IJIO
(*International Journal of Industrial Optimization*)
ISSN 2723-3022 (online) | 2714-6006 (printed)
Vol. 7, No. 1, 15-26
DOI: https://doi.org/10.12928/ijio.v7i1.11576

Semantic brain tumor segmentation from 3D MRI using u^2-net with custom dilated and residual u-block

Elvaret [a], Habibullah Akbar [b,*], Nanna Suryana Herman [c], Marwan Kadhim Mohammed Al-Shammari [d]

[a] Informatics Engineering, Faculty of Computer Science, Esa Unggul University, Indonesia
[b] Master of Computer Science, Faculty of Computer Science, Esa Unggul University, Indonesia
[c] Faculty of Nursing, University College MAIWP International, Malaysia
[d] Computer Center, University of Baghdad, Baghdad, Iraq
* Corresponding Author: habibullah.akbar@esaunggul.ac.id

ARTICLE INFO

Article history
Received: September 15, 2024
Revised: November 27, 2025
Accepted: January 20, 2026

Keywords
Computational Efficiency;
Magnetic Resonance Imaging;
ReSidual U-Block (RSU);
Semantic Segmentation;
Three Dimension (3D);

ABSTRACT

Segmentation of brain tumors in volumetric medical images is challenging due to the complexities of the tumor structure, the types, and the heavy-weight 3D data processing. In contrast, 2D-based segmentation methods on the slice data reduce the amount of information due to the anisotropic shape of the tumors and lead to poor segmentation results. This study proposes a 3D network structure combining ReSidual U-Block (RSU), custom dilated block, and U^2-Net for automatic segmentation of brain tumors from MRI images, namely 3D RSU U^2-Net+. The RSU and custom dilated block are embedded and joined in the nested U-Net structure to obtain multi-resolution features and global information, enhancing segmentation accuracy while reducing computational overhead. The proposed method outperformed the segmentation results of the standard U-Net, on brain tumor data in the medical segmentation Decathlon (MSD) dataset. The proposed model achieves an average validation soft dice loss of 0.1320 and dice score coefficient of 78% and intersection over union of 64% for testing. Although having 3 times parameters, the model requires less GPU time (397.7 minutes) than U-Net (433.6 minutes), demonstrating improved computational efficiency resulting from the effective use of residual and dilated blocks. Moreover, the model achieves 75.4% average sensitivity and 99% specificity for edema, enhancing, and non-enhancing tumors. These experimental results show that the 3D RSU U^2-Net+ has been able to outperform the U-Net. However, the model's performance on non-enhancing tumors remains relatively lower compared to other tumor types, indicating on opportunity for further optimization.

1. Introduction

Early diagnosis of gliomas (tumors) within the brain can alleviate adverse events and promote better treatment such as chemotherapy and radiation. Studies by [1] has shown that patients with pilocytic astrocytoma (grade I) have an excellent prognosis with a 96% survival rate at 10 years (with surgical intervention) while low-grade astrocytoma (grade II) has a median survival time (MST) of 5.6 years. Then, patients with anaplastic astrocytoma (grade III) have an MST of 1.6 years while

glioblastoma (grade IV) has a poor prognosis with an MST of only 0.4 years. Inspection of brain tumors using magnetic resonance imaging (MRI) or images can improve diagnostics and treatment planning. Segmentation and measurement of the tumors within these images require expertise and experience. Doctors can make mistakes due to time-consuming manual procedures and fatigue [2]. To improve disease diagnosis and treatment, automated segmentation based on machine learning has attracted lots of attention [3]. There are several challenges for brain tumor segmentation such as complex characteristics due to its variation of shape, size, and location, classification ambiguities [4] and computationally demanding.

The characteristics problem can be categorized as features and modalities. Previous studies have focused on hand-crafted feature selection before feeding into classifiers. Unfortunately, hand-crafted features are only suitable for local details and fail to capture the global structure of the targeted object of tumors [5]. Many studies have switched to convolutional neural network (CNN) architectures [6]. However, the nature of convolutional neural networks (CNN) is limited to local context information [7]. To establish long-range dependencies and global context information, residual connections, also known as skipping connections ([8] used the term skip layer), are used to join features between layers. Moreover, segmentation of medical images is challenging because most medical image data consists of 3D volumes [9]. CNNs have advantages as they can extract multiple image features automatically instead of hand-crafted feature selection. Some studies can be found in [5], [10]–[12].

Implementation of training CNNs on a standard computer can take a significant amount of time. A study comparing the training of deep learning-based classification of MRI contrast on both CPU and GPU showed that training ResNet18 on a single CPU laptop (2 GHz Quad-Core Intel Core i5) for 30 epochs took 199 minutes. However, using a ResNet50 model on a GPU (NVIDIA Tesla P100) for the same 30 epochs reduced the training time to just 10 minutes [13]. For addition using GPU, train TwoPathCNN for only a few minutes and its performance achieves 0.58, 0.81, and 0.72 dice scores for enhancing, completing, and core tumor respectively [14]. Therefore, it is better to utilize a graphical processing unit (GPU) for training deep learning models [15].

Comparing brain tumor segmentation algorithms is difficult because the literature contains different datasets, tumor types, and disease stages. To alleviate these problems, researchers are encouraged to work based on standard datasets such as Brain Tumor Segmentation (BraTS). BraTS dataset contains three types of tumors that require semantic pixel labeling (semantic segmentation). Firstly, edema, a swelling of the brain tissue surrounding the tumor. This is a natural response from the body due to the presence of the tumors and may be caused by increased intracranial pressure. Edema can be seen as an area with increased signal in the MRI T2-weighted image and FLAIR (Fluid Attenuated Inversion Recovery), indicates fluid leaking into the surrounding tissue. Secondly, enhancing tumors with gadolinium enhancement is part of brain tumors that show increased activity when using gadolinium contrast material in MRI images. This increased activity indicates that this area is active and may contain tumor cells that are more aggressive or active. Lastly, non-enhancing tumors do not show increased activity when using a contrast material such as Gadolinium in MRI images. This means that non-enhancing tumors do not take up contrast material strongly and are not seen more clearly in Gadolinium-enhanced images [16], [17].

From 2012-2018, various algorithms such as Random Forest, ResNet, DenseNet, U-Net and DeepSCAN have attempted to tackle BraTS challenge. According to [3], the best models at the time were developed based on U-Net/U^2-Net architecture and variational autoencoder (VAE). Although U-Net-based architectures achieve good attention in the literature, their nesting variants are vulnerable to overfitting [18]. The VAE approach achieves the performance of 0.7664, 0.8839, and 0.8154 average dice scores for enhanced tumor core, whole tumor, and tumor core respectively [19]. On the other hand, U-Net architecture is the most popular for medical image segmentation challenges [20]. U-Net follows the encoder-decoder architecture pattern. The encoder progressively down-samples the feature maps, and the decoder up-samples them while concatenating them with the corresponding features from the encoder. The encoder aims to capture local details of the tumor features while the decoder captures the global structure. U-Net is a network of 23 convolutional

 10.12928/ijio.v7i1.11576

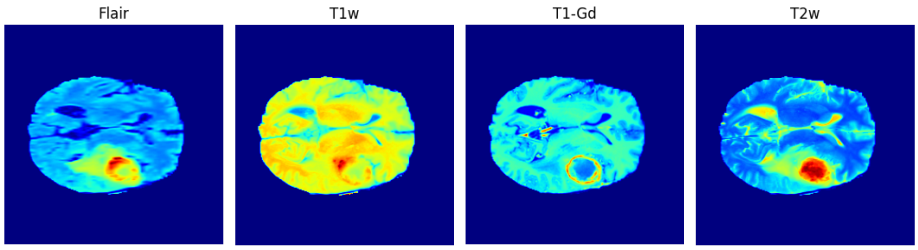
layers and was originally developed for 2D medical images [21]. However, U-Net has been adopted to handle 3D medical images as well but only to achieve a performance of around 79% dice score [22].

The contribution of this study paper is twofold: the advancement of U-Net structures to handle 3D magnetic resonance images and faster computational time. The proposed model is faster despite having more parameters than U-Net due to several architectural and operational optimizations. It uses ReSidual U-Block (RSU) to train deeper networks allowing gradients to flow more easily through the network. Different dilation rates are utilized to capture multi-scale features without increasing the number of parameters. Moreover, it has intricate skip connections, combining up-sampling and concatenation of features at multiple scales.

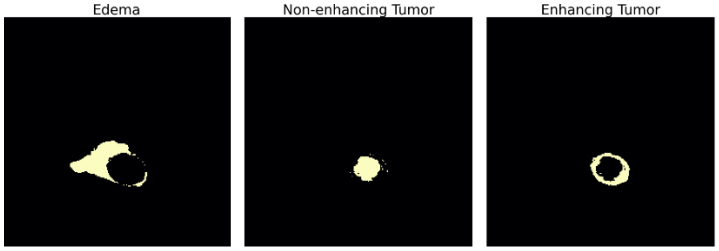
2. Method

2.1. Dataset

We used the medical segmentation decathlon dataset. This dataset presents a challenge in identifying targets that are complex and located in various positions [20]. The dataset contains 750 MRI (only 484 have ground-truth labels) images from patients with glioblastoma (grade IV) or lower-grade glioma. Each MRI image set consists of 4 imaging modalities or channels: T1 (T1-weighted), T1-Gd (T1-weighted with gadolinium contrast), T2 (T2 weighted), and FLAIR (Fluid Attenuated Inversion Recovery) including 3 ground truth labeling pixel-wise for edema, non-enhancing tumor, and enhancing tumor. These multiple modalities help differentiate tissue characteristics, making the dataset particularly valuable for assessing the effectiveness of advanced segmentation models. A sample of these channels is given in Fig. 1.



(Top)



(Bottom)

Fig. 1. Images of the four channels of the first patient at slice 75 (top) and corresponding labels (bottom)

Each image has a size of 240 x 240 x 155 x 4 and a mask with the size of 240 x 240 x 155 which aligns with our model's multi-scale feature extraction approach. By leveraging the distinct contrast across modalities, our model can refine spatial representations and enhance segmentation accuracy. The data format used in the dataset is Nifty (neuroimaging informatics technology initiative with suffix. nii). The dataset contains three targets of tumor sub-regions: edema, enhancing tumor, and non-enhancing tumor.

2.2. Preprocessing

We extract 5 sub-volumes from each MRI image within the dataset with a criterion each sub-volume contains at least 10% of the tumor. The 10% threshold was selected based on practical considerations and initial observations to balance computational efficiency with sufficient tumor representation. While alternative thresholds (e.g., 5% or 15%) could be explored, maintaining a consistent preprocessing approach ensures fair model comparisons and prevents unnecessary computational overhead. Additionally, ensuring a minimum tumor presence helps the model focus on learning relevant tumor features rather than biased toward classifying large non-tumor regions. In total, there are 1000 sub-volumes generated from 200 MRI images. These sub-volumes are split using a ratio of 0.8:0.1:0.1 for training, validation, and testing data. In addition, the pixel values within the images are normalized based on mean and standard deviation.

2.3. Residual U-Blocks (RSU) and U^2-Net

ReSidual U-Block (RSU) is similar to residual block (ResNet) by inserting shortcut connections between layers [23]. However, RSU uses U-Net structure to replace the convolution layers. By nesting the U-Net structure, [24] developed a U^2-Net structure. It has 11 cubes that are filled by the RSU structure. In the encoder, RSU blocks have a decreasing number of layers by feature mappings for its encoding process. At the bottom parts, the RSU is dilated to capture larger receptive fields. Finally, the decoder parts up-sample the maps and fuse them with a concatenation operation and then to a sigmoid function to the final probability map.

Each RSU block uses a 3D version of U-Net as we work with 3D MRI data. It uses 3D convolution to extract features from volumetric data and 3D up-sampling to reconstruct the original dimensions. The nested U-Net structure with RSU will enhance computational efficiency and segmentation accuracy by reducing redundant operations through multi-scale feature extraction, facilitating better information flow, and preserving both fine-grained and high-level contextual features, resulting in high performance with minimal resource utilization. Furthermore, [25] adds more pathways between the encoder and decoder to reduce semantic gap between convolutional layers.

2.4. Proposed Network

The proposed network architecture is given in Fig. 2 namely 3D RSU U^2-Net. The architecture consists of 6 RSU U-Net+ blocks and 3 residual custom dilated blocks (see Fig. 2). The input is volumetric images with dimensions of 160 x 160 x 16 and the output has the same dimension as the input. Fig. 3 (a) shows RSU U-Net+ blocks contain U-Net whose decoders are connected to RSU from encoder parts via re-designed skip pathways with less dense connections compared to [25]. In the encoding process, the convolution layer (Conv) has a kernel size of 3 x 3 x 3 followed by Batch Normalization (BN), and is activated using ReLU function. After two convolutional processes, the features are down-sampled using max pooling. After the encoding phase, the bridge connects the encoder to the decoder.

Then, the decoding process performs 3D deconvolution to increase image resolution and combine feature maps from the encoder parts. The residual custom dilated blocks aim to capture a larger area of the receptive fields with dilation up to 8 pixels (see Fig. 3 (b)). The custom dilated block optimizes the model architecture by expanding the receptive filed without increasing parameters, enabling capture of multiscale contextual information for precise boundary detection [24]. The kernel number (KN) begins with the number of 32 and doubles each convolutional layer with the maximum number of 512.

10.12928/ijio.v7i1.11576

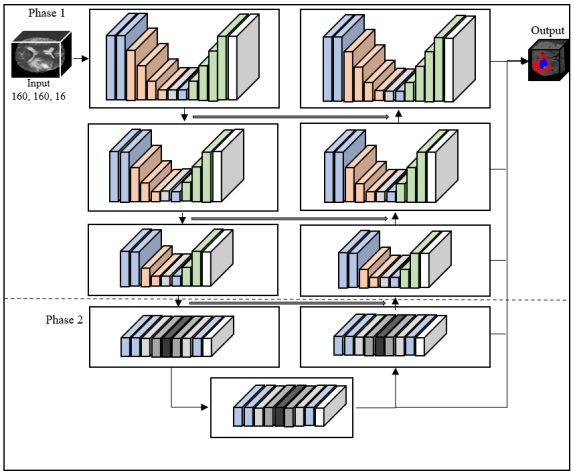
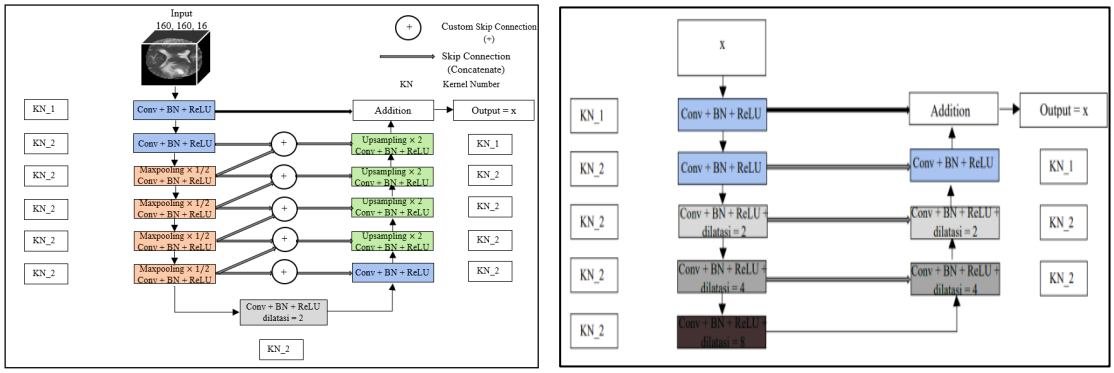


Fig. 2. (a) RSU U-Net+ block and (b) residual custom dilated block



(a) (b)

Fig. 3. (a) RSU U-Net+ block and (b) residual custom dilated block

2.5. Evaluation

To evaluate the performance, we use GPU T4 with 15 GB of VRAM on the Google Collaboratory environment. There are 6 metrics i.e., soft dice loss (SDL), computing time, dice score coefficient (DSC), IoU (Intersection over Union) or Jaccard index, sensitivity or recall, and specificity. The SDL and computing time metrics are used for training and monitoring the model while the remaining 4 metrics are used for both the training/validation and testing phases. The computing time is used to measure the training process using the Adam optimizer. The soft dice loss measures the overlap between two volumetric images: the predicted segmented volume and the ground truth mask. Equation (1) shows the SDL formula.

$$Soft\ Dice\ Loss = 1 - \frac{1}{3}\sum_{c=1}^{3}\frac{2\times\sum_{i=1}^{N} p_i \times g_i + \varepsilon}{\sum_{i=1}^{N} p_i^2 + \sum_{i=1}^{N} g_i^2 + \varepsilon} \quad (1)$$

where:

p_i : the model's estimated output value for each individual pixel i
g_i : the actual ground truth label for every pixel i
N : the total summation of pixels contained within the 3D volume
ε : a smoothing constant $(1e-5)$ included to avoid zero-division errors
c : the target classes (1,2,3) denoting edema, non-enhancing tumor, and enhancing tumor respectively.

Besides soft dice loss, the training is also evaluated using computational time.

Equation (2) and Equation (3) show DSC and IoU metrics are used for model evaluation. The dice score coefficient measures the overlapping portion between the predicted and ground truth areas

and is divided by the combined number of elements in both sets. This metric illustrates how well the predictions fit the actual shape and position of the identified 3D brain tumor structures. Meanwhile, IoU is calculated as the ratio of the shared overlapping region to the union of the predicted and reference masks. Elevated IoU scores indicate a superior volumetric concordance between the model's output and the ground truth. For the DSC and IoU formula:

$$DSC = \frac{1}{3}\sum_{c=1}^{3}\frac{2\times\sum_{i=1}^{N}p_i\times g_i+\varepsilon}{\sum_{i=1}^{N}p_i+\sum_{i=1}^{N}g_i+\varepsilon} \quad (2)$$

where:

p_i : the model's estimated output value for each individual pixel i
g_i : the actual ground truth label for every pixel i
N : the total summation of pixels contained within the 3D volume
ε : a smoothing constant $(1e-5)$ included to avoid zero-division errors
c : the target classes (1,2,3) denoting edema, non-enhancing tumor, and enhancing tumor respectively.

$$IoU = \frac{1}{3}\sum_{c=1}^{3}\frac{\sum_{i=1}^{N}(p_i\times g_i)+\varepsilon}{\sum_{i=1}^{N}(p_i+g_i-p_i\times g_i)+\varepsilon} \quad (3)$$

where:

p_i : the model's estimated output value for each individual pixel i
g_i : the actual ground truth label for every pixel i
N : the total summation of pixels contained within the 3D volume
ε : a smoothing constant $(1e-5)$ included to avoid zero-division errors
c : the target classes (1,2,3) denoting edema, non-enhancing tumor, and enhancing tumor respectively.

Equation (4) and Equation (5) show the sensitivity (recall) and specificity. Sensitivity evaluates the proposed model's ability to correctly classify true positive regions relative to the total actual positive areas, whereas specificity evaluates its performance in recognizing true negative regions among all negative instances.

$$\text{Sensitivity} = \frac{True\ Positive}{True\ Positive + False\ Negative} \quad (4)$$

$$\text{Specificity} = \frac{True\ Negative}{True\ Negative + False\ Positive} \quad (5)$$

3. Results and Discussion

The training was performed using a patch-based approach whereby the original image (resolution of 240 x 240 x 155) is preprocessed into sub-volumes (resolution of 160 x 160 x 16). An example of a sub-volume extracted from a 3D MRI image is given in Fig. 4.

Table 1 shows the experimental results for the training and validation phase whereby the U-Net has 16.318.307 parameters, and the 3D RSU U^2 Net+ has 51.490.695 (more than 3 times). In general, the best performance was given by 3D RSU U^2-Net+. Although it has a triple times parameter, the proposed network is more efficient than the standard U-Net during the training phase. Moreover, the training result shows that the 3D RSU U^2 Net+ performs better on SDL, Val SDL, DSC, and IoU

10.12928/ijio.v7i1.11576

metrics, achieved when the learning rate is 0.00001 and epochs is 35. We have experimented with other parameters such as the learning rate (lr) and epoch. However, the results for other values do not improve the model performance. Better Val SDL shows that the proposed network can handle imbalanced data better than the U-Net. Although the proposed model learns better than U-Net for all training metrics, it has a slightly decreased effectiveness during validation (Val DSC and Val IoU).

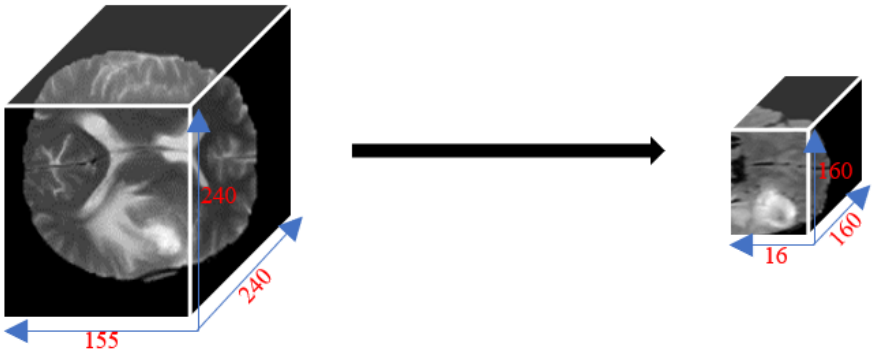


Fig. 4. A sample of sub-volume extracted from the original 3D MRI image

Table 2 and Table 3 show the experimental results for the testing phase. The proposed model has better DSC and IoU values indicating better segmentation results that measure the similarity and overlapping between the predicted areas and the ground truth. It achieves DSC of 0.7779 and IoU of 0.6439 which is better than U-Net. In classifying the three types of class (edema, non-enhancing tumor, and enhancing tumor), the proposed model has significantly better average sensitivity (0.0415 better) but slightly less average specificity (only 0.001 worse). Thus, the proposed model is better at segmenting correct areas according to the labels.

Table 1. The Results of SDL, DSC, IoU, and GPU Time for Training Models. Val means validation and min is abbreviation of minutes.

Model	lr	epoch	SDL	Val SDL	DSC	Val DSC	IoU	Val IoU	GPU Time (min)
U-Net	0.00001	35	0.0993	0.1357	0.8269	**0.8045**	0.7105	**0.6790**	433.6
3D RSU U2-Net+	0.00001	35	**0.0814**	**0.1320**	**0.8497**	0.8007	**0.7420**	0.6735	**397.7**
3D RSU U2-Net+	0.00001	13	0.1303	0.1693	0.7732	0.7536	0.6365	0.6147	-
3D RSU U2-Net+	0.0001	13	0.1309	0.1324	0.7815	0.7897	0.6487	0.6586	-
3D RSU U2-Net+	0.001	13	0.1632	0.1526	0.7415	0.7395	0.5978	0.5944	-

Table 2. The Results of Testing Models for DSC and IoU Metrics

Model	DSC	IoU
U-Net	0.7768	0.6424
3D RSU U^2 Net+	**0.7779**	**0.6439**

Table 3. The Result of Sensitivity and Specificity for Class Labels

Label	U-Net		3D RSU U^2 Net+	
	Sensitivity	Specificity	Sensitivity	Specificity
Edema	0.8566	0.9841	0.8690	0.9851
Enhancing Tumor	0.7262	0.9970	0.7991	0.9956
Non-enhancing Tumor	0.5551	0.9951	0.5942	0.9927
mean	0.7126	**0.9921**	**0.7541**	0.9911

Table 4. Sensitivity dan Specificity

Label	Patient	Sensitivity	Specificity
Edema		0.9610	0.9891
Enhancing Tumor	**10**	0.7511	0.9761
Non-enhancing Tumor		0.5666	0.9956
Edema		0.7527	0.9785
Enhancing Tumor	**57**	0.9461	0.9961
Non-enhancing Tumor		0.5292	0.9783
Edema		0.8398	0.9871
Enhancing Tumor	**100**	0.9206	0.9791
Non-enhancing Tumor		0.7742	0.9935
Edema		0.8240	0.9782
Enhancing Tumor	**478**	0.9558	0.9798
Non-enhancing Tumor		0.9379	0.9972
mean		0.8285	0.9857

Table 4 shows the sensitivity (0.8285 on average) and specificity (0.9857 on average) for patients 10, 57, 100, and 478. The sensitivity and specificity of the proposed model attained more than 96%. However, it suffers from sensitivity to recognize enhancing tumor (0.7511) and non-enhancing tumor areas (only 0.5666). The segmentation result of patient 57 has the best sensitivity and specificity on enhancing tumor while less sensitivity on non-enhancing tumor. For patient 100, the results show high specificity for all classes (edema, enhancing tumor, and non-enhancing tumor) i.e. higher than 92%, but slightly less sensitivity on edema (0.8398) and non-enhancing tumor (0.7742). In contrast, the segmentation results for patient 478 show good performance for all classes.

The upper part of Fig. 5 shows that the proposed model displays visual similarities between predicted and ground-truth areas while the lower part displays the 2D slice with three different viewpoints, namely sagittal plane (left-to-right), coronal plane (front-to-back), and transverse plane (top-to-bottom). The red color is used to highlight edema, blue is used to highlight enhancing tumor, and green is used to highlight non-enhancing tumor. For patient 10, the predicted segmentation areas are very similar to the ground-truth areas. For patients 57, 100, and 478 there are good visual similarities for edema and enhancing tumor areas, but more non-enhancing tumor areas than the ground-truth. Moreover, there are a few spots in edema areas for patient 100 but not detected by the model. Similar segmentation results can also be seen in the sagittal, coronal, and transverse planes whereby most of the edema areas can be segmented correctly according to the ground-truth images.

In general, these results indicate that the model is effective for detecting edema and enhancing tumor areas, but has limitations in recognizing non-enhancing tumors properly. This difficulty arises from non-enhancing tumor poorly defined boundaries, intensity similarity to normal tissue, and small patchy distributions. While 3D RSU U^2 Net+ shows improved performance across all classes (particularly +7.3% for Enhancing Tumor), its modest +3.9% gain for non-enhancing tumor suggests architectural enhancements alone cannot fully address these inherent challenges. The high specificity (>0.99) across all classes indicates both models effectively avoid false positives in healthy tissue. However, non-enhancing tumor marginally lower specificity (0.9951 vs 0.9970 for Enhancing Tumor) reveals persistent difficulties in ambiguous border regions.

10.12928/ijio.v7i1.11576

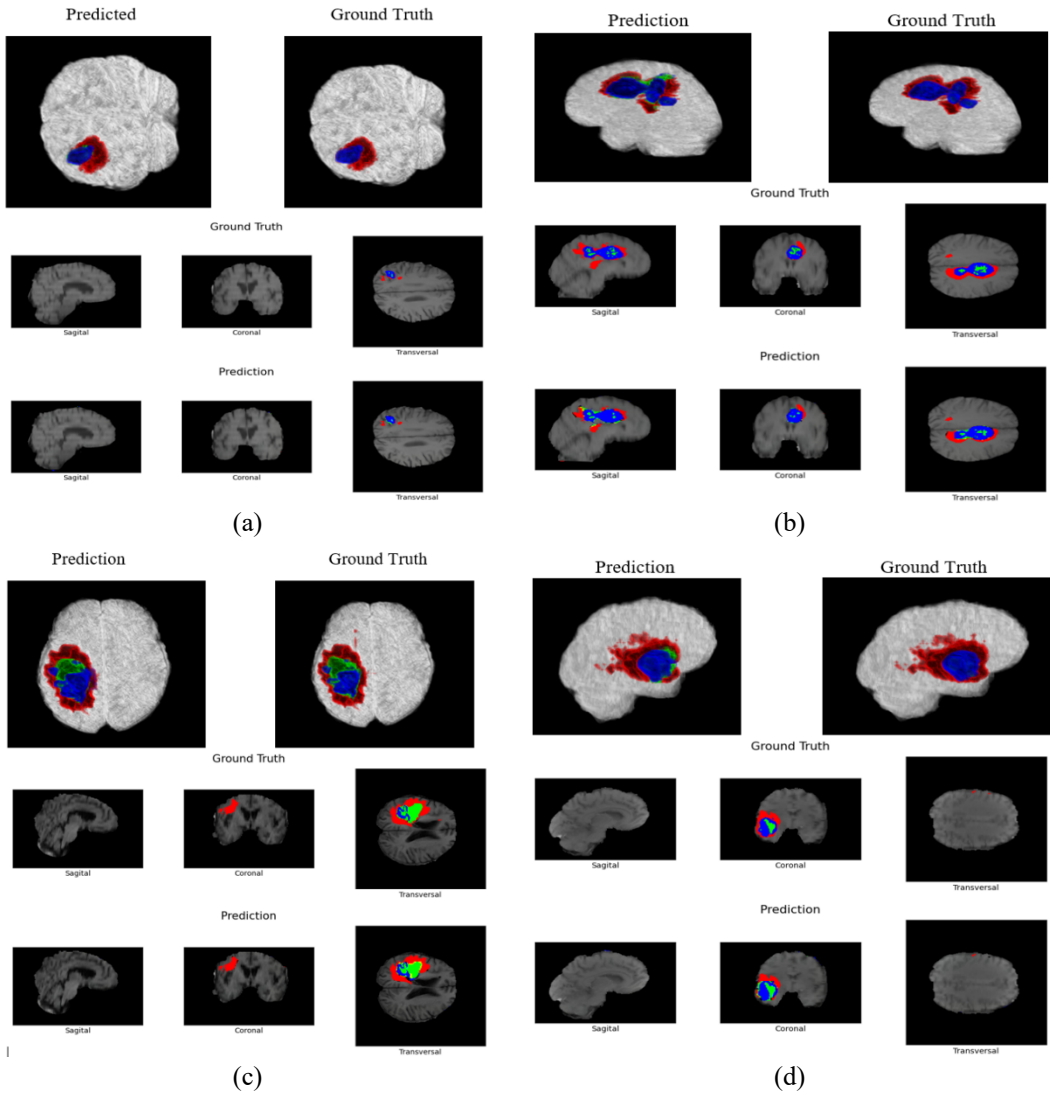


Fig. 5. Visualization of 3D Segmentation results: (a) patient 10, (b) patient 57, (c) patient 100, and (d) patient 478.

Table 5. Dataset Class Proportions

Label	Training	Validation	Testing
Edema	55.01%	63.97%	59.04%
Enhancing Tumor	23.51%	20.22%	18.05%
Non-enhancing Tumor	21.48%	15.81%	22.91%

Furthermore, the model's lower performance in segmenting non-enhancing tumors can be attributed to the imbalanced class representation across the training, validation, and testing datasets, as shown in Table 5. Non-enhancing tumors represent a smaller proportion of the data, particularly in the validation and testing sets (21.48% in training, 15.81% in validation, and 22.91% in testing). This imbalance is a factor in the model's struggle to accurately segment non-enhancing tumor regions. Additionally, the Soft Dice Loss function, which averages across all classes, tends to be less sensitive to class imbalances in multi-label segmentation tasks. In multi-label segmentation, where multiple classes are predicted simultaneously, regions with fewer non-enhancing tumor pixels contribute minimally to the overall loss, reducing the model's ability to learn their characteristics effectively. Moreover, the visual similarity between non-enhancing tumors and healthy tissues further complicates the segmentation task, making it harder for the model to distinguish between these regions.

To address these shortcomings, incorporating a more adaptive loss function such as Focal Loss could help mitigate the impact of class imbalance by assigning greater importance to challenging classes [26], [27]. In this context, using class-specific weighting strategies or hybrid loss functions can also further enhance model performance. Additionally, applying targeted data augmentation techniques could improve the model's robustness [28] Post processing techniques like conditional random fields (CRF) may further refine segmentation boundaries by incorporating spatial context [29].

4. Conclusion

We presented an architecture of 3D RSU U^2-Net for semantic segmentation of brain tumors from MRI images. The architecture combines ReSidual U-Block and residual custom dilated blocks into a nested U-Net structure. The model can learn global context feature maps more efficiently than U-Net despite having 3 times more parameters. The experimental results show the proposed model improves the U-Net model. It achieved a good dice score coefficient of 78%, sensitivity of 82.8%, and specificity of 98.6%, but low for IoU (only 64.4%). Acceptable visual similarities are shown for recognizing edema and enhancing tumor areas, but less accurately for non-enhancing tumor areas. This research also contributes to the field of medical image segmentation by demonstrating the advantages of combining residual blocks and custom dilated convolutions, providing a basis for future innovations in automated tumor segmentation. Future work may involve the integration of hybrid optimization techniques and more sophisticated attention mechanisms to further enhance the segmentation accuracy. Moreover, the model should be evaluated on different target regions of body parts available in the medical segmentation Decathlon dataset or other medical datasets that could provide further insights.

Author Contribution: All authors contributed equally to the main contributor to this paper. All authors read and approved the final paper.

Funding: This research received no external funding.

Acknowledgment: The authors would like to thank all parties who provided administrative and technical support during the completion of this study. Special appreciation is extended to the institutions and colleagues who facilitated access to computational resources and provided technical assistance during the model implementation and experimental evaluation. The authors also acknowledge the providers of the Medical Segmentation Decathlon dataset for making the data publicly available, which greatly supported this research.

Conflicts of Interest: The authors declare no conflict of interest.

References

[1] H. Ohgaki and P. Kleihues, "Population-Based Studies on Incidence, Survival Rates, and Genetic Alterations in Astrocytic and Oligodendroglial Gliomas," *J. Neuropathol. Exp. Neurol.*, vol. 64, no. 6, pp. 479–489, Jun. 2005, doi: 10.1093/jnen/64.6.479.

[2] M. N. Putri, I. Katili, A. Hariri, T. A. Budiarti, and G. M. Wibowo, "Perbandingan Pegukuran Volume Tumor Brain MRI Menggunakan Teknik Manual Dan Metode Active Contour," *J. Imejing Diagnostik*, vol. 7, no. 2, pp. 94–97, Jul. 2021, doi: 10.31983/jimed.v7i2.7474.

[3] M. Ghaffari, A. Sowmya, and R. Oliver, "Automated Brain Tumor Segmentation Using Multimodal Brain Scans: A Survey Based on Models Submitted to the BraTS 2012–2018 Challenges," *IEEE Rev. Biomed. Eng.*, vol. 13, pp. 156–168, 2020, doi: 10.1109/RBME.2019.2946868.

[4] C. M. Umarani, S. G. Gollagi, S. Allagi, K. Sambrekar, and S. B. Ankali, "Advancements in deep learning techniques for brain tumor segmentation: A survey," *Informatics Med. Unlocked*, vol. 50, p. 101576, 2024, doi: 10.1016/j.imu.2024.101576.

[5] P. Wang and A. C. S. Chung, "Relax and focus on brain tumor segmentation," *Med. Image Anal.*, vol. 75, p. 102259, Jan. 2022, doi: 10.1016/j.media.2021.102259.

[6] A. Barata, H. Akbar, M. Pilliang, and A. Nasihin, "Graphology analysis for detecting hexaco personality and character through handwriting images by using convolutional neural networks and particle swarm optimization methods," *Int. J. Ind. Optim.*, vol. 3, no. 2, pp. 110–120, Sep. 2022, doi: 10.12928/ijio.v3i2.6242.

[7] L. ZongRen, W. Silamu, W. Yuzhen, and W. Zhe, "DenseTrans: Multimodal Brain Tumor Segmentation Using Swin Transformer," *IEEE Access*, vol. 11, pp. 42895–42908, 2023, doi: 10.1109/ACCESS.2023.3272055.

[8] Z. Tabatabaei, A. Colomer, K. Engan, J. Oliver, and V. Naranjo, "Residual block Convolutional Auto Encoder in Content-Based Medical Image Retrieval," in *2022 IEEE 14th Image, Video, and Multidimensional Signal Processing Workshop (IVMSP)*, Jun. 2022, pp. 1–5, doi: 10.1109/IVMSP54334.2022.9816325.

[9] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation," in *2016 Fourth International Conference on 3D Vision (3DV)*, Oct. 2016, pp. 565–571, doi: 10.1109/3DV.2016.79.

[10] F. Isensee, P. Kickingereder, W. Wick, M. Bendszus, and K. H. Maier-Hein, "No New-Net," 2019, pp. 234–244. https://doi.org/10.48550/arXiv.1809.10483

[11] D. V. Gore and V. Deshpande, "Comparative Study of various techniques using Deep Learning for Brain Tumor Detection," in *2020 International Conference for Emerging Technology (INCET)*, Jun. 2020, pp. 1–4, doi: 10.1109/INCET49848.2020.9154030.

[12] M. Kolařík, R. Burget, V. Uher, K. Říha, and M. K. Dutta, "Optimized High Resolution 3D Dense-U-Net Network for Brain and Spine Segmentation," *Appl. Sci.*, vol. 9, no. 3, p. 404, Jan. 2019, doi: 10.3390/app9030404.

[13] N. D. Gai, "Highly Efficient and Accurate Deep Learning–Based Classification of MRI Contrast on a CPU and GPU," *J. Digit. Imaging*, vol. 35, no. 3, pp. 482–495, Jun. 2022, doi: 10.1007/s10278-022-00583-1.

[14] M. Havaei *et al.*, "Brain tumor segmentation with Deep Neural Networks," *Med. Image Anal.*, vol. 35, pp. 18–31, Jan. 2017, doi: 10.1016/j.media.2016.05.004.

[15] E. BUBER and B. DIRI, "Performance Analysis and CPU vs GPU Comparison for Deep Learning," in *2018 6th International Conference on Control Engineering & Information Technology (CEIT)*, Oct. 2018, pp. 1–6, doi: 10.1109/CEIT.2018.8751930.

[16] B. H. Menze *et al.*, "The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)," *IEEE Trans. Med. Imaging*, vol. 34, no. 10, pp. 1993–2024, Oct. 2015, doi: 10.1109/TMI.2014.2377694.

[17] S. Bakas *et al.*, "Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features," *Sci. Data*, vol. 4, no. 1, p. 170117, Sep. 2017, doi: 10.1038/sdata.2017.117.

[18] X. Li, X. Fang, G. Yang, S. Su, L. Zhu, and Z. Yu, "TransU^2-Net: An Effective Medical Image Segmentation Framework Based on Transformer and U^2-Net," *IEEE J. Transl. Eng. Heal. Med.*, vol. 11, pp. 441–450, 2023, doi: 10.1109/JTEHM.2023.3289990.

[19] A. Myronenko, "3D MRI Brain Tumor Segmentation Using Autoencoder Regularization," 2019, pp. 311–320. https://doi.org/10.1007/978-3-030-11726-9_28

[20] M. Antonelli, "The Medical Segmentation Decathlon," *Nat. Commun.*, vol. 13, no. 1, p. 4128, Jul. 2022, doi: 10.1038/s41467-022-30695-9.

[21] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical

image segmentation," *Int. Conf. Med. Image Computing and Computer-Assisted Intervention*, 2015, doi: 10.1007/978-3-319-24574-4_28.

[22] S. F. Ahmed, F. S. Rahman, T. Tabassum, and M. T. I. Bhuiyan, "3D U-Net: Fully Convolutional Neural Network for Automatic Brain Tumor Segmentation," in *2019 22nd International Conference on Computer and Information Technology (ICCIT)*, Dec. 2019, pp. 1–6, doi: 10.1109/ICCIT48885.2019.9038237.

[23] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2016, pp. 770–778, doi: 10.1109/CVPR.2016.90.

[24] X. Qin, Z. Zhang, C. Huang, M. Dehghan, O. R. Zaiane, and M. Jagersand, "U2-Net: Going deeper with nested U-structure for salient object detection," *Pattern Recognit.*, vol. 106, p. 107404, Oct. 2020, doi: 10.1016/j.patcog.2020.107404.

[25] Z. Zhou, M. M. Rahman Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: A Nested U-Net Architecture for Medical Image Segmentation," 2018, pp. 3–11. https://doi.org/10.1007/978-3-030-00889-5_1

[26] N. Abraham and N. M. Khan, "A Novel Focal Tversky Loss Function With Improved Attention U-Net for Lesion Segmentation," in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, Apr. 2019, pp. 683–687, doi: 10.1109/ISBI.2019.8759329.

[27] X. Zheng, Z. Yu, L. Chen, F. Zhu, and S. Wang, "Multi-label Contrastive Focal Loss for Pedestrian Attribute Recognition," in *2020 25th International Conference on Pattern Recognition (ICPR)*, Jan. 2021, pp. 7349–7356, doi: 10.1109/ICPR48806.2021.9411959.

[28] J. Xu, M. Li, and Z. Zhu, "Automatic Data Augmentation for 3D Medical Image Segmentation," 2020, pp. 378–387. https://doi.org/10.1007/978-3-030-59710-8_37

[29] S. Chen, Z. Sedghi Gamechi, F. Dubost, G. van Tulder, and M. de Bruijne, "An end-to-end approach to segmentation in medical images with CNN and posterior-CRF," *Med. Image Anal.*, vol. 76, p. 102311, Feb. 2022, doi: 10.1016/j.media.2021.102311.