Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

LATUP-Net: A lightweight 3D attention U-Net with parallel convolutions for brain tumor segmentation

Ebtihal J. Alwadee [a,b,*], Xianfang Sun [a], Yipeng Qin [a], Frank C. Langbein [a]

[a] School of Computer Science and Informatics, Cardiff University, Cardiff, CF24 4AG, UK
[b] Department of Computer Science, College of Engineering and Computer Science, Jazan University, Jazan, Kingdom of Saudi Arabia

ARTICLE INFO

Keywords:
Brain tumor segmentation
Lightweight
Parallel convolutions
Attention
U-net
Deep learning
Grad-CAM

ABSTRACT

Early-stage 3D brain tumor segmentation from magnetic resonance imaging (MRI) scans is crucial for prompt and effective treatment. However, this process faces the challenge of precise delineation due to the tumors' complex heterogeneity. Moreover, energy sustainability targets and resource limitations, especially in developing countries, require efficient and accessible medical imaging solutions. The proposed architecture, a Lightweight 3D ATtention U-Net with Parallel convolutions, LATUP-Net, addresses these issues. It is specifically designed to reduce computational requirements significantly while maintaining high segmentation performance. By incorporating parallel convolutions, it enhances feature representation by capturing multi-scale information. It further integrates an attention mechanism to refine segmentation through selective feature recalibration. LATUP-Net achieves promising segmentation performance: the average Dice scores for the whole tumor, tumor core, and enhancing tumor on the BraTS 2020 dataset are 88.41%, 83.82%, and 73.67%, and on the BraTS 2021 dataset, they are 90.29%, 89.54%, and 83.92%, respectively. Hausdorff distance metrics further indicate its improved ability to delineate tumor boundaries. With its significantly reduced computational demand using only 3.07 M parameters, about 59 times fewer than other state-of-the-art models, and running on a single NVIDIA GeForce RTX3060 12 GB GPU, LATUP-Net requires just 15.79 GFLOPs. This makes it a promising solution for real-world clinical applications, particularly in settings with limited resources. Investigations into the model's interpretability, utilizing gradient-weighted class activation mapping and confusion matrices, reveal that while attention mechanisms enhance the segmentation of small regions, their impact is nuanced. Achieving the most accurate tumor delineation requires carefully balancing local and global features. The code is available at https://qyber.black/ca/code-bca.

1. Introduction

Brain tumors, particularly gliomas, are among the most lethal forms of cancer due to their inherent complexity and high variability among patients. Gliomas, classified into high-grade and low-grade, consist of different tumor regions, including the enhancing tumor, necrotic core, and surrounding edema [1]. Magnetic resonance imaging (MRI) is the standard method for diagnosing gliomas, and precise segmentation of these tumors is crucial for effective treatment planning. Accurate segmentation helps clinicians differentiate tumor tissue from healthy tissue, which directly influences diagnosis, treatment strategies, and prognosis. However, manual delineation of tumor regions across MRI slices is both time-consuming and labor-intensive, with the accuracy highly dependent on the clinician's expertise and subjective thresholding. This subjectivity, combined with the effort involved, underscores the need for efficient and accurate automatic segmentation techniques for brain tumors [2].

In this work we introduce a novel lightweight deep learning architecture, LATUP-Net (Lightweight 3D ATtention U-Net with Parallel Convolutions). It significantly reduces the computational resources needed while maintaining state-of-the-art brain tumor segmentation performance, as demonstrated on BraTS 2020 [1,3,4] and BraTS 2021 [1]. We incorporate parallel convolutions in the first encoder block of a U-Net architecture, inspired by the inception block [5]. This reduces feature redundancy and parameter count by sharing an initial feature extraction stage across multiple convolutional paths, followed by pooling operations to capture multi-scale spatial features. The design harnesses diverse features efficiently while maintaining a lower parameter count than traditional inception blocks. We further add

[*] Corresponding author at: School of Computer Science and Informatics, Cardiff University, Cardiff, CF24 4AG, UK.
E-mail addresses: AlwadeeEJ@cardiff.ac.uk (E.J. Alwadee), SunX2@cardiff.ac.uk (X. Sun), QinY16@cardiff.ac.uk (Y. Qin), frank@langbein.org (F.C. Langbein).

https://doi.org/10.1016/j.compbiomed.2024.109353
Received 17 May 2024; Received in revised form 21 October 2024; Accepted 31 October 2024
Available online 21 November 2024

an extension mechanism, which generally yields smaller performance improvements, but also does not add substantial computational costs.

Deep learning, particularly convolutional neural networks (CNNs), has revolutionized medical image analysis by providing powerful tools for segmentation, classification, and object detection. CNNs are widely used in medical imaging due to their ability to automatically extract hierarchical features, making them well-suited for complex tasks like brain tumor segmentation [6,7]. Among these architectures, U-Net has become a leading model for medical image segmentation and has demonstrated remarkable performance across various tasks. The U-Net architecture has been widely adopted, receiving over 20,000 citations as of 2023, reflecting its profound impact on medical image segmentation research [8,9]. To meet the increasing demands of clinical applications, U-Net-based models have evolved by enhancing their network structures, incorporating new modules, and expanding to 3D networks [10–12]. However, improving segmentation accuracy often comes at the cost of increased model complexity and parameters [13]. Existing networks require large numbers of parameters and computational resources, which poses challenges for real-world use in resource-limited environments [14]. To address this challenge, lightweight networks have been developed to reduce model parameters and computational resource requirements without compromising performance [6,15]. These models employ strategies such as full convolution and deep separable convolution to achieve high segmentation accuracy while minimizing resource usage, making them more suitable for deployment in resource-constrained environments. In our specific approach we exploit parallel convolutions to reduce computational resources required for training and inference.

In particular, in MRI scans with small tumor lesions, attention mechanisms have been employed to improve results to focus on a specific region and its features. Such integrating can improve the accuracy of small tumor segmentation [16]. Common lightweight attention methods for medical image semantic segmentation include BAM [17], CBAM [18], and Squeeze-and-Excitation (SE) [19]. BAM uses atrous convolution to achieve a larger receptive field, CBAM combines spatial and channel attention separately, and SE focuses on channel attention to address feature loss during convolutional pooling. However, using both spatial and channel attention increases computational complexity. The Squeeze-and-Excitation mechanism is chosen for its efficient feature extraction through channel attention while maintaining computational efficiency, making it suitable for small-scale segmentation tasks [19].

We explore the integration of attention mechanisms within LATUP-Net to improve segmentation of tumor regions. While attention mechanisms effectively emphasize tumor-specific features, our investigation highlights a trade-off: attention may focus too narrowly on these features and overlook broader contextual information, which is also essential for accurate segmentation. By balancing local detail and global context, LATUP-Net achieves performance comparable to state-of-the-art models while drastically reducing computational demands.

In summary, the contributions of our work are:

- We introduce LATUP-Net, an efficient 3D U-Net variant that combines parallel convolutions and attention mechanisms to achieve high segmentation accuracy at a fraction of the computational cost of existing models.
- We demonstrate the effectiveness of parallel convolutions in capturing multi-scale features, resulting in a richer, more efficient representation. They optimize feature diversity and computational efficiency by leveraging a shared initial convolution followed by distinct paths and pooling operations.
- We critically examine the role of attention mechanisms in segmentation, using Grad-CAM [20] and confusion matrix analysis to assess their impact. Our findings show that while attention enhances focus on tumor features, a balanced approach considering both local detail and global context improves overall segmentation performance.

The subsequent sections of this paper are structured as follows: Section 2 presents an overview of previous research conducted on the segmentation of brain tumors. Section 3 explains the LATUP-Net architecture, encompassing its essential elements and their impact on performance. Section 4 provides a comprehensive analysis of the experimental configuration, encompassing the datasets used and the evaluation metrics employed. Section 5 presents and analyses the obtained results. Finally, Section 6 concludes the paper and provides suggestions for further research.

2. Related work

To analyze existing methods for brain tumor segmentation, we consider three perspectives: convolutional neural networks (CNN) and their variants, lightweight models, and attention mechanisms

2.1. CNN and related models

Brain tumor segmentation using CNNs has been extensively studied in the literature, with most methods employing either 2D or 3D convolutions. Initially, 2D models dominated the field, where CNNs processed individual 2D slices derived from 3D MRI scans. However, these 2D slices inherently lack the volumetric context present in full 3D MRIs, leading to the potential loss of important semantic features. This issue is compounded by the fact that the resolution within the plane of 2D slices is often higher than that across slices, and the presence of small gaps between slices further exacerbates the loss of spatial continuity. To capture 3D feature information, 3D CNNs emerged as the preferred approach for analyzing MRI images of brain tumors, addressing the limitations inherent in 2D slice-based analysis [21].

While 3D convolutions make better use of spatial information, they also require more computational power and memory. To address this issue, Chen et al. [22] developed a memory-efficient solution, while preserving most of the volumetric information, by introducing a decoupled 3D U-Net model. It relies on separating a 3D convolution into sequential 2D and 1D convolutions and creating three parallel branches of these separated convolutions, one for each orthogonal view (axial, sagittal, and coronal) for the 1D convolution direction. The suggested model achieved competitive results while demonstrating high efficiency when tested on the BraTS 2018 dataset. While local and global features are necessary for making decisions during segmentation, low-level feature gradients (such as those containing information about boundaries, edges, lines, or dots) converge to zero as one proceeds deeper into the model. To address this, Wang et al. [23] proposed a TransBTS architecture that effectively embeds a transformer into a 3D U-Net model. To begin with, local feature maps are extracted using a 3D CNN encoder. The extracted features are transmitted through a transformer to capture global features. Afterward, the decoder incorporates the local and global features during the upsampling process to produce the segmentation result. Zhu et al. [24] propose a BTS method that combines deep semantic features and edge features for semantic feature extraction and fusion from multimodal MRI. The method uses the Swin Transformer for semantic feature extraction, shifted patch tokenization for training efficiency, and an Edge Spatial Attention Block (ESAB) for feature enhancement. Though both models require more computational resources, it has shown promising results on BraTS 2018 to BraTS 2020, achieving competitive or higher Dice score performance compared with state-of-the-art 3D models.

Traditional U-Net architectures perform exceptionally well on semantic segmentation tasks [8]. Nevertheless, such structures lack strategies to extract global feature information [25,26]. To address this, the inception module [5] and a densely connected module [27] were added to the U-Net architecture by Zhang et al. [28]. Each inception module in the network uses 1 × 1, 3 × 3, and 5 × 5 convolution kernels to acquire multi-scale information. Their method performs admirably in segmenting images of lung tissue, blood arteries, and brain tumors. Meanwhile, the transformers' self-attention mechanism automatically brings in global information but lacks the inductive bias, so it does not obtain sufficient fine-grained features. Thereby, combining transformers and CNNs may leverage the strengths of both. However, this combination often neglects lesion boundaries, which are essential for accurate segmentation. To address this issue, Xu et al. [11] integrated the Swin-T network with a dual-path feature inference module to enhance the original Swin-T network, resulting in improved edge segmentation performance for cranial tumors. Zhu et al. [10] used skip connections to integrate multi-level edge fusion features, derived from the sparse dynamic encoder, into the decoder, therefore improving the transmission of spatial edge information and further refining the network's segmentation performance.

While these approaches improve the representational power of the models, they also increase the number of parameters, heightening the risk of overfitting and limiting effectiveness in scenarios with limited training data. Therefore, given the constraints of low resource funding in many hospitals, it is crucial to strike a balance between processing efficiency and network size through the development of lightweight networks.

2.2. Lightweight models

U-Net variants have demonstrated satisfactory segmentation results for medical images. However, 3D networks require significantly more GPU memory than 2D networks with the same convolutional network structure and depth. Consequently, hardware requirements limit improvements in segmentation. Researchers have proposed a series of lightweight models to reduce network complexity and overcome hardware limitations. By reducing the number of network parameters and achieving highly accurate segmentation, Chen et al. [29] created a dilated multi-fiber (DMF) network that replaces convolutions with dilated convolutions of varying sizes as the fundamental unit. Although dilated convolutions, which modify the convolution's field of view by introducing gaps in the convolutional kernel, can capture features at various scales, they do not necessarily increase the diversity of features captured. Luo et al. [30] proposed a lightweight hierarchical decoupled convolution (HDC) unit by replacing 3D convolutions with pseudo-3D convolutions. However, the model's final segmentation precision is not very good, despite its ability to explore multi-scale, multi-view spatial contexts rapidly with a large reduction in computing complexity. In addition, Magadza et al. [31] utilized depth-wise separable convolutions to reduce computational complexity without sacrificing performance. However, this method cannot handle diverse and fundamental features in multiple, independent directions and orientations.

In a recent study, Zhu et al. [32] proposed a CNN-based model for brain tumor segmentation. Their approach combines three modules that use multimodal, spatial, and boundary information. This method examined the overall spatial aspects of the image allowing it to accurately acquire the tumor's location and how it relates to other tissues in MRI scans. The proposed model proved to be more effective and efficient than other existing state of the arts methods.

In summary, while traditional 3D U-Net models and their variants offer high segmentation accuracy, they are often resource-intensive and prone to overfitting in limited data scenarios. In contrast, lightweight models like the DMF network [29], HDC unit [30], and depth-wise separable convolutions [31], although less precise, provide a viable solution for environments with computational constraints. Our proposed lightweight 3D network addresses these concerns by employing a specific version of parallel convolutions which enhances feature extraction and segmentation performance with significantly fewer parameters, offering a balanced solution between computational efficiency and segmentation precision. We further incorporate an attention mechanism into our model, which addresses the overfitting issue typically associated with complex models.

2.3. Attention mechanism

Traditional U-Nets give equal importance to all features within the feature maps. Given the notable class imbalance in brain tumor segmentation, some features are more crucial than others for accurate results. Attention mechanisms have emerged as effective tools to emphasize these crucial features and downplay less significant ones. Generally, attention mechanisms are bifurcated into two main types: channel attention and spatial attention. Channel attention enables the network to adaptively weigh the importance of different channels based on specific features in the image. This can potentially prioritize channels that are crucial for tumor detection [19]. Spatial attention, instead, fine-tunes the spatial feature maps adaptively, allowing the network to concentrate on specific regions with significant features [33]. In this study, we explore various lightweight attention mechanisms, all recognized for their capacity to enhance model expressiveness and boost overall performance.

3. The LATUP-Net architecture

Here, we explain the components of our LATUP-Net architecture, illustrated in Fig. 1 and Table 1, a lightweight variant of the original U-Net [21] with fewer parameters intended for the semantic segmentation of 3D brain tumors. Moreover, LATUP-Net utilizes multi-scale parallel convolutions (see Section 3.1) and channel attention on multi-modal data fusion (see Section 3.2).

Our encoder consists of three down-sampling blocks with 32, 64, and 128 filters, respectively. Only the first encoder block contains our parallel convolution block. The remaining two encoder blocks consist of a squeeze and excitation attention block [19] followed by two consecutive convolutions with instance normalization [34], and LeakyReLU activation with a negative slope of 0.1. The resulting tensor is passed through a dropout layer at a rate of 0.2. For each encoder block, an identity skip connection is added to map the encoder blocks onto their corresponding decoder blocks. All convolutions have a kernel size of $3 \times 3 \times 3$, and down-sampling is achieved by $2 \times 2 \times 2$ max-pooling to reduce the spatial resolution of the feature maps.

The decoder takes the feature maps of the encoder and doubles their spatial resolution using 3D trilinear up-sampling. It has three up-sampling blocks, each consisting of two $3 \times 3 \times 3$ convolutions with 128, 64, and 32 filters respectively, followed by instance normalization, and LeakyReLU. A squeeze and excitation attention block has been added between the two convolutions. This is followed by a dropout layer, implemented with a rate of 0.2 to mitigate overfitting by randomly deactivating a portion of the neural connections during training. After the dropout layer, an additional $3 \times 3 \times 3$ convolutional layer is incorporated. This layer is pivotal for refining the feature representations post-dropout, ensuring the restoration of spatial dimensions, and enhancing the network's ability to learn detailed, spatially coherent features essential for accurate segmentation.

The last two blocks of the encoder and decoder use an L2 regularizer. Finally, a $1 \times 1 \times 1$ convolutional layer with softmax activation is applied to the output of the decoder, which generates a probability map for each voxel indicating its likelihood of belonging to one of the tumor region classes to be segmented.

The LATUP-Net architecture is intentionally designed to minimize the dependence on complex ensembling or additional computational resources. In alignment with best practices outlined in [9], we focus on delivering an efficient, lightweight model that demonstrates its innovations without depending on confounding performance boosters. This approach ensures fair comparisons and highlights the true impact of our architectural choices. By combining proven techniques, such as multi-scale parallel convolutions and lightweight attention mechanisms, LATUP-Net balances the need for high segmentation performance with reduced computational complexity, making it suitable for resource-constrained environments [35].

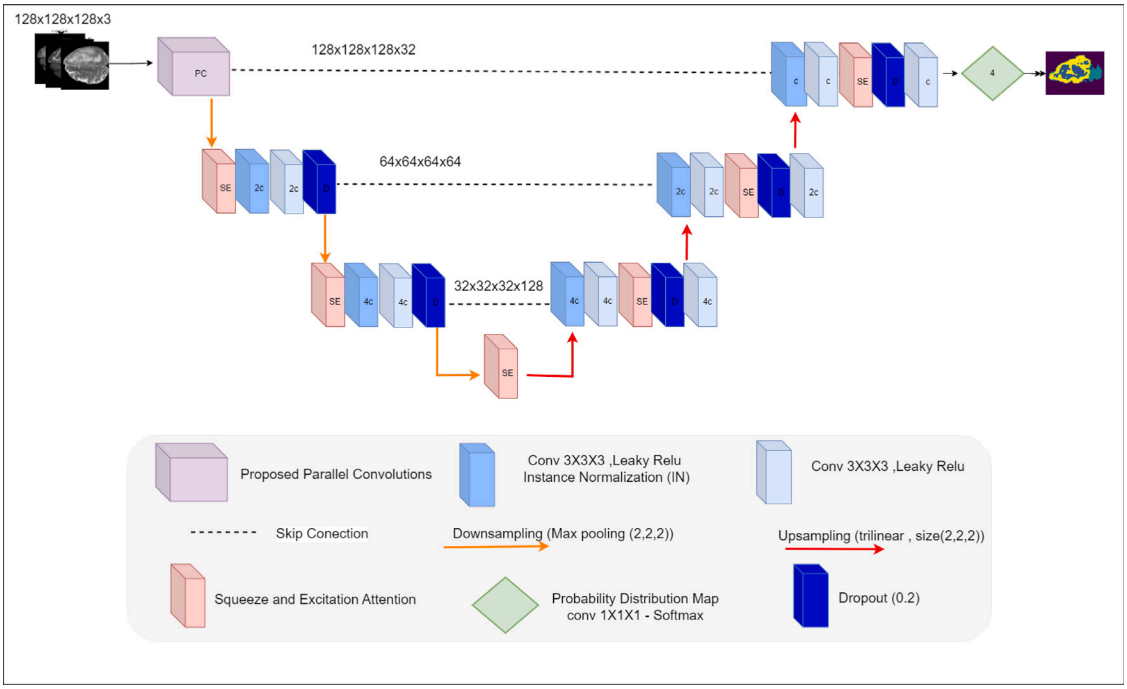
128x128x128x3
PC
128x128x128x32
c
c
SE
D
c
4
SE
2c
2c
D
64x64x64x64
2c
2c
SE
D
2c
SE
4c
4c
D
32x32x32x128
4c
4c
SE
D
4c
SE
Proposed Parallel Convolutions
Conv 3X3X3 ,Leaky Relu
Instance Normalization (IN)
Conv 3X3X3 ,Leaky Relu
Skip Conection
Downsampling (Max pooling (2,2,2))
Upsampling (trilinear , size(2,2,2))
Squeeze and Excitation Attention
Probability Distribution Map
conv 1X1X1 - Softmax
Dropout (0.2)

Fig. 1. The LATUP-Net architecture.

In line with findings by Rajamani et al. [36], integrating an attention block at the network's bottleneck — specifically, an SE block in our case — facilitates the capture of longer-range dependencies within the lowest-resolution activation maps. This adjustment boosts performance with only a slight increase in model complexity.

3.1. Parallel Convolutions (PC)

CNNs have demonstrated efficacy in feature extraction. However, indiscriminate augmentation of network layers might precipitate overfitting and computational overheads [37]. A balance between network depth and width remains paramount. Our strategy employs parallel convolutions with varying kernel sizes, drawing inspiration from the inception model [38]. This design allows the network to capture features at different scales, yielding a more efficient model with enhanced segmentation performance.

To improve the representation power of the network, which is a key factor in improving its accuracy and reliability, parallel convolutional layers can be added to different encoder and decoder blocks. However, adding parallel convolutions to all blocks may result in overfitting due to the large number of learnable parameters and limited training data. Therefore, it is added only to the first block of the encoder to extract the most fundamental and diverse features from the input data. Parallel convolutions can capture these features at different scales and orientations. This way, we improve representation power while reducing the risk of overfitting.

The proposed PC block (see Fig. 2) is designed to process the input through a series of convolutional layers with different kernel sizes, each aiming to capture features at various spatial scales. Initially, the input passes through a shared embedded layer of a single $3 \times 3 \times 3$ convolution, which extracts a preliminary set of features from the input data. Following this, the features are processed in parallel through three distinct paths: one continues directly from the initial $1 \times 1 \times 1$ convolution, another passes through an additional $3 \times 3 \times 3$ convolution, and the third through a $5 \times 5 \times 5$ convolution. Convolutional layers with smaller kernel sizes, such as $1 \times 1 \times 1$ or $3 \times 3 \times 3$, are adept at detecting local patterns like edges and textures. Layers with larger kernels, like $5 \times 5 \times 5$, are suited for identifying broader spatial patterns and hierarchical structures within the data, thereby providing an extended receptive field. Each path then concludes with a max-pooling operation, reducing the dimensionality and computational load of the subsequent layers. The outputs of these parallel paths are concatenated, combining the multi-scale features into a unified feature map that is richer and more informative than what could be obtained from any single path.

This approach contrasts with the inception block [38], which typically includes multiple parallel paths starting from the same input, each with different combinations of convolutions and sometimes pooling operations, without a shared embedded convolution. The inception block aims to capture multi-scale information by applying various-sized convolutions (e.g., $1 \times 1 \times 1$, $3 \times 3 \times 3$, $5 \times 5 \times 5$) in parallel and then merging their outputs. However, each path in an inception block operates independently of the others, without a shared feature extraction stage. This increases the number of parameters and may lead to redundancy, as each path may learn similar features.

Our proposed PC block contributes to making the model lightweight in several ways. Firstly, the shared embedded layer ensures that all paths operate on a common set of features, reducing redundancy and the need for each path to learn from scratch. This decreases the number of parameters compared to having multiple independent paths, as seen in inception blocks [38]. Secondly, by limiting each path to a single convolution and a pooling operation after the shared convolution, the model avoids the parameter growth associated with stacking multiple convolutions in each path. This streamlined approach enables efficient multi-scale feature extraction without the complexity and parameter overhead typically associated with more elaborate multi-path designs. Consequently, this design choice balances capturing diverse spatial features and maintaining a compact, efficient model architecture.

3.2. SE attention block

The attention mechanisms explored in this study include Squeeze-and-Excitation (SE) [19], the Convolutional Block Attention Module (CBAM) [18], Efficient Channel Attention (ECA) [39], and Residual Squeeze-and-Excitation (RSE) [40]. We further introduce a modified variant of SE where the fully connected (dense) layers are replaced by

Table 1
LATUP-Net model architecture details.

Layer	Input shape	Layer type	Stride	Output shape	Parameters
Input layer	(128, 128, 128, 3)	Input	–	(128, 128, 128, 3)	0
Conv3D_1 (enc1_pc_embed)	(128, 128, 128, 3)	Conv3D (32 filters)	(1,1,1)	(128, 128, 128, 32)	2624
Conv3D_2 (enc1_pc_1_conv)	(128, 128, 128, 32)	Conv3D (32 filters)	(1,1,1)	(128, 128, 128, 32)	1056
Conv3D_3 (enc1_pc_2_conv)	(128, 128, 128, 32)	Conv3D (32 filters)	(1,1,1)	(128, 128, 128, 32)	27 680
Conv3D_4 (enc1_pc_3_conv)	(128, 128, 128, 32)	Conv3D (32 filters)	(1,1,1)	(128, 128, 128, 32)	128 032
MaxPooling3D_1 (enc1_pc_1_maxpool)	(128, 128, 128, 32)	MaxPooling3D	(2,2,2)	(64, 64, 64, 32)	0
MaxPooling3D_2 (enc1_pc_2_maxpool)	(128, 128, 128, 32)	MaxPooling3D	(2,2,2)	(64, 64, 64, 32)	0
MaxPooling3D_3 (enc1_pc_3_maxpool)	(128, 128, 128, 32)	MaxPooling3D	(2,2,2)	(64, 64, 64, 32)	0
Concatenate (enc1_pc_concat)	(64, 64, 64, 32)	Concatenate	–	(64, 64, 64, 96)	0
SE Layer_1 (enc2_SE_mult)	(64, 64, 64, 96)	Squeeze and excitation	–	(64, 64, 64, 96)	2304
Conv3D_5 (enc2_conv1)	(64, 64, 64, 96)	Conv3D (64 filters)	(1,1,1)	(64, 64, 64, 64)	165 952
InstanceNorm_1 (enc2_instance_norm)	(64, 64, 64, 64)	Instance normalization	–	(64, 64, 64, 64)	128
Conv3D_6 (enc2_conv2)	(64, 64, 64, 64)	Conv3D (64 filters)	(1,1,1)	(64, 64, 64, 64)	110 656
Dropout_1 (enc2_dropout)	(64, 64, 64, 64)	Dropout	–	(64, 64, 64, 64)	0
MaxPooling3D_4 (enc2_maxpool)	(64, 64, 64, 64)	MaxPooling3D	(2,2,2)	(32, 32, 32, 64)	0
SE Layer_2 (enc3_SE_mult)	(32, 32, 32, 64)	Squeeze and excitation	–	(32, 32, 32, 64)	1024
Conv3D_7 (enc3_conv1)	(32, 32, 32, 64)	Conv3D (128 filters)	(1,1,1)	(32, 32, 32, 128)	221 312
InstanceNorm_2 (enc3_instance_norm)	(32, 32, 32, 128)	Instance normalization	–	(32, 32, 32, 128)	256
Conv3D_8 (enc3_conv2)	(32, 32, 32, 128)	Conv3D (128 filters)	(1,1,1)	(32, 32, 32, 128)	442 496
Dropout_2 (enc3_dropout)	(32, 32, 32, 128)	Dropout	–	(32, 32, 32, 128)	0
MaxPooling3D_5 (enc3_maxpool)	(32, 32, 32, 128)	MaxPooling3D	(2,2,2)	(16, 16, 16, 128)	0
SE Layer_3 (bn_SE_mult)	(16, 16, 16, 128)	Squeeze and excitation	–	(16, 16, 16, 128)	4096
UpSampling3D_1 (dec3_upsample)	(16, 16, 16, 128)	UpSampling3D	(2,2,2)	(32, 32, 32, 128)	0
Conv3D_9 (dec3_conv1)	(32, 32, 32, 128)	Conv3D (128 filters)	(1,1,1)	(32, 32, 32, 128)	131 200
InstanceNorm_3 (dec3_instance_norm)	(32, 32, 32, 128)	Instance normalization	–	(32, 32, 32, 128)	256
Concatenate (dec3_concat)	(32, 32, 32, 128)	Concatenate	–	(32, 32, 32, 256)	0
Conv3D_10 (dec3_conv2)	(32, 32, 32, 256)	Conv3D (128 filters)	(1,1,1)	(32, 32, 32, 128)	884 864
UpSampling3D_2 (dec2_upsample)	(32, 32, 32, 128)	UpSampling3D	(2,2,2)	(64, 64, 64, 128)	0
Conv3D_11 (dec2_conv1)	(64, 64, 64, 128)	Conv3D (64 filters)	(1,1,1)	(64, 64, 64, 64)	65 600
InstanceNorm_4 (dec2_instance_norm)	(64, 64, 64, 64)	Instance normalization	–	(64, 64, 64, 64)	128
Concatenate (dec2_concat)	(64, 64, 64, 64)	Concatenate	–	(64, 64, 64, 128)	0
Conv3D_12 (dec2_conv2)	(64, 64, 64, 128)	Conv3D (64 filters)	(1,1,1)	(64, 64, 64, 64)	221 248
UpSampling3D_3 (dec1_upsample)	(64, 64, 64, 64)	UpSampling3D	(2,2,2)	(128, 128, 128, 64)	0
Conv3D_13 (dec1_conv1)	(128, 128, 128, 64)	Conv3D (32 filters)	(1,1,1)	(128, 128, 128, 32)	16 416
Concatenate (dec1_concat)	(128, 128, 128, 32)	Concatenate	–	(128, 128, 128, 64)	0
Conv3D_14 (dec1_conv2)	(128, 128, 128, 64)	Conv3D (32 filters)	(1,1,1)	(128, 128, 128, 32)	55 328
Conv3D_15 (prob_map)	(128, 128, 128, 32)	Conv3D (4 filters)	(1,1,1)	(128, 128, 128, 4)	132
Total parameters					**3,069,060**

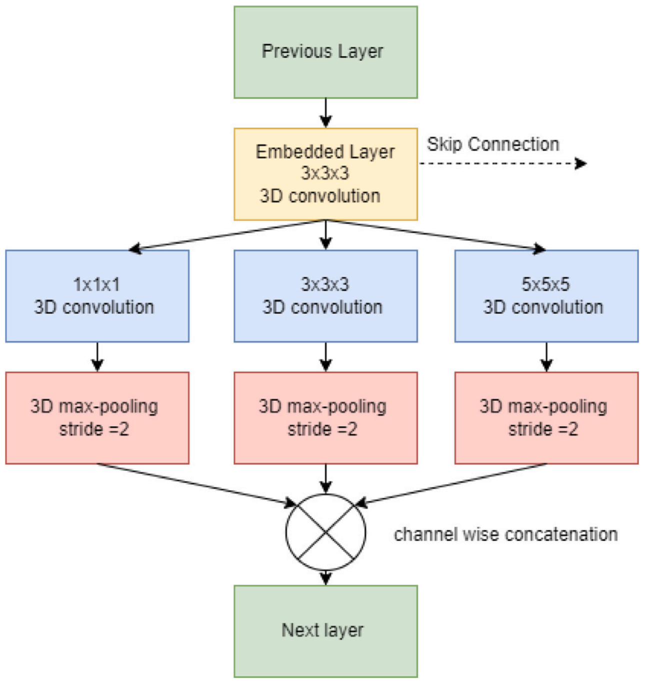

Fig. 2. Proposed parallel convolutions.

a 3D convolutional layer. We also experiment with combined mechanisms such as fusing CBAM and SE. The motivation behind combining CBAM and SE is to leverage the strengths of both. CBAM's ability to focus on pertinent spatial and channel features and SE's capacity to recalibrate channel-wise features may enhance the model's ability to capture complex interdependencies in the data. Another approach is designed to exploit the strengths of convolutions while adhering to the principles of the SE mechanism. Replacing the dense layer in SE with a 3D convolution layer (SE-3D) is aimed at maintaining the spatial information of the input tensor and capturing local spatial correlations, while simultaneously maintaining the ability to recalibrate channel-wise features.

Based on rigorous evaluation of attention mechanisms in Section 5, we incorporate an SE block [19] into our final model, which is illustrated in Fig. 3. This block is recognized for its efficiency and lightweight nature. The SE block is composed of two distinct operations: Squeeze and Excitation. In the squeeze phase, input images of size $H \times W \times D \times C$ are transformed to a $1 \times 1 \times 1 \times C$ format through a Global Average Pooling (GAP) layer, which compresses the spatial resolutions, retaining only channel-centric information for the subsequent excitation operation. The excitation phase employs a series of layers, beginning with a fully connected layer complemented by a reduction factor r. This is then subjected to ReLU activation, succeeded by another fully connected layer, culminating in a sigmoid activation to produce the final output of the Excitation operation. A scaling transformation is executed to assimilate the channel-specific data, yielding an output enriched with channel-level information.

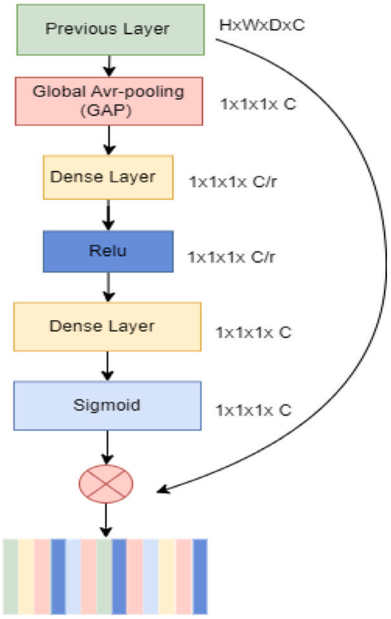

Fig. 3. Squeeze and excitation block.

4. Data and implementation details

4.1. Data and pre-processing

The proposed model is trained and validated using the Brain Tumor Segmentation (BraTS) benchmark datasets: BraTS 2020 [1,3,4] and BraTS 2021 [1]. BraTS 2021, a superset of BraTS 2020, encompasses 1251 patients, including high-grade gliomas (HGG) and low-grade gliomas (LGG). The BraTS 2020 dataset contains 369 patients, of which 76 have been diagnosed with LGG, with the remainder having HGG.

The BraTS 2020 and BraTS 2021 datasets are used for efficiency and sustainability, with initial testing on the smaller dataset allowing for algorithm refinement and model selection before deployment on the larger dataset. It conserves computational resources and time, and aligns with iterative development best practices, where initial testing on a subset of the data can provide quick feedback, critical for early-stage model tuning and optimization [41].

Each dataset consists of 3D scans with 155 individual “slices” or images, each having 240 × 240 pixels. These scans capture four MRI modalities — T2, T1, T1ce, and FLAIR — crucial for brain tumor segmentation, offering distinct insights. T1-weighted images illustrate anatomical structures, distinguishing between gray and white matter. T2-weighted images aid in visualizing edema, by emphasizing water content. T1ce images, with contrast enhancement, focus on blood vessel imaging, a key component in identifying active tumor growth. Conversely, FLAIR images suppress cerebrospinal fluid signals, illuminating anomalies in intensity and subtle lesions indicative of tumor expansion. Ground truth segmentation masks are meticulously annotated by one to four expert neuroradiologists per case. The scan has been segmented into four primary classes: background (BG, Label 0), necrotic and non-enhancing tumor (NCR/NET, Label 1), edema (ED, Label 2), and enhancing tumor (ET, Label 4) as exemplified in Fig. 4. Following common practices in the literature, we classify these into three main tumor regions for segmentation: whole tumor (WT) encompassing NCR/NET, ED, and ET (Labels 1, 2, 4), tumor core (TC) consisting of NCR/NET and ET (Labels 1, 4), and the sole ET (Label 4).

The BraTS datasets have been meticulously preprocessed by their developers, including co-registration to a consistent anatomical template, interpolation to a unified resolution (1 mm^3), and skull stripping. However, MRI scan intensity values often display inconsistencies and may fluctuate due to many factors. To mitigate this, we normalize the intensity range to the interval $[0,1]$ using the min–max scaler [42]. This adjustment not only enhances data consistency but also optimizes it for deep learning algorithms.

Table 2
Hyperparameters for the LATUP-Net model.

Hyperparameter	Value
Input size	128 × 128 × 128 × 3
Batch size	1
Hidden layer activation	Leaky ReLU($\alpha = 0.1$)
Optimizer	ADAM ($\beta_1 = 0.9$, $\beta_2 = 0.999$)
Learning rate	1×10^{-4}
Number of epochs	200
Loss function	Weighted Dice score loss (see Section 4.3)
Dropout	0.2
Regularization	L2 (factor 0.02)
Output layer activation	Softmax
Output size	128 × 128 × 128 × 4

We further crop the images to a standard size of 128 × 128 × 128 voxels, centered on the MRI scans. Preliminary tests suggested that including T1 images with T1ce, T2, and FLAIR only marginally improves segmentation results. Since T1ce is essentially a contrast-enhanced derivative of T1, and T1 mainly contributes to identifying a small fraction of the edema, which FLAIR can effectively detect as well, we chose to exclude T1 from our inputs to conserve computational resources.

4.2. Implementation details

We implemented our network via Keras in Tensorflow 2.15. Computations are executed on a single NVIDIA GeForce RTX3060 12 GB GPU, which is considered a relatively low-end consumer card. For training, we employed the ADAM optimizer [43], setting the learning rate to 1×10^{-4}. Training proceeds with a batch size of 1, a choice primarily dictated by GPU memory constraints. We used a constant number of 200 epochs. To mitigate the risk of overfitting and enhance the model’s generalization capabilities, L2 regularization was applied to the convolutional kernel parameters with a factor of 0.02. This regularization factor was selected based on experimentation during the model selection process, as detailed in the supplementary material.

Leaky ReLU with a leak factor of $\alpha = 0.1$ was used as the activation function for the hidden layers. This value is widely used in deep learning and was chosen based on its effectiveness in similar tasks, such as image segmentation, where it helps regularize the model by allowing gradient flow for negative inputs, contributing to stable training. Studies, including nnU-Net [44] and the original leaky ReLU paper [45], have shown that $\alpha = 0.1$ works well in practice.

During preliminary experimentation, we evaluated the ReduceLROnPlateau callback for dynamic learning rate adjustments. However, observations indicated a predisposition towards overfitting when it was employed. As such, it was excluded from the final training (see Section 5.1.1).

The specific hyper-parameter settings we adopted during model training are detailed in Table 2. Detailed results with a discussion are in Section 5 with additional information about the parameter selection process available in the supplementary material. The source code is available at [35] with final models and analysis results at [46].

4.3. Loss function

Loss function selection is a critical factor in contemporary deep-learning network designs, especially in the field of brain tumor segmentation. Recent studies indicate that no single popular loss function consistently offers superior performance across various segmentation tasks [47].

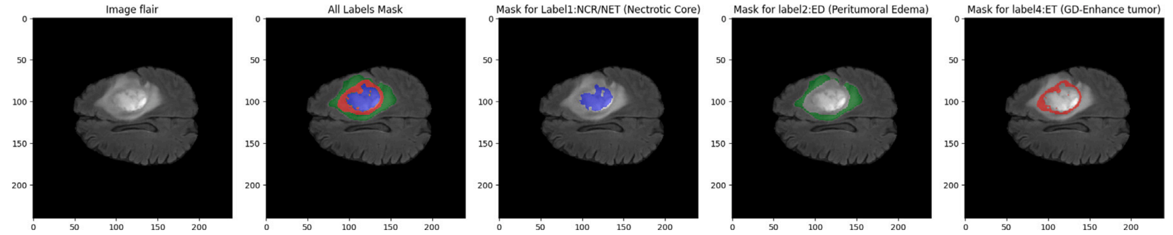
Image flair
All Labels Mask
Mask for Label1:NCR/NET (Nectrotic Core)
Mask for label2:ED (Peritumoral Edema)
Mask for label4:ET (GD-Enhance tumor)
0
50
100
150
200
0
50
100
150
200

Fig. 4. MRI scan of a brain tumor featuring ground truth segmentation masks: blue represents necrotic and non-enhancing tumor areas (NCR/NET, Label 1), green highlights edema regions (ED, Label 2), and red indicates areas of enhancing tumor (ET, Label 4).

Compound loss functions, which combine two or more types of loss functions, have emerged as the most robust and competitive in different scenarios [47]. In our experiments, we aim to enhance segmentation performance and address the severe class imbalance in the BraTS datasets by combining Dice loss with Binary Cross Entropy (BCE) [48] and Dice loss with focal loss [49]. However, based on our experiments, these compounded loss function approaches did not significantly outperform Dice loss alone. Therefore, to boost the segmentation performance and solve the class imbalance problem, the loss function used during the final training process is the Weighted Dice score Loss (WDL).

The Dice score loss for each class i, corresponding to the network output channels BG, NCR/NET, ED, and ET, is

$$DSL_i = 1 - \frac{2\sum_n \left(y_{i,n}^{true} \odot y_{i,n}^{\text{pred}} \right) + \epsilon}{\sum_n \left(y_{i,n}^{true} \right)^2 + \sum_n \left(y_{i,n}^{\text{pred}} \right)^2 + \epsilon}. \tag{1}$$

y_i^{true} and y_i^{pred} represent the ground truth and predicted segmentation masks for class i, respectively; n iterates over all elements of y_i^{true} and y_i^{pred}; $\odot$ signifies point-wise multiplication; and ϵ is a negligible constant introduced to avoid division by zero. In our experiments, we set $\epsilon = 0.00001$. Note that here the network output masks are the original BG, NCR/NET, ED, and ET regions in the ground truth (see Section 4.1).

The WDL weights w_i for class i are computed according to the ENet paper [50],

$$w_i = \frac{1}{\log(C + \frac{c_i}{T})} \tag{2}$$

where $C = 1.02$, c_i is the voxel count for class i, and T is the total count of voxels across all classes. This formula ensures that classes with fewer voxels receive higher weights to balance the loss during the training process. Here we use the WT, TC and ET regions instead of the individual output channels above to compute the weights, and we get $w_{\text{WT}} = 1.64$, $w_{\text{TC}} = 2.55$, and $w_{\text{ET}} = 3.40$.

Overall this gives our Weighted Dice score Loss (WDL),

$$\begin{aligned} \text{WDL} = {} & w_{\text{WT}} \cdot (DSL_{\text{NCR/NET}} + DSL_{\text{ED}} + DSL_{\text{ET}}) \\ & + w_{\text{TC}} \cdot (DSL_{\text{NCR/NET}} + DSL_{\text{ET}}) \\ & + w_{\text{ET}} \cdot DSL_{\text{ET}}, \end{aligned} \tag{3}$$

which is equivalent to

$$\begin{aligned} \text{WDL} = {} & (w_{\text{WT}} + w_{\text{TC}} + w_{\text{ET}}) \cdot DSL_{\text{ET}} \\ & + (w_{\text{WT}} + w_{\text{TC}}) \cdot DSL_{\text{NCR/NET}} \\ & + w_{\text{WT}} \cdot DSL_{\text{ED}}. \end{aligned} \tag{4}$$

In this expression, the dice score loss (DSL) for each class $i \in \{\text{NCR/NET}, \text{ED}, \text{ET}\}$ is computed separately and weighted according to the importance of the corresponding tumor region (WT, TC, and ET) and then summed up to compute the total weighted Dice score loss for the segmentation task. This loss function links the clinical relevance of each tumor region with the network's output channels, ensuring that the segmentation process prioritizes the most clinically significant areas which is crucial for achieving optimal segmentation performance. Using the ENet weights helps in addressing the class imbalance by assigning higher weights to smaller but more important tumor regions.

It is also important to note that while the network output channels include the background class (BG), it is not included in the loss and weights calculation. We found its presence improves the segmentation performance of the model. We also explored the use of different output channels, weights, and manual adjustment of the weights. However, it did not yield satisfactory results.

4.4. *Evaluation metrics*

We measure the effectiveness of the proposed model using the Dice similarity coefficient (DSC), and the 95th percentile Hausdorff distance (HD95). DSC and HD95 are widely accepted as the primary performance evaluation metrics in image segmentation tasks. The DSC quantifies the spatial overlap between the ground truth and the predicted segmentation region. HD95 calculates the 95th percentile of the distances between the points in the ground truth and the predicted set. This is akin to the conventional symmetric Hausdorff distance but reduces the impact of outliers by focusing on the 95th percentile. HD95 in particular indicates the accuracy of boundary prediction, revealing the model's precision in delineating the tumor margins.

The metrics are defined as

$$\text{DSC} = \frac{2|P \cap T|}{|P| + |T|}, \tag{5}$$

$$\begin{aligned} \text{HD95} = \max \Bigg(& \max_{p \in P_{95\%}} \min_{t \in T} \|p - t\|, \\ & \max_{t \in T_{95\%}} \min_{p \in P} \|t - p\| \Bigg), \end{aligned} \tag{6}$$

where p and t are voxel coordinates for the predicted and ground truth regions respectively; P is the predicted region, and T is the ground truth region. The corresponding 95th percentile regions are represented by $P_{95\%}$ and $T_{95\%}$.

We employ different metrics for evaluating our model based on the data partitioning approach used. For model selection (see Sections 5.1, and 5.2), we employ an 80/20 training-testing holdout split, and we report the mean and standard deviation of the per-sample (per-patient) metrics to gain insights into the model's performance under controlled conditions. Once the optimal model (LATUP-Net) is determined, and for comparison to the state-of-the-arts models (see Section 5.5), five-fold cross-validation is applied to consider any data dependencies and ensure a comprehensive evaluation of the model's performance across varied data scenarios. During cross-validation, we calculate the mean and standard deviation of the mean DSC and HD95 across all folds to assess the model's robustness and general performance.

5. Experimental results and discussion

This section covers four main analyses. Firstly, we analyze the performance of some variants of our architecture and the influence of learning rate optimization on the segmentation performance. Then, we investigate which attention mechanism gives the best performance. We also evaluate the effectiveness of the attention mechanism. Finally, we compare the performance of our LATUP-Net to the state-of-the-art on BraTS 2020 and 2021.

Table 3
Comparative analysis of model architectures for brain tumor segmentation using the BraTS 2020 dataset: This table illustrates the mean and standard deviation (indicated by ±) for the per-sample Dice similarity coefficient (DSC) and the 95th percentile Hausdorff distance (HD95). Results are segmented into whole tumor (WT), tumor core (TC), and enhancing tumor (ET) categories, based on an 80/20 train/test set split.

Model	DSC (%)			HD95 (mm)		
	WT	TC	ET	WT	TC	ET
U-Net	83.22 ± 9.29	77.13 ± 18.34	60.65 ± 27.44	18.50 ± 22.17	15.38 ± 22.91	19.34 ± 30.24
Inception	87.53 ± 7.49	80.91 ± 18.64	69.28 ± 29.05	10.87 ± 15.72	6.58 ± 7.78	14.94 ± 29.06
PC	88.13 ± 7.14	84.19 ± 16.74	70.23 ± 28.40	4.99 ± 4.00	5.63 ± 6.71	12.86 ± 26.39
PC + SE	88.52 ± 7.10	83.26 ± 17.18	71.86 ± 27.02	5.98 ± 7.88	5.51 ± 5.20	12.96 ± 26.55
PC + WDL	**89.58 ± 5.70**	**85.35 ± 15.49**	73.21 ± 27.18	**4.78 ± 4.41**	5.25 ± 6.38	11.65 ± 26.33
PC + SE + WDL	88.72 ± 6.33	84.71 ± 15.65	**74.49 ± 25.98**	5.76 ± 4.41	**5.15 ± 6.28**	**10.65 ± 25.33**

5.1. Overall performance analysis

To find the best variant of our proposed model and training, we compare the following architectures:

U-Net: The baseline U-Net model trained using Dice loss.

Inception: Modified U-Net model with inception module trained using Dice loss.

PC: Modified U-Net model with parallel convolutions trained using Dice loss.

PC + SE: Modified U-Net model with parallel convolutions and channel attention trained using Dice loss.

PC + WDL: Modified U-Net model with parallel convolutions trained using weighted Dice score Loss.

PC + SE + WDL: Modified U-Net model with parallel convolutions, and attention trained using weighted Dice score Loss.

The models are initially selected based on a single 80/20 split using the same training hyperparameters (see Table 2) on the BraTS 2020 dataset. Table 3 shows the segmentation results of the test set from these training runs. The performance is assessed by employing two key metrics: per-sample DSC and HD95.

In the Inception model, we replace the first U-Net block with an inception module. This modification improves the segmentation results compared to the U-Net, with DSC improvements of 4.31, 3.78, and 8.63 for whole tumor (WT), tumor core (TC), and enhancing tumor (ET), while reducing the HD95 by 7.63, 8.8, and 4.44 for WT, TC, and ET, respectively. However, despite these performance gains, the Inception model significantly increases memory usage during training, likely due to the module's more complex structure, which combines multiple convolutional and pooling operations.

The PC model, using parallel convolutions, demonstrates superior efficiency compared to both U-Net and Inception. It achieves faster convergence during training and reduces the need for computational resources. Additionally, PC provides substantial improvements in segmentation performance compared to U-Net, with DSC gains of 4.91, 7.06, and 9.58 and HD95 reductions of 13.51, 9.75, and 6.48 for WT, TC, and ET, respectively.

When comparing our proposed parallel convolutions (PC) with the Inception module, PC offers a strategic advantage. By replacing the conventional $1 \times 1 \times 1 \times 1$ convolution with a $3 \times 3 \times 3 \times 3$ convolution, PC achieves a more spatially compact and contextually rich representation, while effectively reducing memory consumption. In contrast, the Inception module with its complex configuration faced a surge in model memory usage during training. Furthermore, our design's judicious positioning of pooling operations optimally condenses feature map dimensions, ensuring efficient memory usage without compromising capturing features. The PC model outperforms the inception model with 0.6, 3.28, and 0.95 DSC increment and 5.88, 0.95, and 2.08 HD95 decrement for WT, TC, and ET, respectively, which proves PC's ability to segment difficult tumor regions such as TC.

The addition of SE attention to the PC model (PC + SE) further enhances segmentation performance, particularly for smaller tumor regions, as evidenced by DSC improvements of 0.39 for WT and 1.63 for ET. However, a slight reduction in TC performance (0.93 decrease in DSC) suggests that while attention mechanisms provide benefits in some areas, they may introduce trade-offs for certain tumor regions. Several lightweight attention modules are compared in Section 5.2.

Notably, our investigations reveal a significant performance enhancement upon employing the WDL with the PC model. Specifically, the DSC for WT, TC, and ET increased by 1.45, 1.16, and 2.98 respectively and HD95 reduced by 0.21, 0.38, and 1.21 respectively.

This improvement underscores the utility of the WDL in case of a segmentation region size imbalance. However, we notice that when adding the attention mechanism, the result of WT and TC decreased slightly in both DCS and HD95, which leads us to check whether attention is needed. Nonetheless, we chose to include SE because there is a 1.28% improvement in the ET segmentation result and enhancements in HD95 for TC and ET. This is further investigated in Section 5.4.

Fig. 5 depicts qualitative comparisons between the various networks in segmenting the distinct tumor regions. The qualitative results demonstrate that our LATUP-Net which consists of PC and SE trained with WDL, outperformed all other models, consistent with the quantitative results in Table 3.

5.1.1. The influence of learning rate decay on the segmentation performance

Training neural networks necessitates careful control of convergence and prevention of overfitting. Traditional models typically utilize learning rate schedulers alongside stochastic gradient descent. However, with the introduction of advanced optimizers such as ADAM, which integrates momentum and regularization, there has been a shift to strategies like `ReduceLROnPlateau`. This approach adjusts the learning rate in response to training plateaus by monitoring validation loss, akin to early stopping criteria [51]. Despite the recognized benefits of `ReduceLROnPlateau` in contemporary optimization scenarios, our experimentation yielded nuanced insights. While training models with and without learning rate decay, a small performance improvement was observed, but the associated learning curves exhibit clear signs of overfitting (see Fig. 6(a)). In contrast, models trained without learning rate decay did not display such overfitting tendencies (see Fig. 6(b)). These observations are likely attributed to our model's already minimal learning rate. Consequently, in our final model configuration, we opted to forgo the learning rate decay to circumvent the observed overfitting tendencies.

5.2. Comparison of different attention mechanisms

In our approach so far, we have used the SE attention mechanism. In this section, we explore whether alternative attention mechanisms (see Section 2.3) may yield improved performance within the confines of our model. For a uniform evaluation, we incorporate all attention mechanisms after the parallel convolutions (PC). It is also important to note that the model is trained using Dice loss and the results presented here are derived from configurations that do not include the weighted

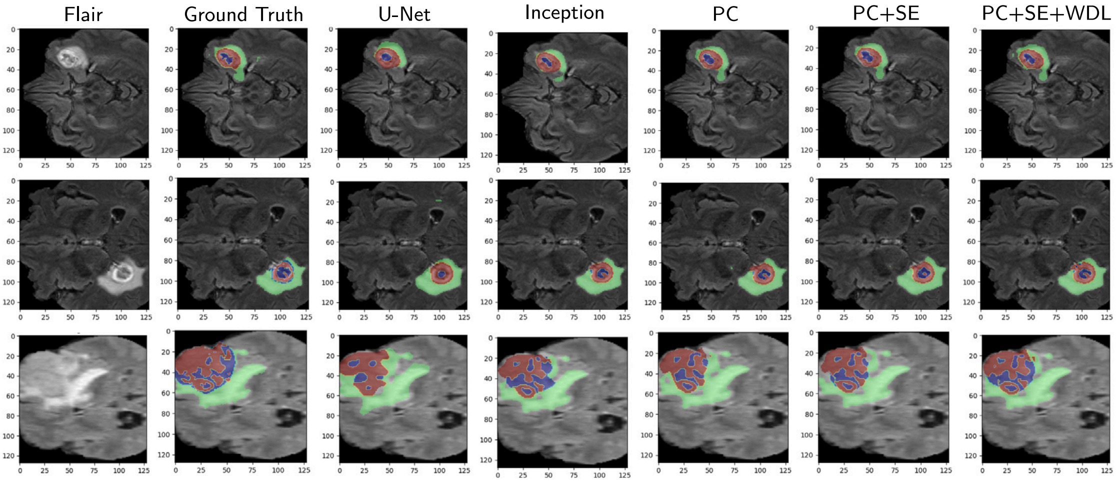

Fig. 5. Segmentation results from the test set that are typical of those produced by the various networks. The results for a single patient from each network are shown in each row. The enhancing tumor is depicted in blue, the necrotic and non-enhancing tumor in red, and the edema in green (after extracting the distinct regions from the partially overlapping segmentation results).

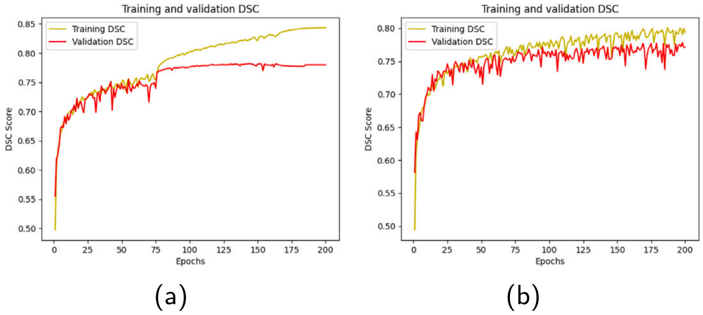

Fig. 6. Modified U-Net with PC: Dice score curves for all tumor regions (a) with and (b) without learning rate decay.

Dice score loss (WDL). This choice was made to isolate the impact of the attention mechanisms themselves, allowing for a clear comparison of their effectiveness in the absence of the WDL's influence.

Table 4 summarizes our findings, detailing the performance metrics of residual SE, CBAM, ECA, and their combinations. While all mechanisms enhance the model's accuracy, SE stands out as the most effective.

CBAM, despite its comprehensive design, might be limited by its complexity, which could result in computational overhead, potentially affecting efficiency, especially in real-time scenarios. On the other hand, while ECA is designed to be lightweight and computationally efficient, it may not capture the intricate inter-channel dependencies as effectively as SE. This trade-off between computational complexity and attention performance suggests that CBAM and ECA may be better suited for specific tasks or resource-constrained environments where accuracy can be sacrificed for speed or memory efficiency.

The SE mechanism's advantage lies in its ability to efficiently highlight enhanced tumors, which is crucial for tasks involving small yet important regions like the tumor core (TC) and enhancing tumor (ET). Additionally, the reduced number of parameters in SE makes it particularly advantageous for large-scale medical image segmentation tasks, where model size and efficiency are important factors. The residual SE mechanism further boosts performance, particularly in terms of Dice similarity coefficient (DSC), as shown by its slightly higher values across all tumor regions.

Analyzing the standard deviation values, the variance in DSC scores appears minimal, which suggests that the choice of attention mechanism might not drastically affect the accuracy. However, in practice, a more streamlined model with fewer parameters (such as PC+SE) may offer competitive or even superior performance due to its balance of effectiveness and computational efficiency.

In contrast, the HD95 metric highlights more prominent differences across the methods, suggesting that attention mechanisms have a varying impact on capturing spatial discrepancies in segmentation, such as outliers or irregularly shaped tumors. For example, CBAM exhibits lower HD95 values for the whole tumor (WT), indicating that its spatial attention might better capture large-scale spatial variations, while SE-based models perform better on smaller, enhancing tumors (ET), where channel-wise attention is more effective at capturing fine-grained details.

Given the observed performance trade-offs, the choice of attention mechanism may depend on the specific task requirements, including segmentation accuracy, computational resources, and the need to balance spatial and channel-wise attention.

5.3. Lightweight model validation

To demonstrate the lightweight nature of LATUP-Net, Table 5 presents a comparison with other models on the floating-point operations (FLOPs) required to evaluate the model once, the number of parameters, and the inference times. Fewer FLOPs indicates a more efficient training process, as less computational power is required. Similarly, a lower number of trainable parameters not only reduces training time but also minimizes memory usage during deployment. Additionally, shorter inference times further highlight the lightweight characteristics of the model. It is important to note that there are various methods for estimating FLOPs. In our study, we used the TensorFlow profiler [52] function from the TensorFlow package, ensuring consistency and accuracy.

Table 4
Efficiency analysis of various attention mechanisms for brain tumor segmentation on the BraTS 2020 test set. This table illustrates the mean and standard deviation (indicated by ±) for the per-sample Dice similarity coefficient (DSC) and the 95th percentile Hausdorff distance (HD95). Results are segmented into whole tumor (WT), tumor core (TC), and enhancing tumor (ET) categories, based on an 80/20 train/test set split. Additionally, the table includes the number of trainable parameters for each model.

Method	DSC (%)			HD95 (mm)		
	WT	TC	ET	WT	TC	ET
PC+SE	88.52 ± 7.10	83.26 ± 17.18	**71.86 ± 27.02**	5.98 ± 7.38	**5.51 ± 5.20**	**12.96 ± 26.55**
PC+CBAM	89.38 ± 6.53	**84.36 ± 15.94**	70.01 ± 29.39	**4.78 ± 4.19**	5.51 ± 5.65	14.12 ± 28.27
PC+CBAM+SE	88.91 ± 6.31	83.87 ± 15.63	70.25 ± 28.33	5.24 ± 5.07	5.58 ± 5.60	12.92 ± 26.42
PC+ (SE-3D)	89.05 ± 6.56	83.91 ± 16.97	69.83 ± 29.63	5.38 ± 6.89	5.72 ± 7.37	13.75 ± 28.14
PC+Residual SE	**89.60 ± 6.50**	84.06 ± 17.38	70.79 ± 29.42	5.40 ± 7.46	5.93 ± 8.16	13.19 ± 26.34
PC+ECA	84.47 ± 6.97	80.12 ± 18.21	63.19 ± 28.15	6.81 ± 4.01	7.08 ± 7.05	15.03 ± 28.13

Table 5
Efficiency comparison of different models based on FLOPs (GigaFLOPs), parameter count (millions), and inference time (milliseconds) measured on an NVIDIA RTX3060 12 GB GPU.

Models	Parameters (M)	FLOPs (G)	Inference time (ms)
U-Net	5.65	23.8	230
Inception	3.65	74.98	333
PC	**3.06**	**15.08**	227
PC + SE	3.07	15.79	**168**
PC + CBAM	3.45	17.35	247
PC + CBAM + SE	3.26	16.08	240
PC + (SE-3D)	3.07	15.79	234
PC + Residual SE	3.07	15.80	238
PC + ECA	3.10	15.80	212

The baseline 3D U-Net, with 5.65 million parameters and 23.8 GFLOPs, is one of the more complex models, particularly in terms of parameters, reflecting its deeper and wider architecture. In contrast, Inception has fewer parameters (3.65 million) but a much higher GFLOP count (74.98), suggesting that while it has fewer parameters, its increased computational demand arises from its more intricate layer configurations.

Compared to Inception, LATUP-Net (PC + SE) achieves a more balanced and efficient design, with 15.79 GFLOPs, 3.07 million parameters, and an inference time of 168 ms, measured using an NVIDIA RTX3060 GPU. It is important to note that inference and training times are highly dependent on the hardware used, which is why these times are not compared with other state-of-the-art models in this study. Despite this, LATUP-Net is far more efficient computationally. The PC-based variants (PC, PC + SE, and PC + Residual SE) maintain a similar parameter range (3.06 – 3.07 million) and low GFLOP counts (15.08 – 16.08), highlighting their efficiency. While models like PC + CBAM and PC + ECA slightly increase GFLOPs, they still maintain efficient inference times. Overall, LATUP-Net provides significant advantages over Inception by delivering better performance with fewer computations, making it a more efficient model, especially in terms of computational complexity and speed.

5.4. Evaluating the effectiveness of the attention mechanism

To understand the impact of the integrated attention mechanism and to determine if it operates as intended in our proposed architecture, we conduct two primary experiments: gradient-weighted class activation mapping (Grad-CAM) visualizations [20] and confusion matrix analysis.

5.4.1. Visual interpretation using Grad-CAM

In investigating the efficacy of attention mechanisms in brain tumor detection, the question arises whether the local context is sufficient for accurate segmentation. Traditional convolutions inherently capture the local context without necessitating attention. Given the qualitative nature of our experiments and the challenges of visualizing every data point, it is crucial to strategically select representative samples for thorough analysis. To interpret and delve deeper into the model's behavior, Grad-CAM is applied to three distinct samples that are the same for both models, with attention (LATUP-Net) and without attention (PC + WDL). These samples are selected based on their loss value with some representative slice from the test set and represent the best-performing, the median, and the worst-performing w.r.t. the WDL for comparative analysis.

Grad-CAM [20] was applied to the layer preceding the softmax activation used for generating the predictions. This layer was particularly selected because it directly contributes to the decision-making process, offering a transparent view into how the model weighs different regions in its segmentation task. Grad-CAM visualizations, as shown in Fig. 7, demonstrate that both models focus primarily on tumor areas. While the attention mechanism intensifies this focus, it does not always lead to the most accurate predictions. In particular, the necrotic region tends to be overemphasized, leading to misclassifications, as necrotic and enhancing regions often share similar textures.

This observation suggests that while attention appears to work as designed by emphasizing certain regions, it may not contribute to improved performance due to its reliance on local features. Grad-CAM further reveals that the segmentation task is heavily driven by these local features, like texture and boundaries, which attention does not necessarily enhance in a meaningful way. Additionally, because Grad-CAM tends to focus on the regions directly involved in segmentation, it does not provide insights into global relationships, such as left–right symmetry of the brain, which the attention mechanism might have been expected to capture.

Overall, the results indicate that the attention mechanism may have limited impact on segmentation performance, as the task is largely dependent on local feature extraction, and the attention mechanism does not provide significant additional context in this particular use case.

5.4.2. Confusion matrix analysis and implications

To provide quantitative evidence supporting the findings from Grad-CAM visualizations, we perform a confusion matrix analysis, as shown in Fig. 8. This matrix details how pixels from one class are misclassified as those from another, offering insights into the LATUP-Net model behavior. It is important to note that there were patients with no enhanced tumor (ET) class in the ground truth. These cases have not been considered for the confusion matrix, as including them would distort the percentages. A primary observation is the misclassification between the necrotic and edema regions, a claim also derived from the qualitative visuals (see Section 5.4.1). The model perceives textural similarities between these areas, further complicated by nested tumor structures. Additionally, the necrotic region is notably vulnerable to misclassification, perhaps due to its texture and position within the tumor's structure. When examining the standard deviation in Fig. 9, combined with the average misclassification percentages, the classification performance variability is clear. This variability might spotlight outliers or deviations in model predictions, with classes like the necrotic region and enhancing region showing significant misclassification variability. The variations imply that certain samples heavily influence averages. In conclusion, the attention mechanism, while offering nuanced insights,

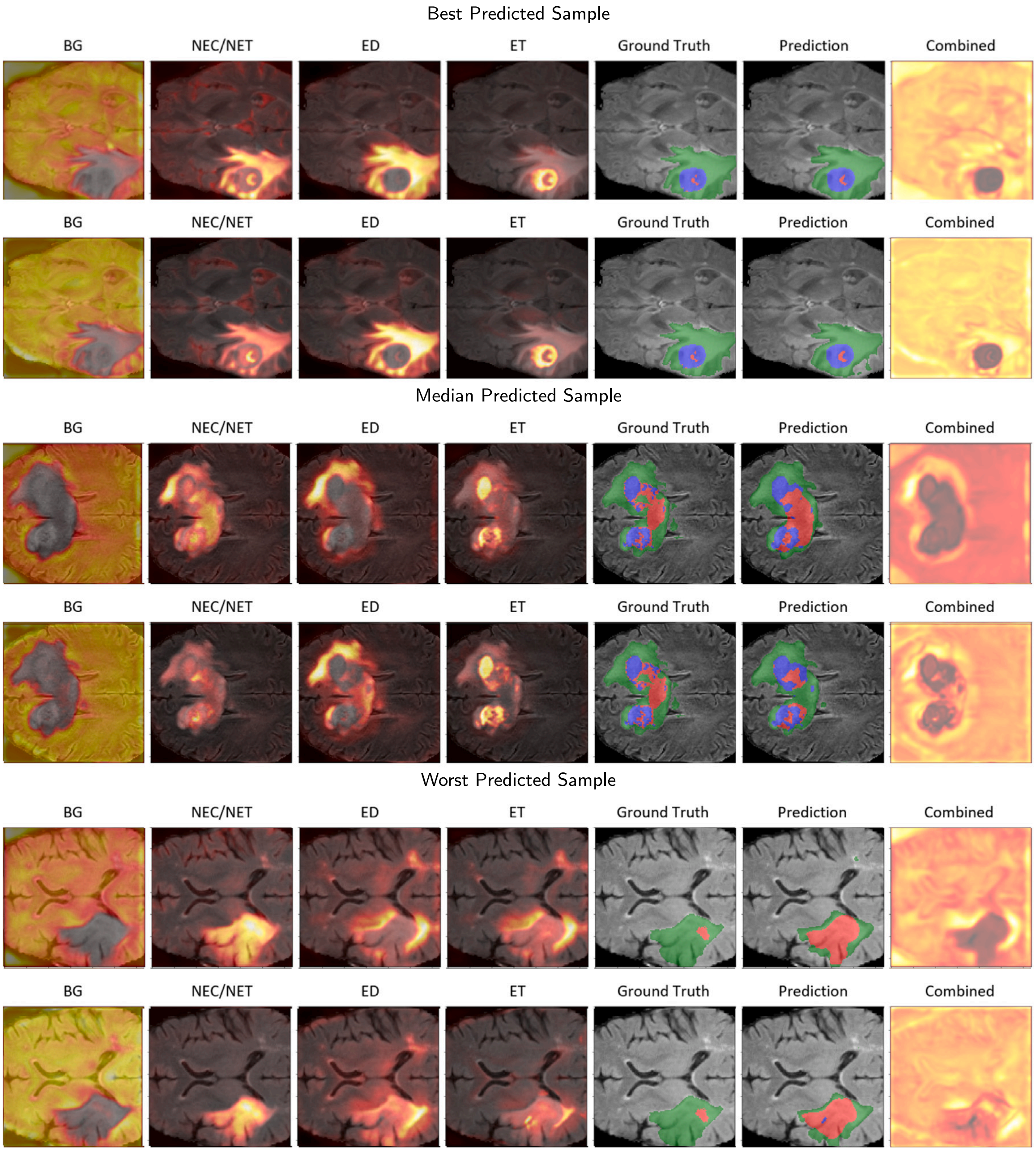

Fig. 7. Visual interpretations of model predictions using Grad-CAM: For each of the three samples selected from the test set, representing the best, median, and worst predicted cases, both the standard and attention-enhanced models. From left to right in each row we show the GradCAM for the BG, NEC/NET, ED, ET output channel, the ground truth and the prediction, and finally the combined GradCAM for all output channels. The first row visualizes the standard model (PC + WDL), and the second row visualizes the attention-enhanced model (LATUP-Net) per case.

can overly prioritize local features, causing the model to overlook important topological relationships recognized by human experts. A balance between local and global contexts seems important.

5.5. Comparison with state-of-the-art models

We perform five-fold cross-validation, with results shown in Figs. 10 and 11. These figures display the per-sample performance of our LATUP-Net model, highlighting the distribution of DSC and HD95 metrics, as well as variability, including outliers. Unlike other studies that report only average metrics, we provide a detailed analysis to demonstrate the model's consistency across test samples.

We then compare the average results of our five-fold cross-validation with other high-performing models, as shown in Tables 6 and 7. For these comparisons, we rely on the evaluation results reported in related publications, a common practice in brain tumor segmentation research. This approach is necessary because the source code for many existing methods is unavailable, and it helps avoid the bias that could be introduced by retraining models.

5.5.1. Comparison with BraTS 2020 results

We compare our LATUP-Net model with various state-of-the-art models on the BraTS 2020 dataset. The evaluation focuses on three key metrics: the Dice similarity coefficient (DSC), the 95th percentile Hausdorff distance (HD95), and the number of model parameters, alongside GFLOPs to reflect computational complexity. For our LATUP-Net model, we present the average performance across five-fold cross-validation to ensure a robust and comprehensive assessment. In contrast, for the state-of-the-art models, we report the results as they

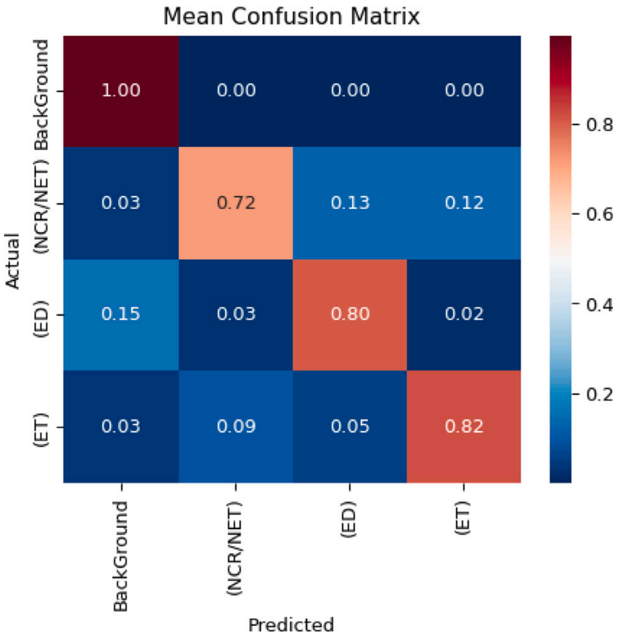

Fig. 8. Confusion matrix illustrating the pixel-wise misclassification rates between different tumor regions, normalized by the number of actual class instances in each row and predicted class. These values are averaged over all samples in the dataset to provide a comprehensive overview of the LATUP-Net model's performance.

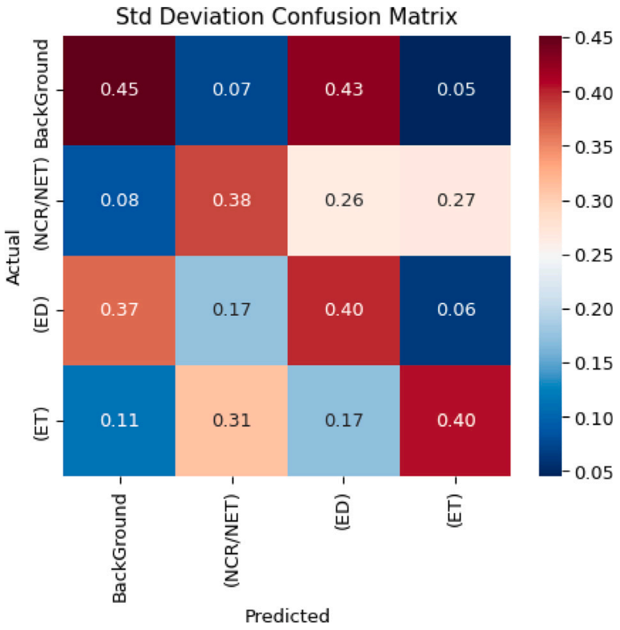

Fig. 9. Confusion matrix of the standard deviation of misclassification rates for each tumor region from Fig. 8 of the LATUP-Net model.

are presented in their respective original publications, which may include their best-split results or cross-validation outcomes. Detailed results of this comparative analysis are compiled in Table 6. Our LATUP-Net model demonstrates significant improvements in HD95 for all tumor regions (whole tumor, tumor core, and enhancing tumor), suggesting accurate predictions of tumor boundaries. This precision is of paramount importance in medical imaging since accurate boundaries can greatly impact clinical decision-making. However, it is worth noting that in the final evaluation of the five-fold cross-validation, the HD95 has been ignored for images that do not have clear boundaries, since having unclear, ambiguous 'edges' from which to measure the distances may mislead the results.

Our model surpasses several others in DSC measurements for whole tumor segmentation. Specifically, it outperforms Raza et al. [53], Ballestar et al. [54], and Messaoudi et al. [55] by 1.81, 4.2, and 7.73 respectively. Additionally, it achieves a notable improvement in efficiency, having only 3.07 M parameters compared to 181.56 [44], 32.99 [23], and 30.47 [53]. Moreover, while Wang et al. [23] model requires 333 GFLOPs, and Raza et al. [53] use 374.04 GFLOPs, our model operates with only 15.79 GFLOPs, highlighting its computational efficiency which is a crucial factor in real-world clinical applications.

When comparing the latest study by Zhu et al. [10], which demonstrates high performance with a DSC of 90.22 for the whole tumor, our model's 88.41 is still competitive. However, LATUP-Net surpasses Zhu's model in HD95, especially in tumor core and enhancing tumor predictions, demonstrating that it excels in boundary accuracy—a critical aspect in segmentation tasks. Although Zhu's model outperforms in DSC across tumor regions, the computational complexity and parameter count are not mentioned. However, by incorporating transformer into the U-Net architecture, it is a moderately sized model compared to LATUP-Net, considering its performance trade-offs.

It is important to note that LATUP-Net's superior performance is not due to common confounding factors, such as ensembling or leveraging additional data. This aligns with the recent recommendations [9], which cautions against practices that artificially inflate a model's performance. Our architecture's innovations — parallel convolution strategies and lightweight attention mechanisms — demonstrate genuine advancements in segmentation tasks without introducing unnecessary complexity or relying on inflated hardware resources. These results underscore the efficacy of simpler, well-configured models in achieving competitive performance in resource-constrained settings.

Although a deeper analysis is needed to fully determine the mechanisms behind this efficiency, the results from the earlier model comparison section provide concrete evidence of the effectiveness of our approach. Specifically, our adoption of parallel convolutions in the first encoder block seems to play a crucial role. This is evident as the PC model has 2.59 million fewer parameters than U-Net, yet achieves rapid convergence during the initial training epochs, leading to shorter training times and reduced computational needs.

Brain tumor segmentation, particularly in the tumor core and enhancing tumor regions, poses significant challenges. While models like nnU-Net exhibit strong results, our LATUP-Net model achieves similar scores. It is critical to differentiate between DSC and HD95 scores. Although our HD95 scores indicate highly accurate boundary predictions, our DSC for the enhancing tumor (ET) is 73.67%, which is not the highest, highlighting an area for improvement in volumetric overlap with the ground truth.

In terms of the number of parameters, our model is remarkably efficient, with only 3.07 million parameters, a stark contrast to nnU-Net's 181.56 million parameters. This efficiency reduces both computational costs and processing time, which is critical for real-time clinical applications, while maintaining comparable or even better performance in some aspects, as seen in the HD95 results. Although Raza et al. [53] has a parameter count closer to ours, their performance does not match ours, underscoring the effectiveness of our architecture.

5.5.2. Comparison with BraTS 2021 results

Upon evaluating our LATUP-Net model on the BraTS 2021 dataset, it continues to show promising results, particularly in segmenting the tumor core, with a DSC of 89.54%, outperforming several models. Compared to Hatamizadeh et al. [62], who also achieved competitive results across tumor regions, LATUP-Net demonstrates a clear advantage in computational efficiency. Our model has only 3.07 million parameters and requires 15.79 GFLOPs, significantly less than Hatamizadeh's model, which requires 61.98 million parameters and 394.84 GFLOPs. This efficiency underscores the suitability of LATUP-Net for practical

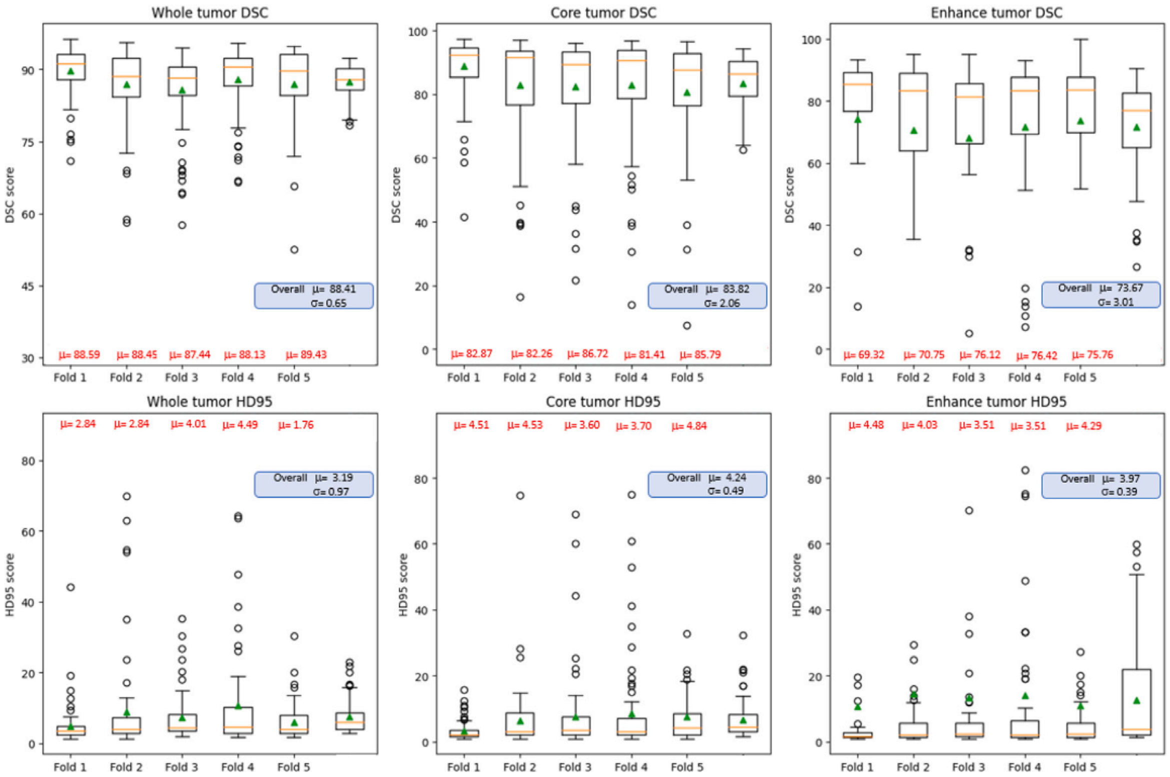

Fig. 10. Boxplots of the DSC and HD95 metrics measured per sample (patient) on the BraTS 2020 five-fold cross-validation results with mean μ and standard deviation σ using the LATUP-Net model. The orange line within each boxplot represents the median of the data. The green triangles represent the mean, and the circles denote the outliers. We also show the average distributions over all five folds. This figure provides a detailed view of our model's consistency and variability across samples.

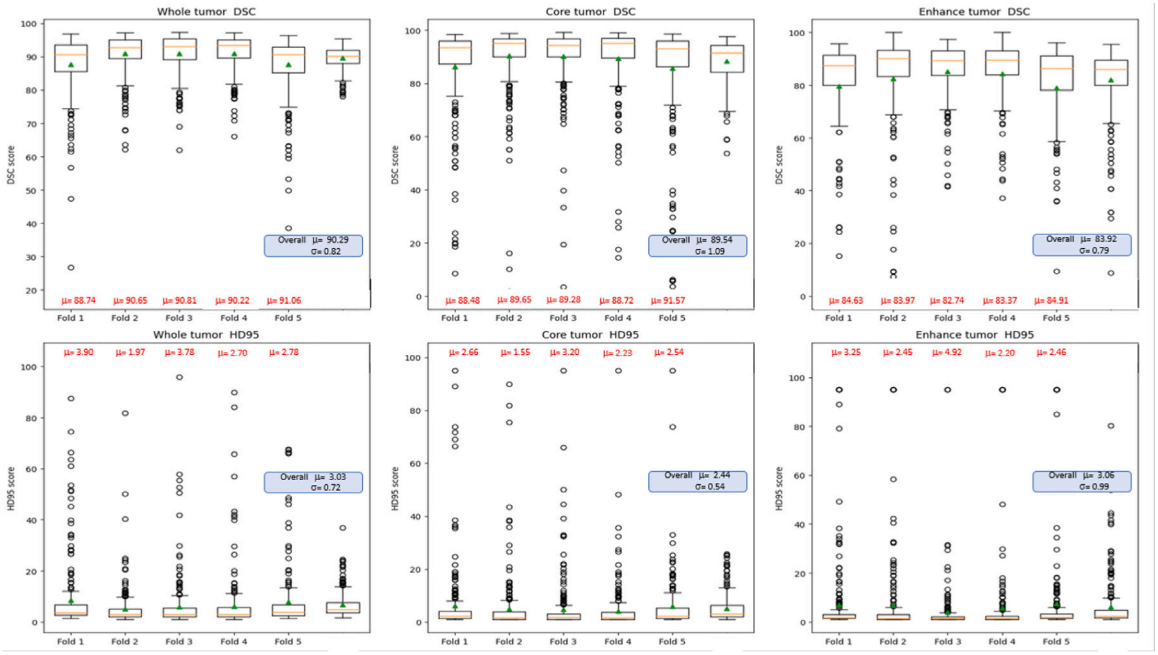

Fig. 11. Boxplots of the DSC and HD95 metrics measured per sample (patient) on the BraTS 2021 five-fold cross-validation results with mean μ and standard deviation σ using the LATUP-Net model. The orange line within each boxplot represents the median of the data. The green triangles represent the mean, and the circles denote the outliers. We also show the average distributions over all five folds. This figure provides a detailed view of our model's consistency and variability across samples.

clinical use, especially in environments where computational resources may be limited.

In terms of HD95, LATUP-Net consistently achieves better results, with scores of 3.03 for the whole tumor and 2.44 for the tumor core. Zhu et al. [10], who achieved top DSC scores for the BraTS 2021 dataset, also produced competitive results in HD95. Nevertheless, LATUP-Net's reduced parameter count and GFLOPs stand out as key differentiators, especially for real-time clinical applications where faster processing times are essential.

Table 6
Comparison of the performance and complexity of different methods on the BraTS 2020 test set. WT – whole tumor, TC – tumor core, ET – enhancing tumor. Results highlighted in red indicate the best result, while those in blue represent the second best. The symbol - indicates values not provided in the original paper (see [10,23,44,53–56]).

Study	WT		TC		ET		Parameters	GFLOPs
	DSC (%)	HD95 (mm)	DSC (%)	HD95 (mm)	DSC (%)	HD95 (mm)		
3D U-Net caseline	83.58	18.50	82.19	15.38	68.76	19.34	5.65 M	23.08
Isensee et al. [44]	88.95	8.498	85.06	17.337	82.03	17.805	181.56 M	–
Tang et al. [56]	89.29	4.62	78.97	10.07	70.30	34.30	–	–
Ballestar et al. [54]	84.21	20.40	75.03	12.92	61.75	48.76	–	–
Wang et al. [23]	90.09	4.96	81.73	9.76	78.73	17.94	32.99 M	333
Messaoudi et al. [55]	80.68	–	75.20	–	69.59	–	–	–
Raza et al. [53]	86.60	–	83.57	–	80.04	–	30.47 M	374.04
Zhu et al. [10]	90.22	4.03	89.20	3.30	82.48	2.29	–	–
LATUP-Net (proposed)	88.41	3.19	83.82	4.24	73.67	3.97	3.07 M	15.79

Table 7
Comparison of the performance and complexity of different methods on the BraTS 2021 test set. WT – whole tumor, TC – tumor core, ET – enhancing tumor. Results highlighted in red indicate the best result, while those in blue represent the second best. The symbol - indicates values not provided in the original paper (see [10,57–63]).

Study	WT		TC		ET		Parameters	GFLOPs
	DSC (%)	HD95 (mm)	DSC (%)	HD95 (mm)	DSC (%)	HD95 (mm)		
Peiris et al. [57]	90.77	5.37	85.39	8.5	81.38	21.83	–	–
Akbar et al. [58]	89.07	11.78	80.73	21.17	78.02	25.8	–	–
Jia et al. [59]	92.53	3.45	87.96	5.86	84.80	14.17	17.91 M	449.79
Li et al. [60]	90.18	6.15	81.61	16.65	76.89	30.21	–	–
Ma et al. [61]	92.59	3.80	87.86	9.20	82.17	21.09	–	–
Hatamizadeh et al. [62]	92.6	5.83	88.5	3.77	85.8	6.01	61.98 M	394.84
Roth et al. [63]	90.6	4.54	83.5	10.11	79.2	16.61	–	–
Zhu et al. [10]	93.10	3.58	90.99	3.27	87.64	2.57	–	–
LATUP-Net (proposed)	90.29	3.03	89.54	2.44	83.92	3.06	3.07 M	15.79

Our model's DSC for the whole tumor and the enhancing tumor are 90.29% and 83.92%, respectively. When we state that these scores are in line with leading models, we specifically refer to the work of Hatamizadeh et al., Jia et al. [59], and Ma et al. [61] (see Table 7).

In summary, the LATUP-Net model offers a fine balance between segmentation accuracy and computational efficiency. Despite its compact architecture, it delivers strong performance, particularly in HD95 metrics, which are crucial for precise tumor boundary delineation. With only 3.07 million parameters and 15.79 GFLOPs, LATUP-Net is highly suited for resource-constrained environments, offering a feasible solution for real-time clinical deployment. While models like Zhu et al. surpass LATUP-Net in certain DSC scores, the efficiency and competitive performance across HD95 metrics highlight the real-world applicability of our model in brain tumor segmentation tasks.

5.6. Limitations and future directions

Our model leverages multi-sequence MRI data for brain tumor segmentation and has demonstrated good performance on the BraTS 2020 and 2021 datasets, particularly when compared to other lightweight models working with similar data. However, we acknowledge that our work was trained on a relatively small dataset of about 1200 patients, which limits our understanding of the full covariance of potential data variations across different centers. This limitation is not unique to our study, as many segmentation models face similar challenges when dealing with multi-center data variations, especially regarding the availability of diverse data sources. We believe that access to larger, more heterogeneous datasets could help address these limitations, but this is an ongoing challenge within the field.

Other key limitations of our study lie in the consistency of model and parameter selection. While we have designed an effective architecture, the process of consistently selecting optimal models and fine-tuning hyperparameters has not been as rigorous as it could be. This challenge is partly due to the availability of computational resources, which restricts us from systematically testing a wider range of models and configurations. A more thorough exploration of model and parameter selection would be ideal, but this would require significant additional resources. As a result, we acknowledge that the model selection process could be improved and we plan to address this in future work.

Future work also includes adapting the architecture to other medical imaging segmentation tasks and refining the balance between attention and convolutional features, particularly to enhance our model's sensitivity to tumor regions and reduce variations in segmentation performance across different regions. Another crucial direction will be to explore the model's robustness when faced with incomplete or missing MRI modalities, a common scenario in real-world clinical settings. Ensuring that LATUP-Net can maintain high segmentation performance even when some modalities are unavailable will enhance its practical applicability in diverse medical environments. Additionally, we are currently studying the robustness and explainability of our model, which are critical for clinical applications [64]. However, this ongoing research extends beyond the scope of the present paper and will be addressed in future studies.

6. Conclusion

In this work, we have unveiled the LATUP-Net network, an enhanced U-Net variant for 3D brain tumor segmentation designed to be lightweight in its computational demand. This model substantially decreases the number of parameters needed while maintaining, and in some aspects surpassing, the segmentation performance of state-of-the-art methods. With 3.07 M parameters, about 59 times fewer parameters than the state-of-the-art nnU-Net with 181.56 M parameters, LATUP-Net underscores an advancement where efficient modeling coupled with parallel convolutions can lead to a significant reduction in overfitting risk and more judicious use of computational resources.

Our model demonstrates an impressive ability to delineate tumor boundaries with high accuracy, as evidenced by its performance in Hausdorff distance (HD95) measurements. These achievements indicate the model's potential utility in clinical settings, where precise segmentations are integral to formulating effective treatment plans. Furthermore, LATUP-Net's lightweight architecture, requiring only 15.79

GFLOPs, makes it particularly suitable for deployment in resource-constrained environments, such as developing countries, where computational resources may be limited.

A pivotal aspect of our research is incorporating attention mechanisms, which refine our model's capability to focus on salient features within MRI scans. Our comparative analysis across different attention mechanisms, such as SE, CBAM, and ECA, reveals that while all contribute to accuracy improvements, SE provides a balance between performance and parameter efficiency, particularly in delineating enhanced tumors. However, the enhancements brought by attention are found to be nuanced. The slight underperformance in Dice score coefficients for enhancing tumor segmentation suggests that attention mechanisms do not unilaterally enhance performance across all regions. This is corroborated by gradient-weighted class activation mapping (Grad-CAM) and confusion matrix analyses. These investigations highlight scenarios where attention mechanisms seem to focus too narrowly on local features, occasionally at the expense of contextual understanding, leading to potential misclassification between regions with texturally similar features. The attention-enhanced model, while showing promise in segmenting small regions, also illustrates that there are instances where traditional convolutions may suffice and that the features they capture can be integral to achieving precise segmentations.

The LATUP-Net model stands as a testament to the possibility of achieving state-of-the-art performance with a fraction of the computational cost, highlighting a promising direction for medical image analysis research and the development of practical, accessible tools for brain tumor segmentation. In real-world clinical applications, however, the dependency on multi-sequence MRI data presents a practical challenge, as some modalities may be unavailable in certain settings. Addressing this limitation is a key consideration for future work.

CRediT authorship contribution statement

Ebtihal J. Alwadee: Conceptualization, Methodology, Software, Investigation, Writing – original draft. **Xianfang Sun:** Supervision, Writing – review & editing. **Yipeng Qin:** Supervision, Writing – review & editing. **Frank C. Langbein:** Software, Data Curation, Supervision, Writing – review & editing.

Declaration of competing interest

We have no conflicts of interest to disclose.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.compbiomed.2024.109353.

References

[1] B.H. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, Y. Burren, N. Porz, J. Slotboom, R. Wiest, et al., The multimodal brain tumor image segmentation benchmark (BRATS), IEEE Trans. Med. Imaging 34 (10) (2014) 1993–2024.

[2] A. Işın, C. Direkoğlu, M. Şah, Review of MRI-based brain tumor image segmentation using deep learning methods, Procedia Comput. Sci. 102 (2016) 317–324.

[3] S. Bakas, H. Akbari, A. Sotiras, M. Bilello, M. Rozycki, J.S. Kirby, J.B. Freymann, K. Farahani, C. Davatzikos, Advancing the cancer genome atlas glioma MRI collections with expert segmentation labels and radiomic features, Sci. Data 4 (1) (2017) 1–13.

[4] S. Bakas, M. Reyes, A. Jakab, S. Bauer, M. Rempfler, A. Crimi, R.T. Shinohara, C. Berger, S.M. Ha, M. Rozycki, et al., Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the BRATS challenge, 2018, arXiv:1811.02629.

[5] H. Li, A. Li, M. Wang, A novel end-to-end brain tumor segmentation method using improved fully convolutional networks, Comput. Biol. Med. 108 (2019) 150–160.

[6] M. Wieczorek, J. Siłka, M. Woźniak, S. Garg, M.M. Hassan, Lightweight convolutional neural network model for human face detection in risk situations, IEEE Trans. Ind. Inform. 18 (7) (2021) 4820–4829.

[7] M. Woźniak, J. Siłka, M. Wieczorek, Deep neural network correlation learning mechanism for CT brain tumor detection, Neural Comput. Appl. (2021) 1–16.

[8] O. Ronneberger, P. Fischer, T. Brox, U-Net: Convolutional networks for biomedical image segmentation, in: Medical Image Computing and Computer-Assisted Intervention–MICCAI: 18th International Conference, Proceedings, Part III 18, Springer, 2015, pp. 234–241.

[9] F. Isensee, T. Wald, C. Ulrich, M. Baumgartner, S. Roy, K. Maier-Hein, P.F. Jaeger, nnU-Net revisited: A call for rigorous validation in 3d medical image segmentation, 2024, arXiv preprint arXiv:2404.09556.

[10] Z. Zhu, M. Sun, G. Qi, Y. Li, X. Gao, Y. Liu, Sparse dynamic volume TransUNet with multi-level edge fusion for brain tumor segmentation, Comput. Biol. Med. (2024) 108284.

[11] Y. Xu, K. Yu, G. Qi, Y. Gong, X. Qu, L. Yin, P. Yang, Brain tumour segmentation framework with deep nuanced reasoning and Swin-T, IET Image Process. 18 (6) (2024) 1550–1564.

[12] L. Wu, S. Hu, C. Liu, MR brain segmentation based on DE-ResUnet combining texture features and background knowledge, Biomed. Signal Process. Control 75 (2022) 103541.

[13] D.R. Sarvamangala, R.V. Kulkarni, Convolutional neural networks in medical image understanding: a survey, Evol. Intell. 15 (2022) 1–22.

[14] J. Beutel, Handbook of Medical Imaging, vol. 3, Spie Press, 2000.

[15] T. Ma, K. Wang, F. Hu, LMU-Net: lightweight U-shaped network for medical image segmentation, Med. Biol. Eng. Comput. (2023) 1–10.

[16] A. Vaswani, Attention is all you need, Adv. Neural Inf. Process. Syst. (2017).

[17] J. Park, Bam: Bottleneck attention module, 2018, arXiv preprint arXiv:1807.06514.

[18] S. Woo, J. Park, J.-Y. Lee, I.S. Kweon, CBAM: Convolutional block attention module, in: Proceedings of the European Conference on Computer Vision, ECCV, 2018, pp. 3–19.

[19] J. Hu, L. Shen, G. Sun, Squeeze-and-excitation networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 7132–7141.

[20] R.R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, D. Batra, Grad-cam: Visual explanations from deep networks via gradient-based localization, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 618–626.

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

[22] W. Chen, B. Liu, S. Peng, J. Sun, X. Qiao, S3D-UNet: separable 3D U-Net for brain tumor segmentation, in: Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 4th International Workshop, Part II 4, Springer, 2019, pp. 358–368.

[23] W. Wang, C. Chen, M. Ding, H. Yu, S. Zha, J. Li, TransBTS: Multimodal brain tumor segmentation using transformer, in: Medical Image Computing and Computer Assisted Intervention–MICCAI: 24th International Conference, Proceedings, Part I 24, Springer, 2021, pp. 109–119.

[24] Z. Zhu, X. He, G. Qi, Y. Li, B. Cong, Y. Liu, Brain tumor segmentation based on the fusion of deep semantics and edge information in multimodal MRI, Inf. Fusion 91 (2023) 376–387.

[25] Y. Liu, X. Zhang, G. Cai, Y. Chen, Z. Yun, Q. Feng, W. Yang, Automatic delineation of ribs and clavicles in chest radiographs using fully convolutional DenseNets, Comput. Methods Programs Biomed. 180 (2019) 105014.

[26] S.L. Oh, E.Y. Ng, R. San Tan, U.R. Acharya, Automated beat-wise arrhythmia diagnosis using modified U-Net on extended electrocardiographic recordings with heterogeneous arrhythmia types, Comput. Biol. Med. 105 (2019) 92–101.

[27] Z. Liu, Y.-Q. Song, V.S. Sheng, L. Wang, R. Jiang, X. Zhang, D. Yuan, Liver CT sequence segmentation based with improved U-Net and graph cut, Expert Syst. Appl. 126 (2019) 54–63.

[28] Z. Zhang, C. Wu, S. Coleman, D. Kerr, DENSE-INception U-Net for medical image segmentation, Comput. Methods Programs Biomed. 192 (2020) 105395.

[29] C. Chen, X. Liu, M. Ding, J. Zheng, J. Li, 3D dilated multi-fiber network for real-time brain tumor segmentation in MRI, in: Medical Image Computing and Computer Assisted Intervention–MICCAI: 22nd International Conference, Proceedings, Part III 22, Springer, 2019, pp. 184–192.

[30] Z. Luo, Z. Jia, Z. Yuan, J. Peng, HDC-Net: Hierarchical decoupled convolution network for brain tumor segmentation, IEEE J. Biomed. Health Inf. 25 (3) (2020) 737–745.

[31] T. Magadza, S. Viriri, Brain tumor segmentation using partial depthwise separable convolutions, IEEE Access 10 (2022) 124206–124216.

[32] Z. Zhu, Z. Wang, G. Qi, N. Mazur, P. Yang, Y. Liu, Brain tumor segmentation in MRI with multi-modality spatial information enhancement and boundary shape correction, Pattern Recognit. 153 (2024) 110553.

[33] A.G. Roy, N. Navab, C. Wachinger, Concurrent spatial and channel 'squeeze & excitation'in fully convolutional networks, in: Medical Image Computing and Computer Assisted Intervention–MICCAI: 21st International Conference, Proceedings, Part I, Springer, 2018, pp. 421–429.

[34] D. Ulyanov, A. Vedaldi, V. Lempitsky, Instance normalization: The missing ingredient for fast stylization, 2016, arXiv:1607.08022.

[35] E. Alwadee, F.C. Langbein, BCa - Brain cancer segmentation python package, version 1.0, software, 2024, URL: https://qyber.black/ca/code-bca.

[36] K.T. Rajamani, P. Rani, H. Siebert, R. ElagiriRamalingam, M.P. Heinrich, Attention-augmented U-Net (AA-U-Net) for semantic segmentation, Signal Image Video Process. 17 (4) (2023) 981–989.

[37] X. Chen, Research on Algorithm and Application of Deep Learning Based on Convolutional Neural Network, Zhejiang Gongshang University, 2014.

[38] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, Going deeper with convolutions, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 1–9.

[39] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, Q. Hu, ECA-Net: Efficient channel attention for deep convolutional neural networks, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 11534–11542.

[40] J. Gu, X. Sun, Y. Zhang, K. Fu, L. Wang, Deep residual squeeze and excitation network for remote sensing image super-resolution, Remote Sens. 11 (15) (2019) 1817.

[41] S. Raschka, Model evaluation, model selection, and algorithm selection in machine learning, 2018, arXiv:1811.12808.

[42] S. Patro, K.K. Sahu, Normalization: A preprocessing stage, 2015, arXiv:1503.06462.

[43] D.P. Kingma, J. Ba, Adam: A method for stochastic optimization, 2014, arXiv:1412.6980.

[44] F. Isensee, P.F. Jäger, P.M. Full, P. Vollmuth, K.H. Maier-Hein, nnU-Net for brain tumor segmentation, in: Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 6th International Workshop, BrainLes 2020, Part II 6, Springer, 2021, pp. 118–132.

[45] A.L. Maas, A.Y. Hannun, A.Y. Ng, et al., Rectifier nonlinearities improve neural network acoustic models, in: Proc. ICML, Vol. 28, 2013, p. 3.

[46] E. Alwadee, F.C. Langbein, BCa segmentation results: LATUPNet, version 1.0. software and data, 2024, URL: https://qyber.black/ca/results-bca-latup.

[47] S. Bakas, H. Akbari, A. Sotiras, M. Bilello, M. Rozycki, J. Kirby, J. Freymann, K. Farahani, C. Davatzikos, Segmentation labels and radiomic features for the pre-operative scans of the TCGA-LGG collection, Cancer Imaging Arch. 286 (2017).

[48] S.A. Taghanaki, Y. Zheng, S.K. Zhou, B. Georgescu, P. Sharma, D. Xu, D. Comaniciu, G. Hamarneh, Combo loss: Handling input and output imbalance in multi-organ segmentation, Comput. Med. Imaging Graph. 75 (2019) 24–33.

[49] T.-Y. Lin, P. Goyal, R. Girshick, K. He, P. Dollár, Focal loss for dense object detection, IEEE Trans. Pattern Anal. Mach. Intell. 42 (2) (2018) 318–327.

[50] A. Paszke, A. Chaurasia, S. Kim, E. Culurciello, Enet: A deep neural network architecture for real-time semantic segmentation, 2016, arXiv:1606.02147.

[51] T. Vo, P. Dave, G. Bajpai, R. Kashef, N. Khan, Brain tumor segmentation in MRI images using a modified U-Net model, in: 2022 IEEE International Conference on Digital Health, ICDH, 2022, pp. 29–33.

[52] T. Authors, Profiler guide, 2024, https://www.tensorflow.org/guide/profiler. (Accessed 15 October 2024).

[53] R. Raza, U.I. Bajwa, Y. Mehmood, M.W. Anwar, M.H. Jamal, dResU-Net: 3D deep residual U-Net based brain tumor segmentation from multimodal MRI, Biomed. Signal Process. Control 79 (2023) 103861.

[54] L.M. Ballestar, V. Vilaplana, Brain tumor segmentation using 3D-CNNs with uncertainty estimation, 2020, arXiv:2009.12188.

[55] H. Messaoudi, A. Belaid, M.L. Allaoui, A. Zetout, M.S. Allili, S. Tliba, D. Ben Salem, P.-H. Conze, Efficient embedding network for 3D brain tumor segmentation, in: Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 6th International Workshop, BrainLes 2020, Part I 6, Springer, 2021, pp. 252–262.

[56] J. Tang, T. Li, H. Shu, H. Zhu, Variational-autoencoder regularized 3D MultiResUNet for the BraTS 2020 brain tumor segmentation, in: Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 6th International Workshop, BrainLes 2020, Part II 6, Springer, 2021, pp. 431–440.

[57] H. Peiris, Z. Chen, G. Egan, M. Harandi, Reciprocal adversarial learning for brain tumor segmentation: a solution to BraTS challenge 2021 segmentation task, in: International MICCAI Brainlesion Workshop, Springer, 2021, pp. 171–181.

[58] A.S. Akbar, C. Fatichah, N. Suciati, Unet3D with multiple atrous convolutions attention block for brain tumor segmentation, in: International MICCAI Brainlesion Workshop, Springer, 2021, pp. 182–193.

[59] H. Jia, C. Bai, W. Cai, H. Huang, Y. Xia, HNF-NetV2 for brain tumor segmentation using multi-modal MR imaging, in: International MICCAI Brainlesion Workshop, Springer, 2021, pp. 106–115.

[60] Z. Li, Z. Shen, J. Wen, T. He, L. Pan, Automatic brain tumor segmentation using multi-scale features and attention mechanism, in: International MICCAI Brainlesion Workshop, Springer, 2021, pp. 216–226.

[61] J. Ma, J. Chen, NnUNet with region-based training and loss ensembles for brain tumor segmentation, in: International MICCAI Brainlesion Workshop, Springer, 2021, pp. 421–430.

[62] A. Hatamizadeh, V. Nath, Y. Tang, D. Yang, H.R. Roth, D. Xu, Swin UNTER: Swin transformers for semantic segmentation of brain tumors in MRI images, in: International MICCAI Brainlesion Workshop, Springer, 2021, pp. 272–284.

[63] J. Roth, J. Keller, S. Franke, T. Neumuth, D. Schneider, Multi-plane UNet++ ensemble for glioblastoma segmentation, in: International MICCAI Brainlesion Workshop, Springer, 2021, pp. 285–294.

[64] E. Alwadee, X. Sun, Y. Qin, F. Langbein, Assessing and enhancing the robustness of brain tumor segmentation using a probabilistic deep learning architecture, in: ISMRM 2024 Conference Proceedings, International Society for Magnetic Resonance in Medicine (ISMRM), 2024, URL: https://submissions.mirasmart.com/ISMRM2024/Itinerary/ConferenceMatrixEventDetail.aspx?ses=D-173, Abstract 4526, Session 47.

Ebtihal J. Alwadee received her B.Sc. (Hons) in Information Systems from King Khalid University, Saudi Arabia, in 2013 and her M.Sc. in Computer Science from California State University, Fullerton, USA, in 2020. She is currently a Ph.D. candidate at Cardiff University, with research interests in visual computing, healthcare, deep learning, medical image segmentation, and explainable AI.

Xianfang Sun received a Ph.D. degree from the Institute of Automation, Chinese Academy of Sciences, in 1994. He is currently a Senior Lecturer at the School of Computer Science, Cardiff University, UK. His main research interests include computer vision, computer graphics, pattern recognition, and artificial intelligence.

Yipeng Qin received a B.Sc. degree in electrical engineering from Shanghai Jiao Tong University, China, and a Ph.D. degree from the National Centre for Computer Animation (NCCA), Bournemouth University, UK. He was a Postdoctoral Research Fellow with the Visual Computing Center (VCC), King Abdullah University of Science and Technology (KAUST), Saudi Arabia. He is currently a Lecturer at the School of Computer Science and Informatics, Cardiff University, UK. His current research interests include deep learning, computer vision, computer graphics, and human–computer interaction (HCI), with a focus on generative modeling and visual content creation.

Frank C. Langbein received his Mathematics degree from Stuttgart University, Germany, in 1998 and a Ph.D. from Cardiff University, UK, in 2003. He is currently a senior lecturer at the School of Computer Science and Informatics, Cardiff University, where he is a member of the Visual Computing Research Section and leads the Quantum Control Research Group. He co-leads Qyber, a research network in quantum control, geometry, medical diagnosis, and machine learning. His research interests include control, machine learning, and geometry applied in quantum technologies, visual computing, geometric modeling, and healthcare. He is a member of the Institute of Electrical and Electronics Engineers (IEEE) and the American Mathematical Society (AMS).