Comparative Analysis of Deep Learning Architectures for Brain Tumor Segmentation

Huyen Quang Tran, Hieu Pham, Phuc Ho, Hanh Thi Minh Tran
The University of Da Nang - University of Science and Technology, Danang, Vietnam
huyenquangtran2002@gmail.com, phamdoanminhhieu2508@gmail.com, hoduyphuc304@gmail.com, hanhtran@dut.udn.vn

***Abstract*—Medical image segmentation is critical for diagnosis, treatment planning, and patient monitoring. In recent years, UNet-based approaches have significantly advanced the performance of automated segmentation tasks. However, many existing medical image datasets are developed for tasks other than brain tumor segmentation, leading to inconsistencies in model training and evaluation across studies. This work investigates the performance of six segmentation models - UNet, ResUNet, DeepResUNet, Recurrent UNet (DRU and SRU), and Duck-Net - for brain tumor segmentation. These models are fine-tuned on a public annotated brain tumor segmentation dataset from Figshare to evaluate these models for the brain tumor segmentation task. Model performance is evaluated using standard metrics, including the Dice similarity coefficient and Intersection over Union (IoU). A comprehensive analysis is conducted to compare accuracy and robustness across architectures, with the aim of identifying the most suitable model for brain tumor segmentation in clinical applications. The dataset and code are publicly available online at https://github.com/NegiKirin/Brain-Tumor-Segmentation.**

***Index Terms*—Brain tumor segmentation, encoder-decoder models, UNet-based models**

I. Introduction

Brain tumor segmentation from magnetic resonance imaging (MRI) presents a particularly challenging case, characterized by heterogeneous tumor morphology, variable contrast, and indistinct boundaries between healthy and diseased tissue. UNet [1] was the first CNN model for image semantic segmentation with end-to-end training procedure. Since then, many UNet models have been proposed for Medical Image Segmentation. UNet with a standard encoder-decoder model for any deep learning segmentation model. Several variants of UNet have been proposed to enhance its representation capability [2]–[5]. While residual connections have been introduced in ResUNet [2], deeper residual blocks have been employed in DeepResUNet [3], and recurrent blocks have been integrated in RecurUNet [4]. Other approaches involve architectures with midscope, widescope, and separated blocks, such as Duck-Net [5]. In addition, several methods aim to improve the ability of UNet to capture global contextual information by combining convolutional neural networks (CNNs) with Transformer-based models, including Attention-UNet [6], UNETR [7], UNETR++ [8], and Swin-UNet [9].

Although numerous studies have demonstrated the efficacy of individual architectures on brain tumor datasets, direct evaluations under a unified framework are scarce. This study addresses this gap by fine-tuning five representative UNet–based models, which are UNet, ResUNet, DeepResUNet, DRU/SRU, and DUCK-Net, on a standardized, publicly accessible brain tumor MRI dataset from Figshare [10]. This study aims to enhance generalization capability and evaluate the segmentation performance of different models on a single dataset. Segmentation performance is assessed using the Dice similarity coefficient and Intersection over Union (IoU), with experiments designed to evaluate both accuracy and robustness across tumors. This comprehensive analysis provides insights into each architecture's relative merits and trade-offs, guiding the selection of optimal models for clinical deployment in neuro-oncological imaging.

The rest of the paper is organized as follows. Section II describes the methodology; Section III reports and analyzes the experimental results; and Section IV concludes with a summary of the findings and suggestions for future research directions.

II. Methodology

In this study, six deep learning models were trained and evaluated for brain tumor segmentation. The models included UNet, a foundational architecture widely used in medical image segmentation; ResUNet, a UNet variant augmented with residual connections to strengthen feature learning; DeepResUNet, an extended ResUNet design emphasizing deeper feature extraction; Recurrent UNet, an improved UNet variant engineered to boost segmentation efficiency under resource constraints; and DuckNet, a novel architecture optimized for segmentation through its distinctive dual-path design. All models were trained on the same dataset and assessed using the Dice similarity coefficient and Intersection over Union. A comparative analysis was conducted to identify the optimal model for brain tumor segmentation, thereby underpinning the advancement of AI-based medical applications.

A. UNet

UNet [1] is a widely recognized deep learning architecture for image segmentation, originally inspired by the work on fully convolutional networks (FCNs) for segmentation tasks [11]. In the FCN framework, successive pretrained convolutional layers (e.g., VGG16 [12], and GoogLeNet [13]) are employed for feature extraction. The extracted multi-scale feature maps are then upsampled to produce the final prediction mask. However, FCNs often suffer from poor localization accuracy because the upsampling process directly reconstructs the mask from coarse feature representations. UNet addresses this limitation by introducing a symmetric expanding path that

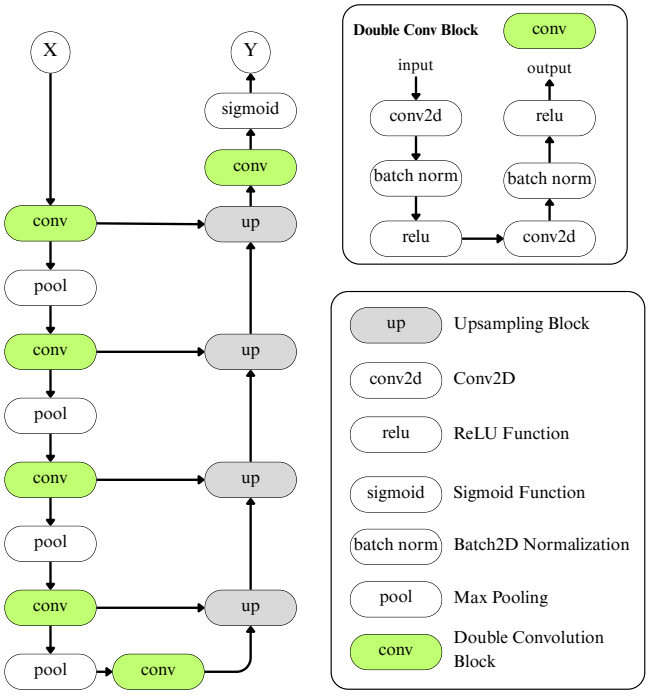

Fig. 1: The UNet architecture

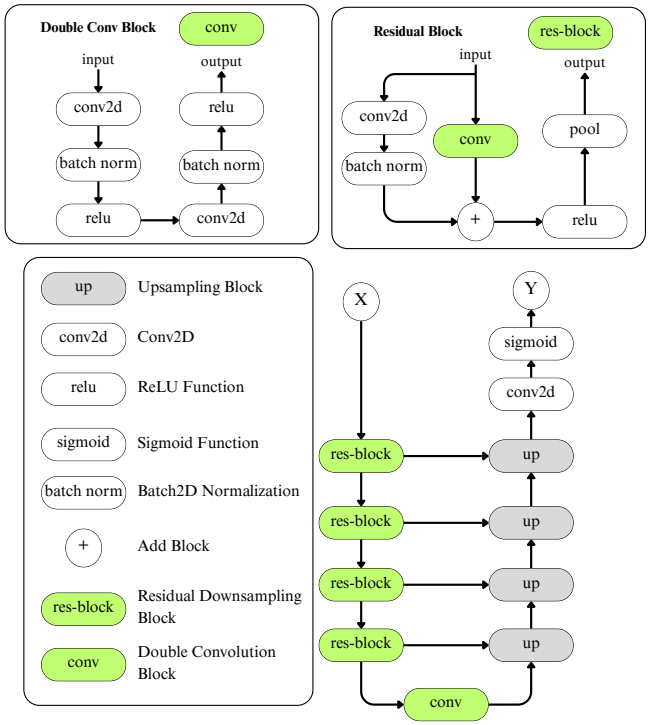

Fig. 2: The ResUNet architecture

complements the contracting path. Specifically, high-resolution features from the contracting path are concatenated with the corresponding upsampling layers in the expanding path, enabling the network to recover spatial detail and generate more precise segmentation masks. The overall architecture of UNet is illustrated in Fig. 1.

B. ResUNet

ResUNet [2], an enhanced variant of UNet, is designed to improve feature extraction and alleviate the vanishing gradient issue in deep architectures by incorporating residual blocks. The original U-shaped layout is preserved, but conventional convolutional layers are enhanced with residual connections within the convolution block as illustrated in Fig.2. Each residual block consists of an identity branch and a double convolution path comprising two stacked convolutional layers. The outputs from both branches are summed, passed through a ReLU activation function, and then processed via a max pooling layer for spatial down-sampling. This design allows deeper network construction while maintaining gradient flow, enhancing the model's segmentation accuracy. The decoder part of ResUNet is identical to UNet's decoder.

C. DeepResUNet

DeepResUNet [3] extends ResUNet by employing deeper downsampling and upsampling pathways to enhance feature representation and segmentation performance. The architecture of the DeepResUNet is illustrated in Fig. 3. In the encoder path, the conventional double convolution blocks are replaced with Pre-Activated Double Convolution Blocks, where batch normalization and activation functions are applied before the convolution operations. Moreover, Pre-Activated Residual Downsampling Blocks are utilized to encode image data more effectively, allowing for improved gradient flow and deeper semantic extraction. In the decoder path, spatial information is reconstructed using Pre-Activated Residual Upsampling Blocks, which serve as upsampling units while maintaining consistency in feature distribution through skip connections. This architectural refinement enables DeepResUNet to capture more complex structural patterns within medical images, which is particularly beneficial for tasks that require high segmentation precision, such as brain tumor delineation.

D. Recurrent UNet

Recurrent UNet [4], an enhanced variant of UNet, was designed to improve medical image segmentation efficiency under constrained computational resources. A recurrent mechanism is integrated to refine predictions across multiple iterations, with recurrent units such as Dual-Gated Recurrent Unit (DRU) and Single-gated Recurrent Unit (SRU) embedded within the encoder–decoder architecture. In the encoder, feature extraction is performed by convolutional blocks, and DRU/SRU modules are employed to retain information through successive passes. In the decoder, spatial details are reconstructed via global skip connections from the encoder, and recurrent units are used to maintain contextual consistency. Through this iterative refinement process, segmentation accuracy and generalization are enhanced, rendering the Recurrent UNet particularly well-suited for complex medical segmentation tasks under limited-resource conditions. Figures 4 and 5 show the architecture of DRU and SRU, respectively. The DRU was integrated

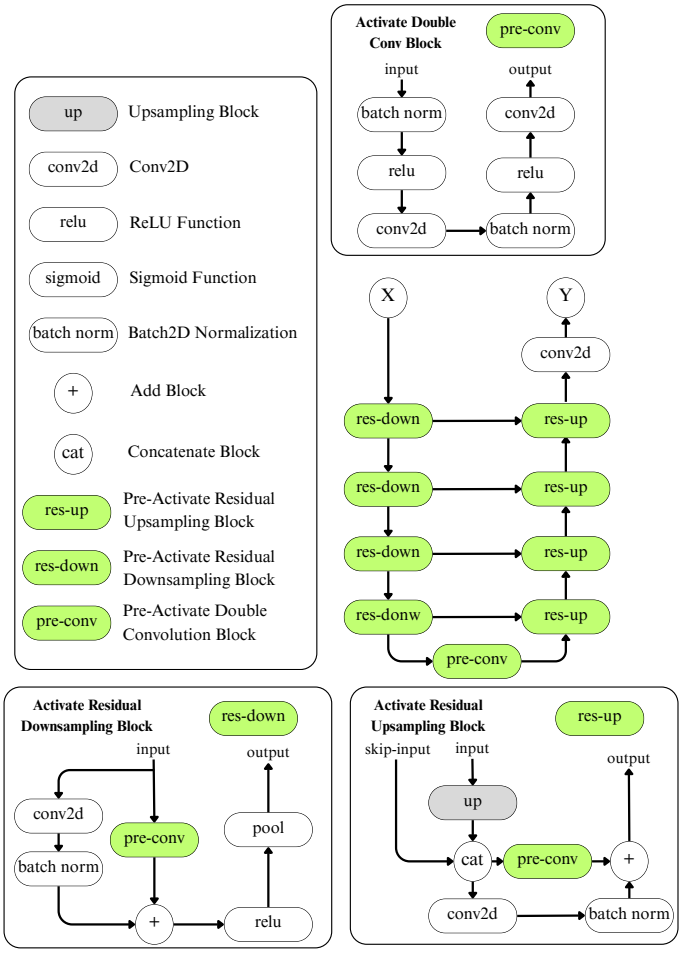

Fig. 3: The DeepResUNet architecture

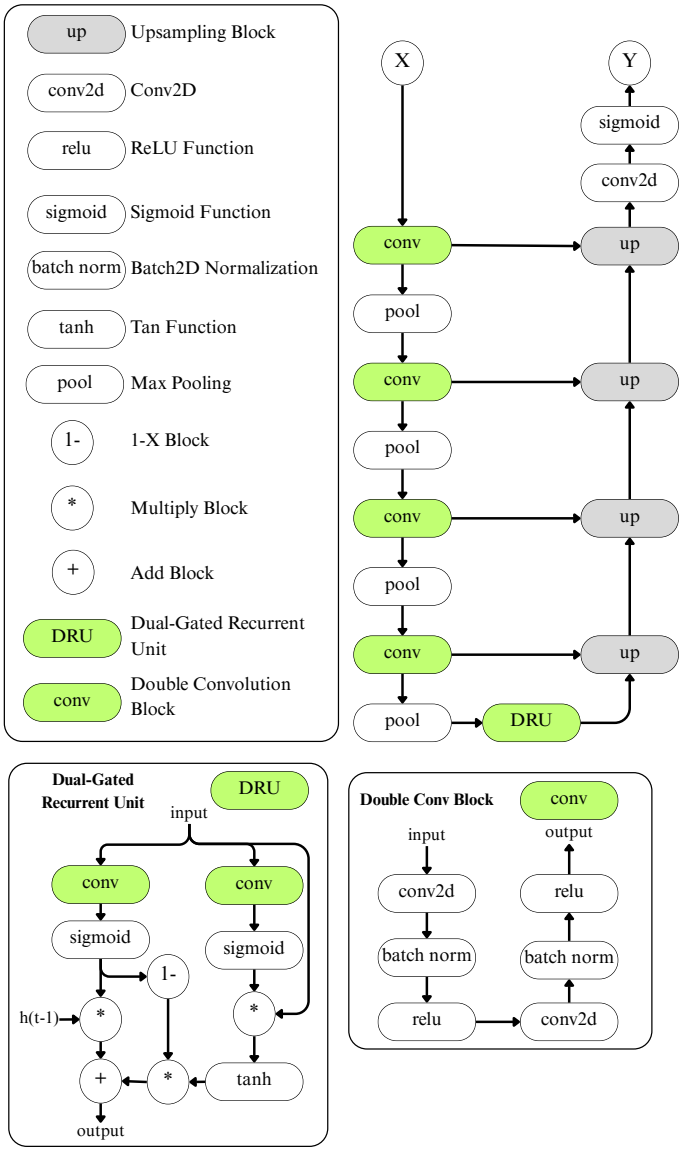

Fig. 4: The architecture of the Dual-Gated Recurrent Units

into the Recurrent UNet architecture to enhance contextual information processing across iterative passes. In the encoder, DRU modules were placed in deeper layers to manage hidden states, retaining salient features while filtering out irrelevant data. Within the decoder, DRU units were positioned near the output to refine segmentation predictions by combining skip-connected feature maps with recurrent state information. Through this integration, complex feature learning was optimized, and overall segmentation accuracy was improved. The SRU has the same architecture as DRU but without a reset gate acting on its input. This modification simplifies the design by eliminating one encoder–decoder network from the recurrent unit. Therefore, SRU has significantly faster training times and lower computational demands than DRUs and other traditional gated units.

E. DUCK-Net

DUCK-Net (Dense UNet with Custom Kernel) [5] was developed as an enhanced UNet variant for medical image segmentation under resource constraints. Dense connectivity and custom convolutional kernels were combined into specialized "DUCK blocks" within the encoder–decoder framework to enable multi-scale feature extraction. In the encoder, DUCK blocks comprising depthwise, pointwise, and atrous convolutions were applied alongside residual connections to capture features across varied kernel sizes. Spatial details were reconstructed through global skip connections and additional DUCK blocks during decoding. Each block integrates wide-scope/mid-scope convolutions, separable convolutions, batch normalization, and ReLU activations to minimize parameter count and accelerate convergence. Dense inter-block connections were employed to enhance feature reuse and alleviate vanishing gradients. Finally, a 1×1 convolutional layer was used to generate binary or probabilistic segmentation masks. DUCK-Net was shown to achieve efficient, high-accuracy performance in medical image segmentation. Details of the DUCK-Net architecture are described in Fig. 6.

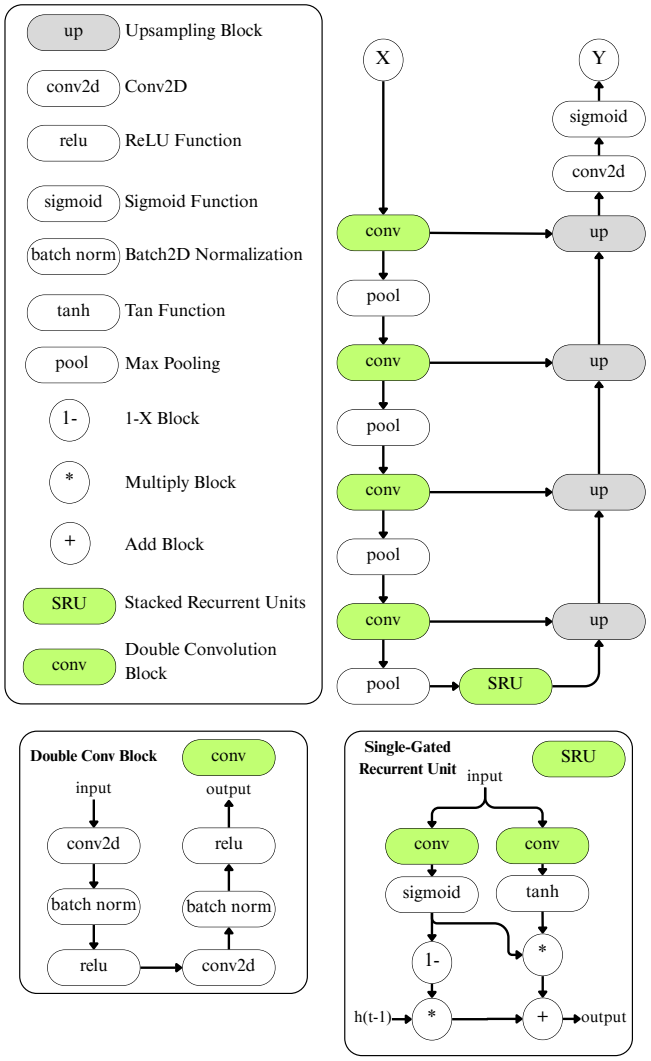

Fig. 5: The architecture of the Single-gated Recurrent Units

III. Experimental results and discussions

A. Dataset

The Brain tumor dataset employed in this study is publicly available on the Figshare platform [10]. It comprises a curated collection of magnetic resonance imaging (MRI) scans specifically acquired to support research in brain tumor analysis and the development of machine learning–based segmentation algorithms. In total, the dataset contains 3,064 image–mask pairs. Each MRI slice is a single-channel grayscale image with a resolution of 512 × 512 pixels. The dataset is partitioned into 80% for training and 20% for testing. For each slice, a binary mask of identical dimensions is provided, where pixels with a value of 1 represent tumor tissue and pixels with a value of 0 correspond to background or healthy anatomy.

B. Experiment Settings

All models were trained using a unified configuration to ensure fairness in performance comparison. Each model was trained for 150 epochs with a batch size of 16, except for DUCK-Net, which was limited to 100 epochs due to its faster convergence and higher computational demand.

For all methods, the initial learning rate was set to a maximum of $1 \cdot 10^{-5}$ and decayed gradually to a minimum of $1 \cdot 10^{-6}$, governed by the Cosine Annealing LR scheduler. This dynamic adjustment facilitated smoother convergence across training. A dropout rate of 0.3 and a weight decay factor of $1 \cdot 10^{-4}$ were applied to mitigate overfitting. The Dice Loss function was used to quantify segmentation error, making it particularly suitable for the frequent class imbalance observed in medical imaging tasks. Model optimization used the AdamW algorithm, known for its robust learning rate adaptation and strong generalization performance.

The performance of the segmentation models was assessed using the Dice similarity coefficient and Intersection over Union (IoU), widely adopted metrics in medical image segmentation. IoU is commonly used in computer vision benchmarks such as object detection or semantic segmentation. In addition, Dice [14] is widely used in medical image segmentation, as it is particularly suitable for cases where the foreground occupies a relatively small region compared to the background. Therefore, these two metrics are well-suited to evaluate brain tumor segmentation models. Dice and IoU are defined as in equations 1 and 2, respectively.

$$Dice = \frac{2TP}{2TP + FP + FN} \tag{1}$$

$$IoU = \frac{TP}{TP + FP + FN} \tag{2}$$

In these equation, True Positive (TP) refers to the number of pixels correctly predicted as foreground; False Positive (FP) denotes the number of pixels incorrectly predicted as foreground when they actually belong to the background; False Negative (FN) represents the number of pixels incorrectly predicted as background when they actually belong to the foreground.

C. Comparative Results

The Dice score and IoU for all six segmentation models are summarized in Table I. As can be seen on this table, DuckNet achieves the highest performance ith a test Dice of 0.8575 and IoU of 0.7561. DeepResUNet ranked second, with a test Dice of 0.8047 and IoU of 0.7235 using approximately 32 million parameters, demonstrating the advantage of deep residual blocks in capturing complex tumor characteristics. ResUNet and UNet yielded competitive results as well.

Interestingly, the lightweight architectures DRU and SRU achieved relatively lower accuracy compared to the larger models but remained competitive, with test Dice scores of 0.7954 and 0.7825, respectively. Despite their smaller parameter counts (12.7M – 13.9M), their results indicate a favorable trade-off between efficiency and accuracy. These models are particularly promising for deployment in resource-constrained environments where computational and memory costs are significant considerations.

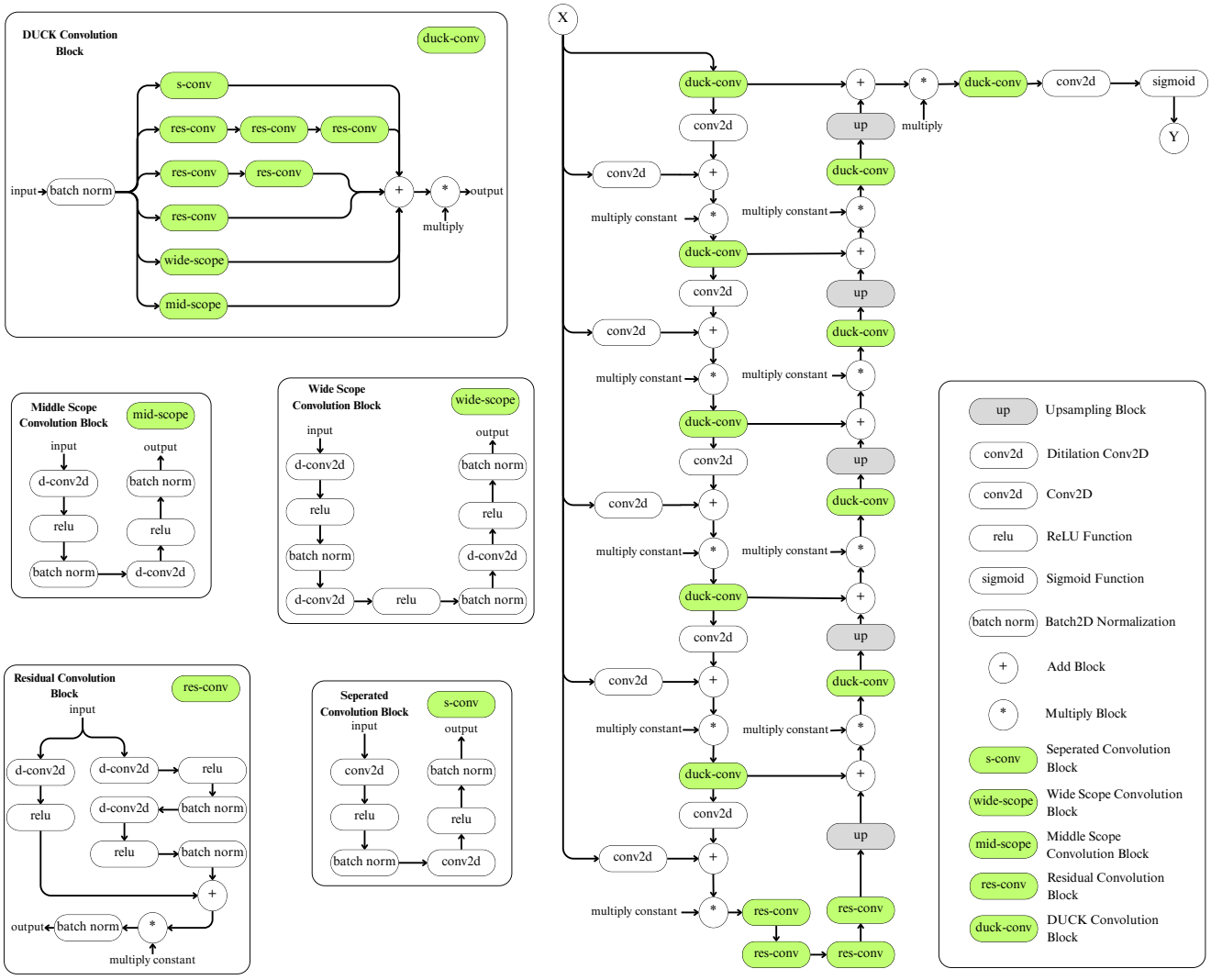

Fig. 6: The DUCK-Net architecture

Model	Parameters	Train		Test	
		Dice	*IoU*	*Dice*	IoU
UNet	31.3M	**0.9833**	0.9577	0.8050	0.7232
ResUnet	31.5M	**0.9833**	**0.9674**	0.8062	0.7233
DeepResUnet	32.6M	0.9792	0.9597	0.8047	0.7235
DRU	13.9M	0.9065	0.8301	0.7954	0.6699
SRU	**12.7M**	0.9260	0.8630	0.7825	0.6536
DUCK-Net	38M	0.9753	0.9520	**0.8575**	**0.7561**

TABLE I: The comparison result on the Brain Tumor Dataset

Fig. 7 presents example predictions of brain tumor segmentation across axial, sagittal, and coronal MRI slices using six models compared against ground truth. DUCK-Net consistently achieved the most accurate and robust segmentations across different orientations, demonstrating superior generalization. However, most models showed difficulty in handling small tumors and poorly defined boundaries. DUCK-Net and DeepResUNet demonstrated better preservation of fine structural details through extensive skip connections and residual pathways, although their relatively large model sizes may limit applicability in resource-constrained environments. In contrast, lightweight models such as DRU and SRU offer notable computational efficiency but require further refinement to achieve segmentation accuracy comparable to that of larger architectures.

IV. Conclusion

Experimental results show that DUCK-Net achieved the best performance, followed by DeepResUNet and ResUNet, both of which demonstrated strong segmentation capabilities. Notably,

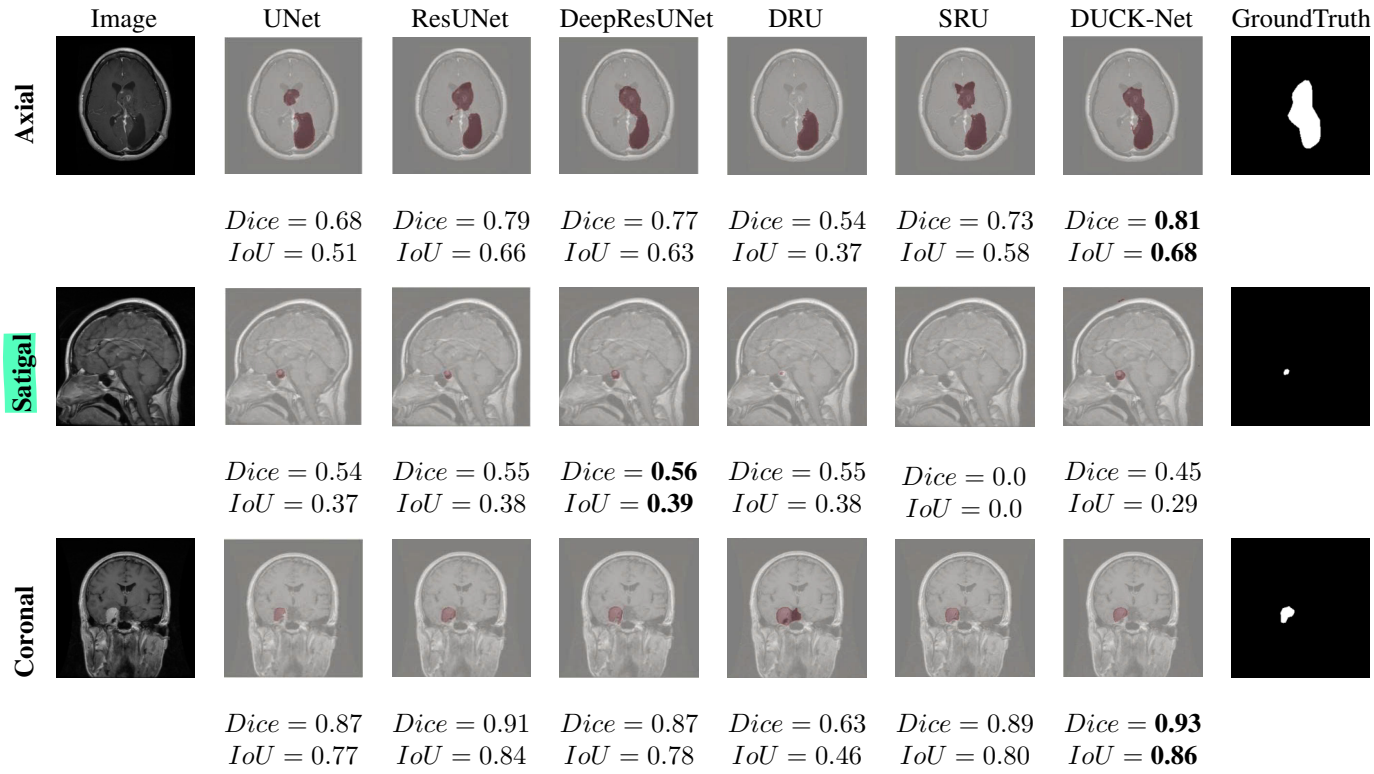

Fig. 7: Example predictions on brain tumor segmentation dataset

lightweight models such as DRU and SRU contain significantly fewer parameters, making them promising candidates for deployment in resource-constrained environments.

These findings underscore the importance of selecting appropriate network architectures for specialized medical segmentation tasks and provide critical insights to guide future research. Potential directions include further optimization of these efficient architectures and exploration of novel methods to enhance the accuracy and efficiency of brain tumor segmentation models.

REFERENCES

[1] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[2] L. Xiang, Y. Li, W. Lin, Q. Wang, and D. Shen, "Unpaired deep cross-modality synthesis with fast training," in *International Workshop on Deep Learning in Medical Image Analysis*. Springer, 2018, pp. 155–164.

[3] Z. Zhang, Q. Liu, and Y. Wang, "Road extraction by deep residual u-net," *IEEE Geoscience and Remote Sensing Letters*, vol. 15, no. 5, pp. 749–753, 2018.

[4] W. Wang, K. Yu, J. Hugonot, P. Fua, and M. Salzmann, "Recurrent u-net for resource-constrained segmentation," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 2142–2151.

[5] R.-G. Dumitru, D. Peteleaza, and C. Craciun, "Using duck-net for polyp image segmentation," *Scientific reports*, vol. 13, no. 1, p. 9803, 2023.

[6] 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.

[7] A. Hatamizadeh, Y. Tang, V. Nath, D. Yang, A. Myronenko, B. Landman, H. R. Roth, and D. Xu, "Unetr: Transformers for 3d medical image segmentation," in *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, 2022, pp. 574–584.

[8] A. M. Shaker, M. Maaz, H. Rasheed, S. Khan, M. H. Yang, and F. S. Khan, "Unetr++: Delving into efficient and accurate 3d medical image segmentation," *IEEE Transactions on Medical Imaging*, 2024.

[9] A. Hatamizadeh, V. Nath, Y. Tang, D. Yang, H. R. Roth, and D. Xu, "Swin unetr: Swin transformers for semantic segmentation of brain tumors in mri images," in *International MICCAI brainlesion workshop*. Springer, 2021, pp. 272–284.

[10] J. Cheng, "brain tumor dataset," 4 2017. [Online]. Available: https://figshare.com/articles/dataset/brain_tumor_dataset/1512427

[11] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3431–3440.

[12] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[13] 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*, 2015, pp. 1–9.

[14] R. R. Shamir, Y. Duchin, J. Kim, G. Sapiro, and N. Harel, "Continuous dice coefficient: a method for evaluating probabilistic segmentations," *arXiv preprint arXiv:1906.11031*, 2019.