JTE HCM-UTE
ISSN: 2615-9740

JOURNAL OF TECHNICAL EDUCATION SCIENCE
Ho Chi Minh City University of Technology and Engineering
Website: https://jte.edu.vn
Email: jte@hcmute.edu.vn

SepU-Net MRI Segmentation Algorithm Using Depthwise Separable Convolution and Pointwise Convolution Integrated U-Net

Phong Mai Hong*, Lam Mai Thanh, Lam Nguyen Ngo
Ho Chi Minh City University of Technology and Engineering, Vietnam

*Corresponding author. *Email: 20119192@student.hcmute.edu.vn*

ARTICLE INFO

Received: 25/04/2025
Revised: 20/06/2025
Accepted: 05/01/2026
Published: 28/02/2026

KEYWORDS

Depthwise separable convolution;
Medical image segmentation;
Lightweight neural network;
Computational efficiency;
Brain tumor segmentation;
Lightweight architecture.

ABSTRACT

Accurate segmentation of brain tumors in MRI remains challenging due to the computational demands of conventional deep learning models. We present SepU-Net, a lightweight convolutional neural network that employs Depthwise Separable Convolutions and efficiently channel attention to reduce model complexity by 69.3% compared to standard U-Net (2.39M vs. 7.76M parameters).Evaluated on both BraTS2020 and BraTS2021, SepU-Net achieves high accuracy (0.9938 and 0.994), mean IoU (0.842 and 0.8318), and Dice coefficients (0.846 and 0.8325), with only minor declines on the more heterogeneous BraTS2021 dataset. Notably, SepU-Net delivers a 32.34% improvement in Tumor Core segmentation and a 10.14% gain in Enhancing Tumor segmentation over U-Net, while maintaining strong precision (0.994/0.9942) and sensitivity (0.9921/0.9915) across datasets. SepU-Net requires only 8.3 GFLOPs per inference, 65% fewer than U-Net and operates efficiently on embedded devices with a memory footprint of 1.4GB. These results validate its ability to balance accuracy and efficiency, enabling real-time segmentation in clinical settings. Future work will integrate attention mechanisms and extend the architecture to 3D for enhanced spatial context.

Doi: https://doi.org/10.54644/jte.2026.1890

1. Introduction

Accurate segmentation of brain tumors in magnetic resonance imaging (MRI) is crucial for diagnosis, treatment planning, and monitoring patient progress [1], [2]. Manual delineation by specialists is not only time-consuming but also subject to significant inter-observer variability, especially when tumor boundaries are complex or ambiguous. As a result, there is an increasing demand for automated, reliable, and consistent segmentation solutions [3], [4].

Convolutional neural networks (CNNs) [5], particularly the U-Net architecture, have become the standard for medical image segmentation due to their ability to extract multi-scale features and preserve spatial information through skip connections [6], [7]. However, conventional U-Net models are parameter-heavy and require substantial computational resources and memory, posing challenges for deployment on resource-constrained devices such as mobile workstations or embedded systems in clinical settings [4], [8], [9].

To address these limitations, recent research has focused on developing lightweight U-Net variants that balance segmentation accuracy with computational efficiency. Among these, SepU-Net stands out by replacing all standard convolutions with Depthwise Separable Convolutions and integrating the Efficient Channel Attention (ECA) mechanism [10]-[12], substantially reducing the number of parameters and computational cost while maintaining competitive segmentation performance. Additionally, SepU-Net optimizes skip connections and regularization techniques, further improving model stability and generalization [9], [13]. Unlike previous studies that typically evaluate on a single dataset, this work extends the assessment of SepU-Net to both the BraTS2020 [14] and the more heterogeneous BraTS2021 [15], [16] datasets, which better reflect real-world clinical diversity. SepU-

Net is systematically compared with the conventional U-Net and other lightweight models in terms of segmentation accuracy, computational efficiency, training speed, and memory usage [7], [13]. In this study, the Result show that SepU-Net maintains high performance across major tumor regions with strong Dice scores, while reducing parameters by 69.3% and GFLOPs by 58.5% compared to the original U-Net [16], confirming its practical potential for deployment in resource-limited clinical environments

1.1. BraTS 2020 Data and Pre-processing

This study utilizes the BraTS2020 dataset for brain tumor segmentation, specifically focusing on two selected MRI modalities as input channels and their corresponding manual segmentation labels as ground truth [14], [17]. These labels delineate three crucial tumor sub-regions: Necrotic and Non-Enhancing Tumor Core (NCR/NET), Peritumoral Edema (ED), and Enhancing Tumor (ET) [18].

The dataset is provided in NIfTI format (.nii.gz), containing 3D image volumes. We use the NiBabel library to convert these 3D images into 2D slices for model input [19], [20]. For a deeper understanding of how NiBabel works, let's look at an example in Figure 1. As shown in Figure 2, each subject contains four MRI modalities and a ground truth segmentation mask, which are extracted into separate 2D datasets. Figure 3 presents examples of FLAIR images and corresponding ground truth masks after preprocessing.

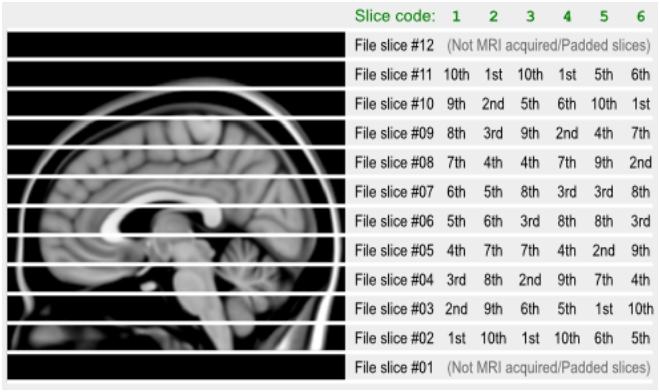


Figure 1. *Illustrate the data preprocessing process using NiBabel, including the steps of intensity normalization, 3D→2D slicing, and histogram equalization [20].*

Figure 1 shows a table with MRI slices, labeled from File slice #12 to File slice #01. Each row represents a slice, with a Slice code indicating its position (e.g., "1st", "2nd", etc.). Some slices are labeled as "Not MRI acquired" or padded slices, meaning they are not actual MRI data but filler data [20].

For example, below

$Let: S_{start} = \#02;\ S_{end} = \#11$ represent the **first and last slice indices** (typically along the Z-axis).

$$\rightarrow N = S_{end} - S_{start} = 10$$

Δt is slice duration(time taken to acquire one slice) ; t_0 *is* offset time(time at which acquisition starts)

N is the number of slices along the Z-axis.

MRI acquisition order (even/odd interleave):

MRI slices may be acquired in interleaved order to reduce interference. The acquisition index K_j for slice j is computed as [21] (1):

$$K_j = \begin{cases} \frac{j}{2}, j\ even \\ \frac{N+1+j}{2}, j\ odd \end{cases} \quad (1)$$

This ordering helps preserve the temporal consistency of slice acquisition.

Slice acquisition time [21]:

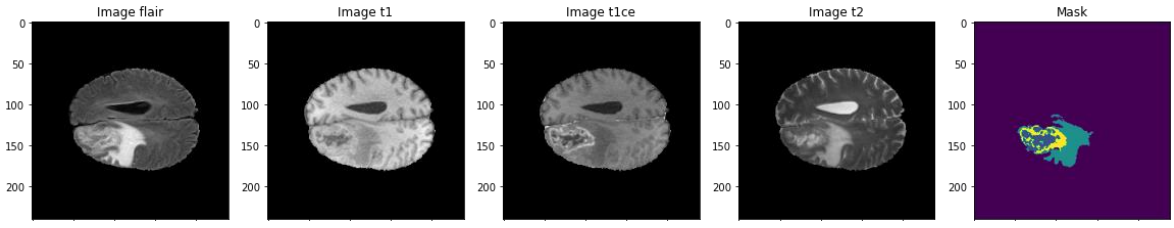
$$t_j = t_0 + (k_j - 1)\,\Delta t \quad (2)$$

Volume shape & axis:

dim= (4,240,240,155)

→ trimmed volume is $240 \times 240 \times N$

→ 4 MRI modalities$(T1, T1ce, T2, FLAIR)$

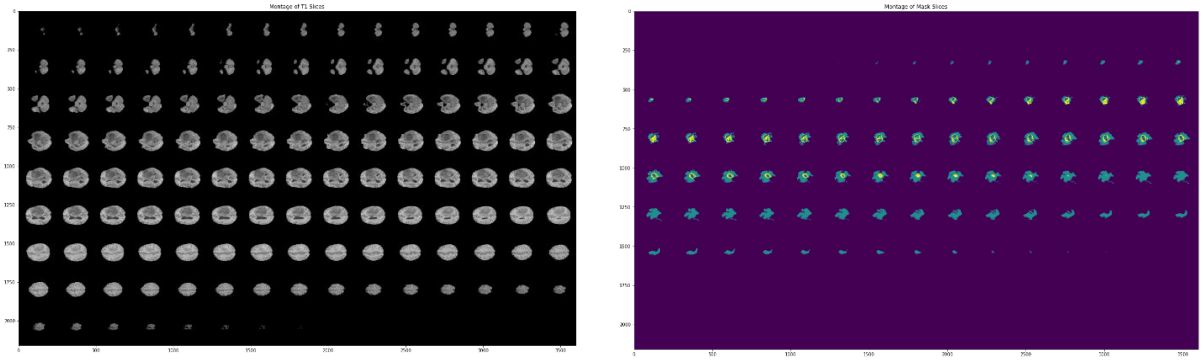


(a) (b) (c) (d) (e)

Figure 2. Different modalities MRI and a Mask channel of MRI's BraTS2020; (a) Image FLAIR; (b) Image T1; (c) Image T1ce ; (d) Image T2 ;(e) Image Mask.

Figure 2 provides four MRI modalities and a ground truth segmentation mask for brain tumor analysis from BraTS2020 datasets [14]: (a) FLAIR is a T2-weighted pulse sequence modified to suppress cerebrospinal fluid signals, thereby highlighting peritumoral edema as hyperintense regions. (b)T1-weighted images deliver soft-tissue (c) Contrast-enhanced T1-weighted (T1ce) images acquired after gadolinium injection accentuate leaky vasculature, enabling precise segmentation of the enhancing tumor sub-region (ET). (d) Standard T2-weighted images depict fluid and pathological changes as hyperintense signals, supporting edema visualization but with lower specificity compared to FLAIR.

We extracted all the 2D MRI data from the 3D dataset for BraTS 2020. For this study, two modalities (FLAIR and T1ce) used for training and testing were selected as input channels, along with the ground truth segmentation. The Figure 3 presents all slices from an example FLAIR modality of the dataset; the remaining modalities are displayed in the same way.



a) Total FLAIR slices are pre-processed. *b) Total Ground truth slices are pre-processed.*

Figure 3. *Two sets of FLAIR and Ground Truth in a Brain.*

1.2. The main Architecture of SepU-Net

Based on the U-Net model [7], [22], Depthwise Separable Convolutions [23], and Pointwise Convolutions [24], for MRI image segmentation, we suggest a novel architecture termed SepU-Net, which has brought significant improvements in speed and complexity (the proposed model has a considerably lower number of parameters compared to equivalent models), while also enhancing the overall accuracy of the model. This section begins by providing a detailed explanation of the Depthwise Separable Convolutions described in Figure 4, which are used to create the SepU-Net architecture and reduce the number of parameters in the model. We will go into detail about the three primary parts of the SepU-Net architecture, which are Down Block, Up Block, and Bottleneck Block. Finally, the SepU-Net architecture's specifics will be discussed.

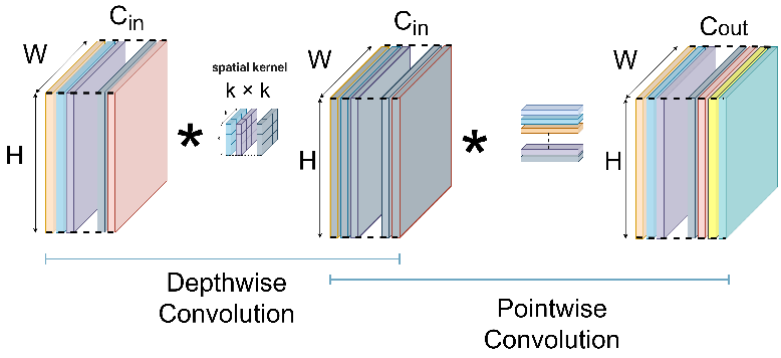


Figure 4. *Description of Depthwise Separable Convolution.*

Depthwise Separable Convolution is an optimal operation that minimizes the number of parameters while preserving comparable performance [23], [24]. The number of filters in a traditional convolution is equal to the number of output feature map channels, and the depth of a filter is comparable to the number of input channels. As a result, the input channels and spatial information are mixed to create each output element. However, as Figure 4 shows, Depthwise Separable Convolution is made up of two architectures: Depthwise Convolution and Pointwise Convolution. A feature map of the same size as the input is produced by Depthwise Convolution, which convolves each input channel independently with its matching filter. To combine the unique information retrieved from the final feature map, Pointwise Convolution is carried out after Depthwise Convolution is finished. Pointwise Convolution is an operation that is identical to a regular convolution with a filter size of 1×1 [25].

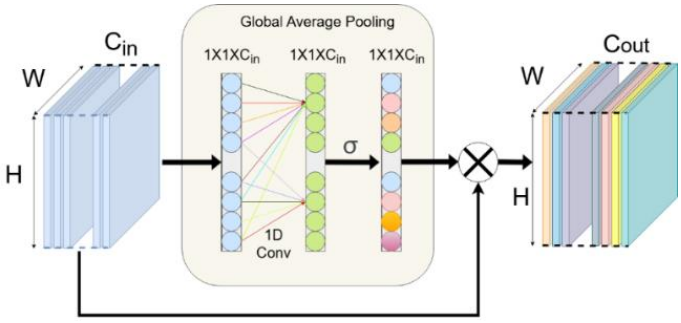


Figure 5. *Overview of the Efficient Channel Attention (ECA) Block.*

Efficient Channel Attention (ECA) block is an extremely lightweight channel-attention module that automatically adjusts the importance of each feature map without significantly increasing computational complexity. Specifically, ECA performs global average pooling and applies a 1D convolution to capture local inter-channel relationships. This allows the model to adaptively emphasize important channels, improving the model's ability to segment complex tumor regions while maintaining computational efficiency. Unlike other attention mechanisms such as Squeeze-and-Excitation (SE) [26], [27] or

Convolutional Block Attention Module (CBAM) [28], the ECA does not reduce the channel dimensions, preserving the full capacity of the feature maps and making it an ideal choice for lightweight models like SepU-Net. First, it applies global average pooling to the input tensor to produce a channel descriptor vector; next, it uses a one-dimensional convolution with kernel size k, which is automatically determined based on the number of channels, to model interactions between each channel and its k neighboring channels; finally, it applies a sigmoid activation to generate attention weights that amplify or attenuate each channel's response. By restricting connections to only k local channels and avoiding any reduction in channel dimensionality, ECA preserves a direct correspondence between inputs and attention weights while incurring minimal parameter overhead, making it ideal for integration into lightweight CNN architectures (Figure 5) [29], [30].

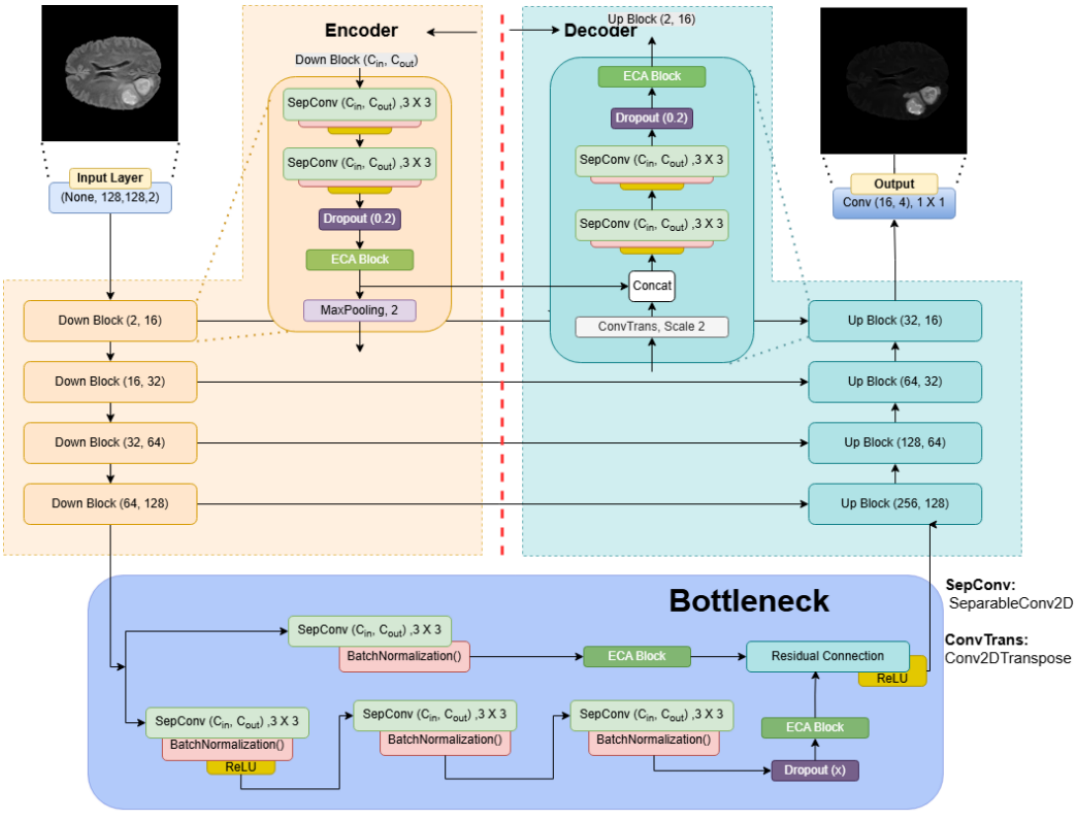


Figure 6. *The SepU-Net architecture consists of three main building blocks: the Down Block, the Bottleneck Block, and the Up Block.*

By replacing traditional convolutions with Depthwise Separable Convolutions, SepU-Net reduces the number of parameters by 69.3%, making it highly efficient in terms of both memory usage and computational cost, which is essential for real-time segmentation in resource-limited environments. The architecture consists of three main building blocks: Down Block, Bottleneck Block, and Up Block (show as Figure 6).

Down Block: Each Down Block contains two Depthwise Separable Convolutions (kernel 3×3, padding='same'), each followed by a ReLU activation function [31], then Dropout layer (rate = 0.2) to reduce overfitting [32]. and finally, a MaxPooling2D operation (2×2) to halve the spatial resolution [33]. The ECA module recalibrates channel responses immediately after these convolutions. A 2×2 max-pooling layer then halves the spatial dimensions, and the pre-pooled tensor is stored for skip connections, preserving high-resolution details for the decoding path [34], [35].

Bottleneck Block: The Bottleneck Block adopts a residual dual-branch design: the main branch comprises three separable convolutions interleaved with Batch Normalization and ReLU (post-first normalization) plus dropout (rate=0.2), while the shortcut branch performs a single separable

convolution and Batch Normalization. ECA is applied to the main branch output before element-wise addition with the shortcut branch, and a final ReLU activation refines the fused features, ensuring effective information flow despite network slimming.

Up Block: Each Up Block begins with a Conv2DTranspose layer (kernel 2×2, stride 2) for upsampling [35]. concatenates with the corresponding encoder feature map via skip connection [15], [22], and applies two Depthwise Separable Convolutions (kernel 3×3, padding='same', ReLU). The ECA module at the end of this block refines the concatenated feature map by emphasizing salient channels, completing the decoder stage with minimal latency. SepU-Net processes 128×128-pixel two-channel inputs (T1ce and FLAIR) and predicts multi-class segmentation masks for Necrotic Core, edema, and enhancing tumor. It is trained with the Adam optimizer (learning rate=0.001) using categorical cross-entropy loss and dropout regularization [14]. Segmentation performance is quantified by accuracy, means Intersection-over-Union (mIoU), and Dice coefficient.

2. Experiments and Results

Having established the SepU-Net architecture with its lightweight design principles, we now evaluate its performance through comprehensive experiments on the BraTS2020 and BraTS2021 datasets.

Database: The study utilizes the BraTS2020 dataset [14], [18], which contains 3D multimodal MRI scans divided into 221 training volumes, 74 validation volumes, and 74 test volumes. These 3D volumes are converted into 2D axial slices, resulting in 285,975 slices for training, 11,500 for validation, and 11,500 for testing. This preprocessing step ensures compatibility with standard 2D convolutional neural network architectures while preserving spatial and contextual tumor features.

Training Produce: In this experiment, we focused on designing a SepU-Net architecture from an existing U-Net architecture [36], which excellently won the gold medal at BraTS2020 on Kaggle. In the original SepU-Net architecture, we achieved positive improvements, and it can be integrated into low-configuration devices; details can be seen in Table 1 and Table 2.

Table 1. *Comparative Performance and Computational Efficiency of U-Net and SepU-Net on BraTS2020 Dataset.*

Metric/Parameter	**U-Net** [36]	**SepU-Net**	**Relative Difference**
Parameters	7,759,908	2,385,110	↓ 69.3%
GFLOPs	~20	~8,3	↓ 58.5%
Test Loss	0.0179	0.0183	↑ 2.23%
Accuracy	0.9941	0.9938	↓ 0.03%
Mean IoU	0.8312	0.8420	↓ 1.3%
Dice Coefficient	0.6512	0.8460	↑ 29.91%
Dice - Necrotic Core	0.6559	0.8680	↑ 32.34%
Dice - Edema	0.7894	0.8212	↓ 4.03%
Dice - Enhancing Tumor	0.7457	0.8513	↑ 10.14%
Precision	0.9944	0.9940	↓ 0.04%
Sensitivity	0.9927	0.9921	↓ 0.06%
Specificity	0.9981	0.9979	↓ 0.02%
Inference Time (per epoch)	28 s	28 s	~ 0%

Based on the results in Table 1, SepU-Net demonstrates substantial improvements in computational efficiency compared to the conventional U-Net architecture. Specifically, SepU-Net achieves a 69.3% reduction in trainable parameters (2,385,110 vs. 7,759,908) and requires 58.5% fewer GFLOPs per

inference (8.3 vs. 20), making it highly suitable for deployment on resource-constrained devices. In terms of segmentation performance, SepU-Net delivers a 29.91% higher overall Dice Coefficient (0.8460 vs. 0.6512) and a 32.34% improvement for Necrotic Core segmentation (Dice: 0.8680 vs. 0.6559). Notably, SepU-Net achieves a 10.14% increase in Dice for the Enhancing Tumor region (0.8513 vs. 0.7457) and a 4.03% increase for Edema segmentation (0.8212 vs. 0.7894). Both models achieve comparable accuracy (0.9941 for U-Net, 0.9938 for SepU-Net), mean IoU, precision, sensitivity, and specificity, with only marginal differences (all less than 0.1%). The training time per epoch is nearly identical (28 seconds for both models), further confirming SepU-Net's efficiency.

To verify the stability and generalizability of SepU-Net, we conducted additional evaluation on the BraTS2021 dataset [16], which was provided by the Radiological Society of North America (RSNA), American Society of Neuroradiology (ASNR), and Medical Image Computing and Computer Assisted Interventions (MICCAI). Note that, unlike BraTS2020, the BraTS2021 dataset includes only testing data and does not include training data. This distinction may partially explain the slight decrease in segmentation performance of SepU-Net when compared to the results on BraTS2020, where the model was evaluated on both the training and testing data.

Table 2. *Comparative Performance and Computational Efficiency of SepU-Net on BraTS2020 and BraTS2021 Datasets with 11,500 slices for testing and evaluation.*

Group	Metric/Parameter	SepU-Net (BraTS 2020)	SepU-Net (BraTS 2021)
General Metrics	GFLOPs	~8.3	~8.6
	Test Loss	0.0183	0.023
Segmentation Performance	Accuracy	0.9938	0.994
	Mean IoU	0.842	0.8318
	Dice Coefficient	0.846	0.8325
Other Metrics	Precision	0.994	0.9942
	Sensitivity	0.9921	0.9915
	Specificity	0.9979	0.988

Table 2 presents the segmentation performance and computational efficiency of SepU-Net on both the BraTS2020 and BraTS2021 datasets, using 11,500 slices for testing and evaluation. In terms of computation, the GFLOPs of SepU-Net increased slightly from 8.3 (BraTS2020) to 8.6 (BraTS2021), indicating that the model maintains high computational efficiency despite a minor increase in computational demand. Although Test Loss increases slightly from 0.0183 (BraTS2020) to 0.023 (BraTS2021), this reflects the higher complexity and increased variability of the BraTS2021 dataset.

Regarding segmentation performance, Accuracy remains very high on both datasets, with 0.9938 on BraTS2020 and 0.994 on BraTS2021, showing that SepU-Net maintains stable performance when handling a more complex dataset. However, segmentation metrics such as Mean IoU and Dice Coefficient show a slight decrease when moving from BraTS2020 to BraTS2021, with Mean IoU dropping from 0.842 to 0.8318, and Dice Coefficient decreasing from 0.846 to 0.8325. These changes reflect the higher difficulty and diversity of BraTS2021 compared to BraTS2020, resulting in a small decrease in segmentation performance. Despite these changes, Precision and Sensitivity remain high, with Precision at 0.994 and Sensitivity at 0.9921 (BraTS2020) and 0.9915 (BraTS2021), showing that the model continues to accurately classify and detect tumors across both datasets. Specificity drops slightly from 0.9979 (BraTS2020) to 0.988 (BraTS2021), reflecting a small decrease in the model's ability to classify true negatives when dealing with the more complex BraTS2021 dataset.

Summary: Based on the results presented in Table 2 and Table 3 SepU-Net stands out as a highly efficient and stable model for medical image segmentation, offering significant improvements over the traditional U-Net architecture. The model achieves remarkable computational efficiency, drastically

reducing the number of trainable parameters and the required computational resources (GFLOPs), making it highly suitable for resource-constrained environments. Despite these reductions, SepU-Net shows superior segmentation performance, especially in key tumor regions, such as the Necrotic Core, Enhancing Tumor, and Edema, providing notable improvements over U-Net. Additionally, SepU-Net maintains a stable and reliable performance across both BraTS2020 and BraTS2021 datasets, which vary in complexity. On both datasets, SepU-Net demonstrates strong segmentation results with high accuracy, precision, and sensitivity, showing its robustness in handling different levels of dataset difficulty. Even with slight decreases in some metrics when transitioning to the more complex BraTS2021, the model remains highly effective, proving its stability and versatility across diverse and challenging datasets. Overall, SepU-Net offers a highly efficient and stable solution for segmentation tasks, with clear advantages over U-Net in terms of computational cost and consistent performance across multiple datasets.

Table 3. *Comparison of SepU-Net and Lightweight Architectures on BraTS2020.*

Model	Parameters	GFLOP	Dice (ET)	Dice (NCR/NET)	Dice (TC)	Main Advantages	Limitations
SepU-Net (Proposed)	~2.39M	~8.3	0.8513	0.8420	0.8460	Lightest, fast, mobile/edge-friendly	Lower performance on ET/TC vs. SGEResU-Net, EfficientNetB4
SGEResU-Net [13]	~3.5M	~15	0.8331	0.9164	0.8685	Residual + attention, high segmentation	More complex, requires stronger hardware
EfficientNetB4-Unet [37]	~20M	~129	0.851	0.903	0.8795	Highest accuracy, multi-scale attention	Very heavy, not suitable for mobile or embedded
LIU-Net [9]	~1.5M	~8	0.8646	0.9027	0.9092	Good balance, Inception blocks	ET performance lower than SGEResU-Net
MBDResU-Net [8]	~3M	~12	0.846	0.912	0.846	Multi-branch, dilated conv, multi-scale	Increased computational complexity

The Table 3 presents a side-by-side comparison of SepU-Net and four other lightweight segmentation models on the BraTS2020 dataset. Despite having the fewest parameters (~2.39 M) and one of the lowest computational costs (~8.3 GFLOPs), SepU-Net achieves Dice scores of 0.8513 (ET), 0.8420 (NCR/NT) and 0.8460 (TC). By contrast, SGEResU-Net attains a 2.96 % higher Dice on the enhancing tumor region (0.8765 vs. 0.8513) but requires nearly 50 % more parameters (~3.5 M) and almost twice the computational power (~15 GFLOPs). This corresponds to roughly a 40 % reduction in memory footprint for SepU-Net and allows deployment on mobile or edge devices with extended battery life. Such a favorable trade-off between accuracy and efficiency is crucial for time-sensitive clinical scenarios, such as intraoperative tumor localization, where rapid on-site inference is required.

This trade-off between performance and computational efficiency highlights SepU-Net's design philosophy, which prioritizes embedded systems overachieving the highest accuracy in every segmentation task. In comparison, other architectures like SGEResU-Net, EfficientNetB4-Unet, and LIU-Net leverage advanced components such as residual connections, attention mechanisms, and multi-scale feature extraction to achieve higher Dice scores and more robust segmentation, especially for small or complex tumor regions. However, these models are more demanding in terms of hardware and are less suited for mobile or embedded applications.

Therefore, the choice between SepU-Net and other lightweight models should depend on the specific application requirements, particularly whether computational efficiency or segmentation accuracy is the priority. SepU-Net is ideal for environments where computational resources are limited but still requires a strong balance between segmentation quality and performance.

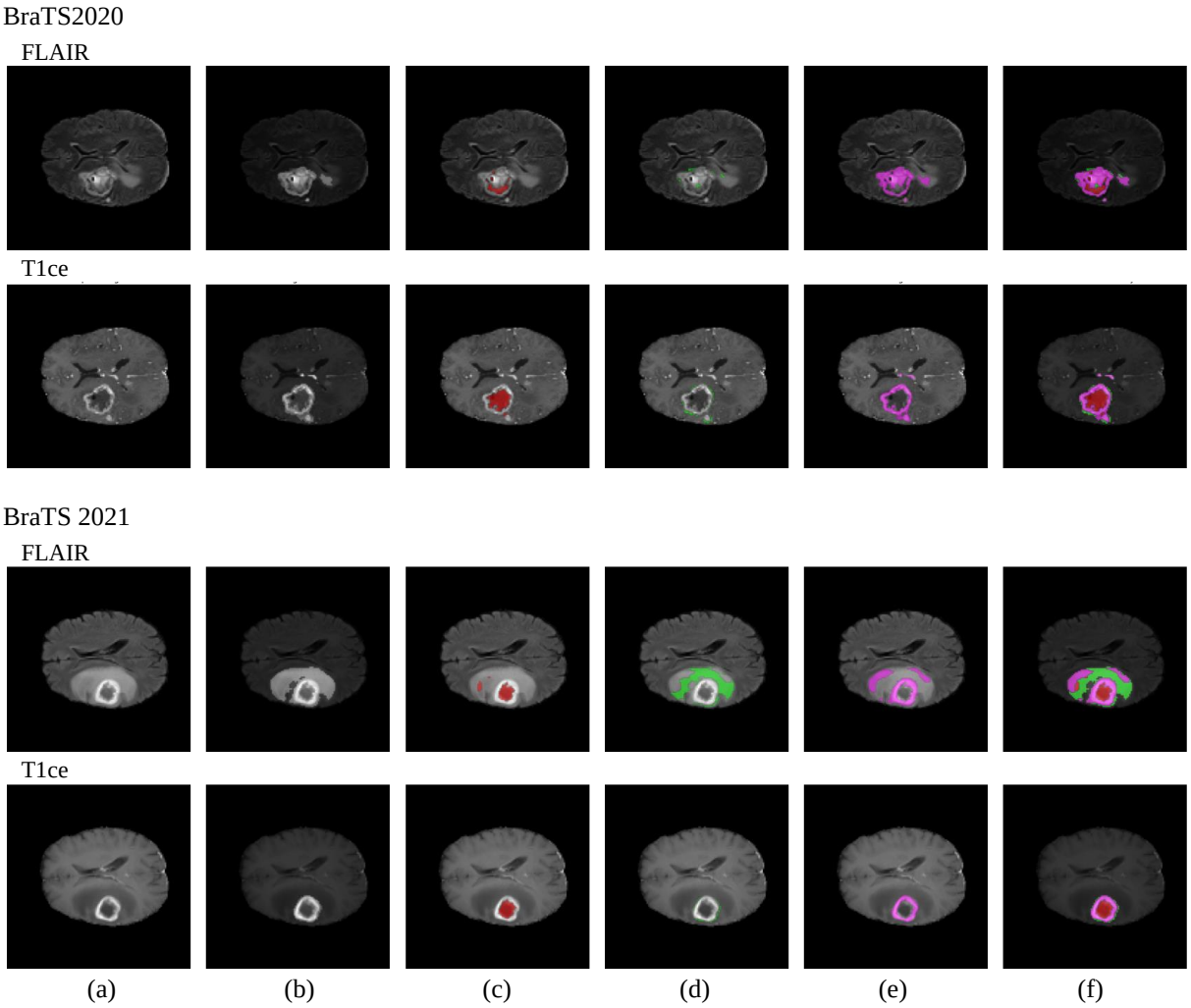


Figure 7. *Presents the multi-class segmentation performance of SepU-Net on representative axial slices from the BraTS2020 and BraTS2021 testing datasets; (a) The original FLAIR or T1ce Images; (b) The Ground Truth; (c) Red zone: Necrotic Core (NCR/NET); (d) Green zone: Peritumoral edema (ED); (e) Magenta zone: Enhancing tumor (ET).*

Result: FFigure 7, presents the multi-class segmentation performance of SepU-Net on representative axial slices from the BraTS 2020 and BraTS 2021 testing datasets. This Figure is divided into the following sections: (a) The original FLAIR or T1ce images, which serve as the input modalities; (b) The ground truth annotations for the tumor regions; (c) The red zone represents the Necrotic Core (NCR/NET), indicating the non-viable central tumor area; (d) The green zone highlights the Peritumoral Edema (ED), corresponding to the swelling and surrounding tissue affected by the tumor; (e) The magenta zone marks the Enhancing Tumor (ET), which indicates the actively growing and contrast-enhanced portion of the tumor. These zones clearly demonstrate how the SepU-Net model accurately segments and distinguishes between different tumor regions across both the BraTS 2020 and BraTS 2021 datasets. By comparing the results from the FLAIR and T1ce modalities, we can observe the complementary roles they play in achieving precise tumor delineation.

3. Conclusions

In this study, we proposed SepU-Net, a lightweight and efficient deep learning architecture for brain tumor segmentation on MRI images. By systematically replacing standard convolutions with Depthwise Separable Convolutions and integrating an efficient channel attention mechanism, SepU-Net achieves a substantial reduction in model parameters and computational cost while maintaining strong segmentation performance. Through comprehensive experiments on both the BraTS2020 and the more challenging BraTS2021 datasets, SepU-Net demonstrated its robustness and generalizability. Specifically, the model maintained high accuracy, mean IoU, and Dice coefficients across both datasets, with only minor declines observed on BraTS2021 due to its increased heterogeneity and complexity.

The comparative results confirm that SepU-Net consistently delivers reliable segmentation of all tumors sub-regions, including Necrotic Core, edema, and enhancing tumor. while offering significant advantages in computational efficiency compared to conventional U-Net and other lightweight models. These findings highlight SepU-Net's suitability for deployment in resource-constrained clinical environments, where both accuracy and efficiency are critical. While the model's performance remains strong, the slight reduction in segmentation metrics on BraTS2021 suggests that further improvements could be achieved by incorporating more advanced attention mechanisms or extending the model to 3D architectures. Future work will focus on these directions to further enhance SepU-Net's ability to handle complex and diverse clinical data.

Overall, SepU-Net provides a practical and effective solution for automatic brain tumor segmentation in real-world medical imaging scenarios, supporting more accessible and timely diagnosis and treatment planning, especially in settings with limited computational resources.

Future work will focus on integrating advanced attention mechanisms into SepU-Net to enhance feature representation and improve segmentation performance for challenging regions such as peritumoral edema. Additionally, extending the architecture to fully 3D implementations is a promising direction to better capture volumetric context and spatial continuity, which are critical for comprehensive brain tumor analysis.

Acknowledgments

I am profoundly grateful for the support and assistance provided by the Faculty of International Education and the Ho Chi Minh City University of Technology and Engineering (HCM-UTE) in the development of this initiative. This endeavor is a component of the 2025 project that is funded by HCMUTE, Vietnam. Special thanks are extended to my tutor, Mr. Nguyen Ngo Lam, for his essential instruction and input, which greatly influenced this study. I also appreciate the HCMUTE teachers whose foundational guidance made this research possible.

Furthermore, I acknowledge the valuable contributions of the scientific community whose prior research and published works in the field of brain tumor segmentation and deep learning provided essential context and inspiration for this study. The insights and methodologies from previous studies have been instrumental in shaping the direction and rigor of this research.

Conflict of Interest

These authors promise no conflicts of interest. Financial and personal relationships did not influence this inquiry. Ho Chi Minh City University of Technology and Engineering (formerly Ho Chi Minh City University of Technology and Education), Vietnam, funded solely execution and development, not results or conclusions.

Data Availability Statement

The complete implementation of SepU-Net, including training scripts, evaluation code, and pre-processing pipelines Hyperparameter Settings, is available at:
https://github.com/MaiHongPhong1902/3D-MRI-lab

REFERENCES

[1] M. Havaei *et al*., "Brain tumor segmentation with deep neural networks," *Med. Image Anal.*, vol. 35, pp. 18–31, 2017, doi: 10.1016/j.media.2016.05.004.
[2] L. Zhao *et al*., "MM-UNet: A multimodality brain tumor segmentation network in MRI images," *Front. Oncol.*, vol. 12, Art. no. 950706, 2022, doi: 10.3389/fonc.2022.950706.
[3] J. K. Ruffle *et al*., "Brain tumour segmentation with incomplete imaging data," *Brain Commun.*, vol. 5, no. 2, 2023, doi: 10.1093/braincomms/fcad118.
[4] P. Li *et al*., "mResU-Net: Multi-scale residual U-Net-based brain tumor segmentation from multimodal MRI," *Med. Biol. Eng. Comput.*, 2024, doi: 10.1007/s11517-023-02965-1.
[5] L. Alzubaidi, J. Zhang, and A. J. Humaidi, "Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions," *J. Big Data*, vol. 8, Art. no. 53, 2021, doi: 10.1186/s40537-021-00444-8.
[6] B. Hou and S. Guan, "Brain tumor segmentation using deep learning: A review," *J. Comput. Electron. Inf. Manag.*, vol. 16, 2025, doi: 10.54097/31ag9n29.
[7] Y. Zhao and L. Lin, "A lightweight U-Net for medical image segmentation," in *Proc. PIERS*, 2024, pp. 1–5, doi: 10.1109/PIERS62282.2024.10618503.
[8] L. Shen *et al*., "MBDRes-U-Net: Multi-scale lightweight brain tumor segmentation network," *arXiv preprint*, arXiv:2411.01896, 2024, doi: 10.48550/arXiv.2411.01896.
[9] G. E. S. Shahid *et al*., "LIU-NET: Lightweight inception U-Net for efficient brain tumor segmentation," *PeerJ Comput. Sci.*, 2025, doi: 10.1109/CW58918.2023.00012.
[10] A. G. Howard *et al*., "MobileNets: Efficient convolutional neural networks for mobile vision applications," *arXiv preprint*, arXiv:1704.04861, 2017.
[11] M. Sandler *et al*., "MobileNetV2: Inverted residuals and linear bottlenecks," in *Proc. IEEE CVPR*, 2018, pp. 4510–4520, doi: 10.1109/CVPR.2018.00474.
[12] K. Avazov *et al*., "Dynamic focus on tumor boundaries: A lightweight U-Net for MRI brain tumor segmentation," *Bioengineering*, vol. 11, Art. no. 1302, 2024, doi: 10.3390/bioengineering11121302.
[13] D. Liu *et al*., "SGEResU-Net for brain tumor segmentation," *Math. Biosci. Eng.*, 2022, doi: 10.3934/mbe.2022073.
[14] Awsaf, "Brain tumor segmentation 2020 dataset," Kaggle, 2020. [Online]. Available: https://www.kaggle.com/datasets/awsaf49/brats20-dataset-training-validation
[15] S. Zabihi *et al*., "SepUNet: Depthwise separable convolution integrated U-Net," in *Proc. IEEE ICIP*, 2021, pp. 2503–2507, doi: 10.1109/ICIP42928.2021.9506283.
[16] S. Gore, "Brain tumour segmentation and analysis using BraTS dataset with improvised 2D and 3D U-Net models," *Research Square*, 2023, doi: 10.21203/rs.3.rs-2791706/v1.
[17] S. Bakas *et al*., "Advancing TCGA glioma MRI collections with expert segmentation labels and radiomic features," *Sci. Data*, vol. 4, Art. no. 170117, 2017, doi: 10.1038/sdata.2017.117.
[18] S. Bakas *et al*., "Identifying the best machine learning algorithms for brain tumor segmentation," *arXiv preprint*, arXiv:1811.02629, 2018.
[19] A. M. Winkler, "The NIfTI file format," 2012. [Online]. Available: https://brainder.org/2012/09/23/the-nifti-file-format/
[20] M. Brett *et al*., "NiBabel: Access a cacophony of neuroimaging file formats," 2012. [Online]. Available: https://nipy.org/nibabel/nifti_images.html
[21] A. M. Winkler, "Brainder," 2012. [Online]. Available: https://brainder.org/2012/09/23/
[22] H. Dong *et al*., "Automatic brain tumor detection using U-Net," in *Med. Image Underst. Anal.*, Springer, 2017, pp. 506–517, doi: 10.1007/978-3-319-60964-5_44.
[23] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proc. IEEE CVPR*, 2017, pp. 1251–1258, doi: 10.1109/CVPR.2017.195.
[24] B. S. Hua, M. K. Tran, and S. K. Yeung, "Pointwise convolutional neural networks," in *Proc. IEEE/GVF*, 2017.
[25] D. Haase and M. A. Daniel, "Rethinking depthwise separable convolutions," *arXiv preprint*, arXiv:2003.13549, 2020, doi: 10.48550/arXiv.2003.13549.
[26] J. Hu *et al*., "Squeeze-and-excitation networks," in *Proc. IEEE CVPR*, 2018, pp. 7132–7141, doi: 10.1109/CVPR.2018.00745.
[27] N. Klingler, "Squeeze-and-excitation networks: A performance upgrade," 2024. [Online]. Available: https://viso.ai/deep-learning/squeeze-and-excite-networks/
[28] S. Woo, J. Park, J. Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," *arXiv preprint*, arXiv:1807.06521, 2018.
[29] Q. Wang *et al*., "ECA-Net: Efficient channel attention for deep convolutional neural networks," in *Proc. IEEE CVPR*, 2020, pp. 11534–11542, doi: 10.1109/CVPR42600.2020.01155.
[30] H. Yang *et al*., "RS-YOLOX: A high-precision detector for object detection in satellite remote sensing images," *Appl. Sci.*, 2025, doi: 10.3390/app12178707.
[31] S. R. Dubey, S. K. Singh, and B. B. Chaudhuri, "Activation functions in deep learning: A comprehensive survey and benchmark," *arXiv preprint*, arXiv:2109.14545, 2022, doi: 10.48550/arXiv.2109.14545.
[32] N. Srivastava *et al*., "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, pp. 1929–1958, 2014.
[33] K. Matoba, N. Dimitriadis, and F. Fleuret, "The theoretical expressiveness of max pooling," *arXiv preprint*, arXiv:2203.01016, 2022, doi: 10.48550/arXiv.2203.01016.
[34] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. MICCAI*, 2015, pp. 234–241, doi: 10.1007/978-3-319-24574-4_28.
[35] K. Machida, I. Nambu, and Y. Wada, "Transposed convolution as alternative preprocessor for brain–computer interface using EEG," *Appl. Sci.*, 2023, doi: 10.3390/app13063578.
[36] R. Rastislav and K. Sayed, "3D MRI brain tumor segmentation using U-Net," Kaggle, 2024. [Online]. Available: https://www.kaggle.com/code/khaledsayedaaaaa/3d-mri-brain-tumor-segmentation-u-net-acc-99
[37] R. Preetha *et al*., "Brain tumor segmentation using multi-scale attention U-Net with EfficientNetB4 encoder for enhanced MRI analysis," *Sci. Rep.*, 2025, doi: 10.1038/s41598-025-94267-9.

Phong Mai Hong is currently an undergraduate student in Computer Engineering at Ho Chi Minh City University of Technology and Engineering (formerly Ho Chi Minh City University of Technology and Education), Ho Chi Minh City, Vietnam. His academic interests include medical image processing, deep learning, and lightweight neural network architectures for embedded and clinical applications.
Phone number: 0865243215. Email: 20119192@student.hcmute.edu.vn. ORCID: https://orcid.org/0009-0003-9741-5092

Lam Mai Thanh received his bachelor's degree in computer engineering from Ho Chi Minh City University of Technology and Education (HCMUTE) (currently Ho Chi Minh City University of Technology and Engineering), Ho Chi Minh City, Vietnam. His academic interests include medical image processing, deep learning, and intelligent systems.
Phone number: 0865243215. Email: 20119137@student.hcmute.edu.vn. ORCID: https://orcid.org/0009-0003-0896-6213

Lam Nguyen Ngo is currently a lecturer at the Faculty of Electrical and Electronics Engineering, Ho Chi Minh City University of Technology and Engineering (formerly Ho Chi Minh City University of Technology and Education). He received his bachelor's and master's degree in radio and electronics engineering from the Ho Chi Minh City University of Technology, Vietnam in 2000 and 2004 respectively. His research interests include wireless communication, data communication, digital signal processing, and computers.
Email: lamnn@hcmute.edu.vn. ORCID: https://orcid.org/0009-0002-6580-0175