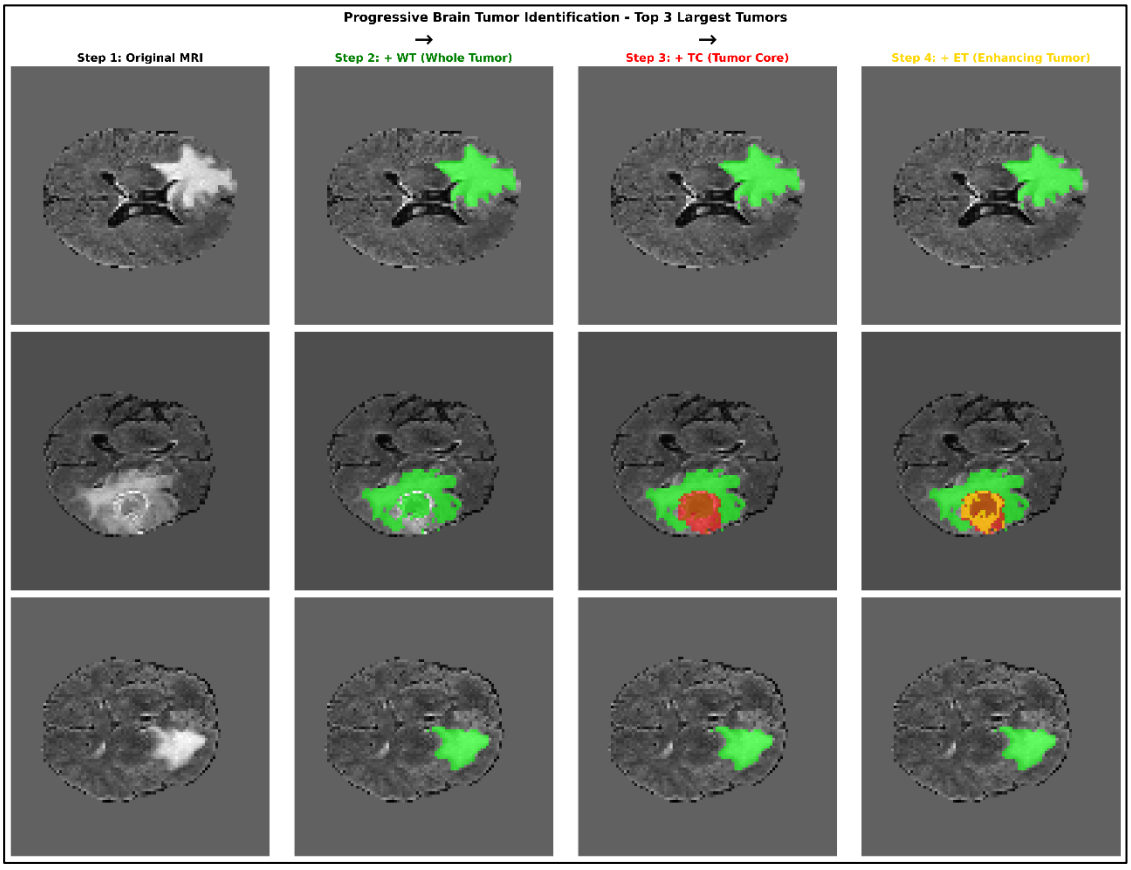

Figure 6. Insight of Progressive Identification Done by the Proposed Approach from WT through TC and ET

5.2 Quantitative Analysis

Table 2 represents the quantitative analysis of the performance, which shows the proposed architecture performs accurately compared to well-known segmentation architectures. The proposed architecture calculates dice score similarity with a combination of Dice loss and Focal loss, achieving values of 0.8406,0.9157 and 0.868 for ET, WT and TC, respectively. Theses dice scores are significantly higher than existing approaches like nnUNET [11], HNF-NetV2[27] and HM-UNET[28], which shows that dealing with the boundary error and less segmentation overlap is taken care of by the proposed approach.

Table 2. Comparison of Dice Scores for Tumor Segmentation

Model	**Dice Score**		
	ET (Enhancing Tumor)	TC (Tumor Core)	WT (Whole Tumor)
ResAtt-NASFPN (Proposed)	0.8406	0.868	0.9157
nnUNET[11]	0.798	0.857	0.911
HNF-Netv2[26]	0.787	0.855	0.913
MMTSN[28]	0.764	0.801	0.882
EEN-UNET[29]	0.654	0.680	0.841
HM-UNET[27]	0.886	0.736	0.782
DMF-Net[25]	0.748	0.748	0.871
Deep-UANET[22]	0.673	0.704	0.861

The result focuses on capturing complex tumor boundaries by removing segmentation outliers across all the categories using the proposed approach. Figure 7 shows that some of the existing approaches perform with a high Hausdorff95 of around 47.62 mm for TC and 26.57 mm for ET. In comparison, the proposed approach evaluates accurately and balances HD95 through overlap and border matrices. Results of the proposed approach achieve 4.57 mm/6.75 mm /6.85 mm for ET/WT/TC, respectively.

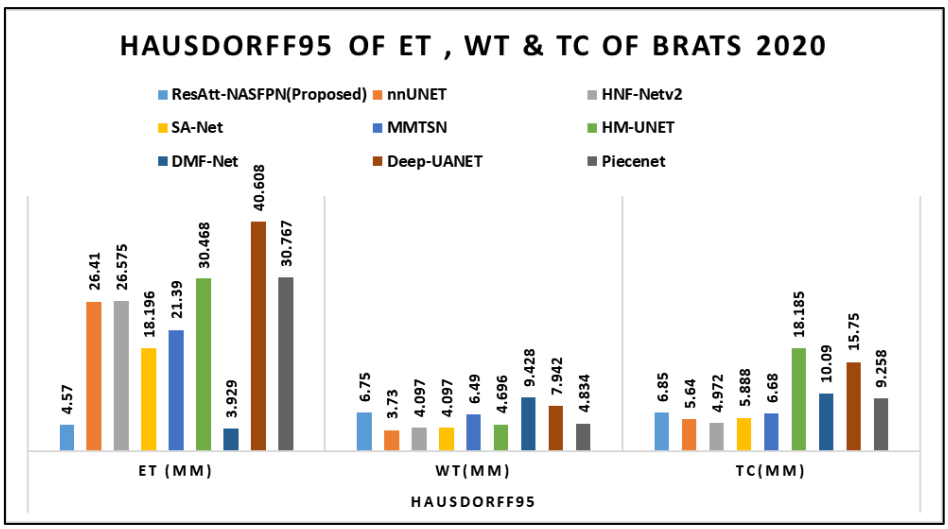

Figure 7. Hausdroff95 Comparison to Determine Pixel-Wise Similarity with the Most Advanced Architecture in Terms of ET, WT, and TC. SLICE – Wise Analysis

From the BraTS2020 dataset, the largest brain tumors are progressively shown in Figure 8. It represents the slice-by-slice visualization from top to bottom using different MRI depths. The results shown represent the sub-regions- the whole tumor, tumor core and enhancing tumor

– using green, yellow and red colors respectively. This demonstrates the consistent and anatomical meaning of segmentation as tumor spatial axial slice. In an MRI scan, the identification of morphological information and sub- areas is very complex; therefore, anatomical distribution, robust generalization, and precise volumetric delineation are required. At each stage, the slice depth and pixel-level quantification are examined accurately. After analysis of slices, the proposed approach achieves higher accuracy in central slices. Observations indicate that the accuracy of ET is 0.35 at the initial level; on the center slice it achieves an accuracy of 0.89 for the same, and at the final slice it again goes around 0.45. These variations are clearly noticeable in the figure.

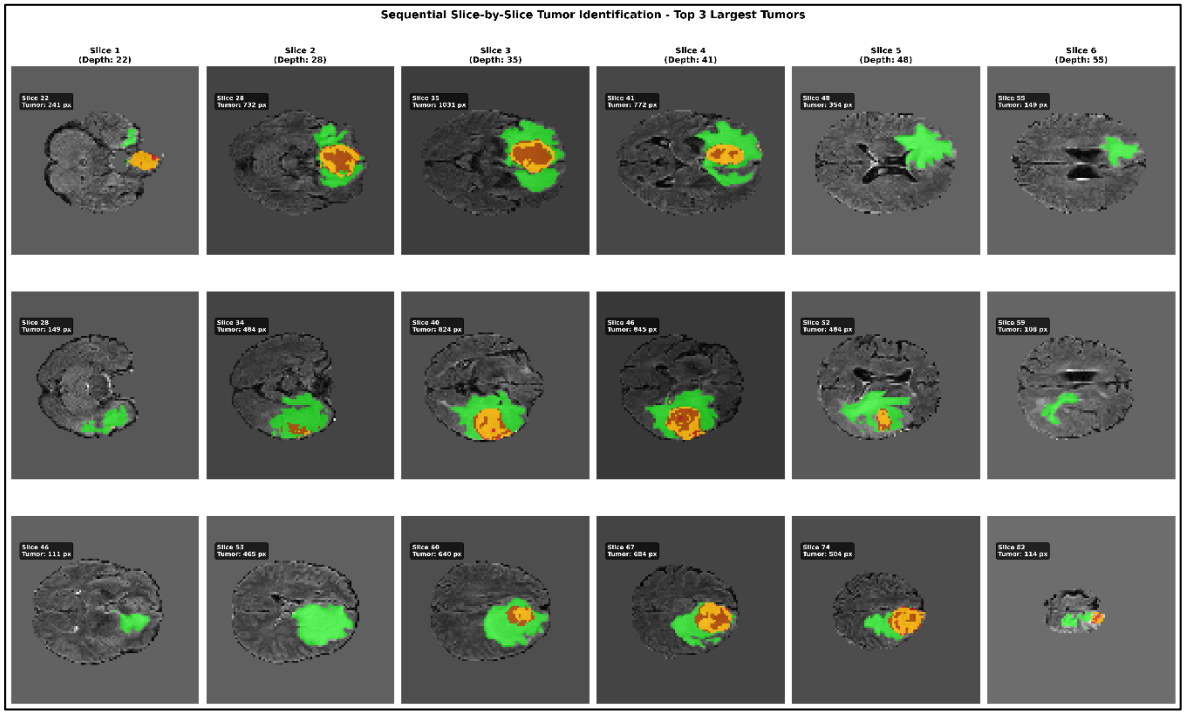

Figure 8. Sequential Slice-Wise Representation of Identifying Different Levels of Tumor by the Proposed Approach.

6. Component based Analysis

With the BraTS2020 dataset, the results demonstrate how gradual architectural changes with different modifications and loss functions impact segmentation performance, as shown in figures 9 and 10. The UNET3D baseline obtained Dice scores of 0.7705 for ET, 0.8204 for WT, and 0.7801 for TC, while its Hausdorff95 distances were 10.11 mm for ET, 16.15 mm for WT, and 15.17 mm for TC. Upon the addition of CBAM blocks, the model's Dice scores improved to 0.7945 for ET, 0.8512 for WT, and 0.8142 for TC, while the Hausdorff95 metrics decreased to 8.02 mm for ET, 12.54 mm for WT, and 11.24 mm for TC, demonstrating feature attention and improved boundary accuracy.

After applying an attention gate on the decoder side ResAtt-NASFPN achieved a precision of 0.814, which indicates suppression of false positive predictions. Additionally, a specificity of 0.9984 represents that the decoder side attention gate successfully minimizes the incorrect background when identifying the classes.

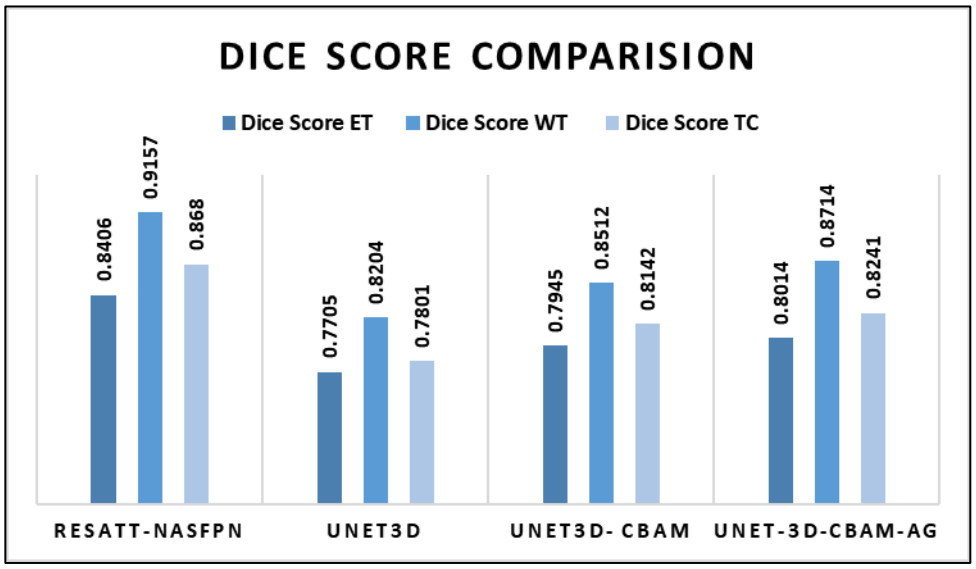

Figure 9. Result of Dice Score Similarity Comparison with Different Components

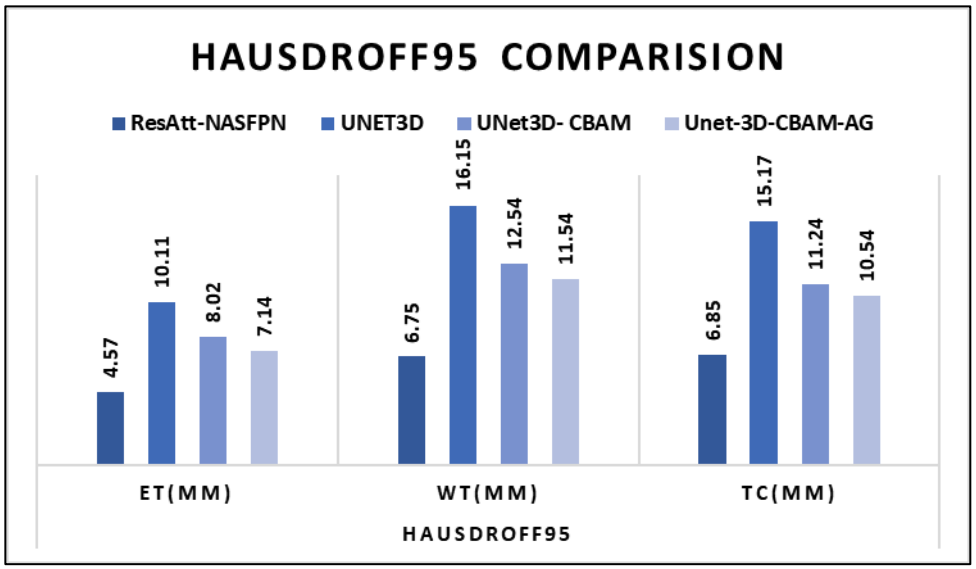

Figure 10. Result of HD95 Similarity Comparison with Different Components

Cascaded attention modules showed an additive effect when adding the attention gates on top of the CBAM setup, which further increased the Dice scores to 0.8014 (ET), 0.8714 (WT), and 0.8241 (TC) and the Hausdorff95 to 7.14 mm (ET), 11.54 mm (WT), and 10.54 mm (TC). Combining Dice and Focal loss, the proposed model reached the highest scores of 0.8406 (ET), 0.9157 (WT), and 0.8680 (TC) and the lowest Hausdorff95 of 4.57 mm (ET), 6.75 mm (WT), and 6.85 mm (TC), thereby confirming the importance of sophisticated attention mechanisms and efficient loss function design in achieving state-of-the-art performance in brain tumor segmentation. The segmentation masks exhibit consistent spatial extent and shape across axial, sagittal, and coronal views, indicating that the model effectively captures volumetric context rather than relying on plane-specific features. The multi-view results are shown in Figure 11.

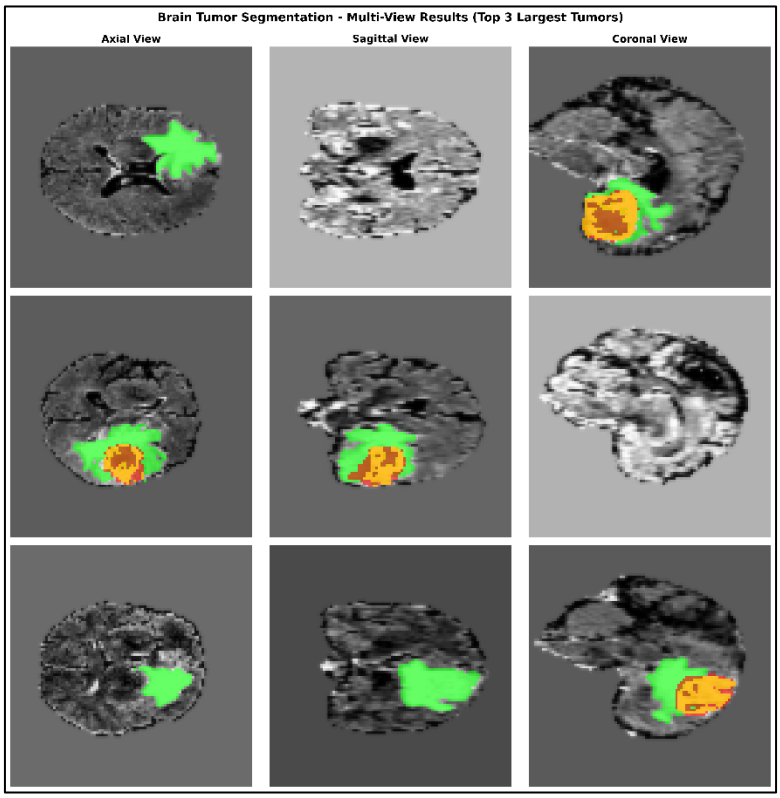

Figure 11. Multi-planar Visualization of Segmentation Results in Axial, Sagittal, and Coronal Views

7. Conclusion

The ResAtt-NASFPN approach offers automatic operation and feature selection for complex regions utilizing an enhanced neural architecture search-based FPN. To ensure better feature selection, a residual block is used on the encoder side for gradient flow. The CBAM module helps with the recalibration of features. Using parameters, the architecture achieves Dice similarity coefficients of 0.8406 ± 0.05 (ET), 0.868 ± 0.06 (TC) and 0.9157 ± 0.03 (WT), on the BraTS2020 validation set, which is 34% lower than the fixed-convolution baselines but with higher segmentation accuracy. The identification of computational primitives through a distribution learning approach that accomplishes architectural search automatically aims at the morphological heterogeneity of glioma sub-regions across MRI modalities (T1, T1ce, T2, FLAIR) with no acknowledgment of the neurologist manually. Medical segmentation generally employs a modular design in which parameters are only fine-tuned as opposed to full retraining when transfer learning is opted for.

References

[1] Jiang, Bin, Hongmei Liu, Dongling Sun, Haixin Sun, Xiaojuan Ru, Jie Fu, Siqi Ge, and Wenzhi Wang. "Mortality Due to Primary Brain Tumours in China and Detection Rate in People with Suspected Symptoms: A Nationally Representative Cross-Sectional Survey." World Journal of Surgical Oncology 19, no. 1 (2021): 71.

[2] Huang, Mengxing, Jie Zou, Yu Zhang, Uzair Aslam Bhatti, and Jing Chen. "Efficient Click-Based Interactive Segmentation for Medical Image with Improved Plain-ViT." IEEE Journal of Biomedical and Health Informatics (2024).

[3] Noor, Rida, Abdul Wahid, Sibghat Ullah Bazai, Asad Khan, Meie Fang, Syam MS, Uzair Aslam Bhatti, and Yazeed Yasin Ghadi. "DLGAN: Undersampled MRI Reconstruction Using Deep Learning Based Generative Adversarial Network." Biomedical signal processing and control 93 (2024): 106218.

[4] Ting, Hsienchih, and Manhua Liu. "Multimodal Transformer of Incomplete MRI data for Brain Tumor Segmentation." IEEE Journal of Biomedical and Health Informatics 28, no. 1 (2023): 89-99.

[5] Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. "U-net: Convolutional Networks for Biomedical Image Segmentation." In International Conference on Medical image computing and computer-assisted intervention, Cham: Springer international publishing, 2015. 234-241.

[6] He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. "Deep Residual Learning for Image Recognition." In Proceedings of the IEEE conference on computer vision and pattern recognition, 2016. 770-778.

[7] Woo, Sanghyun, Jongchan Park, Joon-Young Lee, and In So Kweon. "Cbam: Convolutional Block Attention Module." In Proceedings of the European conference on computer vision (ECCV), 2018. 3-19.

[8] Oktay, Ozan, Jo Schlemper, Loic Le Folgoc, Matthew Lee, Mattias Heinrich, Kazunari Misawa, Kensaku Mori et al. "Attention U-Net: Learning where to Look for the Pancreas." arXiv preprint arXiv:1804.03999 (2018).

[9] Zoph, Barret, and Quoc V. Le. "Neural Architecture Search with Reinforcement Learning." arXiv preprint arXiv:1611.01578 (2016).

[10] Lin, Tsung-Yi, Piotr Dollár, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie. "Feature Pyramid Networks for Object Detection." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2117-2125. 2017.

[11] Isensee, Fabian, Paul F. Jäger, Peter M. Full, Philipp Vollmuth, and Klaus H. Maier-Hein. "nnU-Net for Brain Tumor Segmentation." In International MICCAI Brainlesion Workshop, Cham: Springer International Publishing, 2020. 118-132.

[12] Menze, Bjoern H., Andras Jakab, Stefan Bauer, Jayashree Kalpathy-Cramer, Keyvan Farahani, Justin Kirby, Yuliya Burren et al. "The multimodal Brain Tumor Image Segmentation Benchmark (BRATS)." IEEE transactions on medical imaging 34, no. 10 (2014): 1993-2024.

[13] Susanto, Evan Kusuma, Handayani Tjandrasa, and Chastine Fatichah. "Data Augmentation Using Spatial Transformation for Brain Tumor Segmentation Improvement." In 2024 international seminar on intelligent technology and its applications (ISITIA), IEEE, 2024. 639-644.

[14] Lin, Tsung-Yi, Piotr Dollár, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie. "Feature Pyramid Networks for Object Detection." In Proceedings of the IEEE conference on computer vision and pattern recognition, 2017. 2117-2125.

[15] Nie, Dong, Jia Xue, and Xiaofeng Ren. "Bidirectional Pyramid Networks for Semantic Segmentation." In Proceedings of the asian conference on computer vision. 2020.

[16] Magadza, Tirivangani, and Serestina Viriri. "Efficient nnU-net for Brain Tumor Segmentation." IEEE Access 11 (2023): 126386-126397.

[17] Liu, Hengxin, Guoqiang Huo, Qiang Li, Xin Guan, and Ming-Lang Tseng. "Multiscale Lightweight 3D Segmentation Algorithm with Attention Mechanism: Brain Tumor Image Segmentation." Expert Systems with Applications 214 (2023): 119166.

[18] Umarani, C. M. ., Gollagi, S. G. ., Bamane, K. D. ., Gupta, P. ., More, S. J. ., & Ankali, S. B. . (2024). Accurate Segmentation of Brain Tumor Image using U-Net Based Self-Attention Mechanism. International Journal of Intelligent Systems and Applications in Engineering, 12(13s), 630. https://ijisae.org/index.php/IJISAE/article/view/4628

[19] Daimary, Dinthisrang, Mayur Bhargab Bora, Khwairakpam Amitab, and Debdatta Kandar. "Brain Tumor Segmentation from MRI Images using Hybrid Convolutional Neural Networks." Procedia Computer Science 167 (2020): 2419-2428.

[20] Nizamani, Abdul Haseeb, Zhigang Chen, Ahsan Ahmed Nizamani, and Uzair Aslam Bhatti. "Advance Brain Tumor Segmentation using Feature Fusion Methods with Deep U-Net Model with CNN for MRI data." Journal of King Saud University-Computer and Information Sciences 35, no. 9 (2023): 101793.

[21] Munir, Khushboo, Fabrizio Frezza, and Antonello Rizzi. "Deep Learning Hybrid Techniques for Brain Tumor Segmentation." Sensors 22, no. 21 (2022): 8201.

[22] Selvaraju, Ramprasaath R., Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. "Grad-cam: Visual Explanations from Deep Networks via Gradient-based Localization." In Proceedings of the IEEE international conference on computer vision, 2017. 618-626.

[23] Xu, Jia Hua, Wai Po Kevin Teng, Xiong Jun Wang, and Andreas Nürnberger. "A Deep Supervised U-attention net for Pixel-wise Brain Tumor Segmentation." In International MICCAI Brainlesion Workshop, Cham: Springer International Publishing, 2020. 278-289.

[24] Huttenlocher, Daniel P., Gregory A. Klanderman, and William J. Rucklidge. "Comparing Images using the Hausdorff Distance." IEEE Transactions on pattern analysis and machine intelligence 15, no. 9 (2002): 850-863.

[25] Zou, Kelly H., Simon K. Warfield, Aditya Bharatha, Clare MC Tempany, Michael R. Kaus, Steven J. Haker, William M. Wells III, Ferenc A. Jolesz, and Ron Kikinis. "Statistical Validation of Image Segmentation Quality based on a Spatial Overlap Index1: Scientific Reports." Academic radiology 11, no. 2 (2004): 178-189.

[26] Ali, Muhammad Junaid, Muhammad Tahir Akram, Hira Saleem, Basit Raza, and Ahmad Raza Shahid. "Glioma Segmentation using Ensemble of 2D/3D U-Nets and Survival Prediction using Multiple Features Fusion." In International MICCAI Brainlesion Workshop, Cham: Springer International Publishing, 2020. 189-199.

[27] Jia, Haozhe, Weidong Cai, Heng Huang, and Yong Xia. "H 2 NF-net for Brain Tumor Segmentation using Multimodal MR Imaging: 2nd Place Solution to BraTS Challenge 2020 Segmentation Task." In International MICCAI Brainlesion Workshop, Cham: Springer International Publishing, 2020. 58-68.

[28] Mallampati, Bhargav, Abid Ishaq, Furqan Rustam, Venu Kuthala, Sultan Alfarhood, and Imran Ashraf. "Brain Tumor Detection using 3D-UNet Segmentation Features and Hybrid Machine Learning Model." IEEE Access 11 (2023): 135020-135034.

[29] Liu, Chenyu, Wangbin Ding, Lei Li, Zhen Zhang, Chenhao Pei, Liqin Huang, and Xiahai Zhuang. "Brain Tumor Segmentation Network using Attention-based fusion and Spatial Relationship Constraint." In International MICCAI Brainlesion Workshop, Cham: Springer International Publishing, 2020. 219-229.

[30] Messaoudi, Hicham, Ahror Belaid, Mohamed Lamine Allaoui, Ahcene Zetout, Mohand Said Allili, Souhil Tliba, Douraied Ben Salem, and Pierre-Henri Conze. "Efficient Embedding Network for 3D Brain Tumor Segmentation." In International MICCAI Brainlesion Workshop, Cham: Springer International Publishing, 2020. 252-262.